Deep learning-enabled studies of galaxy mergers and supermassive black hole evolution

by

Robert W. Bickley
B.Sc., University of Connecticut, 2019

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Physics and Astronomy

Deep learning-enabled studies of galaxy mergers and supermassive black hole evolution

by

Robert W. Bickley
B.Sc., University of Connecticut, 2019

Supervisory Committee

Dr. Sara L. Ellison, Supervisor
(Department of Physics and Astronomy)

Dr. Luc Simard, Departmental Member
(Department of Physics and Astronomy)

Dr. Hossen Teimoorinia, Departmental Member
(Department of Physics and Astronomy)

Dr. Alexandra B. Albu, Outside Member
(Department of Electrical and Computer Engineering)

Abstract

When the smooth evolution of an isolated galaxy is punctuated by a merger event with a companion of similar mass, theory and observations indicate that a metamorphosis will begin. Dramatic changes in the morphologies and kinematics of merging galaxies are thought to funnel gas towards their centres, leading to elevated star formation rates and supermassive black hole (SMBH) accretion rates. The transformation brought about by mergers appears to be the missing link between the two main types of galaxies – blue star-forming spiral galaxies, and red quiescent elliptical galaxies – observed in the Universe. Simulations predict that galaxies are experiencing the most rapid changes immediately after coalescence (when the merging companions are no longer distinct objects), but observational samples of post-merger galaxies predating this work are generally incomplete (small, and possibly not representative of the post-merger class) or contaminated.

In this work, I present the methodological details of an updated post-merger identification effort using a simulation-trained convolutional neural network (CNN a type of automated machine vision tool) to flag galaxies that are very likely to be post-mergers. I present a proof-of-concept feasibility study using mock observations of simulated galaxies (Chapter 2) before applying the CNN to classify real images of galaxies in the low-redshift Universe (Chapter 3). The CNN classification effort is followed by a manual quality control exercise, which finally leads to the identification of large (with some 100s of galaxies each), pure, and defensible post-merger samples from two different imaging surveys: the Canada France Imaging Survey (CFIS), and the Dark Energy Camera Legacy Survey (DECaLS). With the post-merger samples in hand, I also present on the demographics and evolutionary characteristics of post-merger galaxies using multiple astronomical surveys for multi-wavelength characterization.

I find that star-forming post-mergers are elevated by a factor of ~ 2 in their star formation rates relative to star-forming non-merger galaxies (Chapter 4). I also find that active galactic nuclei (AGN; the observable phenomena associated with SMBH accretion) are more common by a factor of 2 – 4 in post-mergers compared to non-mergers, and that those AGN appear to be about twice as luminous as AGN in non-mergers (Chapter 5). Finally, I use new X-ray observations from the extended ROentgen Survey with an Imaging Telescope Array (eROSITA) space mission to verify that AGN are unusually common in post-mergers, and to characterize the strength of the connection between mergers, SMBH, and AGN obscuration (Chapter 6). In each result, I also compare the characteristics of the new post-merger samples to statistically identified groups of galaxy pairs that are presumed pre-mergers. Close galaxy pairs are somewhat more likely to experience elevated star formation, SMBH accretion, and obscuration than their isolated peers,

but the results for galaxy pairs are generally weaker than for post-mergers.

Together, the results of my studies indicate that the amplitude of transformation seen in post-mergers is unique in the low-redshift Universe. Looking forward, I project the viability of future astronomical surveys for post-merger identification, and find something rather unexpected: while next-generation observatories will offer an opportunity for marginal improvement in identifying the remnants of major galaxy mergers, imaging that is already available (CFIS, DECaLS) is well suited to the task (Chapter 7). I therefore posit that the present generation of astronomers studying galaxy mergers can use forthcoming surveys like Euclid and the Legacy Survey of Space and Time (LSST) to answer more difficult and granular questions about the impact of mergers on galaxy evolution.

Table of Contents

List of Tables

List of Figures

Glossary

Accretion disk A flat structure formed by material being accreted onto a supermassive black hole, usually associated with high accretion states.

Accuracy (Of a merger classification model) the number of correct predictions divided by the total number of predictions. Includes all classes, mergers and non-mergers.

Active Galactic Nucleus (AGN) The suite of observational phenomena associated with ongoing or recent supermassive black hole accretion at the centre of a galaxy.

Active galactic nucleus (AGN) feedback The processes by which AGN regulate the growth of galaxies, usually by heating or ejecting gas from galaxies.

Baldwin Phillips and Terlevich (BPT) diagram A diagnostic optical emission line diagram that separates galaxies by their dominant ionization source. Typically used to distinguish between star-forming HII regions, narrow-line regions of AGN, and low-ionization emission regions.

Broad-line active galactic nucleus (BLAGN) An AGN identified via an observationally detected broad emission line.

Broad-line region (BLR) Gas clouds orbiting close to the supermassive black hole, responsible for broad emission lines that can be observed when the active galactic nucleus is relatively unobscured.

Canada-France Imaging Survey (CFIS) An optical imaging survey taken at the Canada-France Hawaii Telescope (CFHT) on Maunakea in Hawaii.

Circum-galactic medium (CGM) The gaseous environment surrounding a galaxy outside of the interstellar medium.

Column density The area density of particles observed along a line of sight, often used to characterize how much material lies between a source and the observer. Usually given in units of atoms per cm^2.

Completeness (Of a merger classification model) the number of correctly-identified mergers divided by the total number of mergers. In this work, completeness usually refers to the post-merger class. The same statistic is sometimes called "recall".

Convolutional Neural Network (CNN) A type of deep learning machine vision model that uses convolution operations to extract visual information from images.

Corona A hot, ionized medium near the supermassive black hole thought to be responsible for exciting ultraviolet photons from the accretion disk up to X-ray energies.

Dark Energy Camera Legacy Survey (DECaLS) An optical imaging survey taken at the Cerro Tololo Inter-American Observatory in Chile, targeting the Dark Energy Spectroscopic Instrument (DESI) footprint.

Decision threshold A criterion used to select a merger sample after predictions have been made by the model. In this work, predictions are a decimal value between zero and one, and different decision thresholds have been used to achieve different proportions of purity and completeness in the final predicted merger sample.

Dex The difference in log scale between two quantities.

Euclid (space telescope) A space-based imaging and spectroscopic telescope sensitive to optical and near-infrared wavelengths.

extended ROentgen Survey with an Imaging Telescope Array (eROSITA) An ongoing space-based all-sky X-ray survey taken with the Spectrum-Roentgen-Gamma (SRG) observatory.

Hyper Suprime-Cam (HSC) An optical imaging camera at the Subaru telescope on Maunakea in Hawaii, used to take the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) imaging survey. The HSC-SSP Wide field studied in this work is often referred to as HSC-W.

Interstellar medium (ISM) The gaseous component in a galaxy.

Legacy Survey of Space and Time (LSST) A forthcoming ground-based survey to be taken at the Vera C. Rubin observatory under construction in Chile.

Low-excitation radio galaxy (LERG) A galaxy with radio emission indicating the presence of an active galactic nucleus, but with the optical appearance of a low ionization state. LERGs are generally thought to be associated with supermassive black holes in a lower accretion state.

Low-ionization emission region (LIER) A region of a galaxy where shocks, weak active galactic nucleus, or other phenomena have ionized the gas. The amount of ionization is lower than what is typically associated with an active galactic nucleus.

Metallicity The proportion of elements heavier than hydrogen and helium measured in a galaxy; chemical enrichment is achieved by stellar nucleosynthesis (stars fusing lighter elements into heavier ones). The proportion of oxygen relative to hydrogen is often used to characterize gas phase metallicity.

Mid-infrared active galactic nucleus (mid-IR AGN) An AGN identified via observationally detected hot dust, which has a distinct power-law-like spectral energy distribution in the mid-IR. It is assumed that the dust has been heated by energetic emission from the supermassive black hole accretion disk.

Narrow-line active galactic nucleus (NLAGN) An AGN identified via observationally detected narrow emission lines, e.g., using optical emission line diagnostics.

Narrow-line region (NLR) Gas clouds orbiting at a greater distance from the supermassive black hole compared to broad-line region clouds, responsible for narrow emission lines. Associated with growing supermassive black holes.

Non-parametric morphological statistics Quantities describing the distribution of light in galaxy images. Non-parametric refers to the fact that the statistics make few assumptions about the data in advance.

Obscuration In the context of active galactic nucleus studies, the extent to which signals from the AGN are hidden from view by gas and/or dust.

Post-merger A galaxy that has recently coalesced after a merger event between its progenitors. Post-mergers are generally morphologically disturbed.

Purity (Of a merger classification model) the number of correctly-identified mergers divided by the number of predicted mergers. The same statistic is sometimes called "precision".

Quiescent (Of a galaxy) not actively forming stars.

Representativeness (Of a merger classification model) the extent to which the merger sample identified by the model encompass the range and distribution of characteristics occupied by the true merger class.

Sloan Digital Sky Survey (SDSS) A landmark imaging and spectroscopic survey taken at at Apache Point Observatory in New Mexico, United States.

Spectral energy distribution (SED) The distribution of photon energies emitted by a source.

Star formation (SF) The process by which galaxies convert gas into stars. The star formation rate (SFR) is usually stated in solar masses per year, and galaxies hosting significant star formation are often called star-forming galaxies (SFGs). Star formation enhancement (the measured amount of excess star formation in a particular galaxy relative to controls) is usually written as ΔSFR.

Supermassive black hole (SMBH) The largest type of black hole, hosted in the centre of a galaxy. Often invoked as a regulating mechanism in galaxy formation.

Supermassive black hole (SMBH) accretion The process by which SMBHs accumulate mass from in-falling material.

Torus A dusty structure encircling the supermassive black hole, identifiable in the mid-infrared when heated by photons from the accretion disk. The dust may not truly be toroidal in its configuration.

Visual classification A process in which people manually inspect and label galaxies. In this work, visual classifications are generally done to identify or confirm galaxies that have completed a merger event.

Wide-field Infrared Survey Explorer (WISE) A space-based infrared telescope that mapped the entire sky multiple times between 2009 and 2011.

X-ray active galactic nucleus (AGN) An AGN identified by an observed excess of X-ray emission, presumed to originate from the corona.

Acknowledgements

I gratefully and respectfully acknowledge that this work was completed on the unceded territory of the Lekwungen speaking Songhees, Esquimalt and W̱SÁNEĆ peoples, as well as that of the Sḵwx̱wú7mesh speaking xʷməθkʷəyəm (Musqueam), Sḵwx̱wú7mesh (Squamish), and səlilwətaɬ (Tsleil-Waututh) Nations, who have stewarded the land for centuries and continue to do so today.

The Canada-France Imaging Survey is conducted on Maunakea in Hawaii. I also recognize and acknowledge with respect the cultural importance of the summit of Maunakea to a broad cross section of the Native Hawaiian community.

I would like to thank my advisor, **Sara Ellison**, for fostering an exceptional research group culture and teaching me to think like an astronomer. I am also indebted to the rest of my academic support structure: **David Patton, Mara Salvato, Hossen Teimoorinia, Leonardo Ferreira, Salvatore Quai, Maan Hani, Connor Bottrell, Shoshannah Byrne-Mamahit, and Scott Wilkinson** for spirited discussion, valuable insights, and lunch breaks of consequence.

I offer my sincere gratitude to **Samir Salim, Andrea Merloni, Jonathan Trump, Justin Hufnagel, and Joanna Woo** for mentorship, manuscript comments, and volunteer classification work.

To my dear friends: **Asya, Nic, Ruth, Rory, Mallory, Ryan, et al.** – thank you for weathering the pandemic with me. Without you, my social life in Victoria would have primarily been the responsibility of a neighborhood cat.

To my family, especially **Jackie, mom, and dad** – thank you for looking out for me, my extremely fortunate start in life, my education, and the knowledge that I can always come to Connecticut if I need anything.

Anne – You are like the ☉, and the kind of person I aspire to be. Thank you for a thousand sanity checks, your perpetual encouragement, and for cooking tomatoes, eggs, and beans with me.

Chapter 1

Introduction

1.1 The role of mergers and coalescence in galaxy evolution

Since the time of Edwin Hubble and the mainstream acceptance of the idea that galaxies are "island universes", distinct from our own Milky Way, astronomers interested in galaxies are faced with a unique set of circumstances under which our research must be conducted. We cannot reach, interact with, or arrange any manual experiments on our subjects. Moreover, since galaxies' lives unfold so gradually relative to our own, most of the data we are able to collect from galaxies have the appearance of being frozen in time. We are in essence trying to determine the rules of a mechanically complicated, multi-player game from still photographs of the players.

Early studies of galaxies in the observational domain were mainly focused on phenomenology, reporting on the diverse characteristics of galaxies: their distances from the Earth, brightnesses on the sky, and their shapes, famously organized by Hubble into a tuning fork configuration (shown in Figure 1.1, reproduced from Hubble 1936). The tuning fork, still in use in many forms (e.g., the questionnaire used for data collection in Galaxy Zoo, a citizen science morphological classification framework organized by astronomers Lintott et al. 2008; Darg et al. 2010), mainly separates galaxies into two main classes: elliptical and spiral. Later iterations (e.g., De Vaucouleurs 1959) would account for additional degrees of complexity, e.g., the fact that many galaxies host both disk-like and elliptical-like (i.e., bulge) components, but the primary sorting mechanism has endured. Stepping beyond phenomenology, Hubble proposed an underlying physical hypothesis for the galaxies, that the simpler elliptical galaxies (which he referred to as "early-type") evolved over time into the more structurally complex spirals ("late-type"). Subsequent information gained about the two galaxy classes would reveal this hypothesis to be basically inverted, with simulations of merging spiral galaxies generally giving rise to elliptical merger remnants (Negroponte & White 1983; Hernquist 1992, 1993; Hopkins et al. 2009; Taranu et al. 2013). Spirals are now known to be dominated by an ordered, rotating stellar disk component,

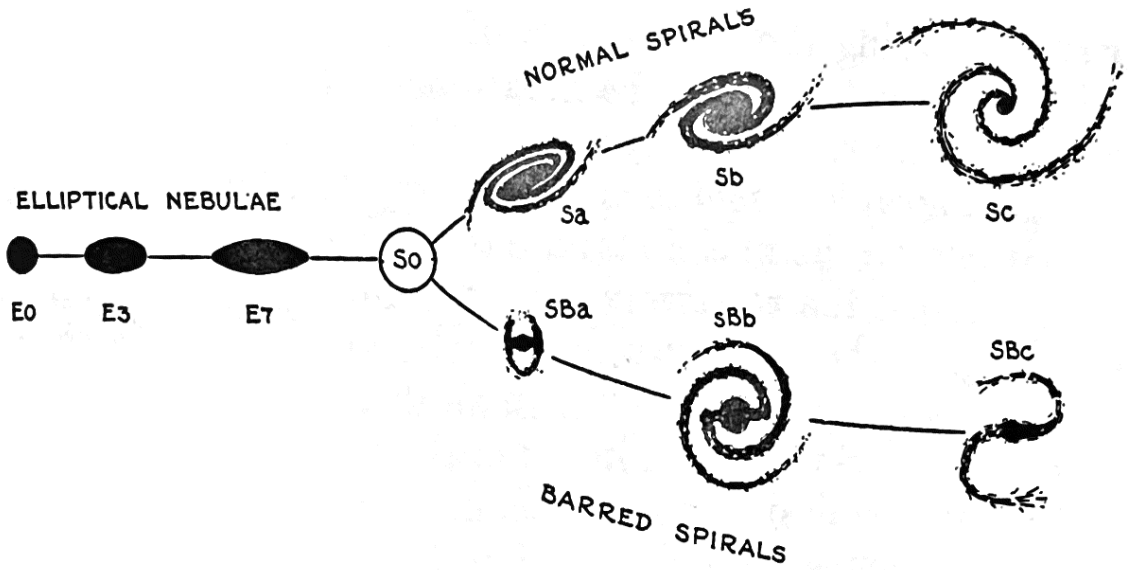

Figure 1.1: Edwin Hubble's galaxy classification "tuning fork", which separates galaxies into two broad categories: ellipticals, which are smooth, spheroidal, and dispersion-dominated, and spirals, which feature arms, a prominent disk, and are rotation-dominated. The distinction between spirals and ellipticals is fundamental to the current understanding of how galaxies evolve over time. The numerical notation for elliptical galaxies indicates the degree of elongation of the spheroid, while the alphabetical notation for spirals (lowercase a, b, c) refers to the flocculence of spiral arms. The inclusion of an uppercase B in the spiral notation indicates whether the galaxy has a bar or not. Reproduced from Hubble (1936), 1958 edition.

while ellipticals are dominated by dispersion. In an evolutionary context thus revised, galaxy interactions and mergers take on new importance.

Later in the 20th century, astronomers used more advanced astronomical instruments and sophisticated data to investigate the question of galaxy formation in greater detail. The search for scaling relations, or strong correlations between two or more galaxy characteristics, led to a highly functional understanding of the ways whole populations of galaxies evolve. Generally, scaling relations only apply to one of the two major galaxy silos (spirals or ellipticals). Important scaling relations for elliptical galaxies include the Faber-Jackson Relation (Faber & Jackson 1976), connecting the masses and luminosities of galaxies, the colour-magnitude relation (Sandage & Visvanathan 1978; Larson et al. 1980), connecting the metallicity[1] of elliptical galaxies to their masses, the Kormendy relation (Kormendy 1977), connecting the sizes and stellar densities of galaxies, and the Fundamental Plane (Djorgovski & Davis 1987), which correlates galaxy sizes, luminosities, and velocity dispersions (of which the Faber-Jackson relation is a projection). To-

[1]In astronomy, metallicity refers to the relative abundance of elements heavier than hydrogen and helium.

gether, these four scaling relations and particularly the extremely tight adherence of elliptical galaxies to the fundamental plane (within the measurement errors at the time of publication for Djorgovski & Davis 1987) indicate that the evolution of isolated elliptical galaxies is strictly regulated by parameters that are well understood: kinetic energy and self-gravitation amongst the populations of old stars that constitute classically "red and dead" elliptical galaxies.

An analogous set of scaling relations have also been discovered for spiral galaxies. The Tully-Fisher relation (Tully & Fisher 1977) parameterizes the correlation between the rotational speed of spirals and their luminosities, which is tight enough to be used as a distance indicator. The Schmidt-Kennicutt relation (Schmidt 1959; Kennicutt 1998), which applies generally for star-forming galaxies, connects the surface density of gas in star-forming galaxies to the surface density of star formation, suggesting the typical efficiency with which cold gas in star-forming complexes (found most often in spiral galaxies) can be converted into new stars. The bifurcation in scaling relations between blue, star-forming, gas-rich spirals and red, passive, gas-poor ellipticals is particularly interesting, and the fact that the same correlations cannot be effectively applied to both morphological types indicates the substantial and lasting impact of galaxy mergers.

1.1.1 Mergers in hierarchical assembly

The modern understanding of galaxy formation and evolution is underpinned by the principle of hierarchical assembly, in which smaller galaxy haloes gravitate towards and merge with one another over time to form larger ones. The first galaxies are now thought to form inside-out (i.e., with stars forming first in the centre) at the dense intersections between filaments of the cosmic web, which trace the initial and evolving dark matter (invisible and likely collision-free matter) and baryon (radiative and collisional matter) distributions in the Universe (White & Rees 1978; Blumenthal et al. 1984). Smooth and gradual accretion of gas via cosmic filaments is thought to account for much of the mass assembled by galaxies over their lifetimes, and to provide the bulk of the gaseous fuel for star formation and supermassive black hole (SMBH) accretion (Madau & Dickinson 2014). This paradigm is expected to be punctuated by numerous minor (with much smaller companions) and fewer major (with companions of a similar mass) galaxy mergers. When galaxies collide, the evolutionary process is accelerated on much shorter timescales (millions rather than billions of years, e.g., Toomre & Toomre 1972; Springel 2000; Di Matteo et al. 2007; Moreno et al. 2019; Patton et al. 2020; Hani et al. 2020), and evidence shows that mergers con-

tribute to a number of observed characteristics of the low-redshift[2] (z) galaxy population, including the canonical bimodality of galaxy colours, morphologies, kinematics, and star formation rates described above (Bower et al. 1992; Schawinski et al. 2014; Renzini & Peng 2015; Nelson et al. 2018).

The relative importance of mergers can be considered in the context of a prototypical infant galaxy. After gas collapses to sufficient density on the order of the Jeans criterion (the density requirement for a cloud of gas to collapse into a star, which depends on the temperature of the gas; Jeans 1902), star formation ensues, and the galaxy forms inside-out. The progenitors of the SMBHs seen in the centres of present-day galaxies likely also form early, with mounting evidence suggesting that the seeds are "heavy" at early times, collapsing directly from gas clouds with $M_\bullet > 10^4 M_\odot$, and perhaps even as massive as $M_\bullet \sim 10^9 M_\odot$ in some cases (Mayer & Bonoli 2018; Natarajan et al. 2024). Early in the galaxy's life, it receives a substantial kick of angular momentum via tidal[3] torques imparted by the surrounding large scale structure (e.g., nearby galaxy haloes; White 1984). Later, as the galaxy's stars and gaseous interstellar medium (ISM) interact frictionally, the peculiar motions of the stars are relaxed, and the galaxy settles over time into a disk – the most relaxed configuration for an collection of particles with a coherent angular momentum vector – in which dense molecular gas complexes facilitate further production of new stars (Tinsley 1980; Gilmore et al. 1989).

Eventually, the galaxy's secular (i.e., unperturbed by major mergers) evolution is likely to come to an end, particularly with the high merger rates predicted at high z and at early times (see Figure 6 in O'Leary et al. 2021). In the case presented here, the merger happens to be gas-rich, with one or both galaxies in possession of a significant gas reservoir. When the galaxy enters the gravitational influence of another galaxy of comparable mass, dynamical friction causes the galaxies to approach each other (Chandrasekhar 1943). At closer separations, the tidal forces affected by the interacting companion disrupt the secular characteristics of the disk (Conselice 2006a; Lotz et al. 2008). As the galaxies orbit one another during their approach, their ordered structures are stretched by tidal torques also exerted by the other galaxy. The forces produce a number of characteristic merger-induced morphological structures: stellar streams, tails, and

[2]Redshift, sometimes z, is a measure of both time and distance in astronomy. Due to the expansion of the universe, distant galaxies are moving away from us, and so their spectra are shifted towards the red by the relativistic Doppler effect. Moreover, galaxies at greater distances are observed to be moving away faster, with increasingly red-shifted spectra (Hubble 1929), so redshift functions as a distance measure. Because the speed of light is finite, one observes galaxies at higher redshift as they were in the past, when the light observed was first emitted. Redshift therefore functions as a metric of time as well (see Hogg 1999 for detailed derivations and useful forms of the relevant equations).

[3]A force that stretches a body in the direction of another body due to the gradient of the gravitational field produced by the latter.

bridges peeling away from and/or connecting the progenitor galaxies, and eventually stellar shells (spherically asymmetric plumes of stars) kicked up by strong torques in the final stages of the interaction (Darg et al. 2010; Janowiecki et al. 2010; Martínez-Delgado et al. 2010; Mancillas et al. 2019; Ferreira et al. 2024). In the final stages of the merger, the gravitational potential changes so rapidly that the stellar orbits in the participant galaxies are dramatically altered in a process called "violent relaxation" (Ridpath 2004). After some billions of years (Gigayears or Gyr hereafter), the thoroughly randomized stellar orbits in the remnant of a spiral-spiral merger are thought to settle into a smooth ellipsoid (Toomre 1978; Schweizer 1982). Similar transformations may be achieved through a number of less dramatic minor merger events (e.g., as simulated by Taranu et al. 2013), but this work focuses on the impact of major mergers.

The stellar components of merging galaxies interact only gravitationally due to the low spatial density of stars, but the kinematics of gas respond differently during the interaction due to its collisional nature. In the pre- and post-coalescence phases of the merger, rotating, non-axisymmetric structures (analogous to the galactic bars seen in some isolated spirals) that are either produced by the companion's tidal forces, or made up of accreted material from the companion, can form. As they orbit the galactic centre, their gravitational influence draws in stars and gas. As the feature rotates, stellar mass builds up in the gravitational potential of the instability. Meanwhile, gas that has fallen into the potential collides with more gas, building up a mis-aligned gaseous over-density on the leading edge of the non-axisymmetric feature. Due to gravitational influence of the stars trailing behind the gas, the gas slows, and must enter an orbit closer to the centre of the galaxy (Mihos & Hernquist 1996; Bournaud 2011). Frequently, these inflows happen very quickly, inducing compressive turbulence (Federrath et al. 2010), and producing energetic shocks where the infall speed is supersonic[4] (Bell 1978). When the gas is sufficiently slowed by the interaction, it can create an over-density in the centre of the galaxy, where it may cause a central starburst (i.e., a short-lived epoch of rapid star formation) and fall onto the galaxy's central SMBH (Barnes & Hernquist 1991; Renaud et al. 2022). Following the moment of coalescence (when the merging galaxies are no longer distinct), several studies have indicated that the most rapid transformations in star formation and SMBH activity are often underway (e.g., Mihos & Hernquist 1996; Springel et al. 2005; Lotz et al. 2008; Ellison et al. 2013; Satyapal et al. 2014; Ellison et al. 2015; Hani et al. 2020; Byrne-Mamahit et al. 2023; Tanaka et al. 2023).

After the rapid changes experienced following coalescence, the newly coalesced galaxy is

[4]In an ideal gas, the speed of sound is proportional to the square root of the gas temperature times the adiabatic index (a thermodynamic property of the gas species) γ divided by the molecular mass. The sound speed describes how quickly information in the form of pressure waves can propagate through a medium. When a mass moves through a medium at a supersonic speed, a discontinuity of pressure, density, and temperature is produced, and the energy produced by the shock must be dissipated. In astronomical shocks, the process accelerates cosmic rays.

expected to move towards quiescence (not actively forming new stars) for several reasons: gas is consumed when it is converted into stars or accreted onto a SMBH, and the elevated rates of star formation and SMBH accretion associated with the post-merger epoch naturally consume this fuel more rapidly than they would in isolated galaxies (Debuhr et al. 2011; Barai et al. 2014; Choi et al. 2014). Star formation and SMBH accretion are also energetic events, and the ionizing radiation produced by bright young stars and/or the luminous SMBH accretion disk could heat the interstellar medium to such a degree that the conditions for star formation cannot be met (Fabian 2012). Gas can be ejected by processes associated with star formation and active galactic nuclei (AGN, the suite of observables driven by the SMBH at the centre of a galaxy). Massive young stars of spectral types O and B survive only for a few millions of years (Megayears or Myr hereafter), after which they explode as supernovae (Weidner & Vink 2010). While supernovae certainly heat the surrounding interstellar medium (ISM), depending on the characteristics of the interstellar medium, these explosions could eject surrounding gas from the galaxy (Larson 1974; Navarro & White 1993; Springel 2000). Moreover, actively accreting SMBHs are known to drive polar jets of ionized material that carry gas into the circum-galactic medium (CGM) at relativistic speeds (Blandford et al. 2019). All of these energetic phenomena are thought to contribute to the truncation of star formation following a significant merger between galaxies (Smethurst et al. 2016; Baron et al. 2017), and evidence for truncated star formation is commonly observed in post-merger galaxies (Ellison et al. 2022).

1.1.2 Crash course: active galactic nuclei

Many of the characteristics of late stage galaxy mergers are thought to be consequences of AGN, since SMBH growth and AGN feedback in simulations (Donnari et al. 2021) and observations (e.g., Goubert et al. 2024) are strongly associated with the truncation of star formation and the transition between active and quiescent galaxies. The mechanisms of SMBH accretion, AGN detection, and AGN feedback will therefore be discussed frequently throughout this work (e.g., in Chapters 5 and 6), thanks to the strength of the connection between mergers and AGN. The detailed physics of AGN are also relevant to the discussion of literature and motivation for the work presented here. In this Subsection, I will therefore summarize the consensus view of AGN (Antonucci 1993; Urry & Padovani 1995), which aims to unify the observationally diverse echoes of SMBH accretion across the entire electromagnetic spectrum under a single model.

The very core of the AGN is the SMBH itself, with a size characterized by the Schwarzschild radius, $r_S = \frac{2GM}{c^2}$, where G is the gravitational constant, M is the mass of the SMBH, and c is the speed of light (Schwarzschild 1916). The radius is defined by the position of the event horizon,

within which the escape velocity is greater than the speed of light. For a supermassive black hole with $5 < \log(M_\bullet/M_\odot) < 10$, the Schwarzschild radius ranges from a fraction of a solar radius $r_S \sim 0.4 r_\odot$ to $\sim$ 200 astronomical units (earth-sun distances). Relative to the physical scales of SMBH accretion and feedback, the SMBH itself is tiny. Observationally, astronomers must search for signs of the SMBH's influence since the black hole is by definition dark.

The energy source powering the phenomena associated with SMBH accretion is the accretion disk, a flat, optically thick structure that can be geometrically thin or thick (Heckman & Best 2014), made up of material falling into the SMBH with a size of $\sim 10^{-2}$ parsecs (hereafter, pc)[5]. As matter falls into the SMBH, gravitational potential energy is converted into heat, and the accretion disk reaches temperatures on the order of $10^4 - 10^5$ Kelvin. The disk radiates as a black-body[6] with a peak wavelength in the rest-frame[7] ultraviolet (UV) or optical (Wien 1978; Carroll & Ostlie 2017). The luminosity of an AGN is governed by accretion rate, with more gravitational potential energy to be accounted for when the black hole is growing rapidly. Figure 1.2, reproduced from Harrison (2014) shows the contributions to the total spectral energy distribution (SED) of an AGN by several phenomena. The dashed blue curve shows the UV-optical contribution of the accretion disk to the SED.

A number of intervening structures can modulate the appearance of an AGN by re-processing UV photons from the accretion disk. The first is the hot corona, where UV photos from the accretion disk are excited to X-ray energies of $\sim 2 - 10$ thousand electron volts (keV) via inverse Compton scattering interactions with relativistic electrons[8] (Mushotzky et al. 1993; Haardt et al. 1994; Wilkins et al. 2016; Chainakun et al. 2019). The corona's contribution to the AGN SED can be seen in cyan on Figure 1.2. The origin of the "soft excess" (pink curve in Figure 1.2) in X-rays with energies $\sim$ 1 keV is still somewhat debated, though recent studies have suggested that it may be the result of a second "warm" (less energetic) coronal component (Boissay et al. 2016; Petrucci et al. 2018; Waddell et al. 2023). Compton up-scattered electrons may undergo an additional up-scattering after being "reflected" off of the accretion disk or dusty torus (described below), resulting in the green reflection curve extending to the highest photon energies shown in Figure 1.2 (George & Fabian 1991). The green emission line shown in the reflection component is an iron emission line, specifically the iron Kα fluorescent line at 6.4 KeV. When photons with

[5]For reference 1 astronomical unit is $\sim 5 \times 10^{-6}$ pc or $\sim 1.5 \times 10^8$ km.

[6]Black-body radiation describes the spectral energy distribution (SED) of an idealized, opaque, non-reflective mass. Under these assumptions, the shape of the SED depends only on the object's temperature.

[7]"Rest frame" refers to the emitted wavelength of light, while "observed frame" refers to the detected wavelength after being cosmologically redshifted. For a given redshift, the relationship between the observed and rest frame wavelengths is given as $z = \frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}}$.

[8]Inverse Compton scattering occurs when a photon encounters a relativistic electron (in the case of AGN, in the coronal plasma). The electron donates energy to the photon, scattering it up to a higher energy.

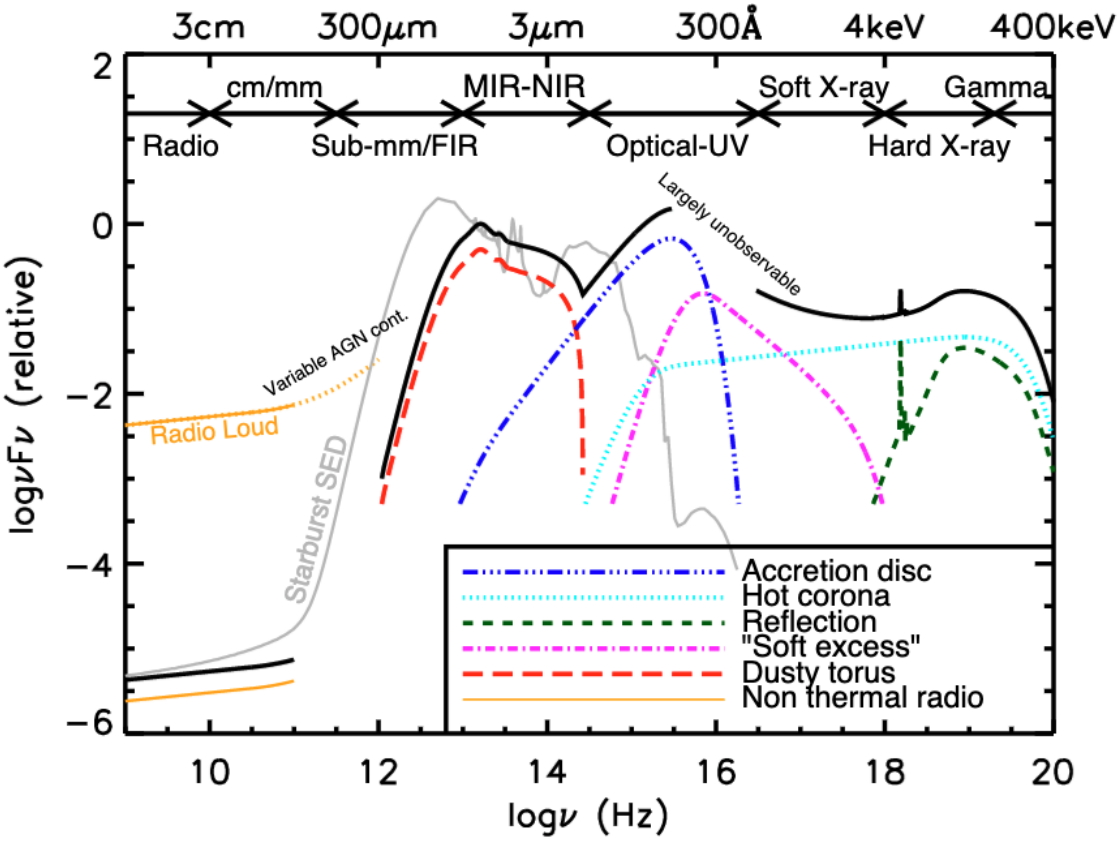

Figure 1.2: Copy of Figure 1.2 from Harrison (2014), which plots the individual contributions (colour-coded curves) of several phenomena associated with AGN to the total SED (black solid curve). The spectrum is also compared to the SED of a typical starburst galaxy (grey). Each of the phenomena represent by the colour-coded SEDs can be explained by the unified model of AGN. A model accounting for the observational diversity of AGN observations is essential in the context of merger studies.

energies in excess of 7.1 keV are absorbed by a neutral iron atom, a K-shell (n=1) electron can be excited up to the L-shell (n=2). In some cases, decaying L-shell electrons can release a photon with 6.4 keV (Fabian et al. 2000).

The next phenomenon seen in many AGN is the ∼pc-scale broad-line region (BLR), which is made up of gas clouds orbiting the SMBH at some thousands of km/s. When energetic UV or X-ray photons from the accretion disk encounter the BLR, they are absorbed, exciting electrons in hydrogen atoms. After a relatively short time – typically assumed to be on the order of minutes to hours (Peterson 2006), based on the characteristics of the hydrogen atom and the typical densities in the BLR, with particle counts on the order of $n_e \sim 10^{9-13}\mathrm{cm}^{-3}$ (Müller & Romero 2020) – the electrons recombine to the hydrogen atom at n>2, before decaying to the n=2 shell. Decay to n=2 results in photon emissions with energies corresponding to the Balmer series; Hα, Hβ, and Hγ (for recombination to n=2 from n=3, n=4, and n=5, respectively). Due to the high velocities with which the BLR is orbiting the SMBH, the emission lines are observed as Doppler-broadened around their central wavelengths. As a result, the BLR's namesake observable – the broad lines – are produced, and can be detected when there is little additional obscuring material or when the AGN is observed at a favourable angle (Ramos Almeida & Ricci 2017; Hickox & Alexander 2018). Direct detection of the BLR (as in a Seyfert 1 AGN or a quasar), the UV / optical blue black-body radiation from the accretion disk, or X-rays from the corona in excess of the expected luminosity from X-ray binaries in young stellar populations (Ranalli et al. 2003; Lehmer et al. 2019) are unambiguous evidence for the presence of an actively accreting SMBH in a galaxy. Additional direct evidence for AGN can be found at radio wavelengths, since synchrotron radiation is generated when relativistic electrons are accelerated by the magnetic field generated by the accretion disk / jet (Blandford et al. 2019). AGN with strong and weak radio components are referred to as "loud" and "quiet", respectively (see the orange curves, Figure 1.2).

Between the above-described "engine" (consisting of the accretion disk, corona, and BLR) and the observer, the unified AGN model accounts for two main interfering phenomena, shown in Figure 1.3. First, many AGN appear to host more diffuse gas clouds at greater distances from the SMBH. These clouds typically orbit about an order of magnitude more slowly (∼hundreds of km/s) than BLR nebulae at distances of 10s–100s of pc from the central engine, and are collectively referred to as the narrow-line region (NLR). In addition to narrow Balmer lines, the NLR also glows with forbidden line emission (denoted by brackets, e.g., [OIII]λ5007, [NII]λ6583). Forbidden lines are so named because the electron decay events responsible for their production can only take place in gas with extremely low densities, where collisions between atoms are unlikely (Osterbrock & Ferland 2006). The ionization and subsequent recombination signature associated with an AGN is distinct from what can be produced by star formation, since the ionizing source

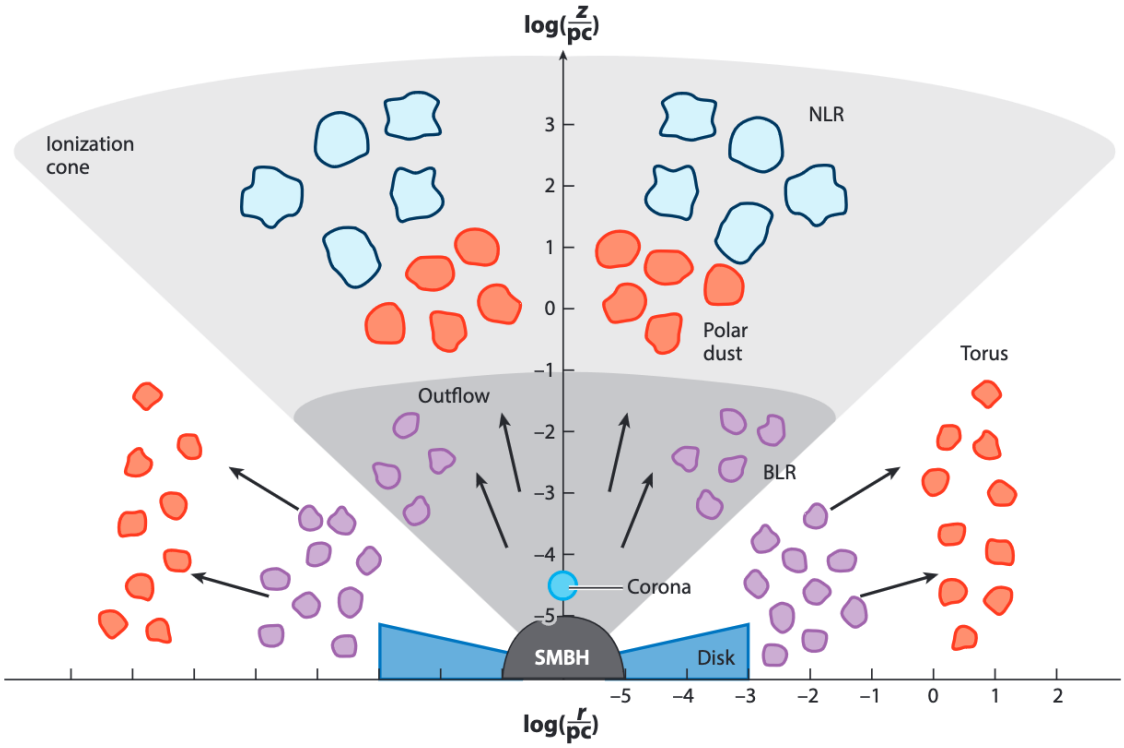

Figure 1.3: A schematic representation of the unified physical model for obscured AGN, reproduced from Hickox & Alexander (2018) and based on the figure published in Ramos Almeida & Ricci (2017). The configuration of the SMBH, accretion disk, corona, broad-line region (BLR, purple) gas clouds, narrow-line region (NLR, light blue) gas clouds, and dusty torus (red) shown in the figure suggest an explanation for the multi-wavelength diversity of AGN seen in galaxies with luminous accreting SMBHs. UV photons from the disk are the primary energy source for the observed phenomena shown in the spectral domain in Figure 1.2, but the detailed structure of the AGN determines how the light is attenuated, processed, re-radiated, and finally observed.

(the accretion disk) is harder[9]. When the NLR obscures the accretion disk / corona / BLR, a galaxy is more likely to be seen as a Seyfert 2 AGN, which can be identified based on the flux ratios of the narrow emission lines (e.g., with the Baldwin, Phillips, Terlevich, or BPT diagram; Baldwin et al. 1981) in the absence of a direct signal from the engine. The BPT diagram can be used to distinguish between galaxies whose nebular emission is dominated by star formation (characterized by the typical emission line ratios seen in star-forming HII regions with ionized atomic hydrogen), low-ionization emission regions (LIERs[10]; i.e., shocks or weak AGN), or the NLR of an AGN. In this work, multiple BPT selection criteria are used. The empirically defined Kauffmann et al. (2003) cut on the diagram is used to select star formation-dominated galaxies in Chapter 4. Since the star formation rates Brinchmann et al. (2004) used in Chapter 4 are estimated using the Hα emission line, it is optimal to ensure that the line itself is dominated by star formation. Meanwhile in Chapters 5 and 6, I use the theoretical "maximum starburst" criterion on the diagram derived by Kewley et al. (2001), which is designed to rule out any galaxies whose nebular emission could possibly be produced by star formation. Composite galaxies with contributions from both star formation and AGN are thought to lie between the two divisions (as shown on Figure 6.6). LIERs are generally separated from true AGN via additional emission line criteria, e.g., using the [SII]λ6717/Hα and [OI]λ6300/Hα line ratios (Kewley et al. 2006b) or a cut in Hα equivalent width (Stasińska et al. 2006; Cid Fernandes et al. 2011; Belfiore et al. 2016). Narrow-line AGN (NLAGN) are an important AGN subclass for galaxy evolution studies, since they are often associated with lower-mass, growing SMBHs (Boroson 2002; Woo et al. 2015). It is worth noting that the diffusivity of the gas in the NLR means that the recombination time is longer than in the BLR, with $n_e \sim 2000$ cm^{-3} and a typical recombination time of ~ 100 years (Peterson et al. 2013). Narrow-line selected AGN samples are therefore somewhat temporally smoothed, with accretion events having occurred over a wider timespan, compared to BLR or X-ray selected samples.

The second interfering phenomenon included in the unified AGN model can be observed at mid-infrared wavelengths ($2 < \lambda < 5$ μm): many AGN hosts show signs of a dusty structure encircling the SMBH at 100pc–1 kpc radii (Ramos Almeida & Ricci 2017). The dust is often referred to as being in a torus (including in this work), but some tension exists in the literature regarding the geometry and smoothness of circum-nuclear dust (Pier & Krolik 1992; Nenkova et al. 2002; Dullemond & van Bemmel 2005; Fritz et al. 2006; Tristram et al. 2007; Stalevski et al. 2012; Assef et al. 2013; Padovani et al. 2017). When dust intervenes, UV photons from the accretion disk heat the dust grains, causing the dust to glow with a thermal, power-law-like SED in the mid-

[9]"Hardness" in terms of astrophysical SEDs refers to the typical photon energies associated with a source, with higher energies being harder and lower energies being softer.

[10]When low-ionization emission regions are near the centre of the galaxy, they are sometimes referred to in the literature as low-ionization nuclear emission regions, or LINERs.

infrared (hereafter, mid-IR). The cooling timescale for the dust grains likely lies between the recombination timescales for the BLR (minutes to hours) and NLR ($\sim$ 10 years) (Ichikawa & Tazaki 2017), so temporal smoothing affects mid-IR-selected AGN samples as well. Critically, there can be degrees of obscuration by both the NLR and dust, such that a galaxy can be double-detected as a narrow-line AGN and power-law dust obscured galaxy (DOG, see Figure 6.8). In other cases, BLR and NLR emission can be blended, or a BLR can be detected in a galaxy that is also a mid-IR power-law DOG. The electromagnetic characteristics of AGN therefore lie along at least four physical axes (accretion rate, strength of BLR emission, degree of NLR obscuration, degree of dust obscuration) before light from the engine exits the centre of the galaxy. The multi-wavelength observability of AGN is further complicated by the fact that visible light is mainly attenuated by dust absorption (e.g., in the torus), while X-ray photons are largely attenuated by gas via Compton scattering[11].

As a result of the multi-wavelength diversity of AGN, different selection methods are prone to varying degrees of incompleteness (Menzel et al. 2016; Agostino & Salim 2019) and studies investigating the merger-AGN connection naturally inherit the biases associated with the AGN criteria used. Narrow-line optical (e.g., Bovy et al. 2011), mid-IR (e.g., Assef et al. 2013) and X-ray selections (e.g., Menzel et al. 2016) only overlap partially. Even with deep X-ray observations, obscuring material leads to narrow-line AGN that are undetectable in the X-ray (which are either intrinsically faint or heavily obscured). In the opposite case, emission from faint narrow-line AGN can be dominated by nebular ionization from star-forming regions, giving rise to a population of X-ray bright AGN with the optical appearance of purely star-forming galaxies.

Galaxy mergers also complicate studies of AGN. Particularly in late-stage galaxy pairs and post-mergers, the same inflows that fuel AGN also increase the amount and covering fraction of obscuring nuclear dusty gas (Ricci et al. 2017; Blecha et al. 2018). At the time of coalescence, competing effects influence the final observability of AGN: if the gas column dominates over the SMBH luminosity, post-mergers will be observed as heavily obscured or Compton-thick (with hydrogen gas column densities of $N_H \sim 10^{24}/cm^2$). If the SMBH is accreting rapidly enough to dominate over the density of obscuring material, however, the dusty gas can be cleared away (increasing the outflow rate from the nucleus by a factor of 10 as in Hopkins et al. 2016), and the observational evidence for SMBH accretion becomes plain (e.g., as modeled in Fabian et al. 2008 and tested in Ricci et al. 2022). Classically, circum-nuclear blowout is associated with late stage galaxy mergers (Sanders & Mirabel 1996). The accretion rate or AGN luminosity and the density of obscuring material therefore govern whether a SMBH will remain obscured (in a state

[11]Compton scattering occurs when an energetic photon donates energy to an electron at rest (rather than a relativistic electron, as in reverse Compton scattering). The photon loses energy and the electron recoils.

of long-lived absorption) or will launch an outflow via radiation pressure, removing obscuring material in the process. The complicated relationship between coalescence, gas inflows, AGN triggering, and observability is occasionally inconvenient, but it also means that studies of AGN in mergers are dual-purpose: in addition to investigating the role of mergers in galaxy evolution, they simultaneously stress-test the leading AGN models under the most extreme conditions in the low-z Universe.

1.2 Simulated and observed merger studies

While the narrative in Section 1.1 serves as a useful primer, it does not fully address the outstanding astrophysical questions in the field of galaxy mergers; i.e., under what circumstances, with what frequency, and to what extent do mergers drive galaxy evolution? In the preamble for this Chapter, I referred to the fact that astronomers studying galaxy evolution must make conclusions about the "rules of the game" based on slow-moving data. Fortunately, the constraints placed on the physics of galaxies by phenomenological studies have allowed astronomers to test their hypotheses directly using computer simulations of galaxy evolution in both idealized (i.e., simplified and generally involving manually chosen configurations of galaxies) and cosmological contexts. In recent decades, results from simulations of galaxy evolution have informed the arrangement of observational studies, which have in turn helped to refine the parameters used by simulators. In this way, theorists and observers are continually pushing one another towards an empirical consensus on galaxy evolution (thanks to observations) that is motivated by physics (thanks to theory).

1.2.1 Predictions from theory

Owing to their dramatic and lasting impact on galaxies, mergers serve as essential laboratories for constraining the processes governing galaxy evolution, and the simulated influence of mergers is therefore a popular topic in the literature. Simulations also benefit from specificity; astronomers can manipulate simulation parameters and out-falling data to make precise and quantitative predictions.

The most straightforward way to simulate a merger event is to instantiate a pair of galaxies on a collision course. Binary simulations (i.e., simulations involving exactly two galaxies) of galaxy mergers have been shown to produce the morphological signs of merger events, including tidal tails, bridges connecting interacting galaxies, and asymmetric shells (Toomre & Toomre 1972; Barnes & Hernquist 1993; Lotz et al. 2008). Binary merger simulations that include recipes for star formation have predicted that intense epochs of star formation follow episodes of morpho-

logical disruption (Mihos & Hernquist 1996; Di Matteo et al. 2007; Cox et al. 2008; Di Matteo et al. 2008; Lotz et al. 2008; Hopkins et al. 2013; Ji et al. 2014; Moreno et al. 2015). When AGN feedback is introduced, simulations also predict that the star formation evolution of the galaxies is regulated by the AGN (Springel et al. 2005; Park et al. 2017; Sivasankaran et al. 2022). Mergers are generally predicted to fuel the growth of SMBHs, and in major mergers with mass ratios (denoted by μ) $\mu \sim 1$, AGN feedback is associated with the truncation of star formation after coalescence when gas is removed and heated. Binary simulations also indicate that the balance between star formation enhancement, black hole growth, and ejective feedback may be responsible for the observed scaling relations between galaxies and their SMBHs (e.g., the $M_\bullet - \sigma$ relation between black hole mass and stellar velocity dispersion; Johansson et al. 2009; Barai et al. 2014; Choi et al. 2014). Other binary simulations have suggested that the gas fractions, orbital parameters, and mass ratios of mergers are important in determining the final appearances (Naab & Burkert 2003; Wuyts et al. 2010) and kinematics (Cox et al. 2006; Di Matteo et al. 2009) of post-mergers.

While binary merger simulations are ideal for constraining the impact of a particular type of merger event, they are not well suited to address broader questions about the influence of mergers on the evolution of the galaxy population. Here, cosmological simulations are favoured. By simulating a large cosmological volume from physically motivated initial conditions, growing a galaxy population from scratch, and evolving it forward to $z \sim 0$ (i.e., present day), simulators can make quantitative and statistical predictions about the role of mergers in producing the galaxy populations observed at low-z (e.g., in the ubiquitous Sloan Digital Sky Survey, SDSS; York et al. 2000; Eisenstein et al. 2011; Blanton et al. 2017). Across multiple cosmological simulations with different implementations, merger events have been shown to affect the morphologies and kinematics of galaxies (Bignone et al. 2017; Santucci et al. 2024; Wilkinson et al. 2024). They are also predicted to enhance star formation, though new stars formed as a result of mergers only account for a small portion of the cosmic star formation budget (Rodríguez Montero et al. 2019; Hani et al. 2020; Patton et al. 2020; Brown et al. 2023). Mergers are also found fueling SMBH accretion in cosmological simulations (Byrne-Mamahit et al. 2023, 2024), but as with star formation, merger-fueled accretion is predicted to represent only a small portion of the overall SMBH growth budget (Steinborn et al. 2018; McAlpine et al. 2020). In some cases feedback from the AGN leads to the eventual truncation of star formation, suggesting that mergers may be a channel for quenching (long-term shutdown of star formation) even though other mechanisms may dominate statistically (Quai et al. 2021; Davies et al. 2022; Zheng et al. 2022; Quai et al. 2023). In general, state-of-the-art cosmological simulations broadly indicate that mergers are responsible for some of the most rapid changes seen in galaxy evolution, but that the global contributions of mergers to star formation, SMBH growth, and AGN feedback are only part of the picture.

1.2.2 Observational tests in galaxy pairs

In order to characterize statistically the impact of mergers in the observational domain, a large sample of galaxy mergers is required. Thankfully, spectroscopic galaxy surveys have made it possible to algorithmically select large samples of galaxy pairs that are likely to merge (Patton et al. 2000; Man et al. 2016; Mantha et al. 2018; Duncan et al. 2019). If two galaxies are close together on the sky (i.e., with a small angular separation) and have similar line-of-sight recessional velocities as observed from Earth (determined by comparing their redshifts), it can be assumed that the galaxies may be pre-mergers destined for coalescence in some Gyrs or less (Barnes 2001; Boylan-Kolchin et al. 2008; Kitzbichler & White 2008; Lotz et al. 2008; Keenan et al. 2014, and see Patton et al. 2024, who find that $\sim$ 85 per cent of IllustrisTNG pairs with separations under 25 kpc will merge within 1 Gyr). Many studies have leveraged this technique to study the physical impact of pair-phase interactions in large statistical samples of pre-merger galaxies. The results of statistical studies of galaxy pairs have indicated that blue colours and detectable star formation activity are more common in interacting galaxy pairs, and that star formation rates in mergers are higher than in isolated galaxies, likely due to an increasing efficiency of star formation from the available gas reservoirs (Larson & Tinsley 1978; Nikolic et al. 2004; Patton et al. 2005; Ellison et al. 2008; Scudder et al. 2012; Knapen et al. 2015; Zuo et al. 2018). It has also been shown that mergers dilute the observed metallicities of interacting galaxies as gas with low [O/H] is funneled to their centres (Kewley et al. 2006a; Ellison et al. 2008; Kewley et al. 2010; Scudder et al. 2012; Bustamante et al. 2020). Consistent with simulations, interacting galaxies are also observed to host AGN more often than isolated galaxies, especially at close separations (Guainazzi et al. 2005; Ellison et al. 2011; Liu et al. 2012; Satyapal et al. 2014; Gordon et al. 2017; Hou et al. 2023), though the picture is complicated somewhat by the competing effects of SMBH accretion and obscuration (Hou et al. 2020; Guainazzi et al. 2021; Ricci et al. 2021).

The impact of ongoing mergers events on galaxy pairs is therefore very well catalogued in the literature, and appears to largely support the physical narrative for gas-rich major mergers presented in Section 1.1.1. Quantitatively, there is also reasonable agreement between the predictions of simulated models (Section 1.2.1) and the results for galaxy pairs. Still, the same models predict that the rate of transformation (the time derivative of star formation, SMBH accretion, AGN feedback, etc.) only peaks after the merger is complete. Studies of galaxy pairs alone therefore do not offer a complete representation of the role of mergers in producing the observed population of galaxies; mergers must be tracked observationally up to and past the time of coalescence.

1.2.3 Observational tests in post-mergers

It is decidedly less straightforward to identify statistically large samples of post-merger galaxies, owing to the fact that they are no longer made up of multiple spectroscopically distinct objects. As a result, conclusions in the literature vary on the importance of coalescence: a connection between coalescence and elevated star formation rates is often found (e.g., in Ellison et al. 2013; Guo et al. 2016; Thorp et al. 2019; Pan et al. 2019; Osborne et al. 2020; Thorp et al. 2022; Tanaka et al. 2023; Reeves & Hudson 2024), but the strength of the signal is highly sensitive to the method used to identify mergers. In cases where the chosen merger identification method yields a contaminated merger sample, the post-coalescence star formation enhancement signal can be diluted or lost (e.g., Pearson et al. 2019b).

A number of studies also find that AGN are found more often *per capita* and with higher luminosities in post-mergers than in non-interacting galaxies (Ellison et al. 2013; Satyapal et al. 2014; Weston et al. 2017; Ellison et al. 2019; Gao et al. 2020; Li et al. 2023b; Comerford et al. 2024), but the connection is shown to be weak or statistically uncertain in other cases (e.g., Cisternas et al. 2011; Kocevski et al. 2012; Schawinski et al. 2012; Villforth et al. 2019; Secrest et al. 2020). It has been suggested that while merger-induced AGN feedback can transform and quench galaxies (e.g., Smith et al. 2018; Ellison et al. 2022; Wilkinson et al. 2022; Li et al. 2023a), mergers are not the primary channel for quenching (Weigel et al. 2017). Others have found evidence for (Rosario et al. 2015; Pierce et al. 2022) or against (Villforth et al. 2014, 2017) an excess of galaxies with disturbed morphologies in samples of AGN, instead addressing the prominence of mergers as a channel for fueling SMBH accretion. Related efforts to quantify the connection between mergers and star formation (e.g., Luo et al. 2014) or the merger-AGN connection (e.g., Lackner et al. 2014a; Hernández-Toledo et al. 2023; Tang et al. 2023; Villforth 2023) use a more holistic approach by selecting broad merger samples that include galaxies at a range of interaction stages, again with mixed results. Still, since galaxies at various interaction stages are mixed together in such studies, the specific signal from post-mergers is diluted even as statistical power increases.

It is unsurprising that the merger identification techniques used in post-merger studies bear so heavily on the results. At the heart of the issue lies an exchange: the identification of a pure merger sample requires that uncertain cases be removed, and the identification of a complete (or representative) post-merger sample mandates that they be included. Ultimately, then, the veracity of any post-merger study is highly sensitive to the combination of completeness and precision inherent to the merger identification method. The main goal of this work is to improve on both of these statistics, and in doing so, approach an accurate measurement of the typical evolutionary impact of merger events.

1.3 Post-merger identification techniques

The state of the art for post-merger identification has changed dramatically in recent years. As the volume and quality of astronomical imaging surveys has increased, approaches to merger identification have compensated via increasing efficiency and sophistication. The essential challenge of post-merger identification is now to develop techniques that are efficient at scale while limiting loss of accuracy, trustworthiness, or explainability.

1.3.1 Visual classification

There is a strong precedent in the literature for post-merger identification by eye. Physical intuition and simulations of galaxy mergers (e.g., like those described in Section 1.2.1) have informed a mutual expectation for the appearance of post-coalescence galaxies. Visual classifications by an expert trained to disentangle the visual characteristics of post-mergers from noise, obstruction, and nominal morphological variation are widely thought to be of formidable quality (e.g., Nair & Abraham 2010; Ellison et al. 2013; Li et al. 2023b). However, inspecting of a survey's worth of galaxy images by eye would represent months of uninterrupted effort. While visual classifications set the standard in explainability – a person can describe their criteria and defend their classifications rhetorically when asked – they are not reproducible, since every person's priors on the appearance of mergers are somewhat different. Owing to their human origin, visual classifications are also highly subjective, and comparisons between the classifications of multiple professional astronomers on a single dataset (e.g., as presented later in Section 2.2.5 or in Kartaltepe et al. 2010, 2015; Robotham et al. 2014; Avirett-Mackenzie et al. 2024) generally reveal a wide diversity of approaches. The explainability and organic comfort (Dwork & Minow 2022) of human classifications would therefore ideally be retained, but newer merger identification techniques have aimed to improve on the efficiency, reliability, and reproducibility.

1.3.2 Non-parametric morphological statistics

By contrast, automated statistical methods of classification (Conselice 2003; Lotz et al. 2004; Pawlik et al. 2016; Rodriguez-Gomez et al. 2019; Nevin et al. 2019), many of which are motivated by physics, require little human input and are computationally inexpensive. Almost instantaneously, the distribution of light in a galaxy image can be collapsed into one or more summary statistics encoding degrees of information about a galaxy's merger status. Concentration (the extent to which a galaxy is centrally bright versus extended), asymmetries (the difference between a galaxy's image and mirror-image), smoothness (presence or lack of clumpy structure), Gini

(the extent to which light is equally or unequally distributed between pixels), and M20 (spatial variance of the brightest pixels), among others, provide clues, albeit ones that are highly sensitive to survey irregularities, resolution, and surface brightness limits (e.g., Ji et al. 2014; Bottrell et al. 2019a; Wilkinson et al. 2024), as to the morphological status of large numbers of galaxies (see also Section 2.2.4). The efficiency, reproducibility, and explainability (insofar as the statistics themselves can be directly explained) of non-parametric morphological statistics has made them a popular choice for merger identification (Kent 1985; Bershady et al. 2000; Strateva et al. 2001; Conselice 2003; Lotz et al. 2004, 2008; Pawlik et al. 2016; Sazonova et al. 2021). Challenges remain, since spurious artefacts like saturated pixels, foreground stars, and blank regions in images typically skew the light distribution in an image in such a way that the statistics are no longer descriptive of the galaxy. Moreover, marginalizing over merger stage, as I aim to do in this work, is challenging with non-parametric morphology, since galaxies at all merger stages are mixed together in parameter space (e.g., with large asymmetries). Efforts to overcome the morphological degeneracy between mergers at different stages often involve higher-order combinations of multiple statistics, like the Gini-M20 plane (Lotz et al. 2004) or multi-parameter fitting methods (e.g., Nevin et al. 2019, or in the proof-of-concept study by Wilkinson et al. 2024) which must be trained (rather than simply calibrated by hand, as for individual statistics). Higher-order combinations of non-parametric statistics take on many of the favourable traits associated with deep learning machine vision methods, since they include a feature extraction component (the measurements taken from the image) and a multi-dimensional statistical fitting component. As a result, they can approach similar accuracies to machine vision tools adopted directly from computer science.

1.3.3 Rise of the machines

Since non-parametric morphological statistics are somewhat limited in their potential to extract merger status from images (see Section 2.2.4), in this work I turn to deep learning machine vision methods. Deep learning is popular in both science and industry for the kind of problem explored in this work – i.e., a physical or essential characteristic of samples (in this case, merger status) can usually be estimated from the data, but the relevant feature combinations are not easily represented by a "handmade" summary statistic. A simple artificial neural network (ANN), like the one illustrated in Fig. 1.4, is a fundamental type of deep learning model. In the case of galaxies, one could imagine a simple neural network that takes multi-band photometry as input, and returns an estimate of a galaxy's redshift (indeed, this is often done to estimate redshifts inexpensively, e.g., in Firth et al. 2003; Collister & Lahav 2004; Bilicki et al. 2018). The input information is fed forward through the ANN graph. Neurons, the fundamental computational building blocks of

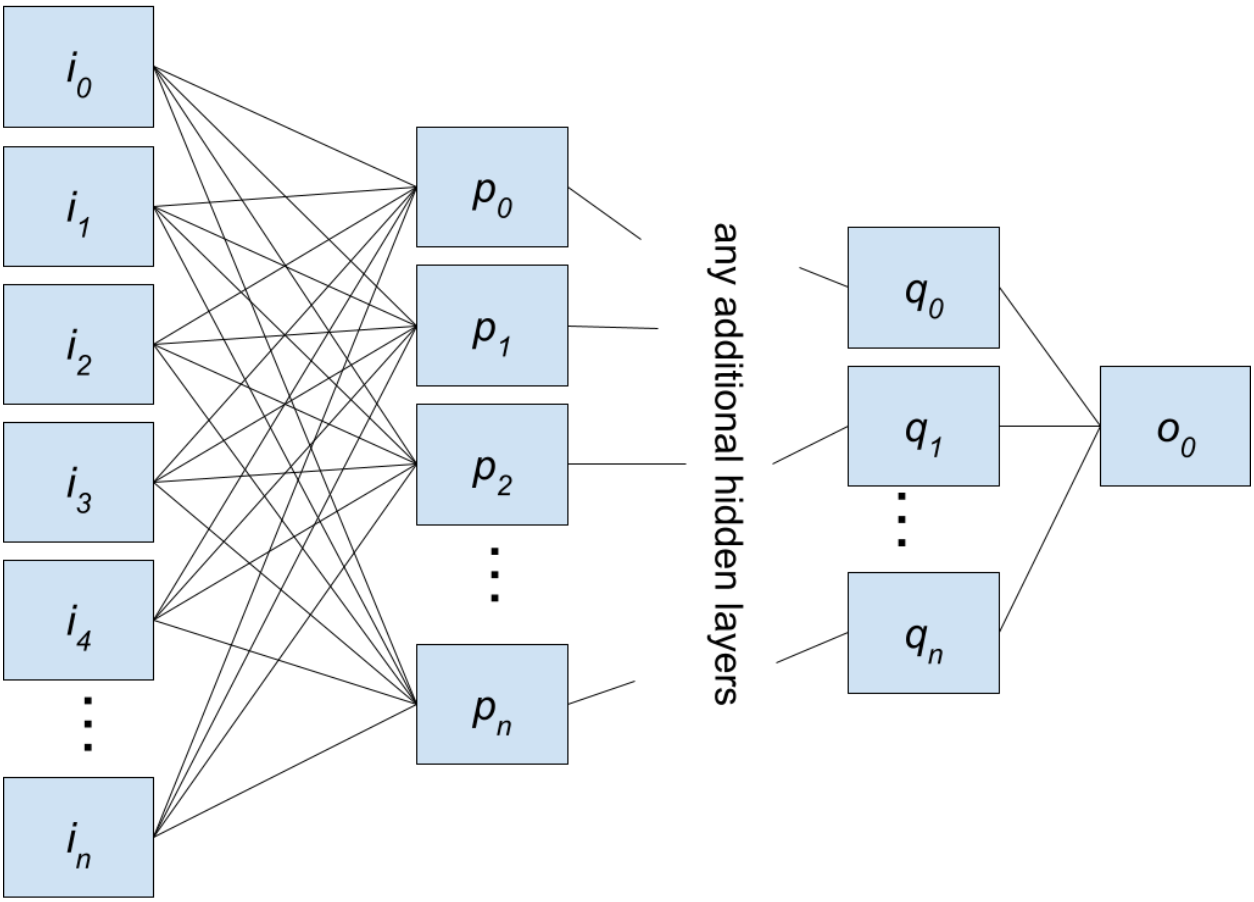

Figure 1.4: A simplistic representation of an artificial neural network (ANN), which takes some number of input parameters i_n, feeds them forward through some number of hidden layers with trainable neurons (p_n, q_n, etc.), and outputs a prediction based on the input, o_0. Signals pass from left to right on the graph, with each neuron feeding forward a weighted combination of the neurons upstream of it.

the ANN, combine the signals from other neurons upstream of them, multiply each input by a factor (called a weight), add a bias, and pass the information downstream, either to the next layer of neurons, or finally, to the model's output layer, where (in this example) the z estimate is made.

The neuron weights in ANNs and other deep learning models are generally initialized at random, or with values that have previously proven useful for a task similar to the one being undertaken. The essential "learning" feature of deep learning models is that they can be trained by iteratively updating their own weights and checking how the updates impact performance. The method of updating the weights as a function of performance is determined by a model's optimizer, an algorithm that operates on the general assumption that the loss (some metric of per-

formance, e.g., the mean squared error for photometric z predictions, or the accuracy of merger status predictions) is a smooth and differentiable function of the model's neuron parameters. The goal of deep learning is to find the minimum loss by descending the gradient of the loss function with respect to the model's parameters. Models alternate between evaluating their performance on data, and updating their weights through a procedure known as back-propagation. In back-propagation, the partial derivative of the loss with respect to a given neuron's parameters is computed, starting at the bottom of the ANN graph. The parameters for each neuron are then updated proportionally to their bearing on the value of the loss function. As its name suggests, back-propagation moves upward through the layers of the graph, checking the value of the loss function with respect to each neuron, until the whole model has been updated. Back-propagation is only possible because it exploits the chain rule for derivatives, with the derivative the loss function with respect to a given ANN layer depending on the derivative with respect to the layer beneath it, and so on. By updating its neurons in this way, the model takes a "step" in n-dimensional parameter space, where n is the number of trainable parameters in the model. In successful steps, the next evaluation step will show that the loss has been reduced. In ideal conditions, the optimizer will eventually find the combination of weights that minimize the model's loss. Unlike shallower multi-parameter methods used to combine non-parametric morphological statistics (e.g., in Nevin et al. 2019), deep learning models are capable of approximating non-linear functions, thereby encoding richer relationships between the input parameters.

Convolutional neural networks (CNNs) are defined by their inclusion of convolution layers, usually near the top of the graph. Convolution layers apply randomized geometric kernels (effectively, small weighted filters which may highlight important features) to images. The information that passes through each of these filters is carried on to the next layer of the graph, where a feature map[12] representing the local relevance of the specific geometry of the filter is produced. Additional convolution layers give CNNs the ability to identify increasingly abstract information about the input image. Frequently, convolution layers are paired with other useful operations: max pooling layers operate like convolution layers, but pass on only the maximum values from the array they receive, leading the network to consider only the most important spatial information. Additional operations, including dropout (in which a proportion of network neurons are randomly dropped, encouraging the network to generalize rather than over-fit) and normalization are often included before the final set of higher-order image features are passed into the network's fully connected layer (FCL), which resembles a simple ANN. In the FCL, the features identified by the convolution process and their relationships to one another are studied by the network. For a binary classification problem, the network's final layer outputs a floating point

[12]The convolution process generates a new 2-dimensional array for each filter.

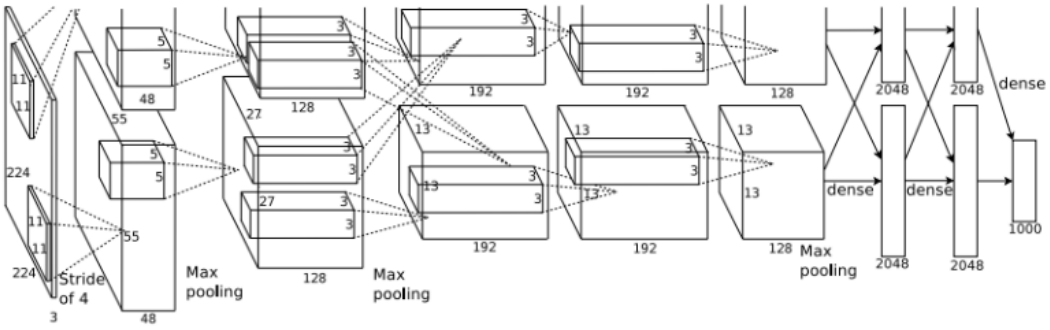

Figure 1.5: Diagram of the architecture of the AlexNet CNN, reproduced from Figure 2 of Krizhevsky et al. (2012).

value corresponding to its prediction. Figure 1.5 shows an example CNN architecture for AlexNet (Krizhevsky et al. 2012), a relatively simple (by current standards) but popular network architecture in the last decade.

Owing to their useful combination of accuracy, reproducibility, and nuance, deep learning models and particularly CNNs have gained favour in the last decade for image classification problems of all kinds, both within and outside of astronomy (Huertas-Company et al. 2015; Domínguez Sánchez et al. 2018; Jacobs et al. 2019; Domínguez Sánchez et al. 2019; Ntampaka et al. 2019; Hausen & Robertson 2019; Huertas-Company et al. 2019). The main benefits of deep learning over shallower[13] machine learning methods is their flexibility, and the power to learn and encode complex features of the data in hidden layers. Even prior to the work presented here, CNNs have shown promise as a means of enhancing the quality of morphological labels assigned to galaxies. In recent years, CNNs have been used to harvest mergers from a range of imaging surveys (e.g., Ackermann et al. 2018; Walmsley et al. 2019; Pearson et al. 2019a; Ferreira et al. 2020; Wang et al. 2020). More recently, as deep learning machine visions methods have become even more popular, they have been applied to more complex problems in galaxy classification, like high-redshift (Ćiprijanović et al. 2020; Ferreira et al. 2020) or temporal (Ferreira et al. in prep; Pearson et al. 2024) predictions of merger stage.

However, neural networks lack an understanding of physics, and so the morphological classifications that fall out of a CNN are only trustworthy insofar as the relevant features in the training data are representative of the problem at hand. For post-merger identification in particular, this means that realistic and diverse examples of post-mergers must be offered to a CNN so that it

[13]Depth in machine learning refers to the number of layers between the input (for these purposes, the image) and output (prediction of merger status). CNNs uses several hidden layers to encode the importance of feature combinations.

can begin to make informed generalizations about the post-merger and non-post-merger classes, and infer merger status from images. Additionally, the CNN must be prepared to encounter a range of features that are explicitly irrelevant to the morphological status of a galaxy: survey-related artefacts and noise, the presence or lack of bright foreground objects, and the presence of neighbour galaxies. Broadly, it is essential that the physical connection between the CNN's input (galaxy images) and output (merger status predictions) be maintained, both during training and when the CNN is used to classify real galaxies.

Importance of accurate training labels

During training, the connection between galaxy images and merger status is determined in part by the quality of the classification labels provided to the CNN. The source of these labels must be well-considered, as their trustworthiness places an inflexible upper limit on the network's ability to learn about the problem, and to make conclusions that are rooted in the physics that give rise to the morphologies visible in galaxy images. Human-provided labels, therefore, are not ideal; regardless of volume or quantity of experience in image classification, biases and strategies employed by human classifiers are frequently detrimental to the purity and overall quality of a training sample. Cosmological simulations, which host manifold populations of galaxies, circumvent the issue of visual biases by virtue of the fact that the ground truth – whether or not a given galaxy has recently coalesced after a merger – is unambiguously accessible. Even deploying a very strict definition of post-merger status, the 100-1 run of the IllustrisTNG cosmological magnetohydrodynamical simulations (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018; Nelson et al. 2019; used later in Chapter 2 of this work) includes more than 2000 post-merger galaxies in a range of environments, derived from a heterogeneous group of progenitors, and visually representative of galaxies in the observed Universe (Rodriguez-Gomez et al. 2019; Eisert et al. 2024). To the extent that the morphological and evolutionary characteristics of the galaxies spawned by IllustrisTNG are faithful to those found in the Universe, a CNN trained on sufficiently realistic images derived from these simulated galaxies will be prepared to identify real post-mergers.

It is also crucial to prepare the network for the diversity of situations in which a post-merger might be found. Due to the ubiquity of galaxy mergers within the framework of hierarchical galaxy assembly and evolution, galaxies of all types, regardless of their initial morphology, dynamical state, mass, or neighbourhood density, might undergo mergers. I seek to account for this fact in training as well, by training the CNN used in this work on post-merger and non-post-merger samples that are matched in redshift and environmental parameters, and which encom-

pass the diversity of morphologies and evolutionary histories of real galaxies. This encourages the CNN towards the conclusion that these are nuisance parameters, and unlikely to be relevant to the classification task at hand.

In addition to the verisimilitude of simulated galaxies themselves, work by Bottrell et al. (2019b) and Huertas-Company et al. (2019) have shown that observational realism, that is, observationally accurate levels of noise, and the presence of observational artefacts in representative quantities, is of paramount importance for the success of a CNN in the context of morphological classification. The connection between merger status and the appearance of a galaxy is generally weakened by observational realism (see Chapter 7), but observational effects are inevitable in the course of experimental astronomy. As such, the ideal strategy is to shift the domain of the training data as near as possible to that of the experimental data in order to prepare the CNN for the task of real-world merger identification. The approach to observational realism used in this work is based on that of Bottrell et al. (2019b), and detailed fully in Section 2.1.3.

In Table 1.1, I present the results of a number of efforts in the literature to classify mergers using deep learning machine vision tools. Several of the studies included in the table use real galaxy images that have been labeled manually *a priori* (e.g., Pearson et al. 2022; Walmsley et al. 2022), while others have selected mergers from simulations for training and evaluation (e.g., Bottrell et al. 2022; Avirett-Mackenzie et al. 2024). In each publication, I have reported the overall fractional accuracy (fraction of galaxies correctly classified in the test set) for the classification task most similar to the one described in this work, favouring binary classification frameworks (merger or non-merger) and post-merger identification in cases where multiple classification tasks are completed. If multiple datasets are used (e.g., in Margalef-Bentabol et al. 2024), I report the accuracy for the dataset most similar to the one used for validation in this work (see Chapter 2). If more than one model is used for the same classification task in a given publication, I report the best results only. For additional detail, I refer the reader to each of the publications cited in Table 1.1, and emphasize that direct comparison is not advised (e.g., the accuracy reported in Ferreira et al. 2024 appears to be lower than in this work, but the Ferreira et al. merger sample is more inclusive and some completeness is exchanged for high purity). Due to the diversity of training datasets, survey quality, observational realism, and models applied to the task of merger identification, the accuracies reported vary significantly. Broadly, accuracy is improved by higher quality imaging data, and the selection of training samples with clearly defined merger categories. The role of image quality in governing the accuracy of merger classifiers is explored in depth later in Chapter 7.

Publication	Training set	Survey	Model	Accuracy
Pearson et al. (2019a)	EAGLE pre- and post-mergers, matched controls	SDSS	Dieleman et al. (2015)	0.65
Wang et al. (2020)	TNG 100-1 pre- and post-mergers, matched controls	KiDS	Simonyan & Zisserman (2014)	0.72
Bottrell et al. (2022)	TNG 100-1 post-mergers, matched controls	None	ResNet38-V2	0.93
Pearson et al. (2022)	Galaxy zoo (Lintott et al. 2008) responses	KiDS	Custom CNN+ANN	0.86
Ferreira et al. (2022a)	TNG 100-1 post-mergers, matched star-forming galaxies	Hubble	Custom CNN	0.80
Walmsley et al. (2022)	Galaxy zoo responses, multiple categories	DECaLS	Zoobot, Tan & Le (2019)	0.88
Domínguez Sánchez et al. (2023)	NewHorizon galaxies with or without visible tidal features	HSC	Custom CNN	0.85
Omori et al. (2023)	TNG 50 pre- and post-mergers, matched controls	HSC	Walmsley et al. (2022)	0.76
Avirett-Mackenzie et al. (2024)	TNG 100-1 post-mergers, matched controls	SDSS	Custom CNN	0.81
Margalef-Bentabol et al. (2024)	TNG 100-1 pre- and post-mergers, non-mergers	HSC	Walmsley et al. (2022)	0.78
Ferreira et al. (2024)	TNG 100-1 pre- and post-mergers, non-mergers	CFIS	Multi-model ensemble	0.84

Table 1.1: The performance reported by several other works for simulation-trained deep learning identification of galaxy mergers. Efforts in the literature use a wide range of training sets; some use real, pre-labeled galaxy images, while others use a simulation-based approach. The data also come from multiple surveys, or use varying implementations of observational realism. The models used for post-merger classification also vary, but custom CNNs with architectures like the one used in this work are popular.

Importances of accuracy and purity

The natural rarity of major mergers (some $\sim$ 1 per cent) in the low-z Universe (Casteels et al. 2014) places the strongest statistical limitation on the success of any tool trained to identify post-merger galaxies. In a Bayesian framework (Bayes & Price 1763), the disadvantage lies in the small prior probability of a given galaxy in the Universe being a merger, since the prior is multiplied by the accuracy of the classifier. As a result, even with a highly accurate classification tool, misclassified non-mergers are likely to outnumber correctly classified post-mergers in the machine-predicted sample (see also Section 2.2.3). The statistical challenge of searching for rare objects is described by a convenient form of Bayes rule:

$$\Pr(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)} \tag{1.1}$$

Where A can be thought of as an intrinsic state, and B is a prediction of that state. $\Pr(A|B)$, therefore, is the posterior probability of some intrinsic state A given a prediction B, $P(B|A)$ is the likelihood of the prediction B given the intrinsic state A, $P(A)$ is the prior on the state A, $P(B|\neg A)$ is the likelihood of the prediction B when the intrinsic state is something other than A, and $P(\neg A)$ is the prior on the state being something other than A. Within the specific context of post-merger identification, one can quantify the posterior probability of post-merger status (A) given a post-merger prediction (B) of any origin, whether from a human classifier or a machine. Figure 1.6, which plots the expected purity of a predicted merger sample as a function of classifier accuracy, shows how $\Pr(A|B)$ scales with $P(B|A)$. I show the curves for posterior probability (which determines the purity of predicted post-merger samples) using two different Bayesian priors on the incidence rate of post-mergers in the Universe, one observational (Patton et al. 2000), and one simulated (from IllustrisTNG). Extremely high accuracies are required to identify post-merger samples that are reasonably pure, and broadly, the accuracies presently reported in the literature (see Table 1.1) are not sufficient. It is therefore essential to take additional steps to account and correct for contamination in the predicted merger samples.

In this work, impurity is mitigated in two ways. First, the CNN makes predictions of merger status using a continuous variable, hereafter $p(x)$. The variable serves as a probability-like indicator of merger status, with higher $p(x)$ indicating better odds of a true post-merger classification. Applying a cut in $p(x)$, sometimes referred to as a decision threshold, at a high value will reduce the amount of contamination in the machine-predicted post-merger sample. In the interest of identifying merger samples that are trustworthy and rhetorically defensible, I also conduct visual classifications on the merger samples returned by the CNN (see Section 2.2.5), effectively using a hybrid approach that retains the benefits of both automated (efficiency, reproducibility, accuracy)

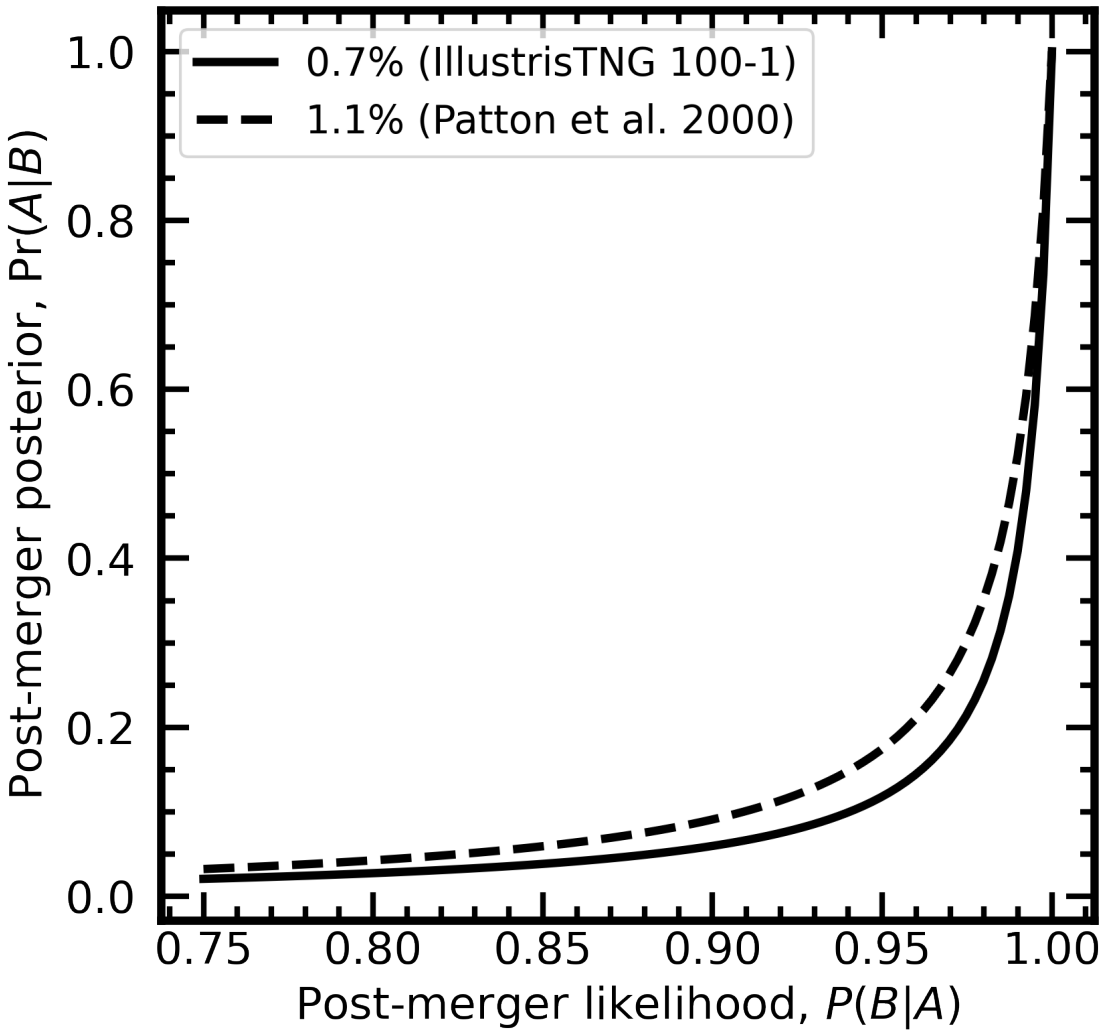

Figure 1.6: The posterior probability (vertical axis) of a predicted post-merger sample from any classifier as a function of likelihood (horizontal axis) shown for two different priors on the natural incidence rate of mergers in the Universe, one empirical (from Patton et al. 2000) and one simulated (from IllustrisTNG). $P(B|A)$ refers to the likelihood of merger status B given some prediction of merger status, A. In the context of a binary classification merger search, the posterior probability is the same as the purity of the predicted post-merger sample, and the likelihood is the same as the completeness on the post-merger class. The plot assumes a decision threshold of $p(x) = 0.5$ is used to distinguish between mergers and non-mergers. Even at high accuracies, the inherent rarity of mergers results in significant contamination. Additional steps (using a cut in merger likelihood, quality control via visual classifications) are used in this work to alleviate the problem.

and visual (trust, explainability) merger classification techniques. In doing so, my classification framework also benefits from the jury principle (Boland 1989), in which outcomes are improved via a combination of decisions by entities with different experience.

1.4 Goals of this work

The remainder of this dissertation details the research I have conducted towards the development, validation, and deployment of a novel, semi-automated framework for identifying large post-merger samples in high-quality astronomical imaging. I will also detail the scientific results obtained via observational characterization of the post-merger samples thus identified. In Chapter 2, I describe the development, training, and extensive validation of a CNN for post-merger identification, including a rigorous comparison against multiple other merger identification techniques. In Chapter 3, I detail the application of the CNN so trained to galaxy images from two surveys, the Canada-France Imaging Survey (CFIS) and the Dark Energy Camera Legacy Survey (DECaLS), and the subsequent visual classification effort. I also describe the overall characteristics of the post-merger galaxies that were identified by the CNN and confirmed visually. In Chapters 4, 5, and 6, I characterize the impact of coalescence after a major merger event on star formation, SMBH accretion, and multi-wavelength AGN observability, respectively. In Chapter 7, I investigate the potential utility of my framework for post-merger identification in legacy, current, and next-generation imaging surveys. Finally, in Chapter 8, I summarize new constraints placed by this work on the role of major mergers in determining the short- and long-term evolutionary trajectory of galaxies. For any cosmology-dependent computations (e.g., luminosity distances, projected separations between galaxy pairs) I assume parameters ($\Omega_{\rm m} = 0.3$, $\Omega_{\Lambda} = 0.7$, $h_0 = 0.7$).

Chapter 2

Developing a convolutional neural network for post-merger identification

Large observational samples of post-mergers are required for statistical characterization of the post-merger epoch of galaxy evolution. In this Chapter, I will describe the efforts leading up to the identification of large post-merger samples in two large imaging surveys: CFIS and DECaLS. First, I will describe the relevant data products and proof of concept project published in Bickley et al. (2021), in which a convolutional neural network (CNN) model was trained and validated on mock observations of galaxies from the IllustrisTNG 100-1 simulation. I will also compare the performance of the model to a number of alternative methods of merger identification, highlighting the strength and flexibility of CNNs for the task.

2.1 Data and methods

The proof of concept study in this Chapter uses two datasets: simulated galaxies for training and validation, and an imaging survey after which I will model the parameters of mock observations. I use the IllustrisTNG 100-1 simulation for the former (see Section 2.1.1), and CFIS for the latter (Section 2.1.2). The data are combined and processed using two main methods, the mock observation pipeline that gives images of the simulated galaxies the appearance of having been observed by CFIS (described in Section 2.1.3) and the CNN that will be trained and evaluated on the images so prepared (Section 2.1.3).

2.1.1 IllustrisTNG 100-1

In order to succeed, the CNN used in this work must be trained on images of post-mergers and non-post-mergers that are realistic and representative of galaxies in the Universe. To acquire galaxy samples meeting these criteria, I turn to the large-box cosmological magnetohydrody-

namical (MHD) simulation IllustrisTNG 100-1 (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018; Nelson et al. 2019).

The IllustrisTNG model is built upon Arepo (Springel 2010), a code for astrophysical simulations. Before the development of Arepo, cosmological simulations could be sorted into two main categories – smooth particle hydrodynamics (SPH), in which systems are modelled using a discrete set of particles without a mesh, and Eulerian hydrodynamics, in which systems are modelled by solving hydrodynamical equations across the surfaces of a mesh (e.g., a Cartesian) grid. Arepo retains the benefits of both SPH techniques (variable spatial resolution) and Eulerian hydrodynamics (e.g., accurate treatment of hydrodynamic shocks) while largely mitigating the disadvantages of both methods. Arepo first solves idealized equations for self-gravitation using point-particles. The positions of the particles are used in each simulation time step to determine the configuration of a 3D Voronoi tessellation (Voronoi 1908). In turn, the coupled hydrodynamic (mass, momentum, energy) and magnetic (average magnetic field) conservation equations are solved using the surfaces of each Voronoi cell as boundaries, in Eulerian fashion.

IllustrisTNG adds to Arepo a number of empirical and heuristic algorithms designed to model essential astrophysical processes (e.g., star formation, the seeding and growth of SMBHs, stellar and SMBH feedback, chemical enrichment by aging and dying stars). Since the spatial scales of star formation, ISM, and SMBH physics are below the grid resolution in IllustrisTNG, their contributions to the energy and momentum balance of the simulation are implemented based on either small-scale theoretical models or observational scaling relations. For example, stars are formed in a Chabrier initial mass function (Chabrier 2003) following the Kennicutt-Schmitt relation (Schmidt 1959; Kennicutt 1998) when gas densities of $n_H \sim 0.1$ cm^{-3} are reached in any Voronoi cells. Haloes in IllustrisTNG will form a SMBH when they reach a mass of $5 \times 10^{10} M_\odot$, and the accretion state is determined by the gas content of neighbouring cells and limited by the Eddington rate of the black hole. The feedback mechanism of the SMBH is determined by the accretion rate and the SMBH mass – for high accretion rates, the SMBH contributes thermal energy to gas in the cells nearby, and for low accretion rates, kinetic winds are launched isotropically from the position of the black hole (Weinberger et al. 2017; Pillepich et al. 2018). A complete list of the sub-grid recipes and how they differ from those in the previous version of the Illustris model are given in Table 1 of Pillepich et al. (2018). The gravitational, MHD, and sub-grid recipes are interdependent, and are therefore computed together for each time step inside the simulation volume, which itself is expanding cosmologically with simulation time. The state of the simulation is saved periodically in the form of "snapshots", and made available to the public by the IllustrisTNG collaboration, with each snapshot corresponding to $\sim$160 Myr.

The results of multiple simulation runs using the IllustrisTNG model have been published,

with a range of volumes and resolutions. For my work, I use IllustrisTNG 100-1, which offers a compromise between the other TNG options (i.e., TNG50-1 with $\sim$ 1/8 the volume and higher resolution, and TNG300-1 with $\sim 9\times$ the volume and lower resolution). The TNG 100-1 volume is cubic in shape with a side length of 110.7 Mpc and the baryonic matter resolution is $1.4 \times 10^6 M_\odot$.

Post-merger and control galaxy selection

Discrete galaxies and their merger histories are identified from the simulation using SubLink (Rodriguez-Gomez et al. 2015), and I use the methodology of Hani et al. (2020) to identify samples of post-mergers and non-post-merger control galaxies. I select galaxies for training and evaluation from simulation snapshots 50 – 99, or $z = 1$ through $z = 0$. I consider only mergers with stellar mass ratios $\mu \geq 0.1$ that occurred at $z \leq 1$ (snapshot 50 of the simulation), and for which the merger remnant is in the stellar mass range 10^{10} – $10^{12} M_\odot$. As such, a galaxy meeting the minimum stellar mass criterion for this work would be resolved with $\sim 10^4$ star particles. In each simulation snapshot, each galaxy is given a post-merger time, $T_{Postmerger}$, denoting the time since the most recent merger along its tree. I then define the post-merger sample as those galaxies whose $T_{Postmerger} = 0$ (i.e., galaxies for which a merger elapsed between the current and previous simulation snapshots). These selection criteria are designed to maximize the volume and diversity of well-resolved post-merger properties and yield a sample of 2332 post-merger galaxies (see Hani et al. 2020 for more details).

I intend to prepare a post-merger identification tool that is capable of distinguishing observed post-mergers, encompassing a range of diverse characteristics, from similarly varied observations of non-post-merger galaxies. I therefore structure my work as a binary classification problem, in which the CNN will identify a given galaxy as either a post-merger or a non-post-merger. To label non-post-mergers correctly, the CNN must be trained on a representative sample of non-post-merger galaxies. My control group identification methodology aims to construct such a sample.

To generate a control group of non-post-mergers, I adapt the observational approach of Patton et al. (2016) to the simulated domain using IllustrisTNG metadata. While many efforts in the literature have matched interacting galaxies to non-interacting controls, Patton et al. (2016) also includes a method for matching controls on environmental parameters that is emulated here. Four unique control galaxies are identified for each post-merger galaxy via a growing parameter search matching in stellar mass, simulation snapshot number (i.e., simulation redshift), the distance in kpc to the nearest and second-nearest neighbour galaxies with $M_\star \geq 0.1 \times M_{host}$ (r_1 and r_2, respectively), and N_2, the number of galaxies within 2 Mpc. Additionally, the same total stel-

lar mass cuts on the post-merger sample are applied to control pool. Control galaxies must not have undergone a merger in the last 2 Gyr (see also Hani et al. 2020). The 2 Gyr lower bound for the control group is a conservative choice, as the morphological signatures of a merger are only expected to last ∼1 Gyr (e.g., Conselice 2006b; Lotz et al. 2008). The continued observability of merger features may be further limited by imaging depth or simulation resolution. Additionally, softening the threshold to 1.5 Gyr does not qualitatively change the results.

The default matching tolerances are 0.1 dex in stellar mass, one simulation snapshot prior to or later than that of the post-merger, and ±10 per cent for r_1 , r_2 , and N_2. A snapshot (i.e., simulation redshift) tolerance is allowed, as each snapshot for $1 \leq z \leq 0$ only corresponds to a Δz of $\leq$0.05, and morphological evolution between adjacent snapshots is therefore limited. Even these relatively narrow parameters usually yield multiple eligible controls, from which one is selected at random. Once a galaxy is selected for use as a control, it is marked as such, preventing it from being reselected as a control for a different post-merger. Disallowing replacement in the training sample ensures that each selected control galaxy is unique. In the case where no eligible controls are found, the tolerances are gradually increased; the snapshot number tolerance is increased by one, the stellar mass tolerance in dex is increased by a factor of 1.5, and the r_1, r_2, and N_2 tolerances are also increased by a factor of 1.5. This is repeated as necessary in four serialized rounds, such that one control is found for each post-merger before repeating the procedure to yield four unique controls for each post-merger. This ensures that the pool of potential control galaxies is not depleted by the post-mergers that find their controls first. In the full control sample, ∼44 per cent are found using the default match tolerances, ∼24 per cent required the tolerances to be softened exactly once, ∼19 per cent required two growths, ∼10 per cent required three growths, and < 4 per cent required more than three growths in parameter space.

Fig. 2.1 shows the distribution of properties of the post-merger and control samples (where applicable, the statistics for post-mergers are repeated four times so that the histograms contain equal numbers of post-mergers and controls). The mock observed redshift distributions (top left-hand panel) match exactly as a consequence of the mirrored treatment of mock observed redshift, detailed in Section 2.1.3, whereas simulation lookback time (top centre panel) is a matched parameter from IllustrisTNG. Fig. 2.1 shows that the matching process has done an excellent job of mirroring the distributions of stellar mass, r_1, r_2, and N_2 between the post-mergers and controls. The similarity of the post-merger and control samples in gas fraction (bottom left-hand panel), calculated as the ratio of gas mass to stellar mass within twice the galaxy's half-mass radius, is a secondary effect of the parameters on which the two are explicitly matched: specifically, gas fraction exhibits a strong negative correlation with stellar mass. Fig. 2.1 also shows the distribution of the stellar mass ratio of the post-mergers' progenitors, and $T_{Postmerger}$ of the controls, where the

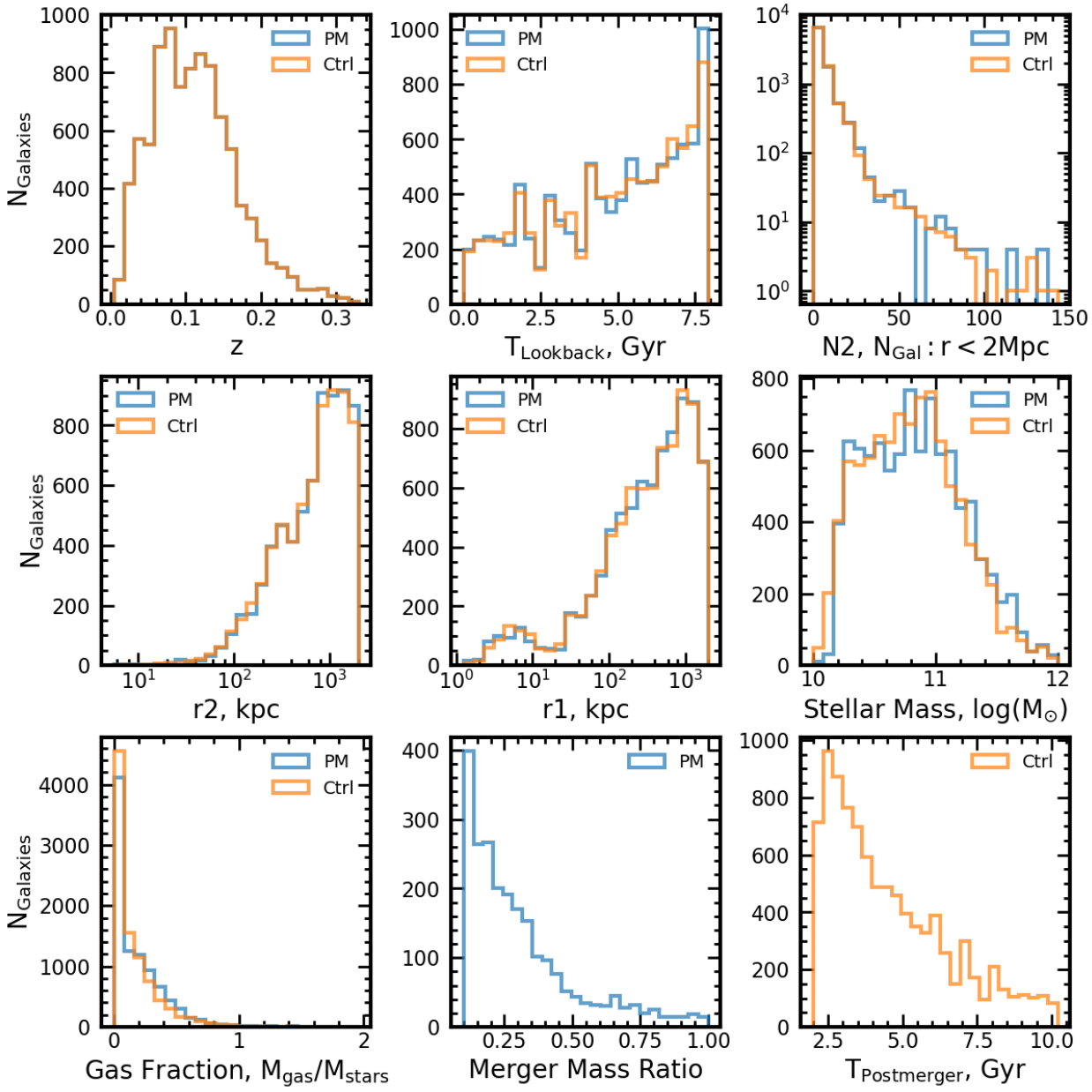

Figure 2.1: The constitutive, environmental, and observational statistics of the post-merger and control samples. Redshift is applied in mirrored fashion on-the-fly for a post-merger, control pair, and so their distributions in the top left panel match exactly. The two samples track each other in lookback time, N_2, r_1, r_2, and stellar mass as a result of the matching procedure. The similarity of the populations in gas fraction is a consequence of the parameters that are explicitly matched. Mass ratio and time-since-merger statistics are also shown for the relevant groups.

latter is zero by definition for the post-mergers and > 2 Gyr for the controls. The majority (85 per cent) of the post-mergers have a mass ratio of m $\leq$ 0.5, and many galaxies in the control sample have undergone mergers of their own within the past 5 Gyr, although the minimum $T_{Postmerger}$ cut-off for control galaxies ensures that no visual signatures of a merger remain.

2.1.2 CFIS

Because the characteristic features of the post-merger phase are relatively faint, morphological merger identification tools must be trained using imaging of adequate depth and resolution (as demonstrated by Bottrell et al. 2019b, Huertas-Company et al. 2019, and Ćiprijanović et al. 2021). CFIS achieves a favourable limiting *r*-band[1] 5σ point-source depth of 24.85 mag, $\sim$0.7-arcsec seeing[2], a pixel scale of 0.187 arcsec, and has strong overlap with the SDSS DR7 spectroscopic galaxy catalogue northwards of 30 degrees declination (see Figure 2.2). CFIS offers a useful combination of imaging quality and volume, with *r*-band imaging that will eventually cover 5000 square degrees of the sky. The depth of the survey is sufficient to capture the low-surface brightness features necessary for merger identification at low-redshift (e.g., Sola et al. 2022). Estimates of the low-redshift merger rate suggest that the final CFIS footprint will include thousands of post-mergers (Lacey & Cole 1993; Lotz et al. 2011; Bluck et al. 2012; Casteels et al. 2014; Rodriguez-Gomez et al. 2015; Martin et al. 2018). CFIS is therefore a good choice for a post-merger search in terms of imaging quality and volume.

CFIS, part of the Ultraviolet Near Infrared Optical Northern Survey (UNIONS) collaboration, is compiled from images taken using the MegaCam wide field optical imager at the 3.6-m Canada-France-Hawaii Telescope (CFHT) on the summit of Maunakea. MegaCam saw first light in 2003, and at the time it was the largest astronomical charged couple device (CCD) mosaic ever built, with a full 1x1 square degree field of view (FOV). The observing pattern employed by CFIS uses three single-exposure visits with FOV offsets in between for optimal astrometric and photometric calibration with respect to observing conditions. The observing pattern also ensures that the entire survey footprint (including areas in the "chip gaps" between MegaCam's multiple CCD segments for a given exposure) will be visited for at least two exposures. After raw images are collected by CFHT, they are detrended (i.e., the bias is removed and the images are flat-fielded using night sky flats) with the software package MegaPipe (Gwyn 2008, 2019). The images are next astrometrically calibrated using Gaia data release 2 (Gaia Collaboration et al. 2016, 2018) as a reference frame. Pan-STARRS 3π *r*-band photometry (Chambers & Pan-STARRS Team 2016)

[1]The CFIS *r*-band is an orange-red visible light filter with its centre at $\sim$ 625 nanometers.

[2]"Seeing" in an observational context refers to the resolution set by the point-spread function (PSF) of the telescope. For ground-based telescopes, seeing is typically dominated by atmospheric blurring.

is used to generate a run-by-run differential calibration across the MegaCam mosaic, and an image-by-image absolute calibration. Finally, the individual images are stacked on to an evenly spaced grid of 0.5×0.5 square degree tiles using Pan-STARRS PS1 stars as in-field standards for photometric calibration.

CFIS includes both *u*- and *r*-band imaging, but I use only the *r*-band throughout this work. While multi-wavelength data could doubtless be used to improve the accuracy of classifiers for merger identification, the practice would give rise to complications when the classifications are used for scientific purposes. Since mergers are known to alter the star formation histories, and thereby, the multi-wavelength colours of galaxies, classifiers could learn to identify mergers based on their colours rather than their morphologies – in essence, learning to identify the aftermath of a merger instead of the direct, morphological evidence. In order to avoid biases brought on by colour information in the training data, I use only monochromatic imaging. While the problem could likely be circumvented with a large enough training set (including large numbers of mergers and non-mergers with a variety of star formation rates and colours), I choose to proceed with caution by excluding colour information.

A map of the on-sky coverage of CFIS data release 2 (DR2), as well as other surveys relevant to this work, is shown in Figure 2.2. The CFIS area overlaps substantially with that of SDSS. While SDSS imaging is not ideal for merger identification, it is used extensively in this work for spectroscopic characterization of mergers after they are identified. I also show the coverage map for eROSITA, the X-ray survey used in Chapter 6, and DECaLS, which I use to identify mergers for the study in Chapter 6. CFIS is now in its fifth data release, but data releases 3-5 were made available after the completion of this work.

2.1.3 Synthetic observation pipeline

Several publications (Bottrell et al. 2019b; Huertas-Company et al. 2019; Ćiprijanović et al. 2021) have studied the importance of observational realism to the reliability of CNN predictions of galaxy merger stage and found that mismatched levels of realism in training and test images would incur a significant penalty in performance. Further, training data with appropriate observational realism were found to be more important to CNN performance than the treatment of either colour or radiative transfer. Because I aim to construct a post-merger identification tool for the specific case of CFIS, I combine the survey's available image data and metadata in order to construct a sample of synthetic training images that patterns itself after the statistical properties and observational qualities of the survey. In addition, I forgo the use of radiative transfer in favour of unprocessed stellar maps as a starting point for the mock observations in order to limit

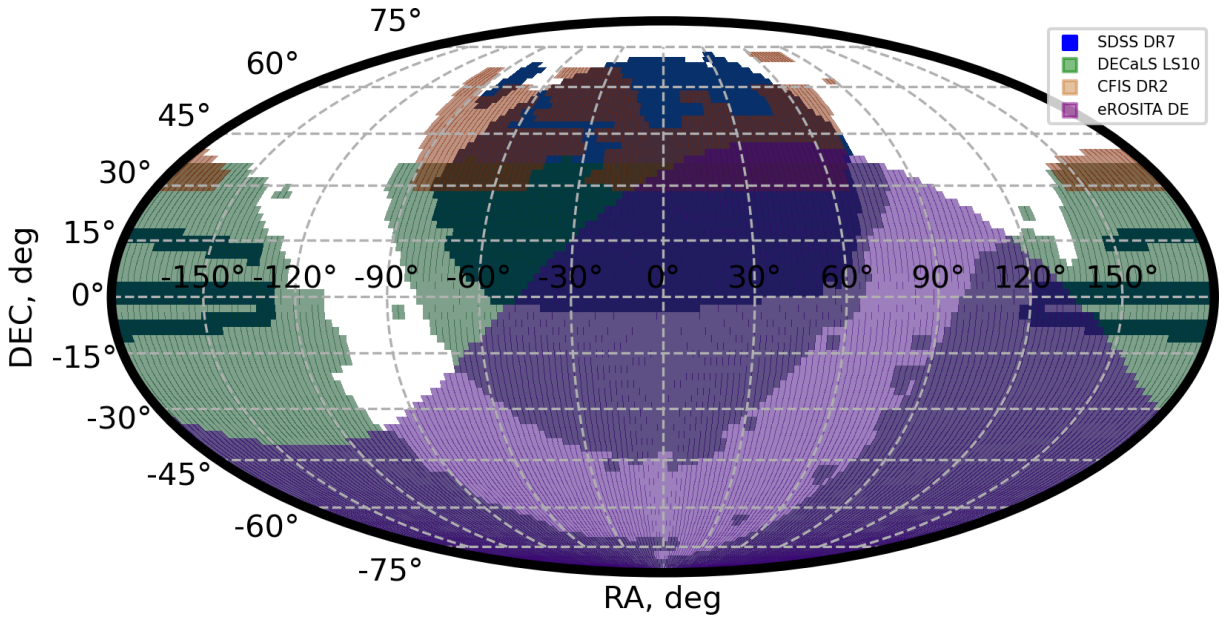

Figure 2.2: Figure illustrating the on-sky coverage of astronomical surveys relevant to this work. SDSS data release 7 (DR7) is shown in blue. Galaxies in SDSS DR7 are selected homogeneously within the area, and optical spectra for each source are publicly available. Dark Energy Camera Legacy Survey (DECaLS) imaging, obtained via the Legacy Survey data release 10 (LS10), and Canada France Imaging Survey data release 2 (CFIS DR2; green and orange, respectively) are each used to obtain deep and high-resolution images of the galaxies catalogued in SDSS DR7. Although SDSS imaging is available, both DECaLS and CFIS better highlight the low-surface brightness features associated with recent merger events. In the next Chapter, details of the merger identification efforts in both CFIS and DECaLS will be presented. In Chapter 6, I use X-ray observations made available through eROSITA. The on-sky coverage of eROSITA (German consortium) is shown in purple.

computational expense as proposed in Bottrell et al. (2019b).

Synthetic observations for this work begin with stellar mass maps rendered from IllustrisTNG 100-1 for each selected galaxy by Bickley et al. (2021) co-author Maan Hani. Each galaxy is observed by four different camera angles, located at the vertices of a tetrahedron that is aligned with the simulation box, with the galaxy at its geometric centre. As a result, the first camera angle looks directly down on each galaxy perpendicular to the top of the simulation box, and the remaining three are inclined upwards from horizontal. These camera angles are chosen to capture distinct projected morphologies for each galaxy so that the CNN will be able to study a greater number of unique examples. Each map is 100 kpc and 2048 pixels on a side. These stellar mass maps are then normalized using the galaxy's r-band absolute magnitude from the Nelson et al. (2018) stellar photometrics tables to produce a pre-cosmology image of intrinsic surface brightness. Single-band photometry ensures that the features learned by the CNN are morphological, and not biased by higher order information such as colour or starburst identification. The top rows of Figs 2.3 and 2.4 show selected stellar mass maps for post-mergers and their controls, respectively.

The pristine surface brightness images are next convolved with RealSim-CFIS, a custom version of the software package RealSim, originally detailed in Bottrell et al. (2019b). The code was first developed to construct synthetic images of simulated phenomena as they might be observed by SDSS. In both versions of RealSim, the noise, resolution, and sky insertion positions for each mock observation are selected to match the statistics of sky brightness, seeing, image artefacts, and projected environment (i.e., crowding) for a catalogue of real galaxies. Small modifications were required to adapt the code for synthetic CFIS observations: CFIS has an angular resolution of 0.187 arcsec pixel^{-1}, a factor of $\sim$2 higher than that of SDSS, allowing for greater preservation of the detail in the original unprocessed images. I also remove the source Poisson noise feature from the original RealSim in the interest of simplicity. A given segment of CFIS is imaged by a unique combination of potentially different CCDs, and building a gain map to compute the Poisson statistics for a mock observation would require reverse-engineering of the gain for each pixel. RealSim would also need to be modified to accept such a gain map. I anticipate that any contribution by Poisson noise would be negligible compared to other simulated sources of noise applied later in the RealSim-CFIS pipeline. Before adding realistic observational effects to an image, it is first converted to SDSS-specific units of flux (nanomaggies) on a pixel-wise basis, with $10^{-0.4(s-22.5)}$, where s is a pixel's surface brightness value in magnitudes per square arcsecond.

The layers of realism applied by the code are as follows:

(i) Redshift dimming. Since future studies of the properties of post-mergers identified by the CNN in CFIS will make use of ancillary data from the SDSS, I choose insertion redshifts

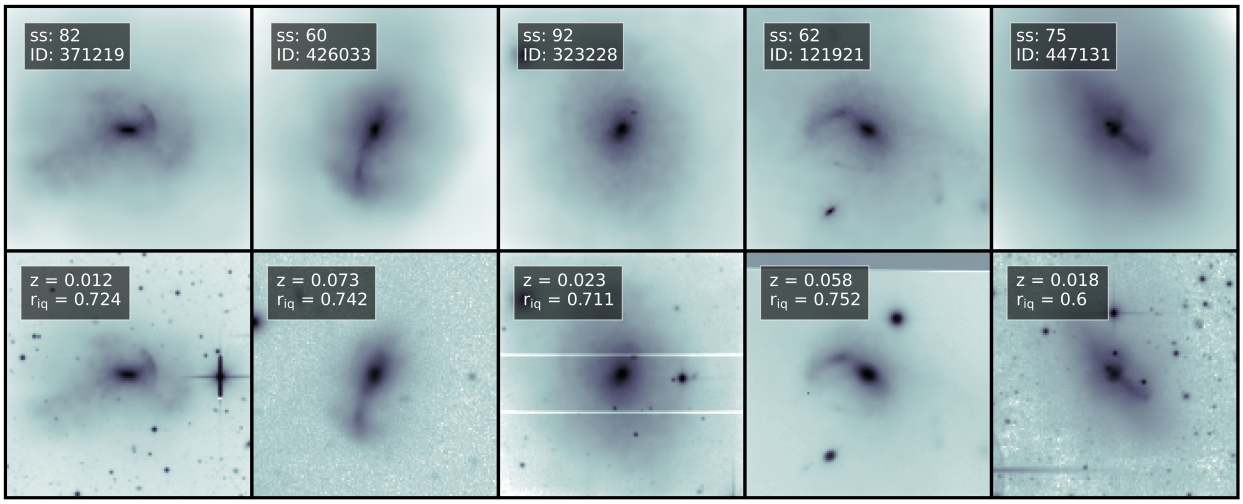

Figure 2.3: Mosaic of 5 randomly selected post-mergers identified in the IllustrisTNG 100-1 run. Top row: stellar mass maps with the legend indicating the snapshot number and galaxy SUBFIND ID. Bottom row: Insertion into an actual CFIS *r*-band image with the legend indicating the insertion redshift and the image quality (in arcseconds) of the original image. Note how artefacts, such as saturated stars and CCD features are retained in the image.

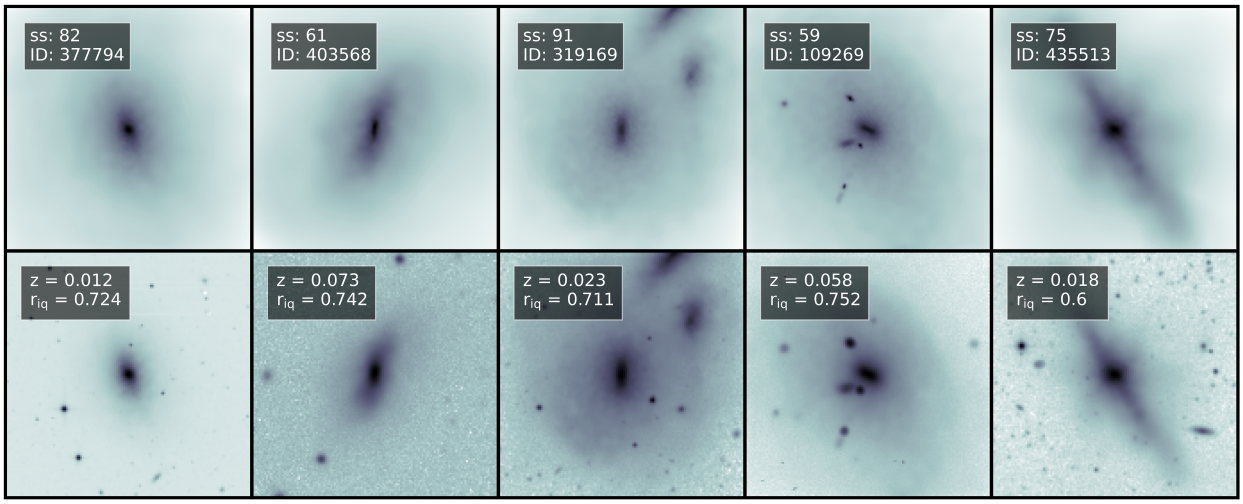

Figure 2.4: As for Figure 2.3 but for control galaxies corresponding to each of the five post-mergers shown in Figure 2.3. Note how the insertion redshifts are identical, as per the mock observation methodology. However, the seeing and image quality can differ between a given post-merger and control, as is usually the case in observational studies.

to match a realistic observational redshift distribution. This redshift distribution is obtained by cross-matching CFIS DR2 object catalogue with the SDSS DR7 with an angular separation tolerance of 2 arcsec. Insertion redshifts are assigned to galaxies in the IllustrisTNG post-merger sample by randomly drawing values from this catalogue (with replacement). I cap the allowed mock observation redshift distribution at $z = 0.5$ in order to rule out any galaxies whose z_{spec} may have been erroneously measured, or any high-redshift quasars that may have been included in the sample. The top left-hand panel of Fig. 2.1 shows the mock observation redshift quantities obtained by drawing at random from the parent redshift distribution thus obtained. Each post-merger and associated control galaxy is assigned identical redshift selections. The first camera angle for a given post-merger and the first camera angle for its matched control galaxy would therefore be mock observed at the same redshift. Once a redshift is selected, the image is realistically dimmed by a factor of $(1 + z)^{-5}$. While cosmological surface brightness dimming accounts for a factor of $(1 + z)^{-4}$ in bolometric surface brightness, an additional factor of $(1 + z)^{-1}$ accounts for the dimming in a given bandpass, since the extent to which the rest-frame spectral energy distribution (SED) of a source is "stretched" increases with z.

(ii) Rebinning. Once the angular size has been calculated from a given redshift and the physical size of the image (100 kpc), the redshift-dimmed image is rebinned to CFIS's actual CCD pixel scale in total flux-conserving fashion.

(iii) Point spread function (PSF). The observational PSF for galaxies in CFIS is dominated by the effects of atmospheric seeing, and galaxies have associated PSF full width at half-maximum (FWHM) values, recorded as r_{iq}, available as a metadata quantity in the CFIS catalogues. CFIS sky image data are separated into regularized 0.5-degree tiles, where each tile is the combination of several individual images. There is a small amount of variation in seeing across each tile resulting from the image mosaic. In order to model a spatially variant PSF, I draw available PSF measurements on the CFIS tile where I intend to perform a mock observation. I next fit a Gaussian function to the resulting distribution, and sample a value from the function at random. Consequently, I obtain a realistic and non-discrete approximation of the survey's atmospheric and instrumental seeing as a function of tile. The PSF "bootstrapping" approach yields similar PSF distributions for the post-merger and control image sets because the underlying statistics are the same, but mirrored distributions are not enforced for post-mergers and controls, as a pair of galaxies matched on their physical properties and redshift are unlikely to be observed in identical conditions.

(iv) CFIS sky backgrounds. Real CFIS skies are added to each image. Since skies appearing in the survey already include statistically representative noise, crowding, foreground objects, and artefacts, there is no further need to add them manually. Following the methods presented in

Bottrell et al. (2017), I employ a sky selection method that makes use of CFIS statistics in order to match the spatial distribution of real galaxies. In order to choose a CFIS tile, and in turn find a suitable survey location for a mock observation, I select a real proxy galaxy from the catalogue of SDSS DR7 galaxies within the CFIS DR2 coverage. The proxy galaxy's tile will be used for the mock observation, and as a result, the CFIS tiles that are more densely populated with galaxies are more likely to be chosen for mock observations. In order to select a specific region of the tile for the mock observation, a large 11-arcmin^2 cutout is generated, centred on the RA and Dec. of the proxy galaxy. Next, SourceExtractor (Bertin & Arnouts 1996) is used to identify an insertion location where the pixel corresponding to the centre of the output image cannot be flagged as a source. This allows for realistic overlap between the IllustrisTNG galaxy and survey objects. The chosen patch of sky is then added to the mock observation image, which is already at the CFIS CCD scale after rebinning. All of the original features and artefacts of the original CFIS image, such as saturated stars, missing sky coverage, and CCD defects, are deliberately maintained in the final mock images through addition (in the case of high-flux artefacts) or masking (for zero-flux artefacts) in order to proportionally expose the CNN to phenomena that it is likely to find in a sample of CFIS images.

In the rare case where the final image is dominated (> 50 per cent) by zero-flux pixels due to lack of sky coverage, the synthetic observation is discarded, and a new observation of the same galaxy is attempted. The bottom rows in Figs 2.3 and 2.4 show the same post-merger and control galaxies as in the upper rows, but with all of the above described observational realism included.

(v) Mock observation scheme. Because of the relative scarcity (2332) of post-mergers in IllustrisTNG[3], I perform four unique mock observations of each of the four camera angles for every post-merger. Each post-merger mock observation (at each camera angle) consists of its own unique random draw from the redshift distribution and sky insertion location. This results in 16 distinct images of each post-merger. Since there are four controls matched to each post-merger, each projected onto four camera angles, only one set of mock observations is generated for each control. The above scheme therefore yields a balanced image data set, with 37312 unique images belonging to each class.

I reserve 10 per cent of the post-merger galaxies, and all of their associated images, and an equal number of control galaxies (2.5 per cent, since the control group contains four times as many galaxies), for testing. The remaining images are shuffled and split in traditional supervised learning fashion between the training set, which the networks study in detail (90 per cent), and the validation set, which is used to check progress from time to time (10 per cent). Different

[3]While a sample of $\sim$ 2000 post-mergers is not small by science standards, it is very small for the purposes of training a CNN.

camera angles and mock observations of the same galaxy produce drastically different images. Therefore, training and validation data are partitioned by individual image and not by galaxy to maximize the training set's morphological diversity and prepare the model for a broader range of possibilities.

(vi) Normalization. Prior to training, all images are normalized such that the minimum pixel value is at 0, and the maximum pixel value is at 1. I normalize images in linear fashion, by subtracting the value of the faintest pixel from the image, and then dividing by value of the brightest pixel. Since the CNN architecture used in this work requires input images of a fixed size, I resize all images to 138 pixels on a side, while maintaining a constant physical width of 100 kpc. At $z = 0.102$, the median simulated redshift of the galaxy population in the union of CFIS DR2 and SDSS DR7, an image with a physical scale of 100 kpc would be 138 pixels across on the MegaCam CCD. Consequently, a fixed size of 138 pixels allows me to minimize the total required amount of resizing. Galaxy images with mock observed $z \leq 0.102$ are downscaled, and those with $z > 0.102$ are upscaled to meet this standard.

(vii) On-the-fly augmentation. Overfitting, the process by which machine learning models memorize the properties of their training data at the expense of general usefulness, is of concern for classification problems. I take numerous precautions against overfitting in this work because all but 10 per cent of the post-merger galaxies appear 16 times in the combined training and validation sets. Each of the 16 appearances is visually unique due to different redshifts and sky insertions. Hence, I use data augmentation to help overcome this concern. Before each training epoch, I apply minor randomized image transformations using the ImageDataGenerator class in Keras (Chollet et al. 2015) – vertical and horizontal shifts, shear transforms, and zooms of at most ±10 per cent, as well as horizontal and vertical flips. In this way, even though the same source images are used in each round of training (epoch), the network never studies the same permutation of an image more than once.

CNN architecture used in this work

The adoption of machine vision techniques in astronomy has grown substantially in ubiquity in the last decade, largely out of necessity – current and next-generation astronomical surveys take spectroscopic, imaging, and photometric data in such volumes that individual study of objects by people has become highly inefficient. Automated categorization and characterization of astronomical objects is essential to the extraction of scientific knowledge from the large datasets available to modern astronomers.

I use a single, custom CNN architecture for merger identification throughout this work. The

CNN is based generally on the structure of AlexNet, and I have made modifications informed by the work of colleagues who have used CNNs for merger identification (e.g., Bottrell et al. 2019b) as well as by my own experience testing various CNNs for this work. More sophisticated CNNs (residual learning frameworks; e.g., He et al. 2015, or EfficientNets; Tan & Le 2019) and other deep learning machine vision models (e.g., vision transformers; Dosovitskiy et al. 2020) were already available at the time when this work was completed, and have grown in popularity since. My main motivation for using a fairly basic CNN is that the scale of the data used in this work (37312 images in each class) is relatively small compared to current training sets used in the machine learning domain (e.g., the ImageNet training set contains over one million images). The accuracy of new deep machine vision models on thousand-class, million-image datasets is therefore only tangentially beneficial to my two-class science case. Moreover, the tests in this Chapter were repeated using an implementation of a ResNet18[4] (He et al. 2015) model. Consistent with expectations, the final accuracies on the test set were similar (within $\sim$ 2 per cent for each class) between the much deeper ResNet and the simple CNN. During training, the ResNet model converges more slowly and slightly overfits the data, with a gap of $\sim$ 5 per cent emerging between the training and validation accuracies by the end of the training schedule. Meanwhile, the dropout implemented in my CNN wholly avoids overfitting on the same data. Even though both models achieve similar final performance, I use the lighter model in the interest of simplicity. It is possible that performance could be meaningfully improved via an extensive tuning effort for a more sophisticated baseline architecture like ResNet. However, such an effort would require substantial amounts of time and computing resources. Indeed, since misclassified galaxies in the test set are found to be mainly "edge cases" (e.g., pre-merger galaxy pairs, see Figure 2.8 below) the results of this Chapter indicate that my merger classification approach is already data-limited, rather than model-limited.

The CNN has four hidden layers, and is built with a combination of open-source software tools in Python[5] using Keras (Chollet et al. 2015) for network construction, training, and data augmentation, and Scikit-Learn (Pedregosa et al. 2011) for partitioning the data into training (studied in detail by the model), validation (used periodically during training to check performance), and test (reserved until training is complete to evaluate the model's performance on unseen data) subsets.

Table 2.1 shows the constitutive Keras layers for the CNN, as well as the number of trainable parameters associated with each. Figure 2.5 shows the model architecture visually as well. In all convolution layers, I use a stride of 1 pixel, and rectified linear unit (ReLU) activation function

[4]The numeral 18 in ResNet18 refers to the fact that the model is 18 layers deep.
[5]www.python.org/

Layer Type	# Parameters	Output Shape
Input	0	(138,138,1)
Convolution 32 Filters Kernel (7,7)	1600	(138, 138, 32)
Max Pooling (2,2)	0	(69, 69, 32)
Dropout 25%	0	(69, 69, 32)
Convolution 64 filters Kernel (7,7)	100416	(69, 69, 64)
Max Pooling (2,2)	0	(34, 34, 64)
Dropout 20%	0	(34, 34, 64)
Batch Normalization	256	(34, 34, 64)
Convolution 128 filters Kernel (7,7)	401536	(34, 34, 128)
Max Pooling (2,2)	0	(17, 17, 128)
Dropout 20%	0	(17, 17, 128)
Convolution 128 filters Kernel (7,7)	802944	(17, 17, 128)
Max Pooling (2,2)	0	(8, 8, 128)
Dropout 20%	0	(8, 8, 128)
Flatten	0	(8192)
Dense	4194816	(512)
Dropout 25%	0	(512)
Dense	65664	(128)
Dropout 25%	0	(128)
Activation, Sigmoid	129	(1)

Table 2.1: The CNN architecture used in this work. Each layer begins with the stock Keras layer of the same name, with any specified hyperparameters detailed in the Layer Type column. The # Parameters column shows the number of trainable network parameters belonging to each layer.

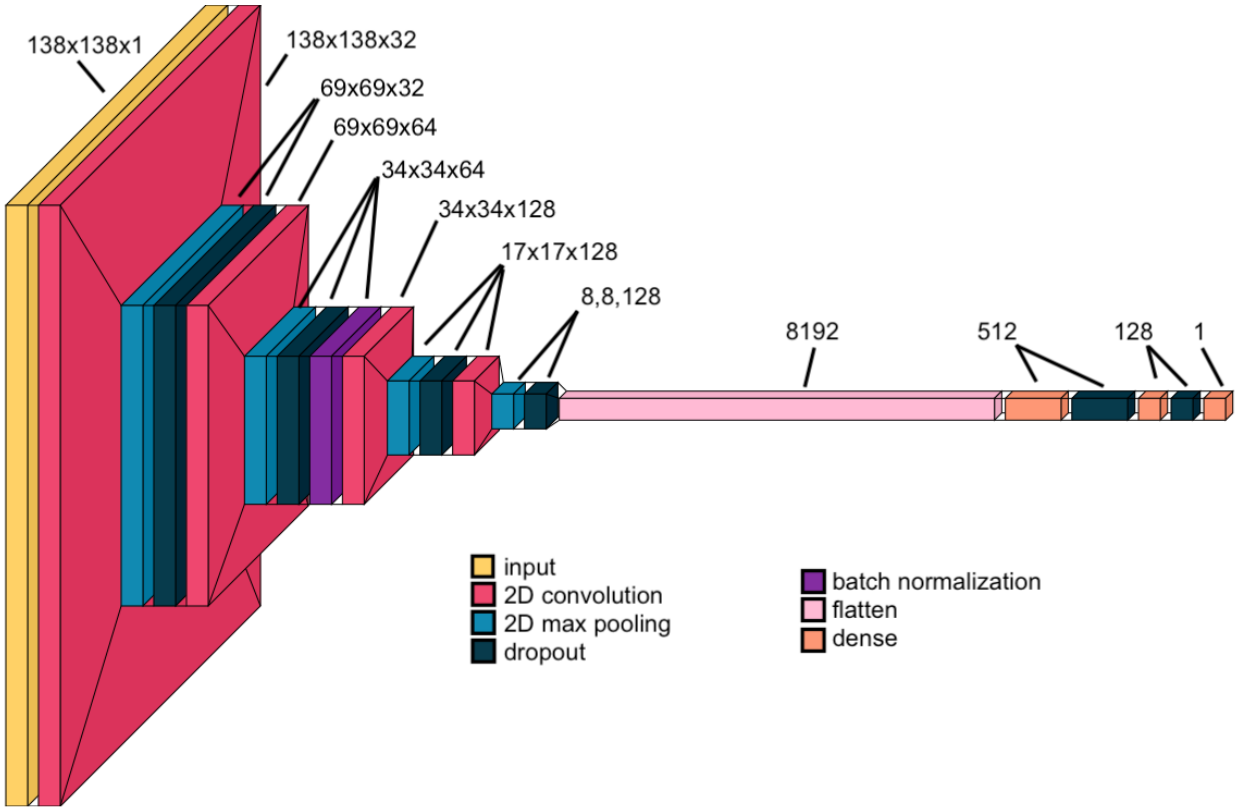

Figure 2.5: A visualization of the custom CNN architecture developed for this work. The visualization was created using VisualKeras (Gavrikov 2020). Each type of layer in the model is assigned a unique colour. The details of the model (dimensions, dropout percentages, parameter numbers) are also shown in Table 2.1.

(Nair & Hinton 2010). The characteristics of the model, including the number of filters applied in each convolution layer, kernel sizes, and dropout percentages associated with each layer, were optimized manually through an incremental iterative search.

A similar search was used for the optimization of training hyperparameters, including the on-the-fly augmentation percentages detailed in Section 2.1.3, an ADADELTA optimizer (Zeiler 2012), and a training batch size of 32 images. The morphological diversity of both the post-merger and control classes introduces significant intrinsic inhomogeneity into the training data. Broadly, the final hyperparameter recipe serves to combat overfitting to any particular subcategory within the data, and encourages the network to generalize.

2.2 Results

Here I present the results of my experiments, first evaluating the CNN's performance by conventional machine learning metrics (Section 2.2.1), and then by studying the stability of that performance as a function of simulation metadata and mock observation quantities (Section 2.2.2). I next produce a mock survey in the style of CFIS using galaxies from IllustrisTNG, and simulate an observational study of star formation enhancement (Section 2.2.3), before comparing my results to other automated methods (Section 2.2.4), and finally to human classifications (Section 2.2.5).

2.2.1 Overall performance of the CNN

In order to evaluate the model's performance by traditional machine learning metrics, I apply the trained model to the reserved test galaxies, including 3728 images each of post-mergers and controls, as detailed in Section 2.1.3. Because this data set does not contain any galaxies that the network has studied in training or referenced in validation, these results roughly correspond to the expected performance of the network should it be deployed to classify unseen data. In other words, one does not have to worry about the memorizing effects of deep neural networks (see also Arpit et al. 2017).

The confusion matrix in Fig. 2.6 shows the performance of the model evaluated on the reserved test galaxies. The model successfully identifies 87 per cent of the post-merger images in the set, and 89 per cent of the controls. Using training and test samples from the EAGLE cosmological simulations (Schaye et al. 2015), Pearson et al. (2019a) report 63 per cent and 67 per cent performance on mergers (including systems that are projected to merge in the next 0.3 Gyr) and non-mergers, respectively, after modulating their model's decision threshold to a position of 0.57. Also using TNG100, Wang et al. (2020) report 76 per cent and 68 per cent for mergers and non-mergers, respectively, with the decision threshold at 0.53. However, the merger defini-

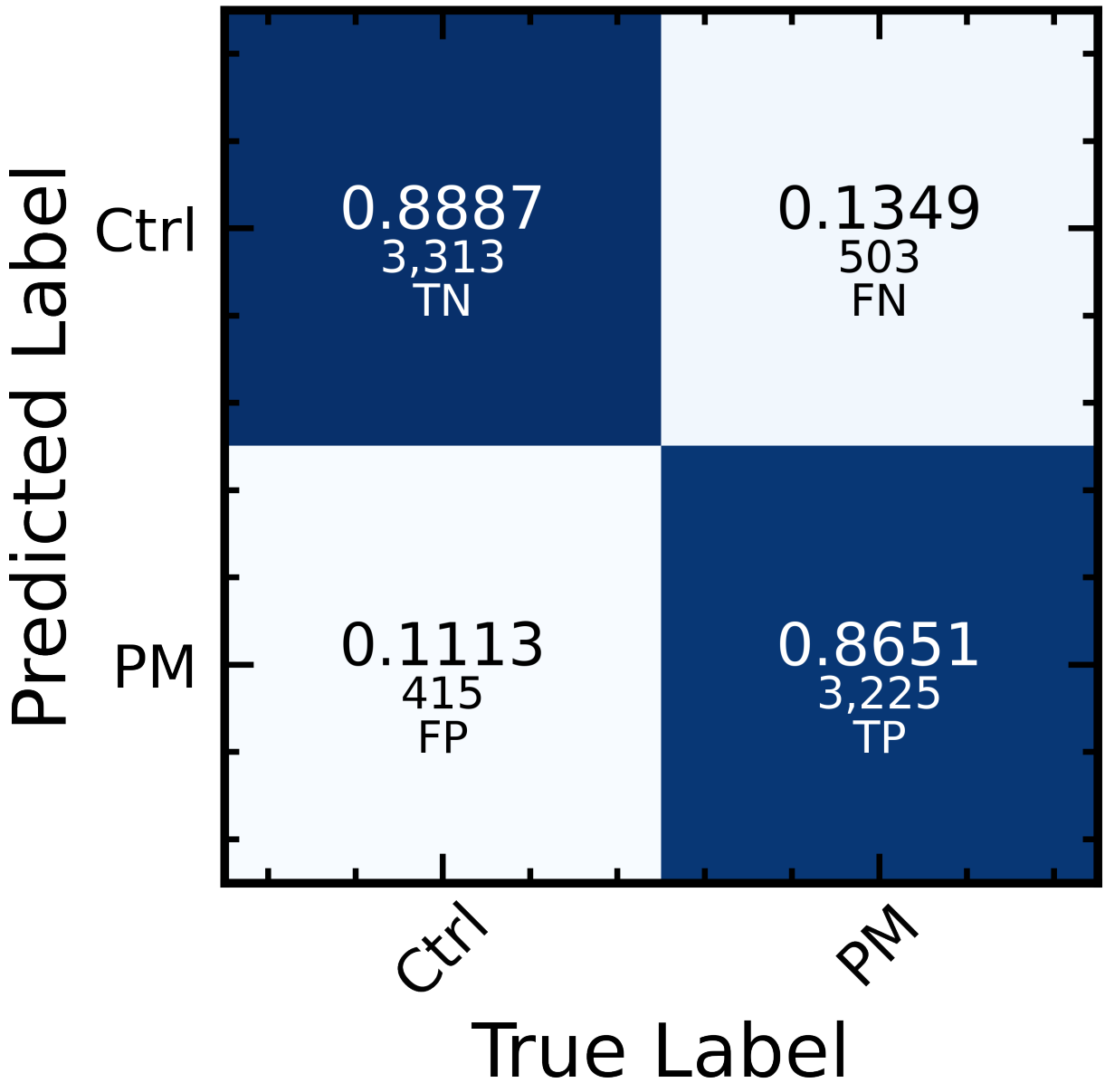

Figure 2.6: Confusion matrix for the trained CNN applied to the never-seen images in the test set. Each quadrant is annotated with the normalized fractional accuracy and the number of galaxies. The total CNN accuracy is 88 per cent.

tion used by Wang et al. (2020) also includes pre-mergers (galaxies that will undergo a merger in the next 1 Gyr), as well as post-mergers that have merged within 500 Myr. My comparison to the literature is therefore not exact. For a more holistic comparison to other machine learning merger classification results, including more recent efforts from the literature, I refer the reader to Table 1.1 in Chapter 1.

The receiver operating characteristic (ROC) curve in Fig. 2.7 shows the false positive rate (i.e., the fraction of controls that are mislabelled as post-mergers) and the true positive rate (i.e., the fraction of post-mergers that are correctly identified) as the model's decision threshold – a decimal prediction of merger status, often referred to in this work as $p(x)$ – is modulated from zero (label everything as a post-merger) to one (label everything as a control). The default threshold for classification is 0.5. The choice of decision threshold is examined later in Section 2.2.3. The dashed diagonal line is characteristic of a hypothetical model with random label assignment by the model. The area under the model's performance curve (AUC) serves as a conventional metric of classification performance, with a value of 1.0 indicating perfect performance; networks that develop a strong grasp of the classification problem at hand are likely to have high AUC scores. Promisingly, the trained model achieves an AUC score of 0.95.

2.2.2 CNN performance trends with galaxy parameters

Having assessed the CNN's overall performance on the test set, I next study its dependence on the galaxies' intrinsic (e.g., stellar mass, gas fraction, and redshift) and environmental properties (e.g., proximity to neighbours). While the reserved test galaxies from Section 2.1.3 are chosen at random, and therefore represent an equitable sampling of the full population, I now reintroduce galaxies previously reserved for training and validation in order to improve the significance of the statistics and drive down the variance of the performance metrics, particularly in regions of a given parameter that are sparsely populated. To this end, I resample the stellar morphologies of the post-mergers and the first round of controls from a new, fifth camera angle, at a vertex in the first octant of a cube with the galaxy at its centre. The resulting stellar mass maps are reprocessed with randomized RealSim-CFIS parameters, assigning new sky locations, observational noise, and mock observation redshift values on an imagewise basis. I also retrain the CNN, folding the previously reserved test galaxies into the training pool, shuffling on an imagewise basis, and partitioning them into test (90 per cent) and validation (10 per cent) data. Allowing the model to study the entire galaxy sample as imaged from the first four camera angles eliminates any image memorization bias when evaluating its performance on the resampled data. The CNN's global performance on the resampled data is the same, 88 per cent accuracy, implying that

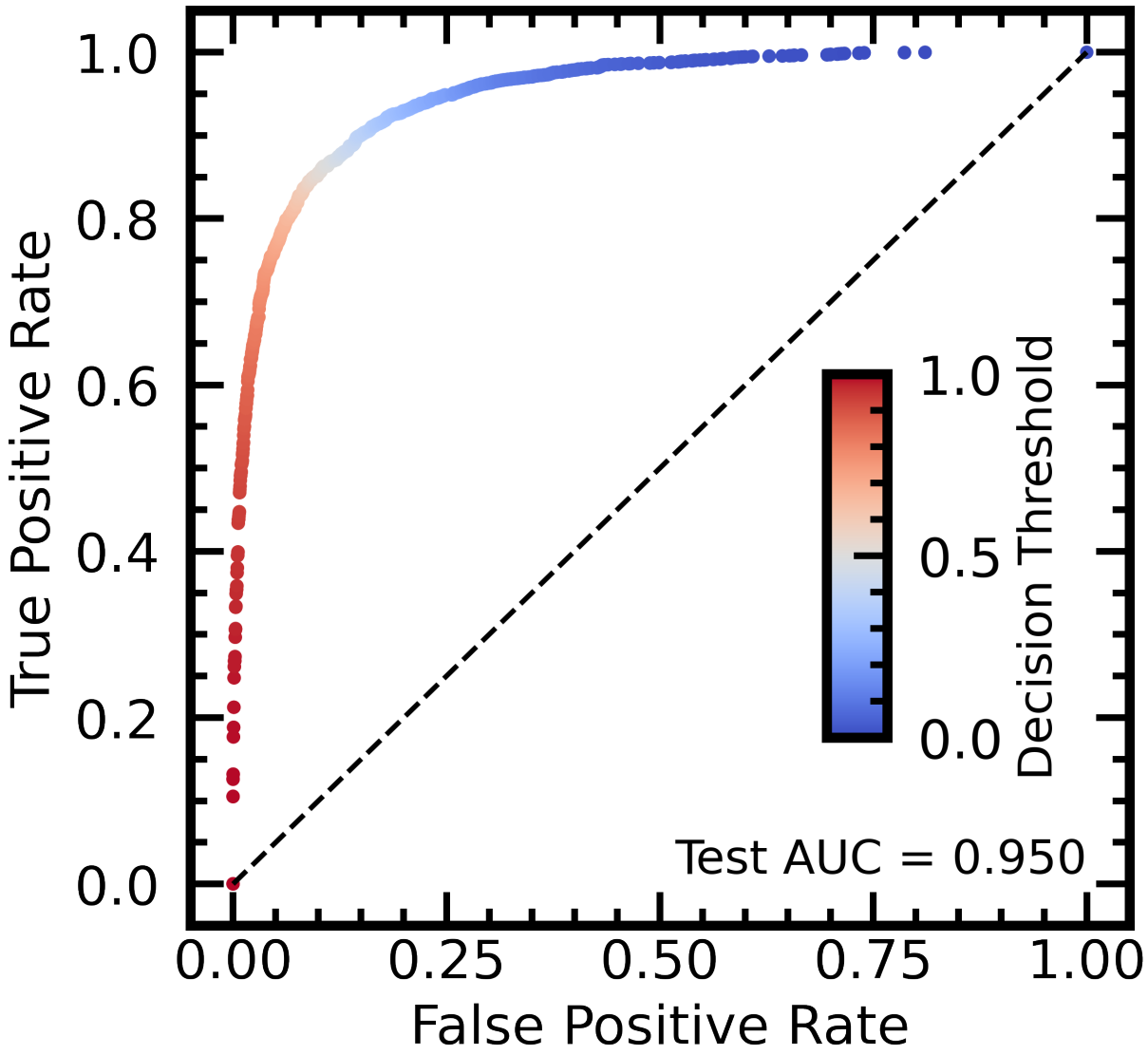

Figure 2.7: The ROC curve with AUC score as the decision threshold is moved from 0 to 1 for the trained model as applied to the test set.

the new projected morphology and mock observation are sufficient to prevent identification via memorization on the galaxies studied in training.

After obtaining model predictions for the resampled data, I bin the constituent galaxies, as well as the model's predictions on them, by a selection of IllustrisTNG metadata quantities. This allows me to study the natural biases imparted on the model by the visual characteristics of the galaxies. Figures 2.8-2.13 show the results of these tests and have the same format throughout. Top panels show the raw number of post-mergers and controls as a function of the property under investigation (e.g., redshift, stellar mass, etc.). These numbers are relevant for performance assessment as one might expect the model to fare more poorly in regimes where it has been exposed to fewer examples in training. The histograms are further divided to show the number of galaxies that are correctly or incorrectly classified, distinguishing between true positives (TP; post-mergers correctly classified as such), false positives (FP; controls that are erroneously classified as post-mergers), true negatives (TN; controls correctly classified as such), and false negatives (FN; post-mergers that are erroneously classified as controls). In the lower panels of Figs 2.8-2.13, I show the fraction of controls and post-mergers that are correctly classified. The horizontal dashed line shows the model's average performance for reference. In all figures, the horizontal error bars show the width of the bins and the vertical error bars are the binomial errors in each bin, $\sqrt{f * (1-f)/N}$, where f is the fraction correctly identified, and N is the number of galaxies in the bin. The pink curve shows the model's average performance on all galaxies in the bin, the blue curve shows the specific performance on post-mergers, and the orange curve shows the specific control (non-post-merger) performance. In the following subsections I describe in more detail the results and conclusions from Figs 2.8-2.13.

Role of environment

I find that a galaxy's simulated environment can strongly influence how it will be classified by my CNN model. Indeed, for networks whose aim is to identify galaxy pairs (e.g., Ackermann et al. 2018; Bottrell et al. 2019b) the presence of a close companion is an essential piece of information. In order to test whether my model's post-merger classifications are affected by the presence of a close companion, I investigate in Fig. 2.8 the fraction of correctly classified galaxies as a function of separation to the nearest neighbour (r_1). Fig. 2.8 shows that when a neighbour galaxy appears within $\sim$50 kpc (the radial extent of the images), the model grows uncertain, assigning post-merger classifications to numerous control galaxies, and yielding below-average total performance. Fig. 2.8 demonstrates that the model retains much of its ability to distinguish between post-mergers and non-post-mergers with a potential pre-coalescence partner down to 10 kpc, be-

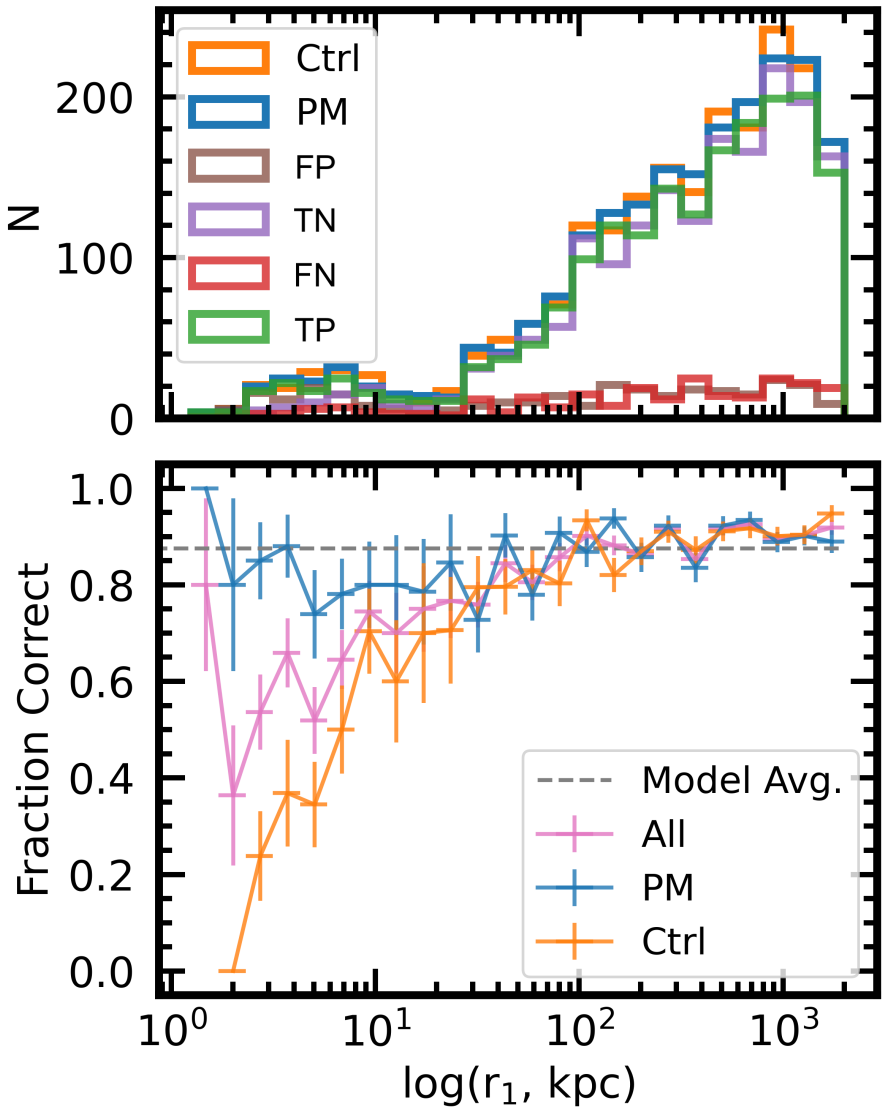

Figure 2.8: The trained model's performance, as a function of separation to the nearest neighbour (r_1). The top panel shows the raw number of post-mergers and controls (blue and orange histograms, respectively), further broken down as correctly and incorrectly classified (FP, brown: controls classified as post-mergers; TN, purple: correctly-classified controls; FN, red: post-mergers classified as controls; TP, green: correctly-classified post-mergers). The bottom panel shows the fraction of post-merger and control galaxy images correctly identified by the model. Horizontal error bars for all points are the bin widths, and vertical error bars are the binomial errors for each bin, $\sqrt{f * (1-f)/N}$ where f is the fraction correctly identified, and N is the number of galaxies in the bin.

low which the visual degeneracy becomes prohibitive. While close pairs are, for my purposes, not necessarily post-mergers, and may be counted as misclassified by the model, in many cases the CNN may still be identifying genuine merger features in pre-coalescence pairs. Still, such close neighbours are rare in both the simulation and the real Universe, and hence should not present a significant source of contamination to the outfalling post-merger sample.

Later in Chapter 3, I indicate that the machine-predicted post-merger samples are sometimes contaminated by galaxies that appear to be in the middle of an interaction. As such, it appears that the CNN's difficulty in distinguishing between post-mergers and galaxies in a close pair persists when the CNN is applied to observational data. The work presented here specifically targets post-mergers, but the mergers at all stages can exhibit the morphological characteristics of an interaction. My model's confusion between pre-merger pairs and post-mergers is therefore unsurprising. However, the distinctive feature of galaxy pairs – that they include two (or in some cases more) bright nuclei – is ultimately a task to which machine vision methods are well suited. For example, Lackner et al. (2014b) used an automated (and machine learning-free) approach of median filtering to identify a highly pure sample of close galaxy pairs with projected separations < 8 kpc; and a similar morphological analysis could be combined with my model to remove close pairs from machine-predicted post-merger samples before further study.

Machine learning also shows promise for separating the two classes; since completing the project detailed in this Chapter, I have contributed to work led by Ferreira et al. (2024) that has significantly improved on this model's ability to separate the two main merger classes. Figure 2.9, adapted from Ferreira et al. (2024) with permission of the author, shows the combined accuracy of a pair of models, one CNN and one SwinTransformer (Liu et al. 2021), for the task of distinguishing between galaxy pairs and post-mergers. The class-wise accuracies for both pairs and post-mergers are above 95 per cent, demonstrating that deep learning machine vision models can be used to separate galaxy pairs and post-mergers when they are trained to do so. The work presented here predates Ferreira et al. (2024), and does not include an automated pair removal step. Since pairs were later manually removed (see details in Chapter 3), the outcomes are philosophically equivalent.

Role of observed redshift

Fig. 2.10 shows the performance of the trained model as a function of insertion redshift. I remind the reader that this is selected at random from the observed SDSS distribution, and is not linked to the simulation redshift. It is perhaps natural to expect that performance may dwindle at higher redshifts, due to loss of spatial resolution and the dimming of faint features. For insertion redshift

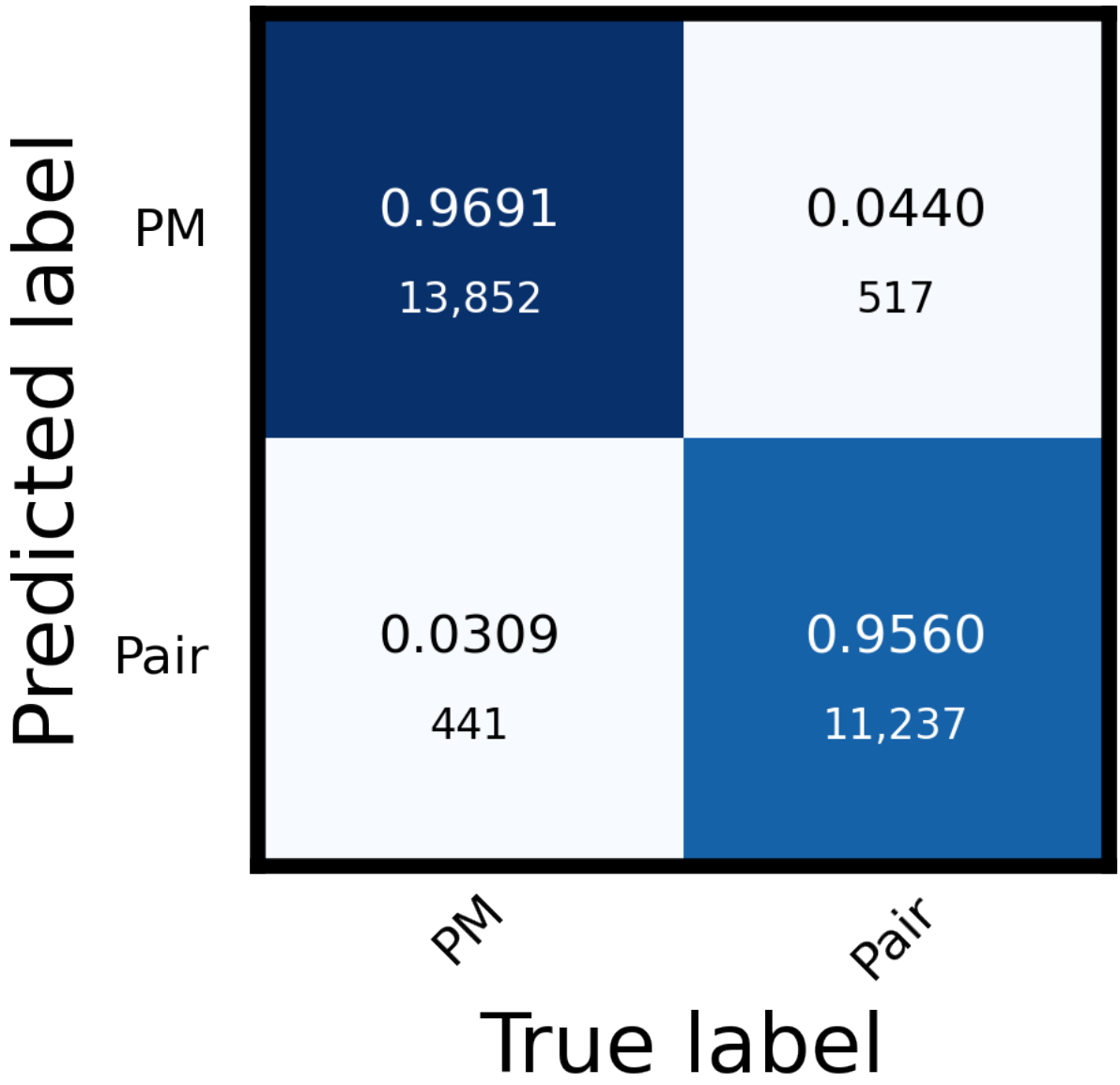

Figure 2.9: A new figure adapted from Ferreira et al. (2024), in which a multi-model merger identification framework is used to separate galaxies in multiple "steps". The first, Step1, distinguishes between mergers and non-mergers. The Step1 performance is stated earlier in Table 1.1. This Figure shows the confusion matrix for Step2, which distinguishes between galaxy pairs and post-mergers using a pair of models, one CNN and one SwinTransformer. The high accuracies on both classes indicate that machine vision deep learning models can distinguish easily between galaxy pairs and post-mergers when they are specifically trained for the task.

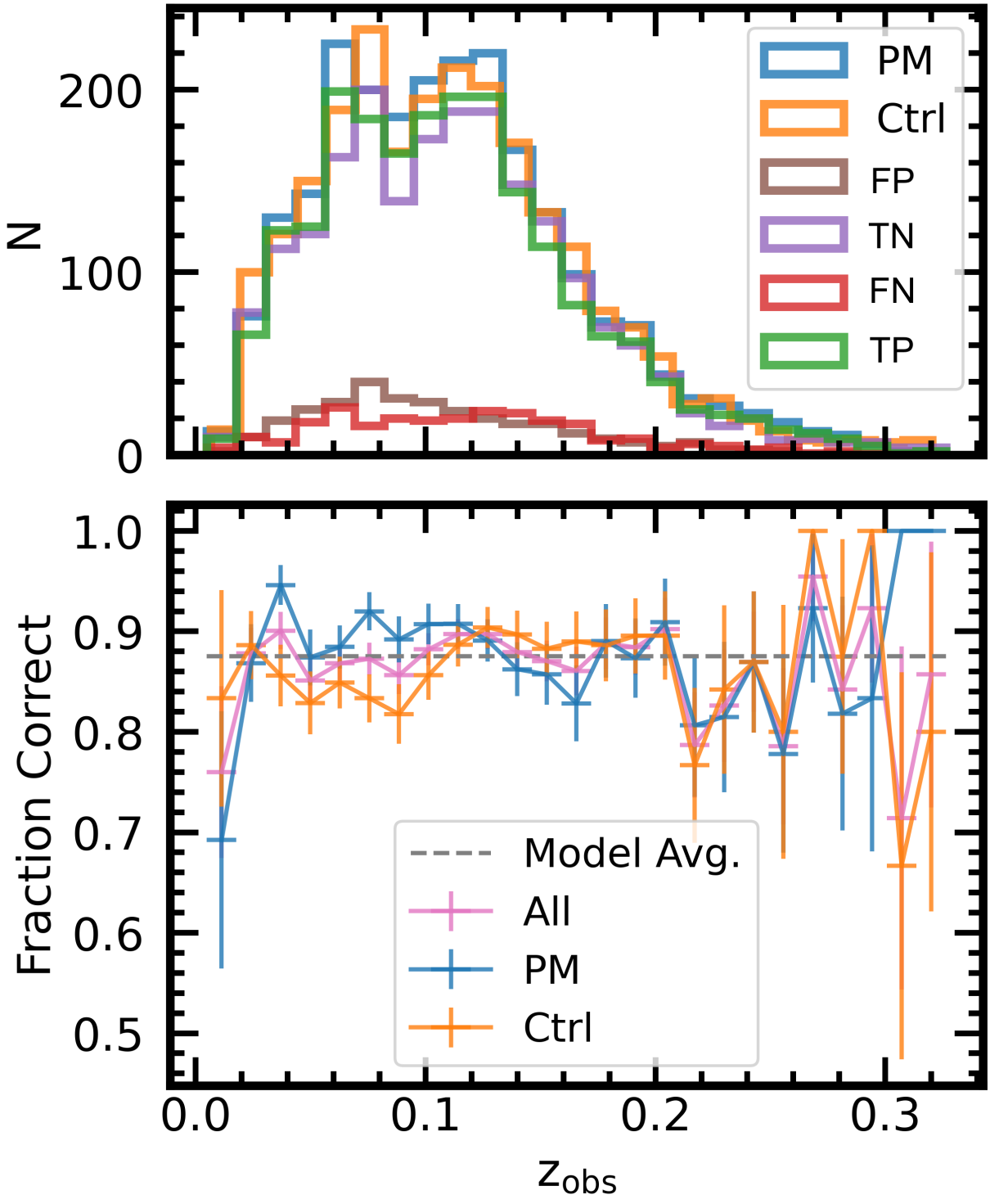

Figure 2.10: As for Figure 2.8 but performance is plotted as a function of mock observed redshift.

values of $\sim$0 – 0.2, where galaxies are better resolved and brighter relative to the sky, performance is consistent with the overall model average, with a slight enhancement in post-merger identification in the lowest mock observation redshift bins. Above $z = 0.2$, degrees of uncertainty are introduced as the resolution diminishes and low surface brightness merger features grow indistinguishable from the background noise. A further contributing factor may also be the relative paucity of training images at $z > 0.2$, as these are rarer in the redshift distribution from which the mock images are generated. I find that artificially enhancing the number of training images with low mock observed redshifts, or indeed, in any sparse tract of parameter space, improves the CNN's performance there. However, with a finite number of training examples, this improved stability comes at the cost of global performance. Using realistic parameter distributions budgets the training material optimally and prepares the CNN to identify galaxies as they are likely to appear in a natural sample.

Role of galaxy mass

Performance with stellar mass (Fig. 2.11) is consistent with the model's average where the bulk of the selected galaxies ($\sim 10^{10.2} - 10^{11.5} M_{\odot}$) is found. Performance drops for the lowest stellar mass bins, and becomes unstable in higher stellar mass bins, where fewer galaxies of both classes exist for the network to study in the training phase. The parent galaxy luminosity function can be modelled by a power law in the low-mass regime, and so a higher number of low-mass galaxies might typically be expected. While they may be numerous, low-mass galaxies in hydrodynamical simulations are also more likely to be impacted by numerical stripping, an effect where stellar mass particles in the outskirts of a low-mass galaxy may erroneously be assigned to a nearby galaxy of greater mass (e.g., Rodriguez-Gomez et al. 2015). The abrupt shelf on the left-hand side of the top-panel histograms in Fig.2.11 can therefore be attributed to my selection criteria, which were developed to minimize the effects of numerical stripping in the sample (see also Hani et al. 2020).

Role of merger mass ratio

Fig. 2.12 shows the model's performance as a function of the ratio of the smaller participant galaxy's stellar mass to that of the larger. The model exhibits minimal volatility in classifying mergers across different mass ratios, with a minor suppression in identifying mergers that are comparatively minor (with a mass ratio of $\leq$0.2), as well as slight instability for mergers near 1:1, of which there are very few examples to study (see blue histogram, top panel of Fig. 2.12). The CNN offers excellent performance for moderate to major mergers (mass ratios $\sim$0.2–0.6), with

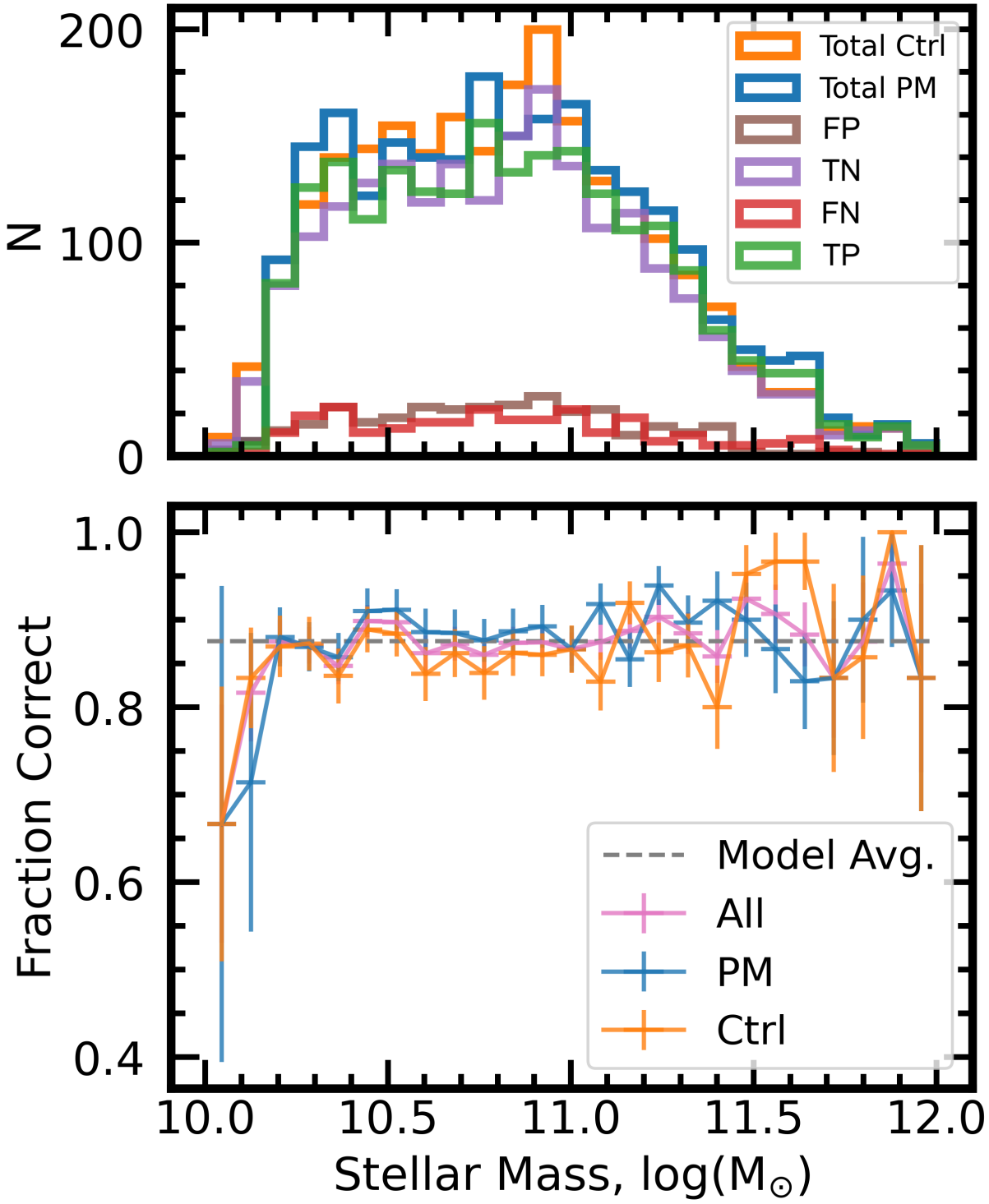

Figure 2.11: As for Figure 2.8 but performance is plotted as a function of galaxy stellar mass.

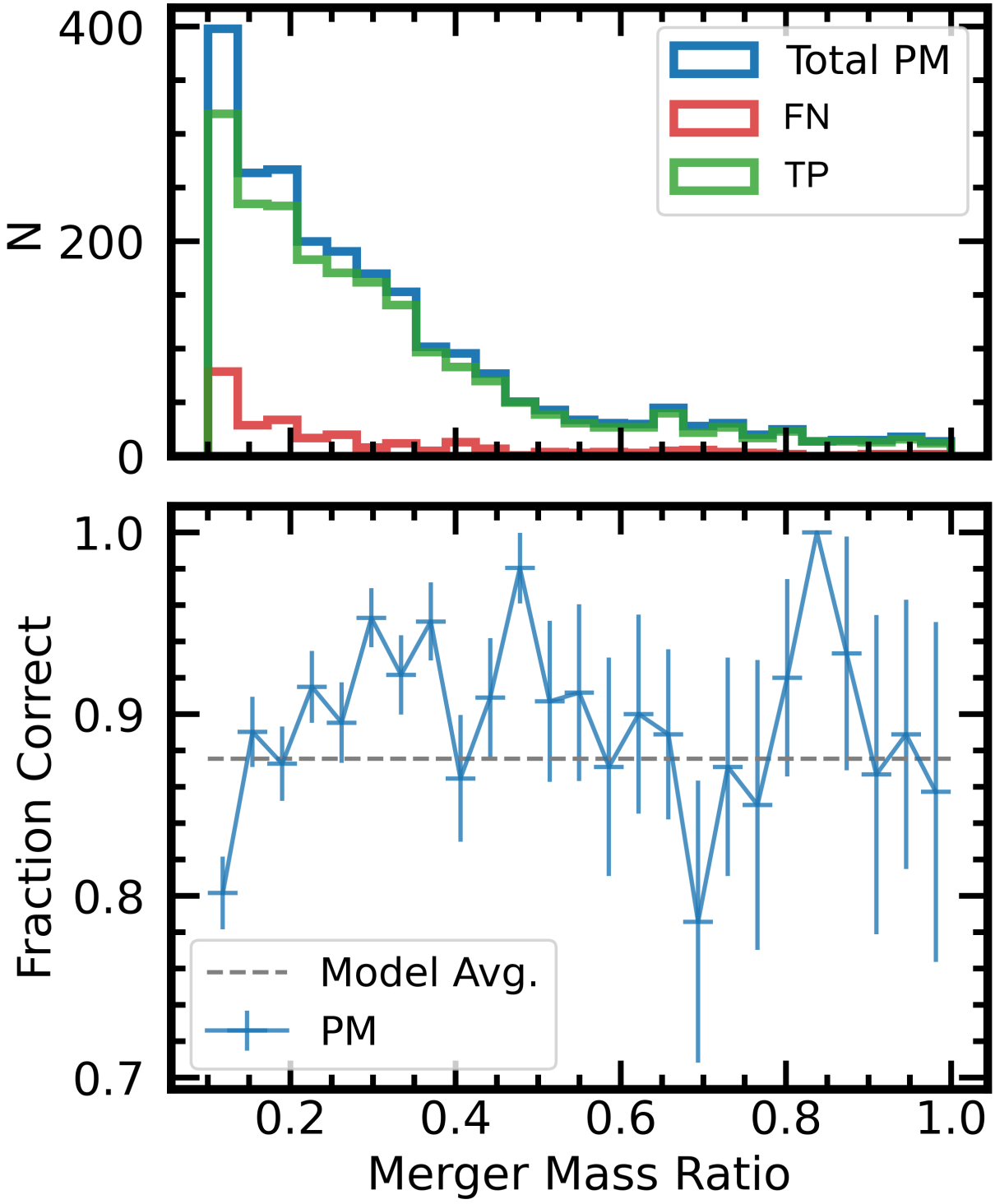

Figure 2.12: As for Figure 2.8 but performance is plotted as a function of merger mass ratio. Post-mergers used in training must have a mass ratio of ≥ 0.1.

minimal scatter in high mass ratio bins with a small number of objects. Because the simulation snapshot timescale is only ∼160 Myr, I can compare to other merger mass ratio studies with the assumption that the galaxies in the post-merger sample have an effective $T_{Postmerger}$ of zero. I do not find a strong link between mass ratio and classification performance as in Nevin et al. (2019), but my finding is consistent with the Lotz et al. (2010a) result that post-merger observability within 0.2–0.4 Gyr is largely insensitive to mass ratio.

Role of gas reservoir

Simulations have previously indicated that gas fraction can affect the observability of the merger phase (e.g., Lotz et al. 2010b; Ji et al. 2014). Fig. 2.13 tests the model's performance as a function of gas fraction. I note that this metric characterizes the gas fraction of the post-merger remnant and not the incoming galaxies, although it should still broadly capture whether (at least one of) the merging galaxies had significant gas reservoirs. Fig. 2.13 shows that in the regime where most of the post-mergers are located (gas fractions less than unity), post-mergers are consistently well classified, even at low gas fractions. Conversely, control galaxies are increasingly poorly classified towards higher gas fractions. I posit that this trend is a secondary effect with stellar mass: nearly all galaxies with gas fractions $M_{Gas}/M_{\star} > 0.3$ belong to the lowest bin of stellar mass as seen in Fig. 2.11. By definition, most galaxies in the sample have typical stellar masses and gas fractions, and so this apparent dive in classification accuracy is insufficient to diminish performance at large.

2.2.3 Mock-CFIS survey

In a matched galaxy sample, like those studied by the CNN in Sections 2.2.1 and 2.2.2, the balance of post-mergers and controls allows the model to enjoy superficially high purity percentages. However, the value of any automated system designed to identify post-merger galaxies lies in its ability to do so within the framework of a large observational sample. One of the most striking distinctions between such a sample and the image data I have used up to this point is the enforced overabundance of post-mergers relative to non-post-mergers. However, mergers are expected to represent only a few per cent of galaxies in the observable (Lacey & Cole 1993; Lotz et al. 2011; Bluck et al. 2012; Casteels et al. 2014), and simulated (Rodriguez-Gomez et al. 2015; Martin et al. 2018) low-z universes. Moreover, the ultimate application of my trained model is the identification of a pure sample of post-mergers in CFIS that will allow me to study the properties of galaxies after their coalescence.

Fig. 2.14 shows the trade-off between purity and completeness (alternatively known as preci-

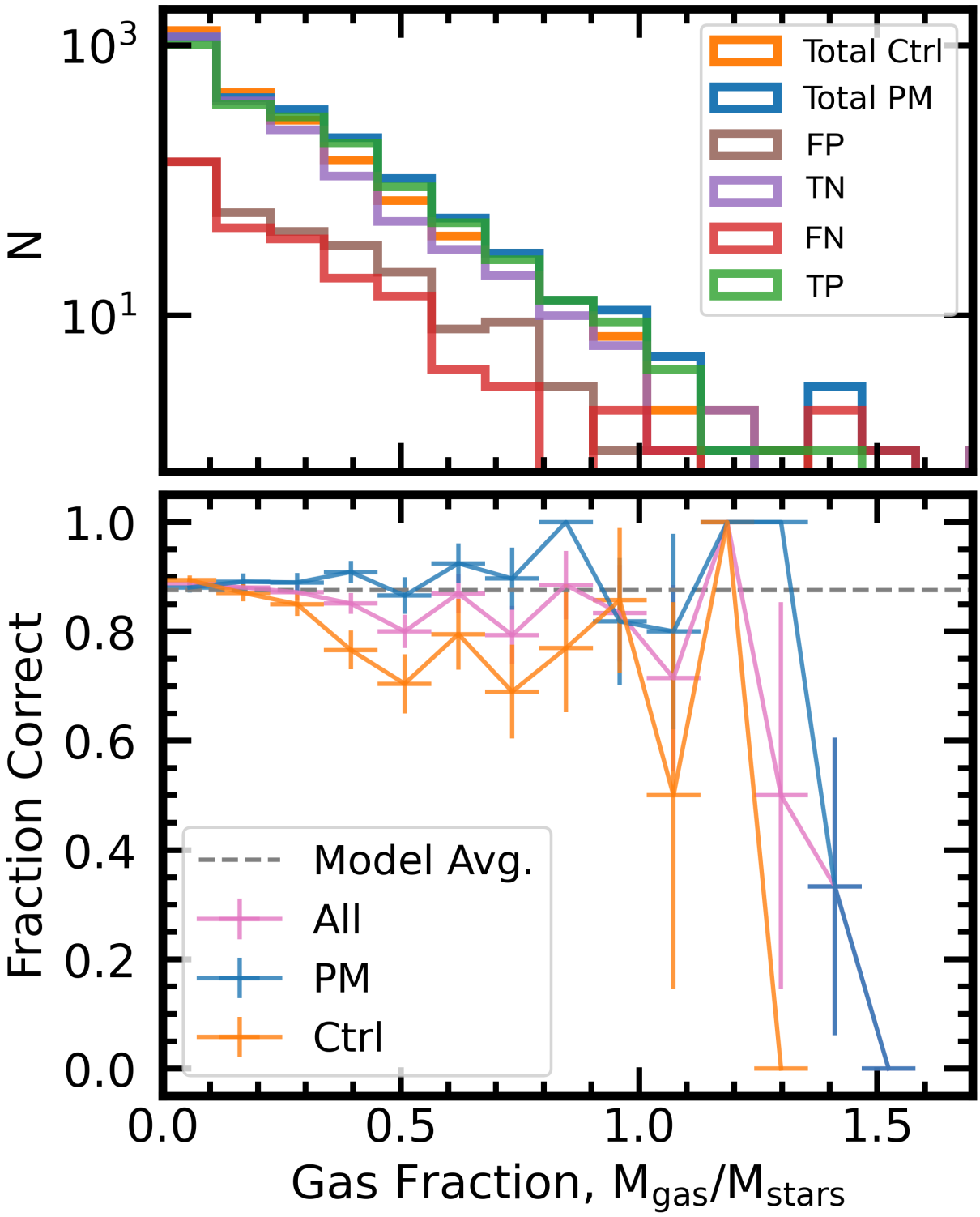

Figure 2.13: As for Figure 2.8 but performance is plotted as a function of the post-coalescence gas fraction. Although the images on which the model is trained contain no gas information, the strong trend is a secondary effect with stellar mass.

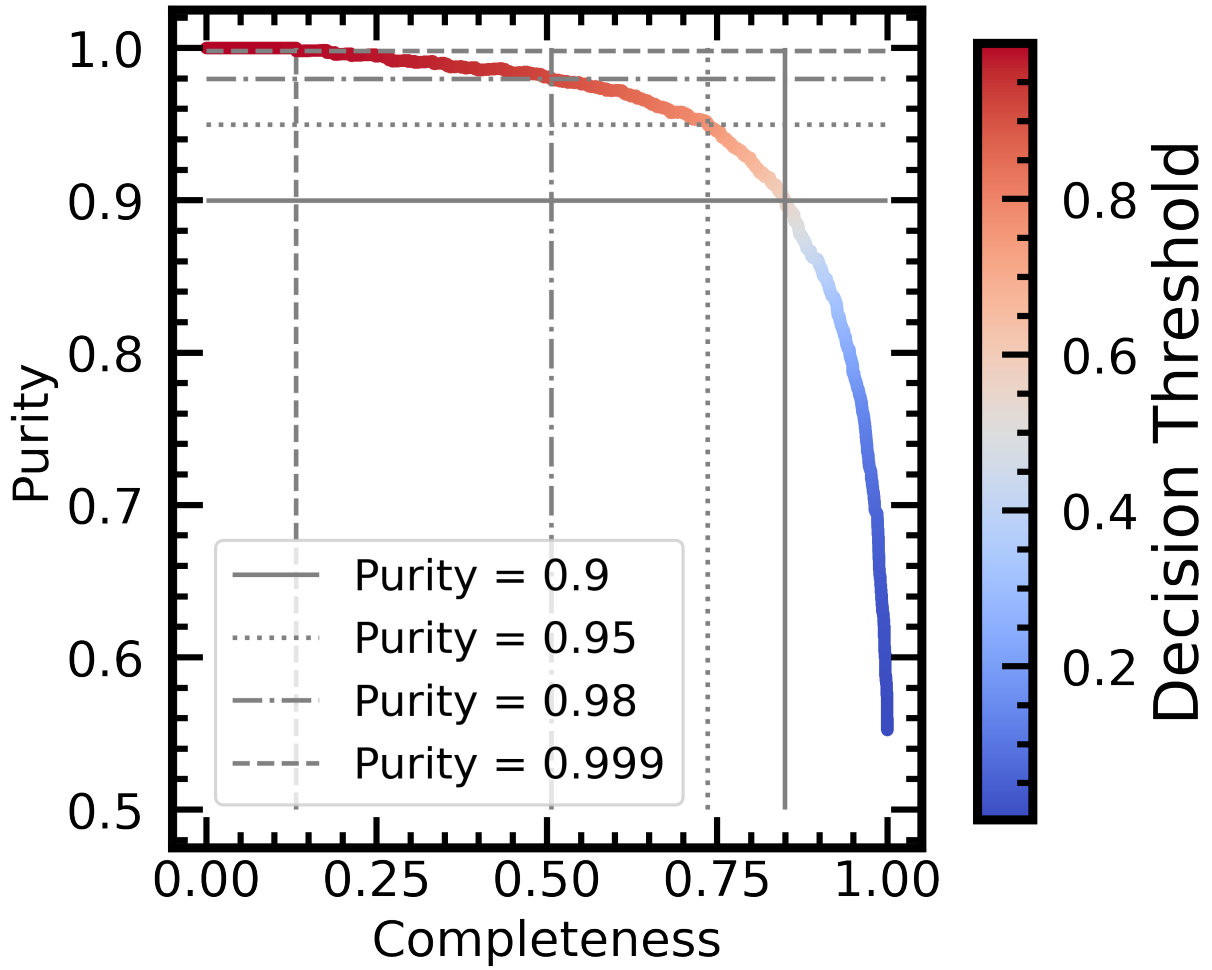

Figure 2.14: The purity-completeness (or precision-recall) curve for the CNN's performance on the test set, which has equal numbers of post-merger and control galaxies. Annotations show the decision threshold setting and completeness cost required to achieve various pre-specified sample purity values.

sion and recall) of the outfalling post-merger sample in the balanced test set (with equal numbers belonging to each class). The default threshold of 0.5 yields a purity of 89 per cent in a balanced data set. In order to explore the impact of the combination of the natural scarcity of mergers with the imperfect identification from any automated (or, indeed, visual) classification method, I prepare and study a mock survey, with post-mergers and non-post-mergers present in quantities representative of TNG100-1.

The Bayesian disadvantage

Provided that the physical models implemented in IllustrisTNG produce post-merger and non-post-merger galaxies that are visually characteristic of low-redshift galaxies in the Universe, the scarcity of post-mergers in the real Universe still poses a challenge to the purity of the outfalling sample, which can be estimated using the form of Bayes rule (Equation 1.1) given in Chapter 1. In this work, I achieve a post-merger identification rate of ∼89 per cent (see Section 2.2.3 below), and so $P(B|A)$, the fraction of correctly identified post-mergers given by $P(A|B)P(B)/P(A)$, is 0.89. $P(B|\neg A)$, the fraction of incorrectly classified non-post-mergers given by $P(\neg A|B)P(B)/(1 - P(A))$, is 11 per cent, or 0.11. The greatest disadvantage lies in the values of $P(A)$ and $P(\neg A)$, the natural occurrence rates of post-mergers and non-post mergers in the Universe (0.007 and 0.993 in TNG100-1; see also Figure 1.6). In total, Bayesian statistics predicts that an object labelled as a post-merger is only ∼5 per cent likely to belong to the post-merger class. Increasing the decision threshold serves to inflate the value of $P(B|A)$, but greatly reduces post-merger completeness. Further (demonstrated later in Section 2.2.3), additional cuts applied during any study of the outfalling sample of post-mergers may unwittingly reduce the sample purity, even after a strict decision threshold has been applied.

The challenge brought on by the rarity of post-mergers does not only apply to automated and machine learning model classifications, however. Later in section 2.2.5, I compare the performance of the machine to that of several people with varying degrees of experience on a small subset of the image sample. Even though people can aim to identify pure merger samples at the expense of completeness, any classifier (human or otherwise) with an accuracy of less than 100 per cent will return results that are significantly contaminated as a result of the minuscule value of $P(A)$.

Survey data preparation

Following the approach of Section 2.2.2, I generate a mock survey containing mergers in a realistic proportion by performing a single synthetic observation of every galaxy in TNG100-1 with a

stellar mass of $M_\star > 10^{10} M_\odot$ from the fifth camera angle, at a vertex in the first octant of a cube with the galaxy at its centre. Since galaxies are randomly oriented with respect to the simulation box, this camera angle is consistent with a random set of orientations, while also projecting different apparent morphology than would have appeared in the original four camera angles used in training. Since this complete data set does not contain matched pairs of post-mergers and controls, individual mock observation redshift values are selected at random from the CFIS DR2 and SDSS DR7 distribution for each object on the fly. In total, the resulting mock survey contains one image each for the 303110 galaxies. Only a small fraction (2332 or 0.7 per cent) of the images are of post-merger galaxies in their first post-coalescence snapshot, while the rest, all galaxies with $T_{Postmerger} > 0$, are counted as non-post-mergers for the purposes of this experiment. The resulting post-merger fraction is consistent with empirical studies that have found merger rates at low-z are only a few per cent (e.g., Patton et al. 2000; Ventou et al. 2017; Conselice et al. 2022). The non-post-merger category in the mock survey is therefore distinct from the control group used earlier.

Evaluating on a set of reserved test galaxies in Section 2.2.1, as well as on a larger resampled galaxy population in Section 2.2.2, established that the chosen combination of the training data, network architecture, and training hyperparameters were sufficient to grasp the nuances of the classification problem without depending on memorization of specific images. For the mock survey, therefore, I will use the same model as in Section 2.2.2, which has studied the entire matched galaxy sample from the first four camera angles, and present its performance on the 303110-image data set just detailed.

Mock survey classification

Fig. 2.15 shows the mock survey confusion matrix for the model trained on all of the images in the original matched data set, with no reserved test galaxies. Pure morphological memorization is unlikely to work reliably, due to the use of a new fifth camera angle in generating the mock survey. Still, the model successfully identifies ∼89 per cent of both the mergers in their first post-coalescence snapshot, and the non-post-mergers, a category that now includes any object of mass $> 10^{10} M_\odot$ not explicitly belonging to the post-merger category. Despite strong fractional completeness of both classes and an AUC score of over 95 per cent, non-post-mergers in the simulation outnumber post-mergers by more than 100 to 1, and so the number of false positives (non-post-mergers identified as post-mergers by the model) outnumber the true positives (correctly labelled post-mergers) almost 14 to 1, resulting in an outfalling post-merger purity of only ∼6 per cent (a quantity in good agreement with that predicted by Bayes rule). Such a low pu-

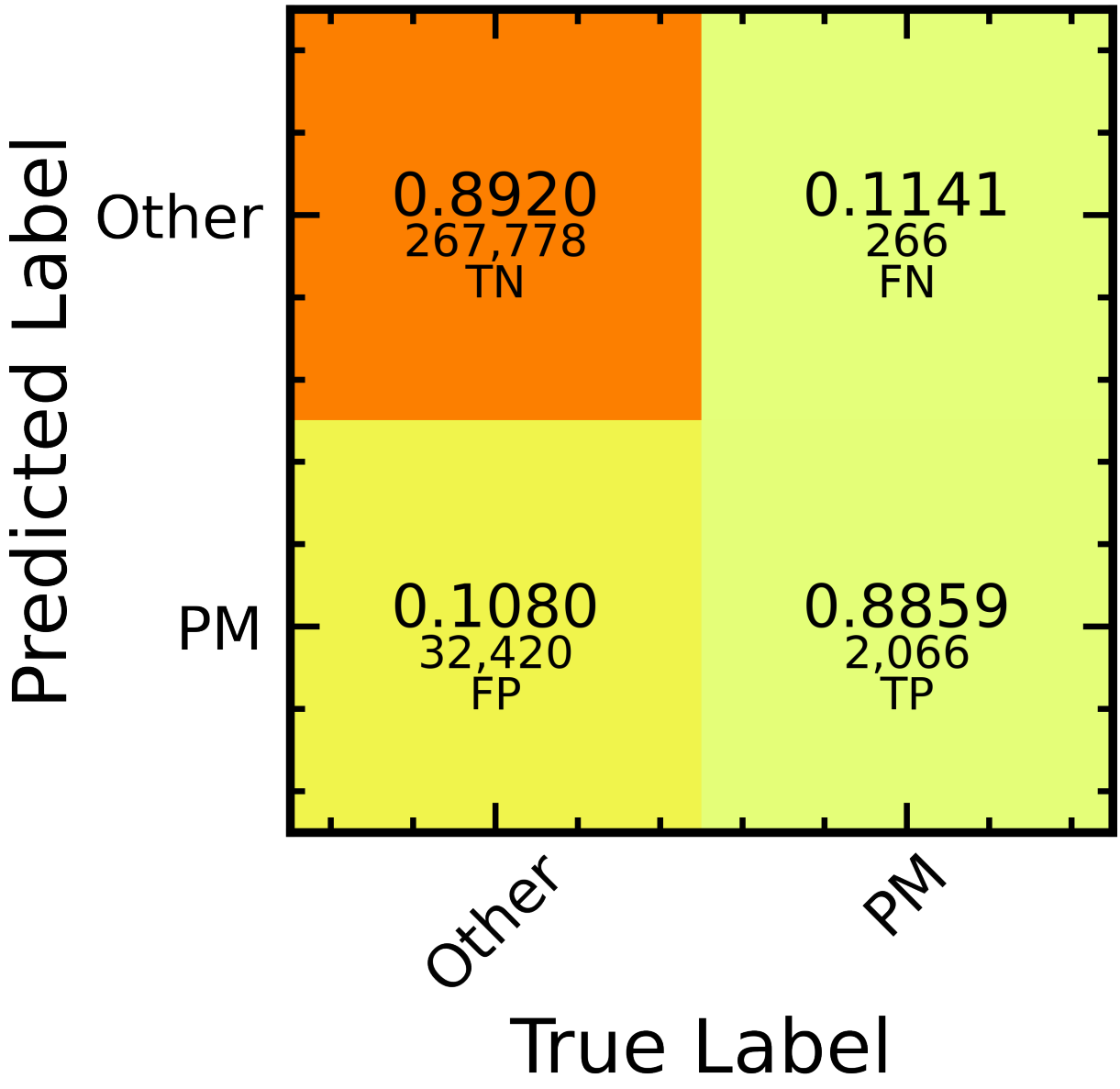

Figure 2.15: Confusion matrix for the fully-trained model's performance on the mock survey images. Each quadrant is annotated with the normalized fractional accuracy and the number of galaxies. Although the overall sample completeness is 89 per cent for post-mergers and controls, the scarcity of post-mergers leads to a relatively small sample.

Threshold	Purity	Completeness	PM Sample Size
0.5	0.06	0.89	34486
0.981	0.5	0.27	2535
0.992	0.6	0.16	1531
0.999	0.8	0.05	439

Table 2.2: Decision thresholds required to yield example purity levels in the mock survey, which contains 303110 galaxies and 2332 actual post-mergers.

rity is obviously problematic if the eventual science goal is to assess the statistical properties of the CNN-identified post-merger sample (e.g., Pearson et al. 2019b). In spite of the disadvantage, however, the model distills the mock survey efficiently, returning a predicted post-merger set that contains 1/10 as many images as the mock survey. Only 0.7 per cent of the mock survey galaxies are post-mergers with $T_{Postmerger}$ = 0 Gyr, but the CNN's predicted post-merger sample is $\sim$9 times as distilled (i.e., the true post-merger fraction in the sample is greater by a factor of $\sim$9, up to 6 per cent). The predicted post-merger sample is also $\sim$10 times as distilled (10 per cent from 1 per cent) in post-mergers with $T_{Postmerger}$ values of $\leq$0.2 Gyr. The $T_{Postmerger}$ statistics of the predicted post-merger sample suggest that the CNN continues to identify legitimate post-merger features for hundreds of Myr after the merger occurs (see also Section 2.2.3). As in Section 2.2.2, I find that a number of non-post-merger galaxies with a nearby neighbour still contaminate the post-merger sample. However, many of these misclassifications are also due to the presence of merger-related features. 46 per cent of false positive galaxies in the mock survey go on to experience a merger within 500 Myr, while only 33 per cent do not experience a merger in the next 2 Gyr. Therefore, while there is a meaningful amount of genuine contamination, the network shows a clear preference towards pre-mergers with merger-induced features compared to galaxies that experience a flyby.

Utility of decision threshold

Fig. 2.16 examines the role of the decision threshold (the "probability" above which an image is classified as a post-merger) on the purity and completeness of the outfalling post-merger sample. I also present selected purity-threshold combinations in Table 2.2. The sigmoid activation function in the CNN's final layer assigns each image a value between 0 and 1, roughly representing the model's certainty in its classification: values very close to 1 correspond to high post-merger certainty, and values near 0 correspond to high non-post-merger certainty. The decision threshold is the cut used to separate post-mergers and non-post-mergers based on the "probabilities"

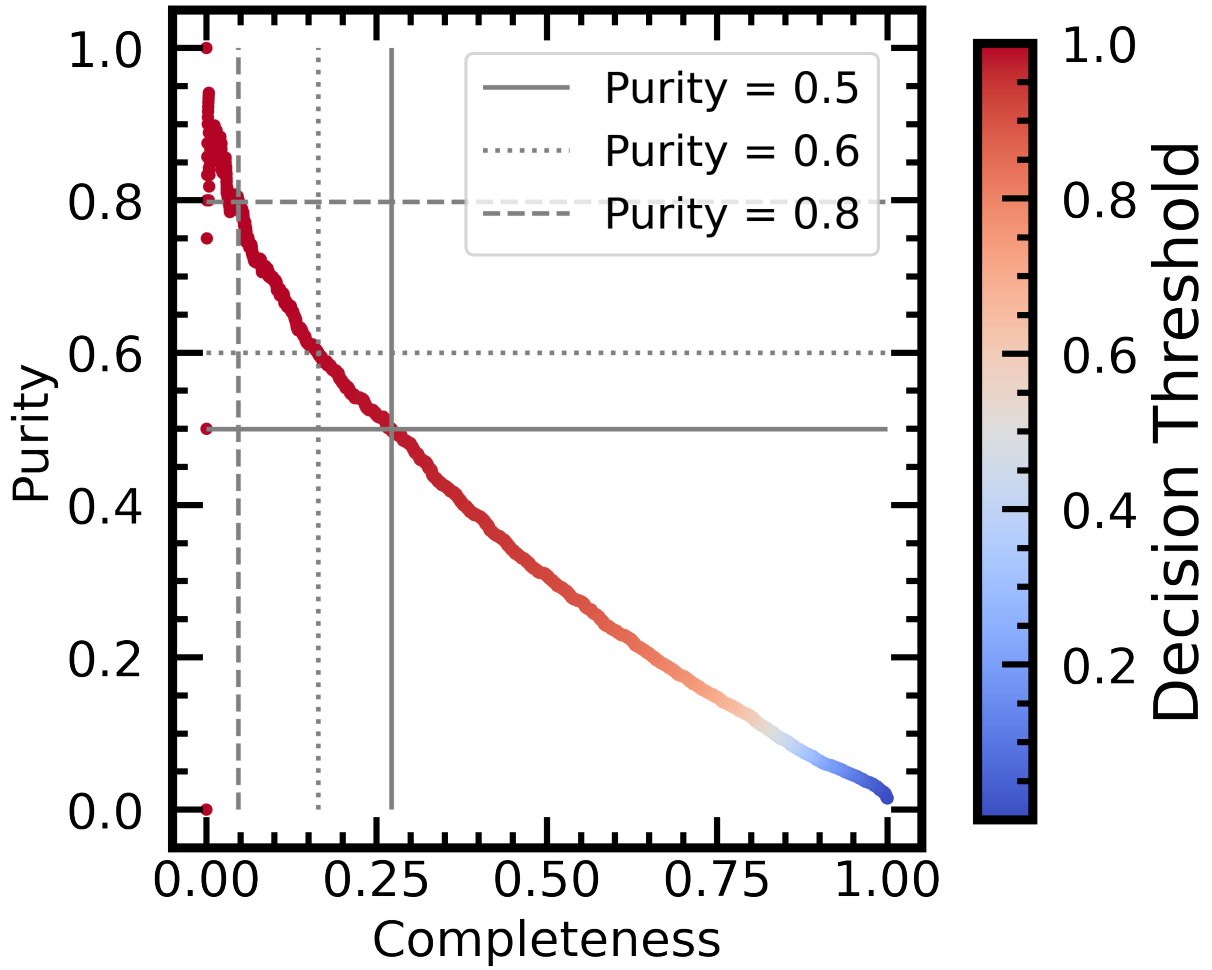

Figure 2.16: The purity-completeness (or precision-recall) curve for the mock survey. For the default decision threshold at 0.5, the purity is very low, approximately 6 per cent. Increasing the threshold as specified in the annotation allows for higher sample purity to be achieved (see also Table 2.2).

assigned by the model.

Calibration curves and the expected calibration error (ECE) metric are often used to assess the practical utility of a network's decision boundary as a true metric of probability, and to evaluate whether recalibration is necessary. Using the model's predictions on the reserved test data from Section 2.2.1, and following the method given in Guo et al. (2017), I construct a calibration curve with 15 quantile bins (i.e., an equal number of samples in each bin of the model's predictions), and calculate an ECE of ∼2 per cent. The ECE for the model is consistent with calibration characteristics for a number of post-correction models detailed in Guo et al. (2017), and so I am able to proceed in using the decision boundary as a metric of post-merger certainty in good faith.

Until now, my analysis has used a default threshold of 0.5. Using a threshold of ∼0.98, I find that reasonably large post-merger samples containing hundreds of galaxies can still be recovered with intermediate purity of 50–60 per cent, and a purity of ∼80 per cent can be achieved when one solely considers galaxies that have been assigned post-merger labels with near-absolute certainty (decision threshold of 0.999). Because there are precious few post-mergers to identify in the survey, however, any significant movement of the decision threshold sacrifices the bulk of the true positive galaxies and may introduce biases into subsequent scientific consideration of the outfalling sample. To alleviate this problem, I posit that a combination of automated classifications and manual inspection (described later in Section 2.2.5 and used to identify real post-mergers in Chapter 3) can be used to identify a defensible and pure sample of post-mergers without applying such an aggressive decision threshold.

Role of time-since-merger

In my analysis so far, I have imposed a very strict definition of "post-merger" on the sample selection, requiring that the merger occurred in the time between the present and the previous simulation snapshot. Additionally, the non-post-merger class in the mock survey is no longer an equal-sized control group as in Section 2.2.1, but rather any simulation object with a stellar mass greater than $10^{10}M_{\odot}$ not explicitly counted as a post-merger. Fig. 2.17 examines the sensitivity of classification accuracy for the non-post-merger galaxies to the times since their most recent mergers. The sensitivity to time-since-merger for the galaxies is strong, particularly for galaxies that last merged within 2 Gyr. Galaxies that belong to the non-post-merger class by my definition and have undergone a merger in the last 0.1 Gyr are ∼60 per cent likely to receive a post-merger label using the default decision threshold (0.5) – i.e., many galaxies are classified as post-mergers due to a real merger that has occurred in the recent past. The CNN is therefore identifying bona fide merger features that are persisting in time. As $T_{Postmerger}$ increases in non-post-merger

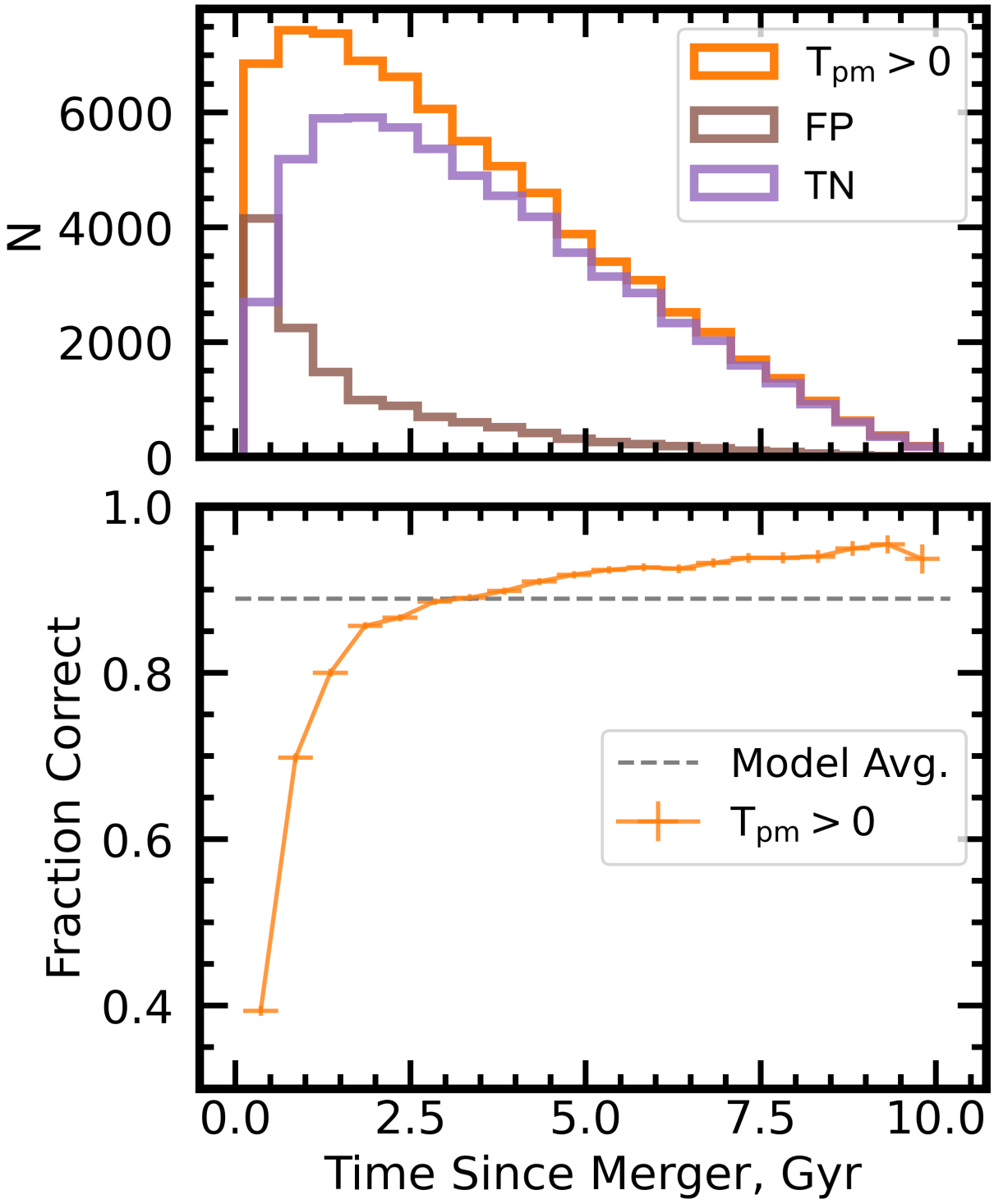

Figure 2.17: Top panel: The number of non-post-merger objects binned by time since merger (orange histogram), and accompanying histograms for the two possible confusion matrix categories (FP, brown: non-post-mergers classified as post-mergers; TN, purple: correctly classified non-post-mergers) into which they could fall, using the default model decision threshold (0.5). Bottom panel: The fraction of correctly-classified non-post-merger objects binned in the same way.

galaxies, the chance of correct classification increases as well. Still, non-post-merger galaxies that have not experienced a merger in the last 2 Gyr stand a chance of being classified as post-mergers, and so the problem of legitimate sample impurity persists, albeit to a lesser extent.

Mock survey star formation enhancement study

In order to assess the impact of sample purity and completeness on my ability to accurately quantify the physical changes incurred by a merger, I investigate the star formation rates (SFRs) of post-mergers in IllustrisTNG. The objective is to compare the true enhancement in SFRs exhibited by the full post-merger sample in TNG (e.g., as previously quantified by Hani et al. 2020) with the recovered SFR enhancement exhibited by the machine-predicted outfalling post-merger sample identified by the CNN.

To compute a given galaxy's SFR enhancement in TNG100-1, I pre-select star-forming galaxies. To this end, I fit a line to the star formation main sequence (SFMS), or the correlation of star formation with stellar mass (e.g., Speagle et al. 2014) in each simulation snapshot, and measure the perpendicular scatter of galaxies about the line in order to apply a snapshotwise cut 2σ below it. The cut in star formation rate removes quiescent galaxies from both the potential post-merger and control pools, in order to facilitate analysis of star formation enhancement within a population of already star-forming galaxies, setting aside the question of merger-induced galaxy revivification. Subsequently, I again follow Hani et al. (2020) and identify post-merger galaxies within the star-forming population as in Section 2.1.1, and control eligible non-post-mergers as having a $T_{postmerger}$ of ≥ 2 Gyr. The resulting post-merger sample consists of 971 galaxies (i.e., 1361 post-mergers are removed from the sample due to inadequate star formation), while the control pool includes 126577. In order to quantify the effects of the merger, I search for control galaxies using two different methods:

(i) matching on stellar mass and simulation lookback time only, and

(ii) matching on r_1 and N_2 in addition to stellar mass and simulation lookback time.

Method (i) mimics an observationally driven approach, where accurate statistics about a galaxy's environment may be challenging to measure due to spectroscopic incompleteness (e.g., Patton & Atfield 2008), while method (ii) seeks to carefully account for nearby neighbours and extended environment in order to study star formation enhancement with as few biases as possible. The default control tolerance for stellar mass is 0.1 dex, and the default control tolerance for r_1 and N_2 are both 10 per cent. Controls must be drawn from the same simulation snapshot, and

so there is effectively no lookback time tolerance[6]. If the default tolerances yield a control pool of more than five galaxies, their median SFR is subtracted from that of the post-merger in question in order to calculate an individual ΔSFR. If there are five or fewer eligible controls, the tolerances are softened as they are in Section 2.1.1. Tolerances are allowed to grow a maximum of four times, in order to maintain reasonable statistical similarity between post-mergers and controls.

Fig. 2.18 shows that my approach yields a "true" value ΔSFR of 0.23 dex for method (i), which included all post-mergers, and 0.21 dex for method (ii), which was able to find suitable controls for 857 of the post-mergers. The green (top panel) and turquoise (bottom panel) histograms in Fig. 2.18 correspond to methods (i) and (ii), respectively.

I then repeat the calculation of ΔSFR, but now for post-mergers that have been identified by the CNN, rather than those selected based on the $T_{Postmerger}$ flag from the simulation itself. I start by considering an extreme case, and use an aggressive decision threshold in order to identify a post-merger sample that is highly pure at the cost of completeness. In order to impress purity upon the post-merger sample, I only consider galaxies that have been labelled as $\geq$0.999 by the model to be eligible post-mergers. Similarly, I only allow galaxies labelled as $\leq$0.001 to be eligible controls. Combining the mock survey labels and the SFMS criterion detailed above results in a star-forming CNN-predicted post-merger sample of 140 galaxies, and a control-eligible pool of 120669. As before, method (i) found suitable controls for the full CNN-predicted post-merger sample, while method (ii) successfully found controls for all but 14 galaxies. Fig. 2.18 shows that although the sample of 140 CNN-predicted post-mergers is in fact only $\sim$49 per cent pure (reduced from the original 80 per cent after removing galaxies with inadequate star formation), both control-matching methods result in median ΔSFR quantities that closely track the simulation's ground truth values: 0.19 dex for both control methods (violet and magenta histograms in Fig. 2.18). Thus, the CNN-based approach recovers the underlying signal, and impurity in the post-merger sample and the low-number statistics associated with the extreme-threshold CNN approach only give rise to a small discrepancy of 0.02–0.04 dex. As for the simulation ground truth, the inclusion of environmental statistics in identifying controls for post-merger galaxies does not appear to have a meaningful effect for the CNN-identified samples. Therefore, even though the CNN is susceptible to the effects of unusually dense environments (see Fig. 2.8), neglecting to control for environment does not strongly impact the results due to the relative rarity of IllustrisTNG galaxies in such environments (see Fig. 2.1).

[6]The same-snapshot requirement for controls used in this section represents a small deviation from Section 2.1.1 in order to avoid applying incongruous SFMS criteria to a given post-merger and its controls.

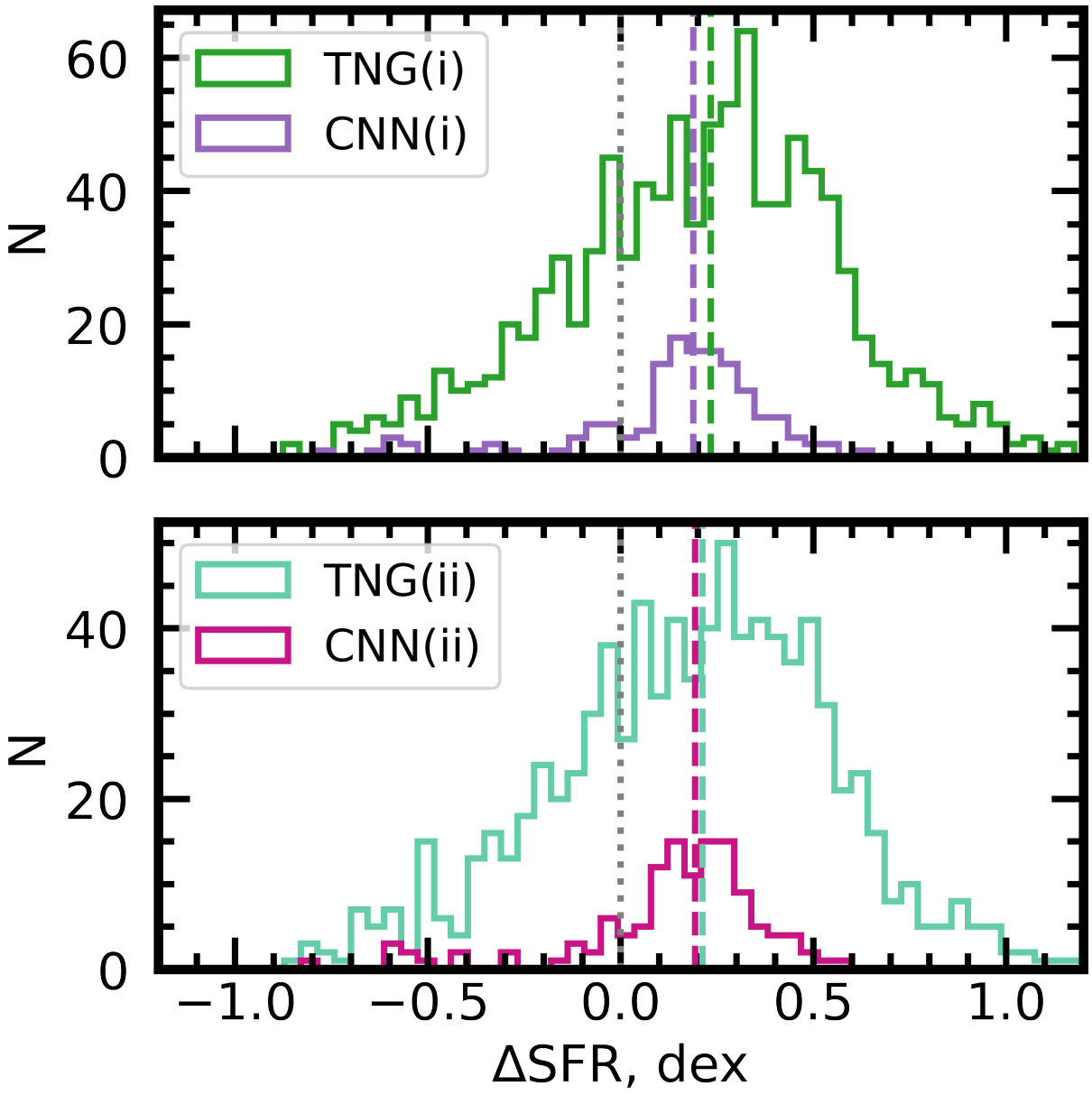

Figure 2.18: Star formation enhancement histograms for control methods (i) and (ii), top panel and bottom panel, respectively. Post-mergers identified using Illustris-TNG metadata with controls matched in stellar mass and redshift only shown in green. Post-mergers for the turquoise curves are identified in the same way, but controls are found using Illustris-TNG environmental parameters (r_1 and N_2) as well. The violet and magenta curves are control-matched in the same way, but are identified by the CNN using extreme decision thresholds for both classes (0.001 for controls, and 0.999 for post-mergers).

Star formation enhancement and purity

While the high-$p(x)$ post-merger sample identified using the CNN for Figure 2.18 was highly pure, it is not practical to sacrifice so many genuine post-mergers in the interest of purity. Next, I investigate the effects of relaxing $p(x)$ to include more galaxies. Fig. 2.19 investigates the role of the decision threshold on the predicted median post-merger ΔSFR in the mock survey. The green line in the top panel of Fig. 2.19 shows how the median ΔSFR of the CNN-identified sample approaches the simulation ground truth (grey dashed line) as the decision threshold is changed. With the binary classification decision threshold in place at 0.5, the result of my mock star formation enhancement study is consistent with little to no ΔSFR (red dashed line) due to high rates of contamination. Fig. 2.19 therefore highlights that the use of the default decision boundary at $p(x) = 0.5$ in the mock survey would miss the statistical SFR enhancement due to high impurity. The second panel shows the Kolmogorov-Smirnov (K-S) statistic (Smirnov 1948) as a function of decision threshold, roughly representing the probability that the CNN-predicted ΔSFR distribution was drawn at random from the simulation's ground truth distribution. The probability remains fairly consistent until an extreme decision threshold (i.e., the same as in Fig. 2.18) is applied, at which point it increases to ∼0.3. Therefore, according to the K–S test, the CNN-identified ΔSFR distribution is most likely to belong to the underlying IllustrisTNG distribution at the highest CNN decision threshold. The final two panels show the purity and completeness of the predicted post-merger sample as a function of threshold, with strictly defined post-mergers ($T_{Postmerger}$=0) represented by the purple line, and galaxies that have merged within 0.2 Gyr shown in pink.

Without leveraging some metric of "probability", false positive galaxies, which are more numerous and less likely to have enhanced star formation than their true positive counterparts, are certain to dominate the statistics. I consider this default-threshold finding to be analogous to the findings of Pearson et al. (2019b), who identify a typical merger-induced ΔSFR consistent with zero. Although the training labels used by Pearson et al. (2019b) are constructed using human-generated responses in Galaxy Zoo, and therefore may benefit from degrees of physical understanding that my CNN may not, I note that even the most minuscule false positive rate will give rise to a highly impure sample from Bayesian statistics, and the non-merger prior in particular (see Section 2.2.3 and Figure 1.6). The lack of an enhancement in SFR in Pearson et al. (2019b) may therefore be a consequence of high impurity in the identified post-merger sample. As the model's post-merger and control thresholds are made more extreme, however, I succeed in enforcing relatively high purity in the sample. The trade-off with an extreme decision threshold is that I also rule out most post-merger galaxies and may introduce accidental biases into the

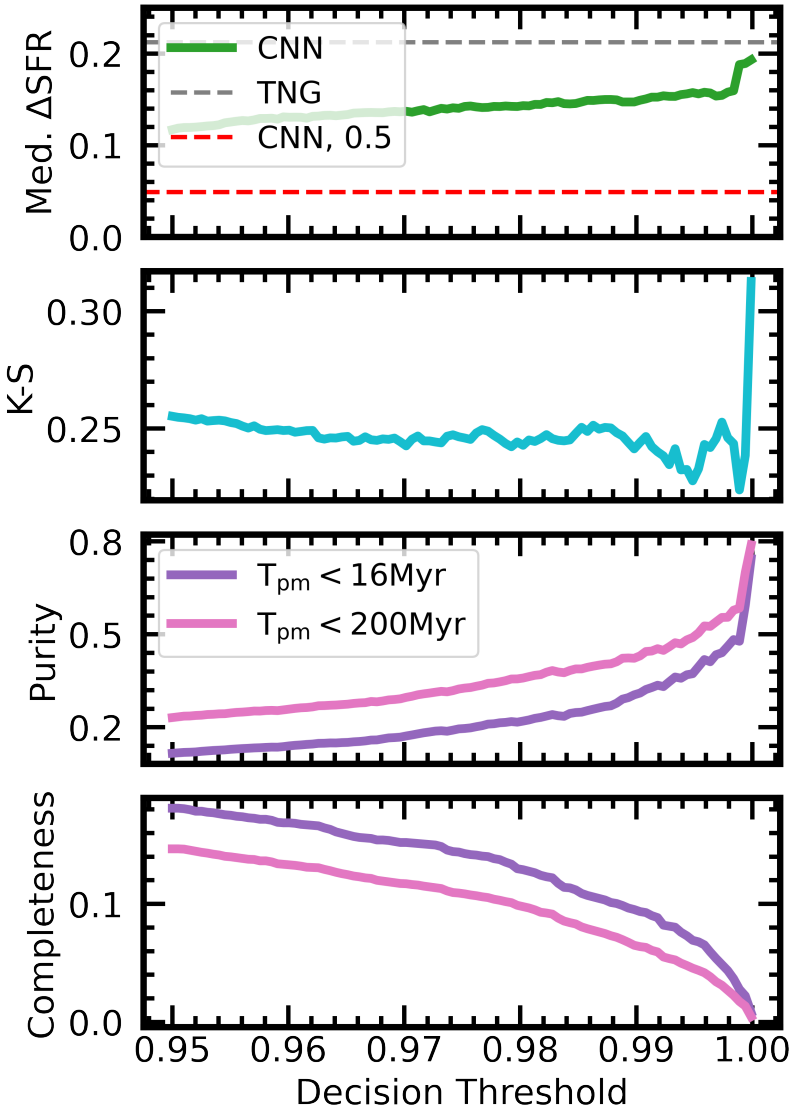

Figure 2.19: The ΔSFR in dex (top panel), K-S statistic (second panel), sample purity (third panel), and post-merger sample completeness (bottom panel) as I move the post-merger CNN decision threshold to the right, while moving the control decision threshold to the left in symmetrical fashion. Horizontal lines are included in the ΔSFR panel, designating the simulation ground truth value (grey) and the CNN result using the natural decision threshold at 0.5 (red). Two purity and completeness curves are plotted, the first of which uses the strict post-merger definition (purple), while the other (pink) uses a more tolerant definition (having merged within 0.2 Gyr.)

post-merger sample. The "slippery slope" of applying cuts in $p(x)$ to identify a pure post-merger sample suggests that another method of quality control is needed, ideally one that is independent from the CNN. In this work, supplemental visual classifications are used to this end (see Section 2.2.5).

2.2.4 Comparison to non-parametric morphological statistics

Before the popularization of machine learning methods in astronomy, a variety of automated, non-parametric morphological indicators (introduced in Section 1.3.2) were developed for merger identification. In this Section, I compare the performance of my CNN against several statistics that have often been used for the goal of merger identification. One popular method uses a cut applied in the Gini-M20 plane, given by Lotz et al. (2008):

$$Gini > -0.14 \times M20 + 0.33 \tag{2.1}$$

where galaxies above the line are counted as mergers. A critical threshold in asymmetry has also been used, as in Conselice (2003):

$$A > 0.35 \tag{2.2}$$

where galaxies with large A values are counted as mergers, or alternatively in shape asymmetry, defined by Pawlik et al. (2016):

$$A_s > 0.2 \tag{2.3}$$

where 0.2 is used as the the lower bound for merger identification in SDSS. The optimal choice for CFIS may be entirely different (e.g., Wilkinson et al. 2022), and I therefore consider a range of possible boundaries in Figures 2.20 and 2.21.

For the definitions and technical implementations for each statistic, I refer the reader to the recent comprehensive overview given in Wilkinson et al. (2024). In addition to their utility for rapid classification of large numbers of images, predictions from non-parametric methods, like the CNN's $p(x)$, can be modulated: the Gini-M20 plane cut-off can be shifted to a more severe position, and the asymmetry and shape asymmetry thresholds can be calibrated for a given data set. To compare the CNN's classification abilities to these other, more direct statistical methods, my colleague Scott Wilkinson processes the RealSim-CFIS images generated for the mock survey (Section 2.2.3) using StatMorph[7] (Rodriguez-Gomez et al. 2019) in order to obtain measurements

[7] statmorph.readthedocs.io/en/latest/

for Gini and M20, as well as asymmetry and shape asymmetry. I then compare the combinations of post-merger purity and sample completeness obtained using non-parametric methods as a function of their respective thresholds to those obtained by the CNN. For all tests that make use of StatMorph, I discard galaxies for which StatMorph raised an error of any kind during analysis. This does not disproportionately impact the post-merger or non-post-merger population; around 26 per cent of the images belonging to each class were flagged. The sample considered hereafter therefore contains 1734 post-merger galaxies according to the strict ($T_{Postmerger} = 0$ Gyr) definition, and 218536 non-post-mergers.

In Fig. 2.20 I investigate the efficacy of Gini-M20 classification for the synthetic images generated for the mock survey. The top panel of Fig. 2.20 shows the distribution of mock survey galaxies in the Gini-M20 plane. Green contours illustrate post-merger fraction, and the dashed line shows the cut (Equation 2.1) proposed by Lotz et al. (2008) to distinguish the locus of post-mergers in the mock sample. Although the relative positions of the dashed line and green contours (top panel of Fig. 2.20) indicate that most of the mergers are above the nominal Gini-M20 threshold, the figure also indicates that shifting this cut upwards could increase sample purity. While post-mergers are unusually abundant in the "spur" sequence in the upper-right of the figure, visual inspection of the "spur" galaxies reveals that foreground stars and other CFIS artefacts are also present in disproportionate quantities. The lower two panels of Fig. 2.20 (for Gini-M20) are analogous to the lower two panels in Fig. 2.19 (for the CNN) and show how the purity and completeness of the post-merger sample change as a function of threshold. As for Fig. 2.19, the lower two panels in Fig. 2.20 show the results for both post-mergers that are in their first post-coalescence snapshot (purple lines) and those that have merged in the last 0.2 Gyr (pink lines). Shifting the Gini-M20 cut upwards leads to increased purity, but consequently the post-merger sample grows increasingly incomplete, leaving fewer galaxies for subsequent study. I therefore find that an IllustrisTNG galaxy's Gini-M20 plane position does bear on its likelihood to have undergone a recent merger, but report weaker performance in both sample purity and completeness compared to the CNN. At the default setting for each method, i.e., the CNN decision threshold at 0.5 and the Gini-M20 cut-off at the position detailed in Lotz et al. (2008), the CNN returns a merger sample that is 6 per cent pure, while the Gini-M20 sample is not meaningfully distilled from the natural merger occurrence rate of 0.7 per cent. The sample identified using Gini-M20 is also much more incomplete, because galaxies of all types, including post-mergers, are most likely to fall below the line (towards the bottom left of the top panel in Fig. 2.20). After modulating both methods to their most extreme thresholds, Gini-M20 can be used to produce an enhanced purity of $\sim$4 per cent, though in a sample that is negligible in size ($<$ 1 per cent complete), while the CNN can produce a sample that is $\sim$80 per cent pure (an order of magnitude more effective than Gini-M20), while

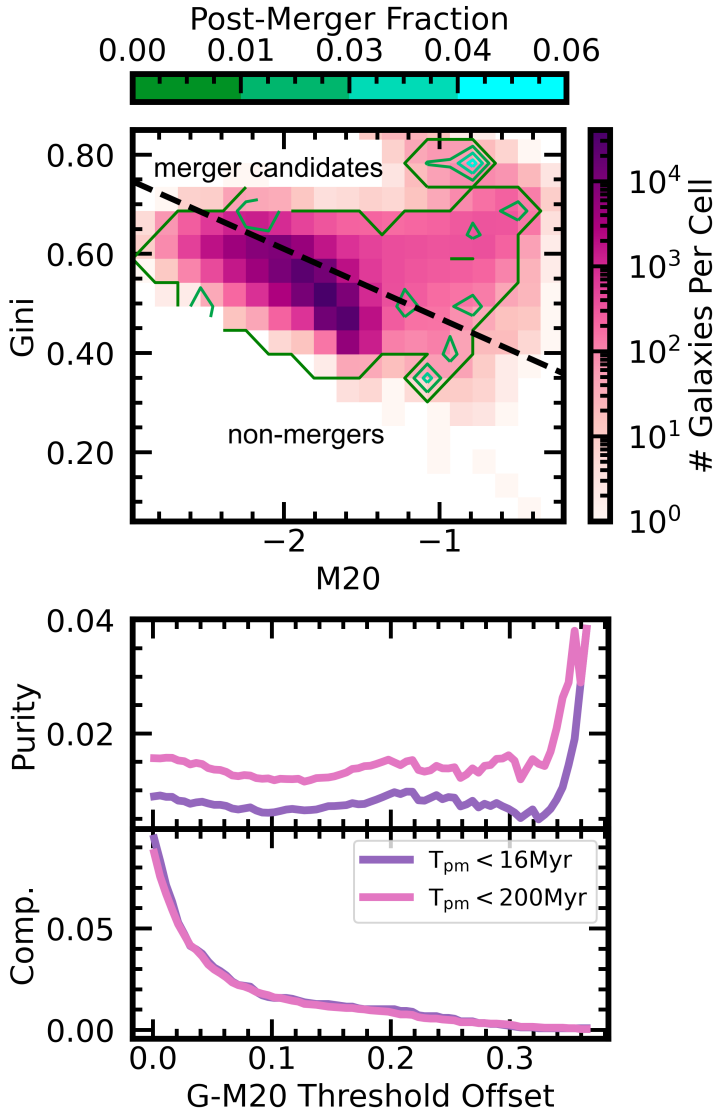

Figure 2.20: Top panel: a 2D histogram (magenta background) of mock survey galaxies and their positions on the Gini-M20 plane, with colour in log scale. The green-blue contours plotted over the histogram show the post-merger fraction. The black dashed line shows the default threshold position, used by Lotz et al. 2008 and later Rodriguez-Gomez et al. 2019. Middle and bottom panels: Purity and sample completeness, respectively, as a function of the vertical shift of the Gini-M20 line shown above. The purple line applies a strict post-merger definition, and only counts galaxies that merged within the last IllustrisTNG snapshot, while the pink line uses a 0.2 Gyr cutoff.

Method	Max. purity	Default completeness
CNN	80%	89%
Gini-M20	4%	9%
A	3%	4%
A_S	1%	47%

Table 2.3: Performance metrics characterizing the utility of the CNN, Gini-M20, asymmetry, and shape asymmetry methods for post-merger identification in the mock survey. The CNN can be used to identify post mergers that are substantially more complete and/or pure than each of the non-parametric morphological statistics.

successfully recovering 5 per cent of the sample. In short, the Gini-M20 method yields a post-merger sample that is a factor of 20 less pure than the CNN, even when pushed to its maximal performance.

Similarly, Fig. 2.21 examines galaxy asymmetries in the mock survey. In the top panel, I show the distribution of asymmetry values (following the application of Conselice 2003 within StatMorph) for post-mergers and non-post-mergers (blue and orange, respectively). Both categories exhibit a similar overall range in asymmetry. In the lower two panels of Fig. 2.21, I again show the purity and completeness of the outfalling post-merger sample as a function of asymmetry threshold. Unlike Gini-M20, the purple line in the middle panel of Fig. 2.21 demonstrates that a simple asymmetry threshold cannot be used to enhance the purity of the strictly defined post-merger sample in the mock survey. As the sample completeness decreases, purity hovers around its natural value of $\sim$0.7 per cent before dropping near $A = 1$ when all post-mergers have been ruled out by the threshold. A temporary enhancement in the purity of the softer-defined post-merger sample is achieved, however, near $A = 1$, though it never exceeds 3 per cent.

In addition to asymmetry (Fig. 2.21), I investigated shape asymmetry, which is also computed by Scott Wilkinson using StatMorph. As with asymmetry, shape asymmetry also yields low sample purity (at most, $\sim$0.8 per cent) that is not improved by varying the threshold. Asymmetry (and shape asymmetry) therefore yields a predicted post-merger sample that is even more impure than that identified using Gini-M20. Visual inspection of the sample suggests that foreground stars or other survey artefacts are frequently responsible for high asymmetries. Since these phenomena do not preferentially affect post-mergers or non-post-mergers, it is reasonable that post-merger purity would not necessarily be enhanced by asymmetry thresholds in excess of $A = 1$.

Table 2.3 shows the maximum achievable purity, as well as the completeness at the "default" setting for the CNN, as well as the three non-parametric morphological statistics studied. For

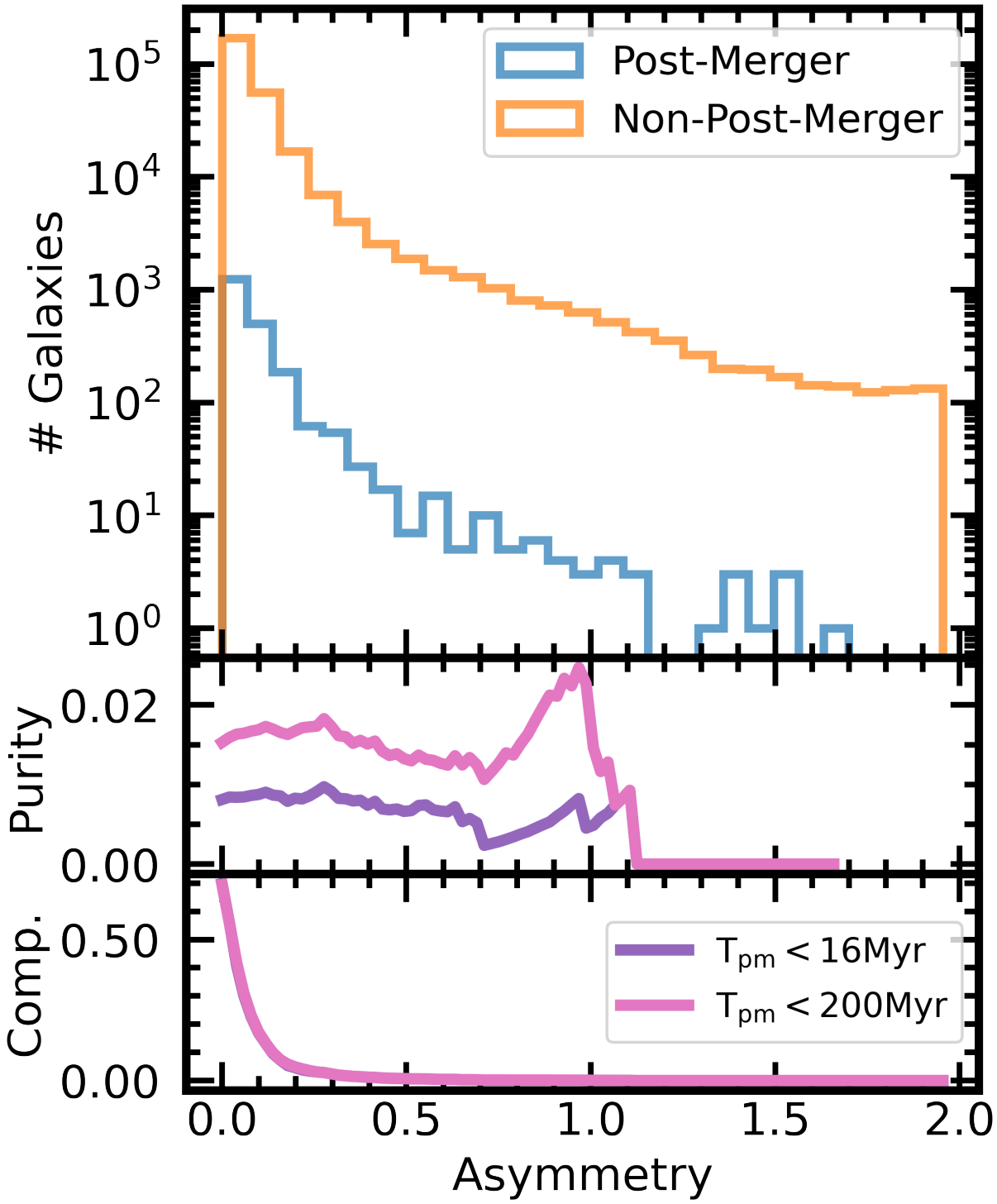

Figure 2.21: Top panel: Log scale histograms of the post-merger and non-post-merger populations in asymmetry as measured by StatMorph. Middle and bottom panels: Purity and sample completeness, respectively, as a function of the asymmetry threshold. The purple line applies a strict post-merger definition, and only counts galaxies that merged within the last IllustrisTNG snapshot, while the pink line uses a 0.2 Gyr cutoff.

the CNN, the default setting refers to the decision threshold at $p(x) = 0.5$, while for Gini-M20, A, and A_S, the default settings are given in Equations 2.1- 2.3 above. I find that the CNN substantially out-performs the non-parametric statistics both in completeness and purity. Studies have found that machine learning combinations of non-parametric morphological statistics are better for merger identification than lone statistics (e.g., Nevin et al. 2019), indicating that greater complexity is required for robust merger detection. Wilkinson et al. (2024) further demonstrate the relevance of complex models by presenting ROC curves for individual non-parametric morphological statistics, as well as linear discriminant analysis (LDA) and random forest (RF) machine learning combinations of the same statistics (see Figure 12 in Wilkinson et al. 2024). In a finding analogous to that of this work, Wilkinson et al. (2024) report better AUC scores for the ROC curves when the statistics are combined using a machine learning method. It is likely thanks to this complexity that a CNN trained with synthetic image data shows much stronger performance than any of the three popular automated merger identification methods studied here, and returns a post-merger sample an order of magnitude more pure. While the CNN benefits from calibration and training on images similar to those studied in the mock survey, it is likely that its success on this task also speaks to a diversity in the morphological characteristics of simulated post-mergers. A trainable, highly dimensional statistical tool like the CNN may be better prepared to characterize the features relevant to post-merger status, and recognize them when applied to new images.

2.2.5 Comparison to visual classifiers

Having demonstrated the CNN's ability to outperform non-parametric methods of post-merger identification in the context of the mock survey, I next turn to visual classification. CNNs identify images in much the same way that a human being does: by learning to recognize features and feature combinations that are relevant to the image's classification. Additionally, trained CNN models require only a fraction of a second to assign a label to an image, while people may require up to several minutes to rigorously classify a single image. Human visual classifications are not without merit, however; human classifications benefit from inherent physical intuition, and in some cases, degrees of relevant training in image classification, all of which lend priors that affect the method by which an individual approaches a classification problem.

In order to compare the performance of the machine to that of a group of people, I organize a visual post-merger classification exercise using a 200-image subset of the data. Given the size of the mock survey, it is not practical to visually inspect the full data set, and so I have selected a small subset for the purpose of a straightforward comparison. The sample is constructed in

pseudo-random fashion so that contributors to my research project can also participate in good faith: I arranged for the test set to contain between 30 per cent and 70 per cent post-mergers. In practice, the visual classification sample contains 112 post-mergers and 88 control images arranged in random order. The participant groups (which includes coauthors of Bickley et al. 2021: RWB, CB, SLE, SW, and the volunteers recognized in the acknowledgments) are provided with no information about the post-merger occurrence rate in the image sample, and are asked to assign either a post-merger or non-post-merger label to each image. The classifications are marked in a standardized text file containing a list of galaxy identification numbers, each corresponding to the file name of a single galaxy image. Images are provided in Flexible Image Transport System (FITS) format so that popular astronomy visualization tools (e.g., SAOimage DS9) can be applied. Because I intend to compare the visual classification results to those of the machine, the participants are instructed to strive for total correctness in classification, as the CNN does by design, and not necessarily post-merger sample purity.

Classifier Statistics

Fig. 2.22 shows the fraction of control galaxies and post-mergers that were correctly identified by the CNN at the default decision threshold ($p(x) = 0.5$), and by the human classifiers. The fractional purity of each classifier's final post-merger sample is also shown. CNN (200) refers to the specific performance of the machine on the 200 selected images, while CNN-avg refers to the CNN's average performance, as detailed in Fig. 2.6. At the time of completing the visual classification exercise, classifiers 1, 2, and 3 are astronomy faculty members, classifiers 4 and 6 are pursuing post-doctoral studies in astronomy, 5 and 7 are graduate students in astronomy, classifier 8 is an undergraduate student in astronomy, and classifier 9 is a mechanical engineer who was given a short lecture on the properties of post-merger galaxies and their manifestations in images, using selected images from the training set that were not included in the classification task as examples.

In general, the total experience with galaxy image classification (i.e., the total number of images classified in an individual's career) seems to have a bearing on each person's performance, but other factors may also be at play. For instance, participants experienced with synthetic images (e.g., C6 and C7) are more likely to have a CNN-like balance of statistics on the test data, while those more familiar with survey observations (e.g., C1 and C5) are more conservative in their assignment of post-merger labels, and exchange post-merger sample completeness for purity. Generally, the CNN recovers a higher fraction (96 per cent) of the post-mergers than does any human participant. Even when instructed to focus on total accuracy, however, several human

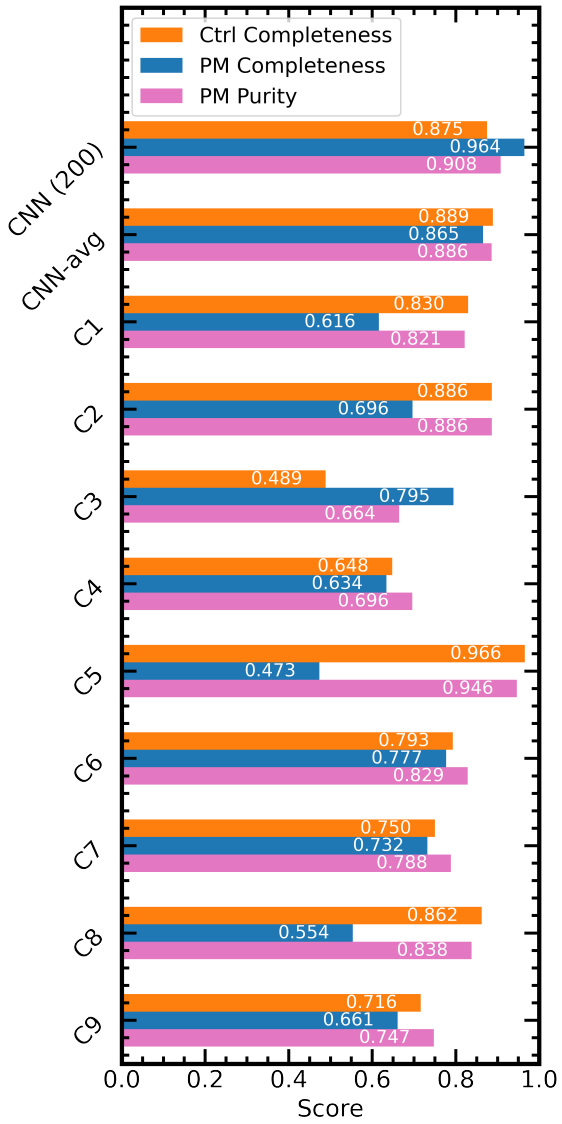

Figure 2.22: The post-merger sample completeness, control completeness, and post-merger purity statistics for various contributors.

classifiers excel in post-merger sample purity, in some cases approaching or even surpassing (e.g., C5) that of the CNN for this image subset.

A hybrid approach

The visual classification experiment can also be leveraged as a test of the combined classification power of a CNN algorithm and one or more individuals. When applied to a large survey, here approximated by the mock survey detailed in Section 2.2.3, the excellent ability of a trained CNN to identify individual post-mergers allows it to generate a predicted post-merger sample that is highly complete and enhanced in purity. While doing so, it simultaneously deprioritizes the visual inspection of a large majority of previously unclassified images. My visual classification experiment suggests that human classifiers are generally not as good at identifying a complete sample, but benefit from a physical understanding and can excel in the identification of pure post-merger samples. All participants were able to enhance the purity of the post-merger sample when considering only galaxies assigned post-merger labels by the machine, in most cases by a few percentage points, and up to 100 per cent in the case of C2. In addition, since the participants were instructed to approach the problem with an interest in total performance rather than purity, an adjustment in the purpose and parameters of future classification tasks would further enhance this effect when applied to survey data.

Since human visual classifiers and the CNN have different expertise and distinct approaches, a method combining labels from the CNN and visual predictions will benefit from the CNN's scalability, efficiency, and accuracy, as well as the intuition and rhetorical capability of the people conducting the classifications. By performing visual quality control on galaxies predicted to be post-mergers by the CNN, I propose a hybrid approach that will benefit from the jury principle (Boland 1989). Put another way, the completeness, purity, and representativeness of the final post-merger sample will benefit from the fact that the CNN and visual classifiers make different mistakes. Critically, this can be done without applying an aggressively high $p(x)$ cut *a priori*, so that the final post-merger sample will not be as biased against mergers whose appearances are somewhat fainter or less dramatic.

2.2.6 Summary

Large samples of pre-coalescence galaxy pairs have facilitated detailed and statistically robust study of the pair phase of galaxy mergers. However, since post-coalescence galaxies are more difficult to identify via automated methods, large samples of post-merger galaxies are missing from the literature. The details of this transformative period in galaxy evolution therefore remain

relatively unstudied.

CFIS will contain relatively deep, excellent quality r-band imaging over $\sim$5000 deg^2 of the sky, making it an excellent data set for the identification of recently merged galaxies. In order to avoid performing labourious visual classifications of a prohibitively large population of CFIS galaxies, I seek to develop a software tool for post-merger sample distillation in order to augment the efficiency of post-merger identification efforts.

In this Chapter, I construct synthetic CFIS observations of simulated IllustrisTNG galaxies in order to study the merits of automated post-merger identification with a CNN. My main findings from this effort are as follows.

(i) I train a CNN on a large sample of synthetic galaxy observations, produced by applying CFIS-motivated observational realism to the stellar mass maps obtained from $0 \leq z \leq 1$ galaxies in the 100-1 run of IllustrisTNG. The final data set contains 37312 images each of post-merger galaxies and control galaxies. I evaluate the model on a sample of post-merger and control galaxies never seen in training. It recovers 88.9 per cent of the control galaxies and 86.5 per cent of the post-mergers galaxies (Fig. 2.6), with an AUC score of 0.95 (Fig. 2.7).

(ii) The model's performance is not significantly impacted by most relevant metadata quantities, including stellar mass and mock observation redshift (Figs 2.11 and 2.10). The model is most prone to misclassification when a galaxy has one or more nearby neighbour(s) within a few tens of kpc (Fig. 2.8). In these cases, the visual characteristics of the morphological disturbances induced by the partner galaxy/galaxies are highly degenerate with those associated with a recent merger, and may be indistinguishable to the network.

(iii) I apply the trained model to a mock survey, containing one image of every $0 \leq z \leq 1$ IllustrisTNG galaxy with stellar mass $> 10^{10}$M$_\odot$. The post-mergers as defined in training, which are re-observed from a new camera angle and with new mock observation parameters, constitute less than 1 per cent of objects meeting this mass criterion. The model correctly labels $\sim$89 per cent of both the post-mergers and the non-post-mergers in the survey (Fig. 2.15). However, from Bayes theorem, the scarcity of post-mergers relative to non-mergers means that the outfalling post-merger sample is only 6 per cent pure when the default CNN decision threshold is used (Fig. 2.16). While legitimate sample impurity exists, I note that many of the contaminating galaxies have undergone a recent merger within 0.2 Gyr (Fig. 2.17).

(iv) I use the CNN's labels and decision threshold to identify post-merger and control samples for a proof-of-concept study of star formation enhancement in the style of Hani et al. (2020). In particular, I investigate how changes in the decision threshold, which affect purity and completeness of the post-merger sample, affect the SFR statistics. An extreme decision threshold scheme of $\geq$0.999 for post- mergers and $\leq$0.001 for controls gives rise to 49 per cent purity in the post-

merger sample, and returns a prediction for the median DeltaSFR in excellent agreement (within 0.02 dex) of the simulation's ground truth for the most conservative control-matching scheme (Figs 2.18 and 2.19).

(v) I compare the CNN's performance on the survey to extant non-parametric statistical methods, and find a linear cut-off in the Gini-M20 plane to be modestly effective in enhancing post-merger sample purity (up to $\sim$4 per cent, with $\leq$1 per cent completeness) in the mock data (Fig. 2.20). However, the CNN outperforms this approach at its default decision threshold, and can identify a highly pure sample ($\sim$80 per cent) with better completeness ($\sim$5 per cent) at its most extreme setting. Using an asymmetry threshold failed to distill the true positive rate in the predicted post-merger sample any higher than 0.7 per cent, the natural occurrence rate of strictly defined post-mergers in IllustrisTNG100-1, but did slightly enhance the density of galaxies with a $T_{Postmerger}$ of $\leq$0.2 Gyr to 2 per cent from a natural density of 1 per cent (Fig. 2.21). Shape asymmetry was ineffective in enhancing the purity of post-mergers in the mock survey. I also consider human visual classifications, and identify a trade-off: visually identified samples can be more pure (up to 95 per cent in the case of Classifier 5) than those returned by the CNN, but are typically more incomplete (47 per cent for the same classifier, versus 96 per cent for the CNN on the same data) (Fig. 2.22).

In Sections 2.2.3 and 2.2.5, I argue that selection biases and sample impurity will be inherent in any effort, automated or manual, to identify post-mergers in a survey with a natural post-merger incidence rate, and that CNN-based post-merger identification is best utilized as a first round of distillation, free from pre-conceptions other than those inherent to the training data. Once a set of CFIS galaxies is processed in this way, a trained person can further improve the quality of the post-merger set through careful inspection.

The image classification techniques developed in this work have been thus far trained and tested on like-generated synthetic observations, e.g., in Section 2.2.3. By training a CNN on a population of simulated galaxies processed with CFIS realism, however, I have simultaneously prepared it for application to CFIS, to the extent that IllustrisTNG galaxies are morphologically representative of those found in the Universe. In the next Chapter, I will apply the model to real CFIS galaxies as part of a hybrid approach comparable to that detailed in Section 2.2.5 in an effort to identify and study a post-merger sample of groundbreaking volume and quality.

Chapter 3

Post-merger identification in CFIS and DECaLS using a CNN

In the previous Chapter, I demonstrated the merit of a convolutional neural network model for post-merger identification in galaxy surveys. The CNN performed well in a proof-of-concept study of merger-induced star formation enhancement carried out on a "mock survey" composed of simulated galaxies convolved with observational realism. In this Chapter, I apply the simulation-trained CNN model to images of real galaxies taken from two astronomical surveys – CFIS and DECaLS. When using any classification tool to identify rare objects (such as galaxy mergers) in astronomical surveys, false positives (non-mergers incorrectly classified as mergers) must be accounted for. To mitigate the biases imposed on subsequent studies of the post-merger galaxies I identify using the CNN, I use the hybrid approach (CNN classification, followed by manual cleaning of the predicted merger sample) proposed in Section 2.2.5. I perform manual inspection collaboratively with two other expert classifiers (co-authors of Bickley et al. 2022 SLE and DRP). Once the method of identifying post-mergers in CFIS and DECaLS has been outlined, I introduce the resulting post-merger samples.

3.1 Data and methods: CFIS

The CNN model introduced in Chapter 2 was prepared for the eventual purpose of identifying a large sample of post-merger galaxies in CFIS. The model's performance could be measured as a function of numerous galaxy parameters in the context of the IllustrisTNG mock survey, since the ground truth (i.e., the true merger statuses of galaxies in the simulation) is known. No such ground truth is available in CFIS, however, so the veracity of the following results depend on an underlying assumption: that galaxies in IllustrisTNG are morphologically similar to those in the low-redshift Universe. In particular, it is important that merger events in IllustrisTNG and in the Universe have a similar influence on the appearance of galaxies. The literature on this topic is promising: for example, Rodriguez-Gomez et al. (2019) found good ($\sim 1\sigma$) agreement in the non-

parametric statistical morphologies of galaxies in IllustrisTNG and real galaxies observed by Pan-STARRS (Chambers et al. 2016). Tacchella et al. (2019), Huertas-Company et al. (2019), Du et al. (2020), and Zanisi et al. (2021) also find that similar feature spaces are shared by IllustrisTNG and real galaxies. More recently, Eisert et al. (2024) used a contrastive machine learning framework to compare IllustrisTNG galaxies and galaxies imaged by Hyper Suprime-Cam (HSC), and reported that the two datasets shared some ∼70 per cent of their parameter spaces. Quantitative analyses across multiple frameworks therefore indicate that a simulation-based classification effort using IllustrisTNG galaxies is reasonable, provided that one conducts it with care.

3.1.1 Simulation-based inference

The merger identification framework used in this work is a type of simulation-based inference (SBI). In general, SBI methods use simulations and machine learning to discover the mapping between underlying (and usually, difficult or impossible to observe) physical information and parameters that are straightforward to observe. In this case, galaxy merger status is the physical information to be extracted, and galaxy morphology is the observable. The connection between galaxy interactions and morphology has been well established for decades (e.g., White & Rees 1978; Lacey & Cole 1993), and so it is reasonable to expect that there exists a strong mapping between the two to be discovered by my model. Indeed, the model's performance in Sections 2.2.1 and 2.2.2 indicates that such a mapping has been found in IllustrisTNG. Intriguingly, galaxy observables appear to encode even higher-order information, e.g., constraints on cosmological parameters (as explored in Villaescusa-Navarro et al. 2022). By comparison, the use of SBI for merger status is fairly safe. The trustworthiness of the results presented here is bolstered by the fact that merger status can often (though not always, as explored in Section 2.2.5) be confirmed by eye. These conveniences allow me to carry out a merger identification effort, using SBI and the CNN to make a preliminary round of classifications, before carrying out visual cleaning of the resulting merger sample.

The flowchart in Figure 3.1 illustrates how my CNN is trained on simulated data, applied to observed galaxies, and combined with visual classifications to produce a large, homogeneously selected, and highly pure post-merger sample. It is this combination of methods that I ultimately use to identify the post-merger sample described in Section 3.1.3.

3.1.2 Preparing the CFIS data for classification

The CNN was trained and previously evaluated on synthetic galaxy images that had been cropped to a physical scale of 100 kpc on a side, upscaled or downscaled to 138×138 pixels (corresponding

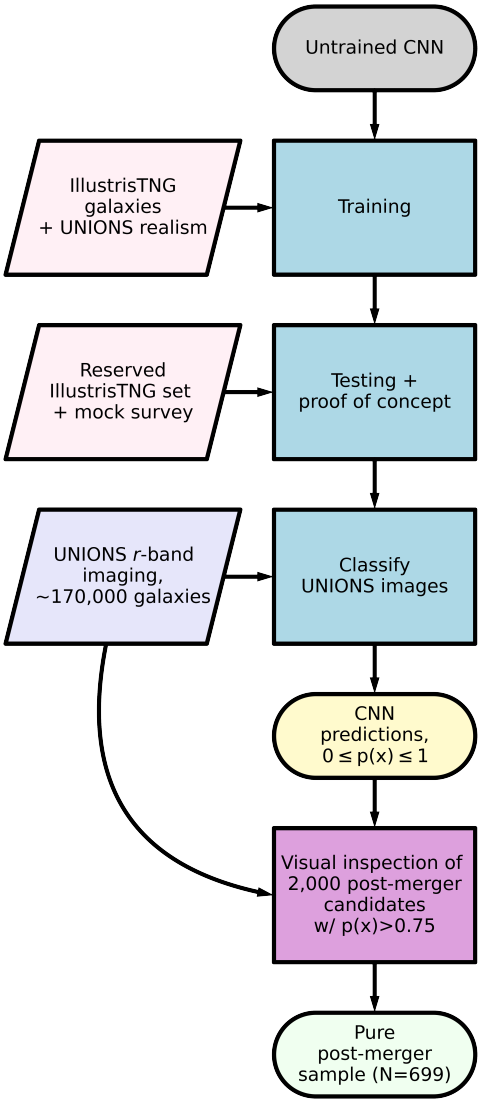

Figure 3.1: A flowchart showing the inputs and outputs (ovals), procedures (boxes), and data sets (parallelograms) described in this work. My CNN is trained and tested on example post-merger and non-post-merger galaxies from the IllustrisTNG simulations before it is used to obtain post-merger classifications on real galaxies in *r*-band UNIONS imaging. The classifications are then inspected visually, and a sample of post-mergers is confirmed and presented.

to the natural pixel resolution of a galaxy imaged by CFIS at the median redshift of the sample), and normalized on a linear scale so that the brightest pixel had a value of 1, and the faintest had a value of 0. In order to prepare galaxies imaged in CFIS DR2 for evaluation in the same way, I therefore require access to an accurate (i.e., spectroscopic) redshift measurement for each object. To this end, I limit my post-merger search to galaxies with $z \leq 0.5$ in the overlap between CFIS DR2 and SDSS DR7, and match objects in the two catalogues with a 2-arcsecond tolerance. The resulting sample contains 168597 galaxies. I use the SDSS-derived spectroscopic redshift to produce CFIS cutouts with the same field of view, pixel size, and normalization characteristics of the images used for training and evaluation in Bickley et al. (2021). The availability of SDSS spectra for post-merger galaxies will also facilitate subsequent spectroscopic characterization of the galaxies' properties (see Chapters 4, 5, and 6).

3.1.3 The CFIS post-merger sample

After preparing the CNN classifier and CFIS galaxy images, the CNN makes a prediction of post-merger status on each image. In this section, I seek to statistically appraise the quality of the CNN's post-merger predictions in an observational context. I then present a catalogue of CNN-predicted post-mergers that also passed rigorous visual inspection by the authors of Bickley et al. (2022) RWB, SLE, and DRP.

CNN-identified post-mergers

I emphasize that true merger status of galaxies in the Universe is unknown to observers. Therefore, in analyzing the results of a post-merger identification effort, robust experimental validation is also elusive, and sample impurity is expected (see Figure 1.6). Nevertheless, I can check for systematic issues by inspecting the global statistical characteristics of the out-falling galaxy samples. Because post-mergers are known to be rare, for example, if a CNN (or any other merger identification tool) found that a survey was comprised mostly of post-mergers, there would be cause for alarm.

Figure 3.2, which shows the stellar mass[1], redshift, and CNN $p(x)$ predictions for the galaxies in CFIS DR2 and SDSS DR7 (top two panels), addresses the systematic approach to validation. The black histogram (top right) shows the log-binned CNN predictions between 0 and 1 assigned to CFIS galaxies with SDSS spectra. As in the mock survey detailed in Chapter 2, the CNN assigns non-post-merger predictions to the vast majority of galaxies when I use the "default" decision

[1]The stellar mass estimate used to produce these figures, as well as the similar figures in subsequent Chapters, are described in Chapter 6 and are useful here because they are available as a rule for the vast majority of galaxies in SDSS DR7.

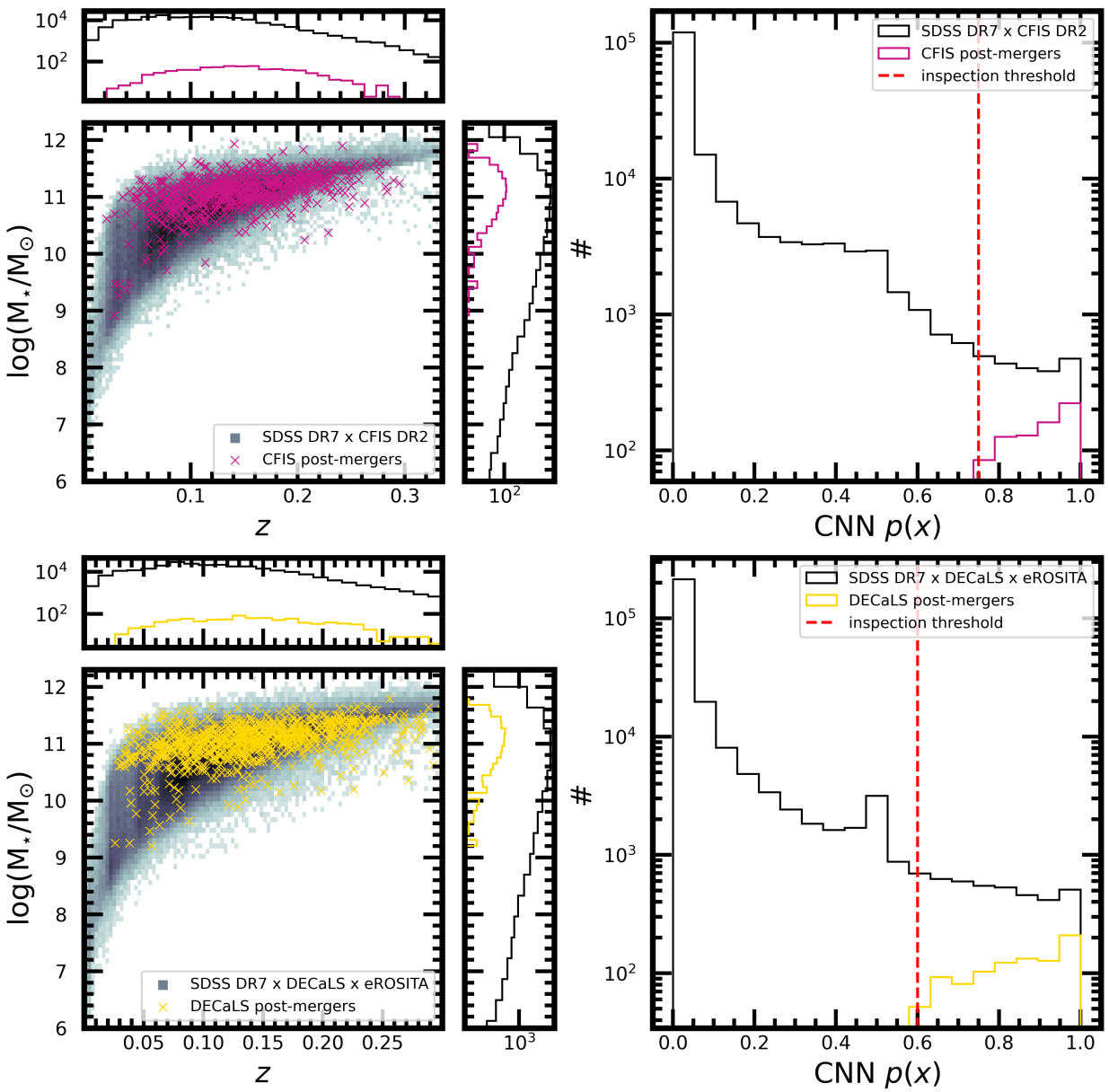

Figure 3.2: A high-level statistical overview of the galaxies involved in the post-merger classification effort in this Chapter. Top left: the stellar mass (from a custom photometric estimate specific to this work, see Chapter 6) plotted against the spectroscopic z from SDSS for galaxies appearing in both SDSS DR7 and CFIS DR2 (greyscale 2D histogram and black auxiliary histograms). I also plot the positions of the visually confirmed CFIS post-merger sample (see Section 3.1.3) in magenta. Top right: a log-scale histogram of the CNN $p(x)$ predictions for the galaxies appearing in SDSS DR7 and CFIS DR2 – most galaxies receive predictions with $p(x) < 0.5$. Visual quality control is conducted on galaxies with $p(x) > 0.75$ (red dashed line), and the visually confirmed post-mergers are shown in magenta. Bottom panels: same as the top panels, but for the post-merger sample identified in DECaLS imaging (see Section 3.2) from the sample of galaxies appearing in the union of SDSS DR7, DECaLS, and the German portion of the eROSITA X-ray survey (see Chapter 6). Visual inspection is conducted in DECaLS imaging for galaxies with $p(x) > 0.6$, and the statistics for the visually confirmed DECaLS post-mergers are shown in yellow.

threshold at 0.5 to differentiate between the two classes. Naturally, more strict (higher) decision thresholds can be used to force even greater selectivity: in Section 2.2.3, I demonstrated that in an unstudied test set of synthetic galaxy images, the network was well-calibrated, meaning that the proportion of true post-mergers in bins of $p(x)$ scaled with the value of $p(x)$. Consequently, if CFIS galaxies are sufficiently visually similar to the training galaxies from IllustrisTNG 100-1, the CNN's numerical post-merger prediction on each CFIS galaxy can be interpreted as probability-like. I therefore count galaxies with $p(x) > 0.75$ as highly likely to be post-mergers in this subsection, and flag them for visual inspection in Section 3.1.3. Although the CNN's $p(x)$ axis represents opposing gradients of completeness (the number of true post-mergers included above some value) and purity (the fraction of galaxies above some value that are true post-mergers), I choose $p(x) > 0.75$ as a convenient criterion for "highly likely" because $\sim$2000 galaxies lie above this value (represented in Figure 3.2 by the dashed red line), and 2000 galaxies can be feasibly inspected by eye for the purposes of this experiment.

Figure 3.3 compares the physical (stellar mass and redshift) and observational (r-band Petrosian magnitude[2]) characteristics of the CNN-predicted post-merger sample (blue histograms) to those of the CFIS galaxy sample appearing in SDSS DR7 (grey histograms). In training, the CNN was shown images of a diverse population of simulated post-merger galaxies, inserted into a realistic range of CFIS galaxy redshifts, survey skies, and observational seeing in order to prepare it to encounter galaxies with similarly varied characteristics in the observations. These efforts were broadly successful; post-mergers were found across the entire observational range of brightness, stellar mass, and redshift.

Still, the predicted post-merger sample is not bias-free. In Section 2.2.3, the network recovered $\sim$ 89 per cent of the post-merger galaxies in a mock survey, designed to mimic the characteristics of the observational sample it would eventually classify in this work. I therefore expect that a non-zero number of true post-merger galaxies in CFIS receive negative (≤ 0.5) classifications from the CNN, and are excluded. Still more true post-mergers may have been excluded from consideration between $0.5 \leq p(x) \leq 0.75$, where I certainly expect a small fraction of post-mergers to lurk. Because images with $p(x) \leq 0.75$ are not inspected, it is difficult to speculate on the cause of their exclusion from the post-merger sample, but it is reasonable to assume that their visual disturbances are not strong enough, due to low surface brightness or low mass. Conceivably, a true post-merger might be obstructed by a CFIS imaging artifact such that its merger status could not be determined. The statistics of the CNN-predicted post-mergers (blue histograms, Figure 3.3)

[2]The Petrosian magnitude is defined based on the ratio of the local surface brightness at the Petrosian radius to the mean surface brightness within the same radius. The Petrosian flux and magnitude are measured inside an aperture defined by the radius.

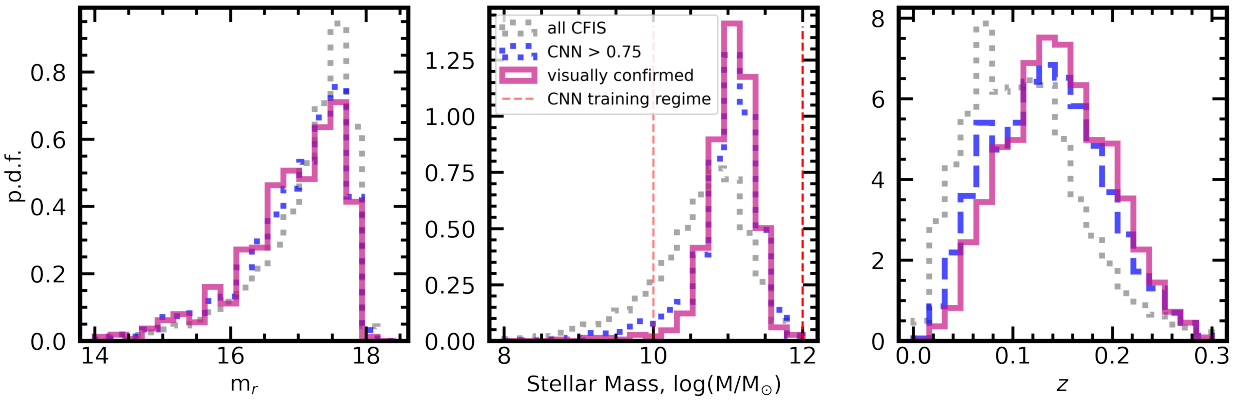

Figure 3.3: The r-band Petrosian magnitudes (left), stellar masses from Mendel et al. (2014) (centre), and redshifts (right) of the sample of CFIS galaxies with SDSS spectra (grey dashed histogram), the CNN-identified sample with predictions > 0.75 (blue dashed histogram), and the 699 visually confirmed post-mergers (magenta histogram, see Section 3.1.3). All three histograms are normalized density distributions, for simplicity of comparison. The three samples are fairly consistent with one another in r-band magnitude, suggesting that the CNN's training on images with a realistic range of imaging signal to noise (S/N) was broadly effective. There is a deficit of very faint galaxies (see $m_r > 17.5$) in the post-merger sample (left panel). Individual inspection of these galaxies reveals that they have relatively low stellar masses, and are less bright at a given distance than their more massive counterparts. The CNN's preference for more massive galaxies (centre panel) is striking, even within the mass range on which the network was trained. Because more massive galaxies are brighter relative to the background, they are more likely to exhibit bright merger signatures if they have undergone a merger. The stellar mass bias is echoed in the redshift distribution (right panel), since the volume-limited nature of the survey means that very massive galaxies are more likely to appear at higher redshift. The cause of this effect is explored later in Figure 3.4.

also reveal that the CNN itself developed biases during its training. While the network correctly identifies a number of post-mergers at very faint apparent magnitudes, there is a deficit in the CNN-identified and visually confirmed samples relative to the sample of CFIS galaxies appearing in SDSS DR7. This bias is likely a natural consequence of the fact that tidal features in brighter post-merger galaxies are more visually distinct (i.e., brighter relative to the background).

The CNN's preference for brighter galaxies manifests most distinctly in the stellar mass statistics of the three samples. At a given redshift, a galaxy with a higher stellar mass is very likely to shine brighter than a lower-mass galaxy. As a result, the CNN population is skewed towards higher masses, peaking near $10^{11}M_{\odot}$ in the middle of the network's training regime (10^{10} – $10^{12}M_{\odot}$). The galaxy sample is volume-limited at high stellar masses, and so its mass bias causes it to prefer galaxies at a higher redshift on average as well. An argument for the brightness-motivated mass bias producing the redshift bias, rather than the other way around, is given in Figure 3.4. The top panel shows the median stellar mass in equal-population bins (i.e., the same number of galaxies fall into each bin) of redshift, while the bottom panel shows the median redshift in equal-population bins of stellar mass. Both panels have a 2D histogram of the full sample of galaxies appearing in CFIS DR2 and SDSS DR7 in the background. The median statistic for the parent sample is shown in grey for both panels, the CNN-identified subset is shown in blue, and the visually confirmed galaxies are shown in magenta. The auxiliary panels show density histograms for each quantity in the same colour scheme.

The top panel of Figure 3.4 shows that in a given redshift bin, the average stellar mass is higher for the CNN-identified and visually confirmed samples than in the parent sample. The network is therefore more likely to select a higher-mass galaxy over a lower-mass galaxy if they both appear at the same z. The preference for more massive galaxies is a natural consequence of the fact that more massive galaxies are more likely to be bright relative to the background at a given distance, and their tidal features are more likely to stand out. The mass bias effect weakens somewhat at higher redshift, where galaxies in the parent sample are more massive on average, and the role of distance in determining the imaged brightness of a galaxy may become more significant.

The bottom panel of Figure 3.4 (to be viewed with the vertical axis as the independent variable) shows that the inverse trend is not true. In all stellar mass bins, the median redshifts of the full, CNN-identified, and visually confirmed samples are in good agreement. Put another way, if two equal-mass galaxies appear at different redshifts in the survey, the network is no more likely to select one than the other. As such, the statistical biases of the sample in mass and redshift can both be comfortably interpreted as a result of each image's signal (galaxy brightness) to noise (sky background, seeing, and any survey artefacts) ratio.

In spite of the CNN's preference for brighter galaxies, its classifications are extremely help-

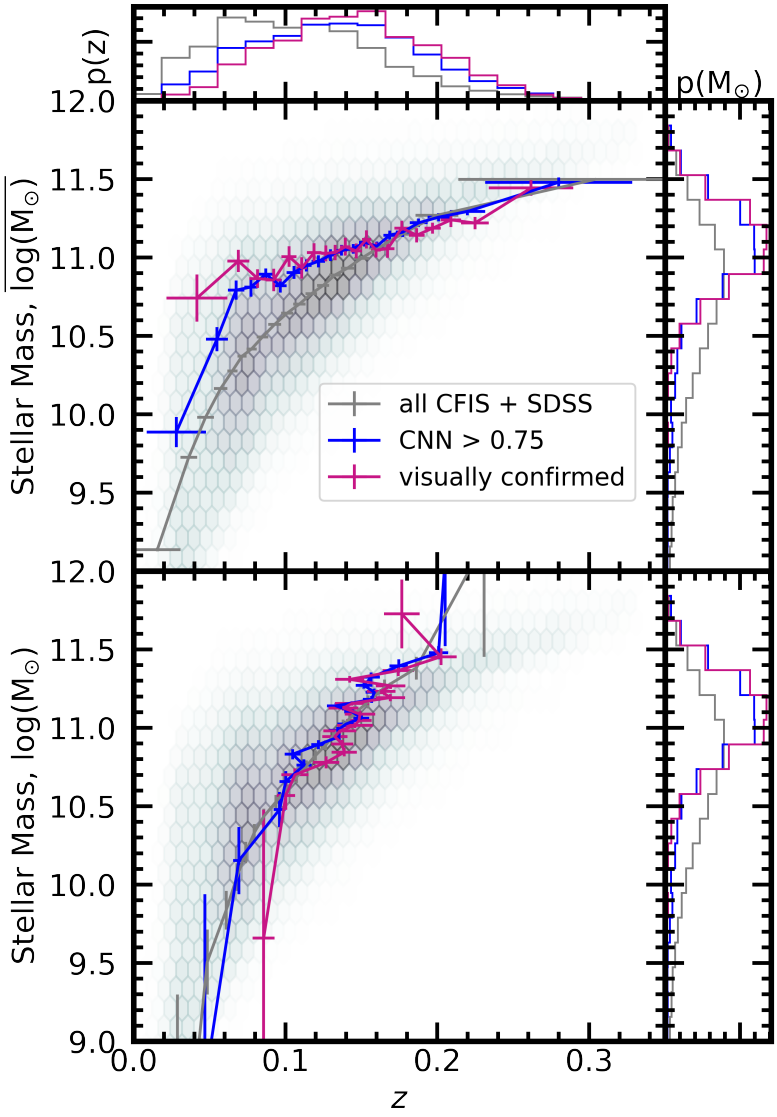

Figure 3.4: The average stellar mass in equal-numbered bins of redshift (top panel) and the average redshift in equal-numbered bins of stellar mass (bottom panel) for the full sample of CFIS DR2 galaxies with SDSS DR7 spectra (grey curve), the sample of galaxies assigned $p(x)$ predictions over 0.75 by the CNN (blue), and the visually confirmed sample (magenta). Both panels are superimposed over 2D histograms of the parent sample in the background. Individual histograms for each statistic are shown in the auxiliary panels. The presence of a redshift-dependent mass bias and the lack of a mass-dependent redshift bias together suggest that the relative brightness of a galaxy in its respective CFIS image may bear on its likelihood to be selected by the CNN. The CNN's preference is likely responsible for some unusual statistical characteristics of the visually confirmed post-merger sample.

ful as a first step in my effort to "distill" a post-merger sample of unique size and purity (i.e., to produce a small sample of high post-merger concentration from a larger sample in which post-mergers are very rare). During training, the weights assigned to the CNN's neurons are determined iteratively as the network studies subsets ("batches" in machine learning training parlance) of the training data. The final weights are chosen based on the moment in the network's training history when it was most successful in classifying post-mergers and control galaxies in the simulated data. Consequently, the network's learned biases may be useful in accomplishing the goal of correctly labelling as many galaxies as possible. The network's preference for brighter and more massive galaxies may be a natural consequence of the fact that the merger statuses of galaxies whose tidal features blend into the background are indeed more ambiguous.

The hybrid classification post-merger sample

After being classified by the CNN, 2000 galaxies given high post-merger $p(x)$ predictions (those belonging to the grey histogram to the right of the red dashed line at $\sim$0.75 on Fig 4.1) are in turn inspected by me, employing a purity-motivated classification philosophy in which galaxies whose post-merger statuses are in doubt are rejected. The visual classifications are performed on 100$\times$100 kpc CFIS monochromatic r-band image cutouts, though the cutouts are provided at their original resolution, since visual classifications do not require a single pixel size. In many cases, SDSS full-colour images of the wider field surrounding the target galaxy are also inspected in order to verify post-merger status specifically (as opposed to a pair-phase interaction). In order to pass inspection, galaxies are required to be fully coalesced, with only one discernible nucleus. They are also required to have visible signatures of a recent merger: stellar streams, shells, and/or rings, often marked by extended and asymmetrical morphologies. Conversely, galaxies are ruled out if they are not fully coalesced, appear to belong to a merging or interacting pair (either in CFIS or by inspection of the galaxy's wider environment via SDSS), or do not have noticeable post-merger morphological characteristics at any contrast setting. Although the CNN's training set includes images with realistic and proportionally-representative CFIS survey artefacts, occasionally predicted post-mergers are obscured by CFIS artefacts in such a way that their merger statuses are indiscernible; these images are also excluded. After this first visual inspection, the systems given positive post-merger labels are inspected jointly with co-authors of Bickley et al. (2022) SLE and DRP until a consensus label is chosen for each galaxy. If one or more of the authors are skeptical about a given galaxy's classification as a post-merger during this review, it is removed from the sample. Around 65 per cent of the galaxies are ultimately removed from the CNN-predicted post-merger sample with $p(x) > 0.75$, but 699 galaxies are unanimously confirmed

as post-mergers by the authors after visual inspection. While this visually confirmed sample is certainly not complete, I argue that it is highly pure. Figure 3.5 shows 36 post-mergers from this visually confirmed sample with logarithmic brightness scaling.

The results of the visual classification exercise are summarized in Figure 3.6. In a series of bins between the lowest and highest CNN $p(x)$ predictions inspected by the authors, I plot the fraction of galaxies for which the visual labels agreed with the network's positive post-merger prediction. The figure reveals a clear positive trend (blue series), suggesting that common criteria are considered by both the CNN and the trained human eye. Even for very high $p(x)$, the agreement fraction is only $\sim$ 50 per cent, highlighting the need for visual inspection of the CNN-identified sample in order to increase purity. Globally, the two classification systems agree that more than a third of the 2000 galaxies inspected are post-mergers. However, the visual classification exercise suggests that the network is less successful in identifying post-mergers outside of its mass training regime in IllustrisTNG, between $10^{10} - 10^{12} M_{\odot}$. The stellar mass criteria were selected to ensure that the simulated training galaxies are well-resolved, with at least $\sim 10^4$ star particles. Perhaps due to this deliberate omission in training, $\sim$ 90 per cent of galaxies with masses $< 10^{10} M_{\odot}$ over the entire range of CNN $p(x)$ are not confirmed as post-mergers. Since the success of a CNN is limited by the quality and range of the training data, it is unsurprising that the network does not make accurate classifications when it is forced to extrapolate in this way. Conversely, it is not a detriment to the purity of the final post-merger sample that galaxies with masses $< 10^{10} M_{\odot}$ are included in the sample, since any correctly classified post-mergers below this mass threshold have undergone the same rigorous visual inspection as their higher-mass counterparts in order to be included.

Galaxies are most commonly rejected during visual classification due to the presence of close double nuclei that only revealed themselves under high-contrast inspection. A number of flocculent barred spiral galaxies with a prominent ring (of approximate Hubble type SBc) are also found to contaminate the CNN-predicted sample. Less commonly, galaxies are rejected due to an obstructive survey artifact, an imaged interacting pair in either the CFIS cutout or in a wider field of view, or a lack of sufficiently strong merger-induced features. If the visual classifications are taken as true, I can quantitatively assess the performance of the CNN as a tool for post-merger sample distillation. If post-mergers constitute $\sim$ 0.55 per cent of the galaxies in the low-redshift Universe (Patton et al. 2000), the CNN-identified sample of galaxies with $p(x) > 0.75$ is $\sim$ 64 times as rich in post-mergers as the original, unprocessed galaxy sample[3]. Pre-filtering by $p(x)$ is therefore an efficient first step in post-merger identification.

[3] 0.55 per cent is the estimate of the merger remnant fraction over a single merger timescale, while many merger timescales have elapsed since z=0.5. In practice, the true post-merger fraction in the sample may therefore be higher.

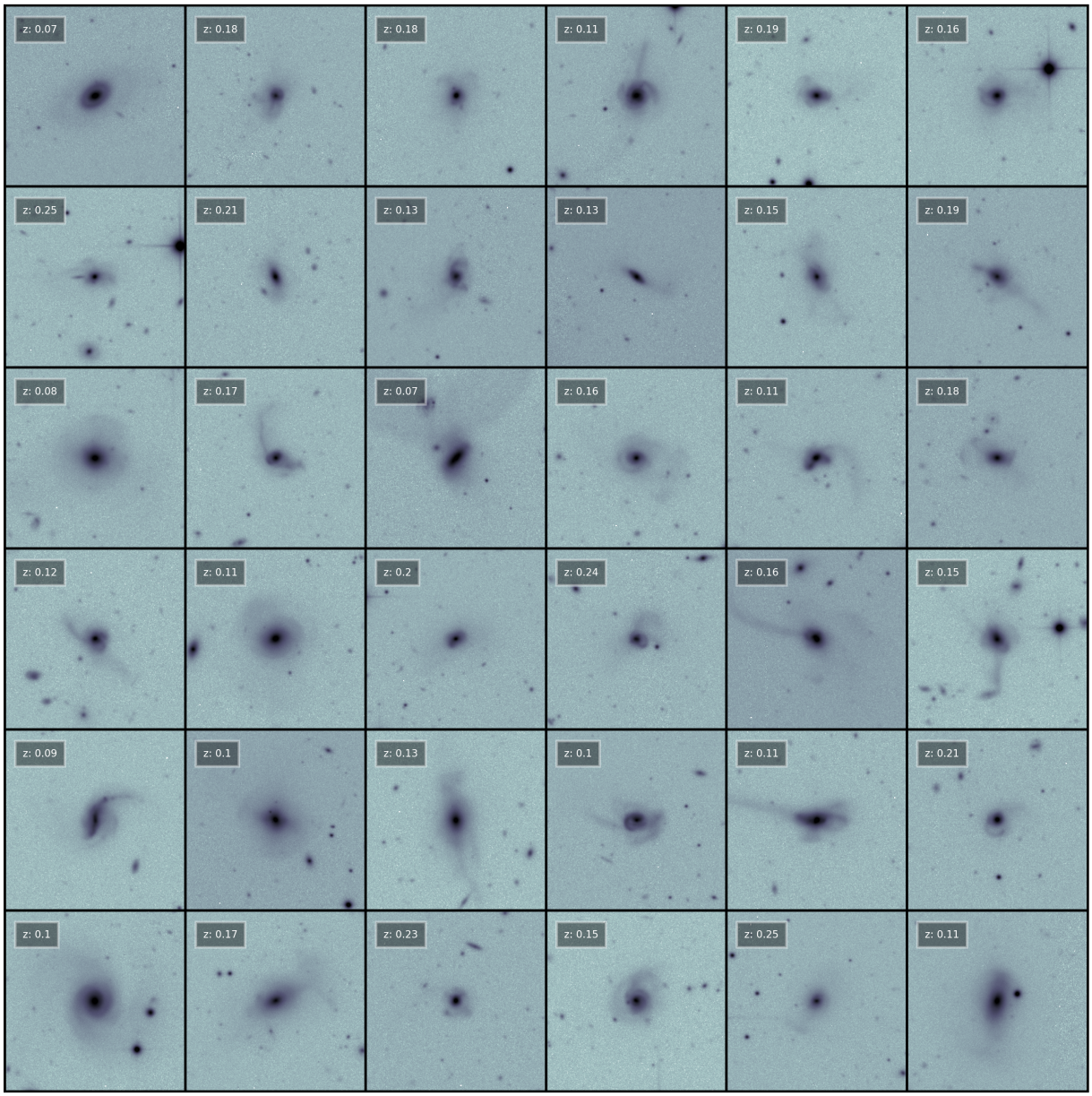

Figure 3.5: A mosaic of 36 randomly chosen post-mergers from the 699-galaxy hybrid (CNN-identified and visually confirmed) post-merger sample, cropped to a physical size of 100 kpc on a side. All galaxies in the sample have distinctive merger-induced morphologies that could not be plausibly produced by any discernible object in the CFIS imaging, or by any spectroscopic companions as identified using SDSS. Furthermore, each of the post-mergers has only a single post-coalescence bright nucleus. Galaxies are shown in log-scale with the contrast adjusted for visual consistency.

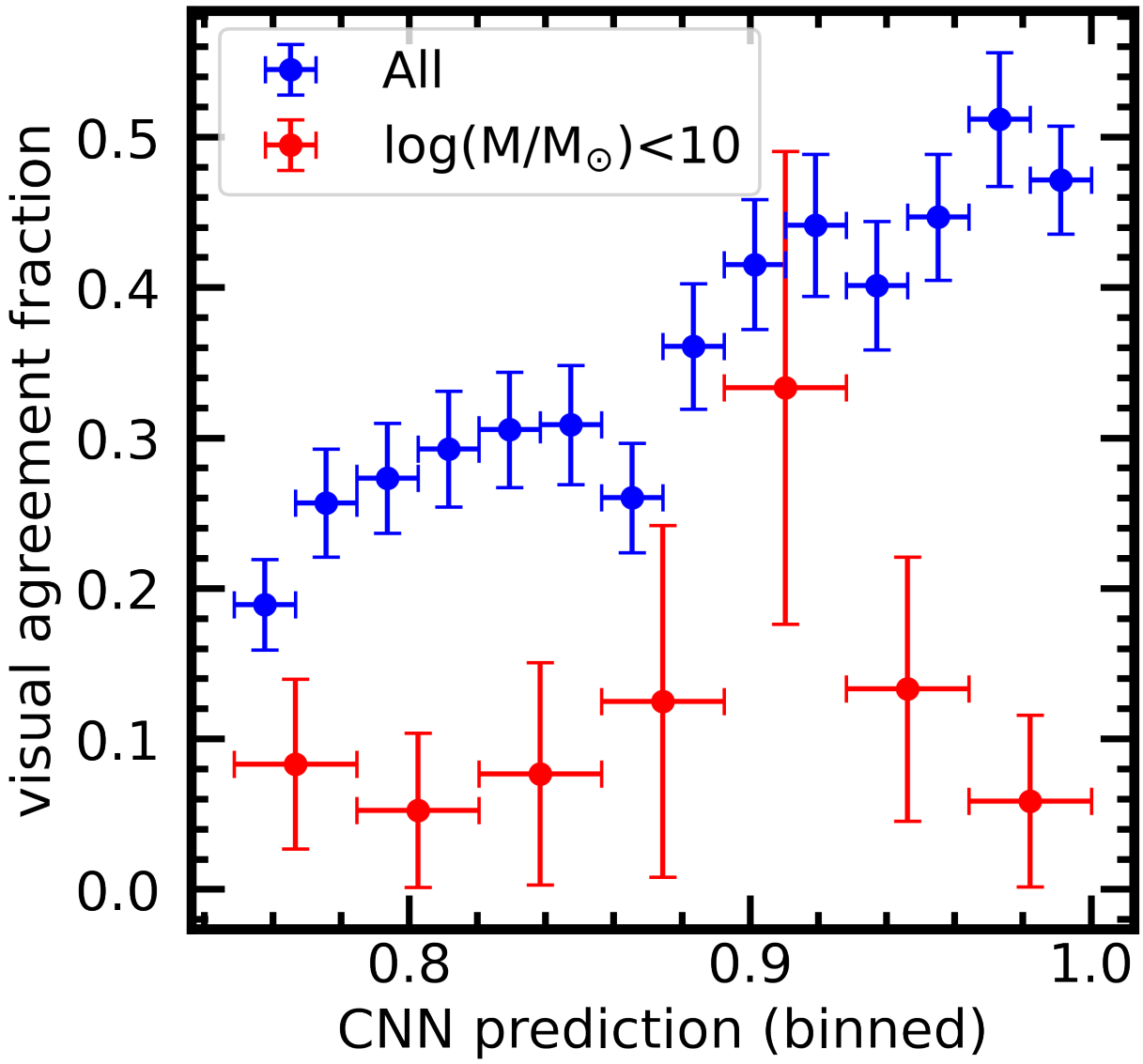

Figure 3.6: The fraction of galaxies labeled as post-mergers during visual inspection in bins of CNN $p(x)$ (blue), and the same for the subset of CNN-predicted post-mergers with masses less than $10^{10}M_{\odot}$ (red). In general, galaxies given very high $p(x)$ values by the CNN are the most likely to have been assigned post-merger labels by the authors. That these quantities trend positively implies that there is a meaningful connection between the criteria for post-merger status for visual inspection and the CNN. The agreement fraction is consistently low for the subset with masses less than $10^{10}M_{\odot}$, suggesting that the CNN has a higher error rate outside of its mass training regime (10^{10} – $10^{12}M_{\odot}$). Vertical error bars are the binomial error in each bin given by $\sqrt{f * (1-f)/N}$ where f is the agreement fraction, and N is the number of galaxies in the bin. Horizontal errors are the bin widths.

After inspection by both the CNN and visual classifiers, galaxies in this work belong to one of three categories:

- galaxies given positive post-merger classifications $p(x) > 0.75$ by the CNN and positive classifications by the visual classifiers (699),
- galaxies given positive post-merger classifications by the CNN and negative classifications by the visual classifiers (1301), and
- galaxies given post-merger classifications $p(x) < 0.75$ by the CNN and which are not inspected visually (166597).

I can also define a more useful sub-category of the third group, for galaxies given classifications of $p(x) < 0.1$ by the CNN, of which there are 131168. Because the CNN is well-calibrated and trained on equally sized samples of post-mergers and non-post-mergers, classifications $p(x) < 0.1$ can be confidently interpreted as negative, particularly due to the natural rarity of post-mergers.

Images of sample galaxies from the three main categories are shown in Figure 3.7. While varying contrast levels are useful in confirming the presence of merger signatures, the confirmed post-mergers in the top row all have features that are visible with consistent, uniform scaling. In the middle row of Figure 3.7 (rejected galaxies predicted to be post-mergers by the CNN), some of the features that deluded the CNN are visible. Notably, while several systems in this category appear to be interacting or pre-coalescence, it is impossible to determine by eye whether these galaxies are truly post-mergers that happen to be experiencing a second interaction. Galaxies that received negative predictions $p(x) < 0.1$ (bottom row) tend to have morphologies consistent with settled disks or undisturbed ellipses.

That nearly two thirds of the post-mergers with CNN $p(x) > 0.75$ are eliminated after visual inspection suggests that the CNN produces a significant number of false positives (non-post-merger galaxies erroneously classified as post-mergers). Having completed visual classifications of all galaxies above this $p(x)$ cut, I can speak qualitatively to the potential reasons for suspected false positive classifications. Most commonly, the network appears to select galaxies with one or more characteristics that mimic the appearance of genuine post-mergers. For example, a number of pre-merger interacting galaxies with extended tidal features are selected by the CNN, but rejected during visual classifications due to being (likely) pre-, rather than post-mergers. Such mistakes could be trained away, by (for example) combining the original CNN with another one designed to identify galaxy pairs (See Figure 2.9; Ferreira et al. 2024). At the time of completing this work, such efforts would have offered limited utility due to the relatively small number of predicted post-mergers and the expense of training additional neural networks. In addition, if

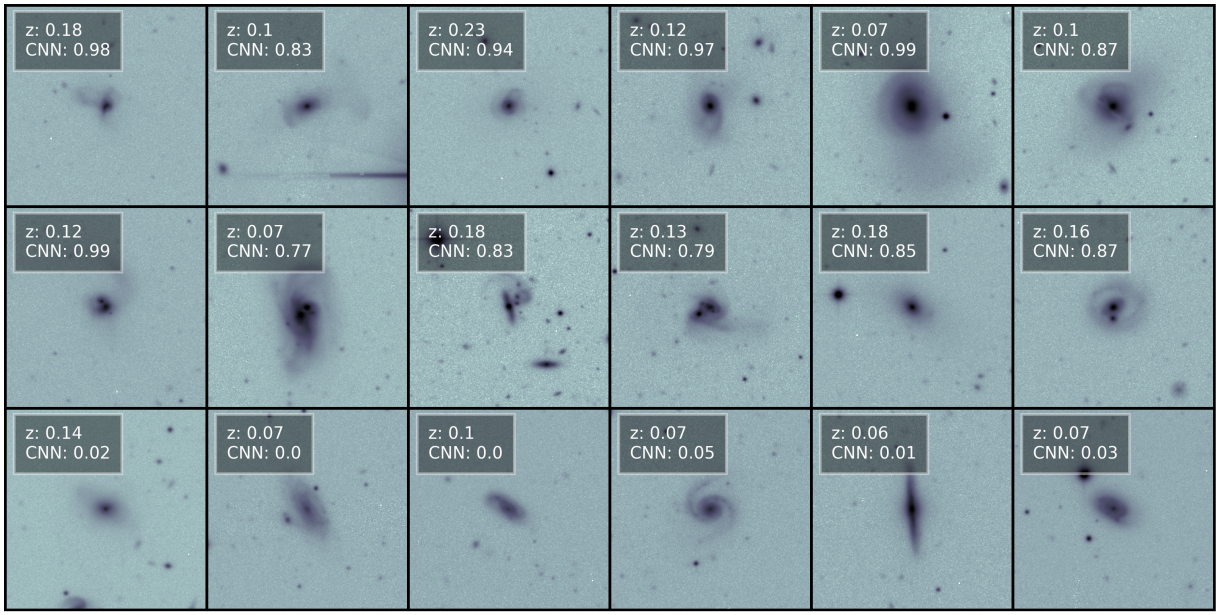

Figure 3.7: CFIS galaxies belonging to each of three categories relevant to this work. Top row: a selection of galaxies with positive $p(x) > 0.75$ post-merger predictions by the CNN, and which are confirmed as post-mergers by eye. Merger features in certain galaxies are sometimes more visible upon closer inspection in CFIS imaging. Middle row: Galaxies with positive post-merger predictions by the CNN which have been ruled out upon close inspection by eye. Many of these are rejected due to a double nucleus or close companion. Bottom row: Galaxies with negative post-merger predictions $p(x) < 0.1$ by the CNN, and which are not inspected visually. These galaxies are eligible to be used as controls in my subsequent studies, which aim to characterize the impact of the post-merger epoch.

the pair classification CNN were any less than 100 per cent accurate in practice, there would be post-merger loss associated with network combination.

Less commonly, the CNN predicts $p(x) > 0.75$ for galaxies of unconvincing post-merger status (e.g., Figure 3.7, middle row, fifth column). While the practical origin of these predictions is unknown, it is possible that a number of post-mergers exhibiting weaker merger characteristics were shown to the network during training, in spite of my prescribed selection of major mergers (with stellar mass ratios of at least 1:10) for training in IllustrisTNG100-1. The orbital parameters for a given galaxy merger can produce a range of visual strengths even at a fixed mass ratio. Any galaxies removed due to unconvincing strength are likely a consequence of the appearance of genuine post-mergers in the training set.

CFIS post-merger characteristics

Statistically, the visually confirmed sample is a reasonably consistent subset of the CNN-identified sample (magenta histograms in Figure 3.3), but biases are necessarily imposed by the visual classification exercise. Disturbed systems with a plausibly interacting companion are always removed from the post-merger sample, even though the presence of a companion does not preclude the possibility of post-merger status. As a result, the visually confirmed post-mergers exist in relative isolation compared to the parent sample. The visually confirmed sample is also likely biased towards more major and disruptive mergers, as a number of CNN-predicted post-mergers are excluded due to the ambiguous (or unconvincing) strength of their tidal features. While the sample of 699 post-mergers is certainly not free of biases, I argue that it is likely representative of the remnants of bright, low-z major mergers. Moreover, the post-merger status of every galaxy in the visually confirmed sample is firmly defensible. The visually confirmed CFIS post-merger sample is available in full as a digital data release with Bickley et al. (2022)[4], and the first ten galaxies in the sample are shown in Table 3.1.

I also compare the characteristics of the visually confirmed, larger post-merger sample to those of the Ellison et al. (2013) post-mergers, which were visually selected from the larger Darg et al. (2010) post-merger catalogue by the author SLE. Though lesser in number (97 in the sample before any experiment-specific cuts are applied), they provide useful context: they represent a highly pure sample as well, albeit one that is likely biased towards mergers of greater visual strength due to their selection from shallower SDSS imaging. Figure 3.8 shows the stellar masses and redshifts of the new CFIS post-merger sample and the Ellison et al. (2013) post-merger sample compared to the sample of SDSS DR7 galaxies appearing in CFIS DR2. Both merger samples span

[4]The full catalogue can be found here: https://academic.oup.com/mnras/article/514/3/3294/6598842#supplementary-data

SDSS DR7 Obj. ID	RA	DEC	z	$M_\star$, $\log(M_\odot)$	SFR, $\log(M_\odot/\mathrm{yr})$	m_r	CNN $p(x)$
587725469589831827	118.57040185	40.12469876	0.0688	11.15	−0.69	15.32	0.868
587725469593501956	125.45850409	47.01624984	0.0748	11.01	−0.62	15.84	0.897
587725471203131595	122.13330213	45.89220975	0.1432	11.17	1.70	17.16	0.999
587725489990402472	258.95204855	54.28300312	0.2154	11.47	−0.38	17.53	0.855
587725491063095592	259.24312269	56.84922519	0.2361	11.45	0.70	17.36	0.955
587725552268738844	118.81820385	45.28471333	0.0503	11.42	0.26	15.36	0.790
587725775067152792	115.14104706	40.56706267	0.1523	10.96	−0.29	17.25	0.863
587725981224468549	123.58884576	49.84016841	0.1691	11.19	−0.08	17.39	0.953
587725981227221163	130.23160138	54.7043356	0.1710	10.83	0.51	17.57	0.978
587725981763371221	127.74461264	53.85535064	0.0631	10.98	−0.61	15.70	0.909

Table 3.1: The first 10 objects in the hybrid visual-CNN CFIS post-merger catalogue. The entire post-merger catalogue is available as digital supplementary material with Bickley et al. (2022). For consistency with , Bickley et al. (2022), I present stellar masses and star formation rates the MPA-JHU (Brinchmann et al. 2004) catalogue for SDSS DR7 in this table and Table 3.2. The Petrosian r-band magnitudes are calculated from SDSS DR7 imaging.

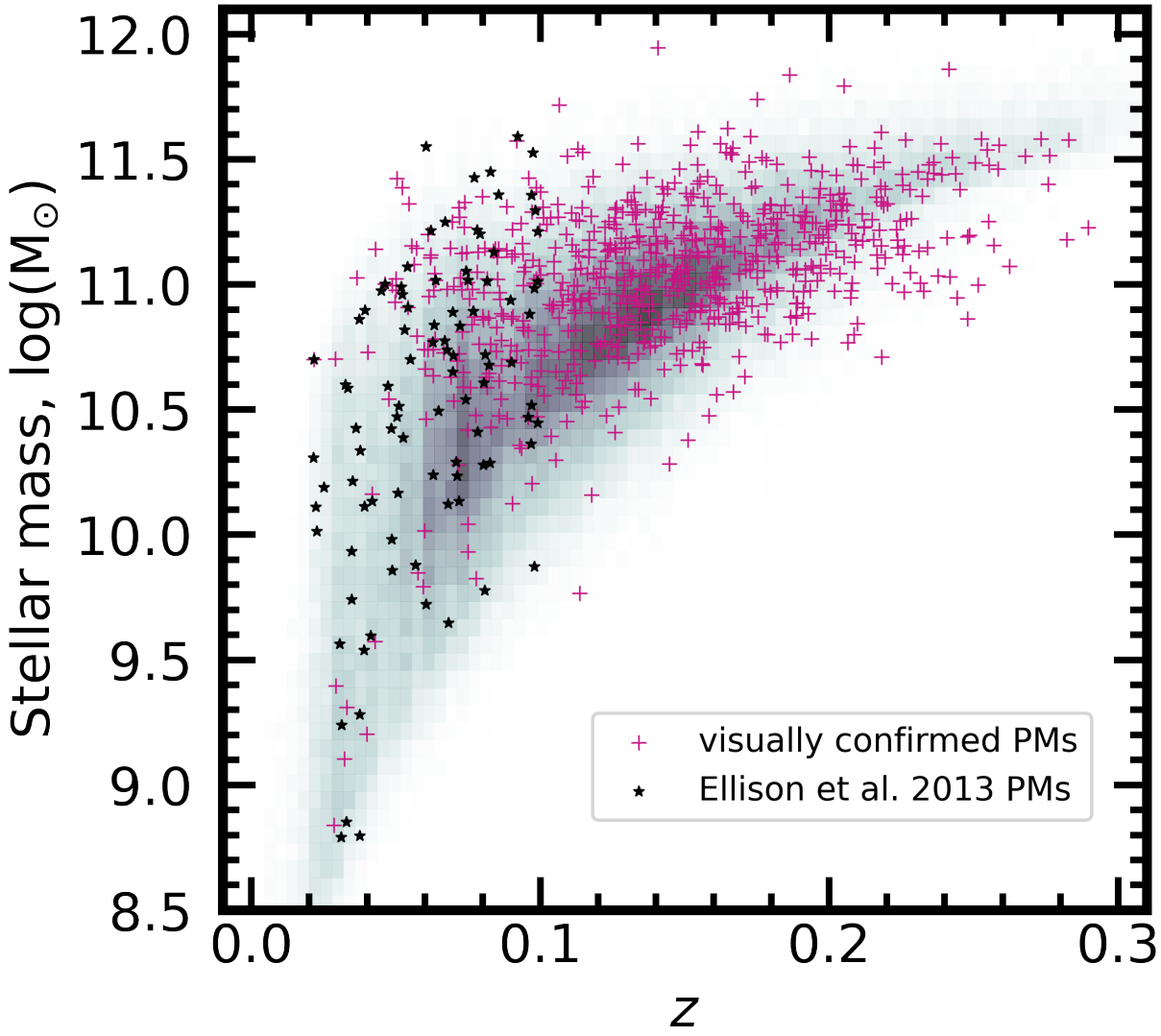

Figure 3.8: The stellar masses and redshifts of the 699 visually confirmed post-merger galaxies (magenta crosses) found in CFIS DR2 and 97 Ellison et al. (2013) post-mergers (black stars) superimposed over the sample of CFIS DR2 galaxies appearing in SDSS DR7. CFIS *r*-band imaging is processed and used as the input for CNN classification as well as for visual inspection in this work.

the dynamic range of CFIS DR2 galaxy stellar masses, but the new visually confirmed post-merger sample identifies more distant post-merger galaxies out to $z \sim 0.3$. The Ellison et al. (2013) post-mergers lie below $z < 0.1$, since the Darg et al. (2010) catalogue from which they were selected had an upper z limit of 0.1. The success of the CNN at identifying galaxies over the full CFIS redshift range likely stems from its training set, which is composed of simulated galaxies inserted at redshifts drawn at random from the real CFIS distribution.

3.2 Data and methods: DECaLS

Later, in Chapter 6, I will present the results of a study of X-ray-bright AGN in the merger sequence, focusing particularly on new analyses made possible by the eROSITA X-ray space mission. The availability of data from the German consortium within eROSITA overlaps poorly with CFIS (see Fig. 2.2). It is also of general interest to construct a post-merger catalogue that extends further south on the celestial sphere. In this Section, I investigate the plausibility of a merger search in DECaLS using the CNN model I have already trained for the task of merger identification in CFIS. A broader and deeper study of the potential of cross-survey inference is also presented in Chapter 7. After demonstrating that post-mergers can still reliably be identified by the model in DECaLS imaging, I also present the results of the post-merger search in DECaLS.

3.2.1 Preparing the DECaLS data for classification

In order to maximize the scientific impact of the post-merger search in DECaLS, I only attempt to classify galaxies meeting the following criteria:

- Available in SDSS DR7 with a redshift, and assigned a spectroscopic class equal to 2 (galaxy), 3 (quasar), or 4 (high-z quasar) by the SDSS spectral classification pipeline.
- $z > 0.005$ to avoid any Galactic objects misclassified by the SDSS pipeline, and $z < 0.3$ to limit the study to the z domain where merger classifications are homogeneously reliable (see Figure 2.10)
- Within the eROSITA (German consortium[5]) footprint with Galactic longitudes between $180° < l < 360°$.

[5]The German consortium has access to half of the all-sky survey, with the Galactic longitudes given in the text marking the boundary.

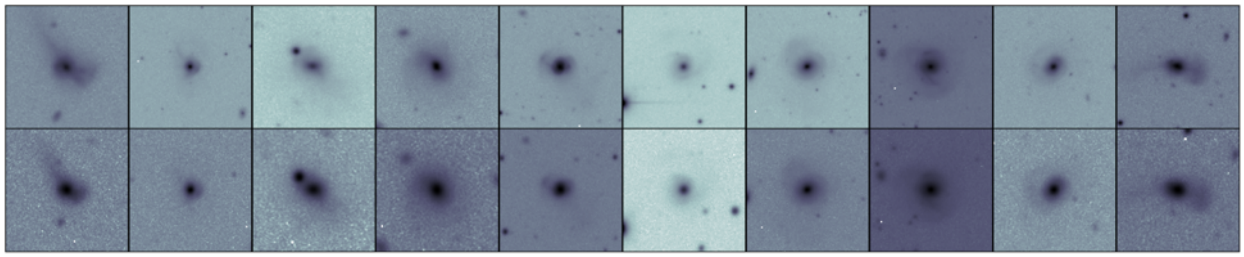

Figure 3.9: Select galaxies from the 699-object sample of post-mergers identified in CFIS imaging, as imaged by both CFIS (top row) and DECaLS (bottom row). While the DECaLS image quality is shallower and lower-resolution, it is generally still possible to see that the galaxies are post-mergers in the DECaLS imaging.

- Dark Energy Camera Legacy Survey (DECaLS, Dey et al. 2019) *r*-band imaging available, taken with the Dark Energy Camera for the Legacy Survey[6]. DECaLS imaging is used in this work to search for the morphological signatures of a recent merger event.

There are 265371 galaxies meeting these criteria. Post-mergers in DECaLS are also identified using the hybrid classification scheme, first using the exact CNN model and weights as in Section 2.2.2 to classify galaxies in DECaLS *r*-band images. DECaLS is not as deep (with a 5σ *r*-band point-source depth of 23.9 mag compared to CFIS's 25.0 mag) and has coarser pixel-scale resolution by a factor of ~ 2 compared to CFIS. The DECaLS pixel scale also has the effect of re-binning the signal and noise in the images, resulting in higher effective surface brightness sensitivity in DECaLS compared to CFIS (by about $\sim$ 40 per cent). It is therefore necessary to test whether the original CNN is appropriate to use on DECaLS imaging. I therefore prepare 100-kpc-square cutouts of the subset of 16025 galaxies (69 of which are post-mergers in the CFIS catalogue in Section 3.1.3) imaged by both DECaLS and CFIS in the *r*-band in the thin stripe of declination between 30° and 32.375° shared by both surveys, normalized both sets of images using the method in Section 2.1.3, and compare the CNN's classifications between the two surveys.

Images of selected galaxies from the CFIS post-merger sample (Section 3.1.3) viewed by both surveys are shown in Figure 3.9, with the CFIS images on the top row, and the DECaLS underneath. The difference in quality between the two surveys is apparent, with the CFIS images being deeper (lower levels of noise relative to the light from the galaxy) and higher resolution. The difference in PSF between the two surveys is also apparent, with point-like sources in the DECaLS imaging being blended outwards about twice as much. Encouragingly, the characteristically disturbed post-merger morphologies seen in each of the CFIS images are generally still visible in their DECaLS counterparts. The qualitative (if not quantitative) similarity in the appearances of

[6] www.legacysurvey.org/dr10/description/

galaxies between the two surveys suggests that it may be reasonable to apply the CFIS-trained model to DECaLS imaging.

The statistics presented in Figure 3.10 informed my approach to the DECaLS post-merger search. Galaxies in the CFIS-DECaLS overlap at large (black histogram, top panel), and post-mergers from the visually confirmed CFIS sample (blue histograms), receive very similar classifications in CFIS and DECaLS. However, a disproportionate number of galaxies receive lower $p(x)$ predictions from the CNN in DECaLS imaging compared to CFIS. In the bottom panel of Figure 3.10, I show that only would only recover $\sim$ 59 per cent of the visually confirmed CFIS post-merger sample if I were to use the same visual inspection criterion, $p(x) > 0.75$, in the DECaLS imaging as I did in CFIS (grey crosshairs). Using a lower CNN $p(x)$ criterion of 0.6 to select post-mergers in DECaLS imaging, I find that I recover $\sim$67 per cent (47 galaxies) of the visually confirmed post-mergers from CFIS imaging in the survey overlap (magenta crosshairs). I also find that a significant number of new visually convincing post-mergers (some 30 galaxies in the overlap) are recovered in DECaLS imaging that are not found by the CNN in the CFIS imaging. Mergers "discovered" in DECaLS and not in CFIS by the same CNN are sometimes due to CFIS imaging artefacts (which are often absent in DECaLS), but in other cases the reason for a given merger's new classification in DECaLS is not visually obvious. While the choice of $p(x) > 0.6$ to identify the sample for visual inspection is arbitrary, my aim is to identify a similarly large sample of visually-confirmed post-mergers in DECaLS compared to CFIS, and the statistics of the cross-survey validation experiment shown in Figure 3.10 indicate that $p(x) > 0.6$ will allow me to do so.

Figure 3.11 shows a selection of CFIS-identified post-mergers that were not recovered in DECaLS imaging (top two rows), galaxies flagged by the CNN for inspection in DECaLS but not in CFIS (middle two rows), and several CFIS-identified post-mergers successfully recovered in DECaLS imaging (bottom two rows). CFIS post-mergers not recovered in DECaLS imaging can be a result of DECaLS imaging artefacts, or the fainter appearance of low-surface brightness structure in the shallower imaging. Still other post-mergers may be missed in DECaLS imaging due to stochastic effects embedded in the hidden layers of the CNN. Some galaxies in Figure 3.11 show no clear reason for having been misclassified in DECaLS. Galaxies in the middle two rows of Figure 3.11 (flagged as mergers in DECaLS imaging but not in CFIS) have a wide range of characteristics. Some (e.g., third row, sixth column) appear to be genuine post-mergers which might have been missed in CFIS imaging due to an artifact or stochastic error by the CNN. Others appear to be pre-coalescence galaxy pairs (third row, fourth column) or other non-post-merger galaxy types, which are excluded later during visual classification. The bottom two rows of Figure 3.11 verify again the conclusion suggested in Figure 3.9: that in many cases, the unique morpholo-

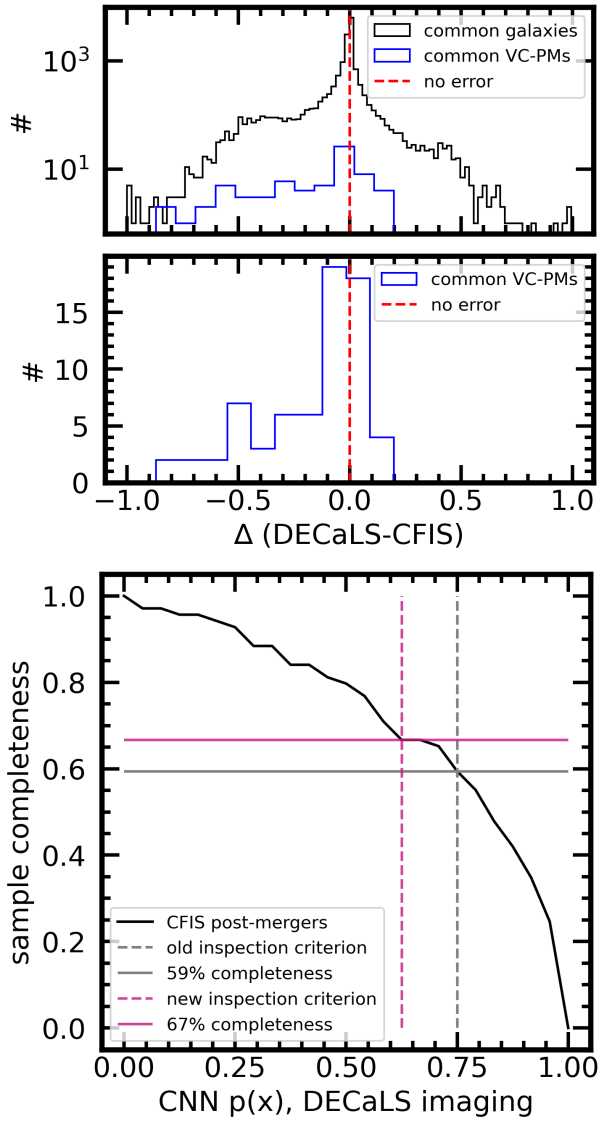

Figure 3.10: Comparing the CNN's predictions on the 16025 galaxies appearing in both DECaLS and CFIS. Top frame, top panel: a log-scale histogram showing the difference between the $p(x)$ values returned by the CNN for individual galaxies appearing in both DECaLS and CFIS. Negative values indicate that the CNN's prediction was lower on the DECaLS image than the CFIS image. The black histogram shows the results for all galaxies in the overlap, while the blue shows the results for galaxies belonging to the post-merger sample identified in CFIS imaging (see Section 3.1.3). While the distribution of differences peaks at zero (i.e., most galaxies receive similar predictions), more galaxies have negative deltas than positive deltas. Post-mergers also are more likely to have negative than positive deltas (highlighted in linear scale, second panel). Bottom frame: the fraction of the visually confirmed CFIS post-mergers (black curve) recovered in DECaLS imaging as a function of CNN $p(x)$. The grey and magenta crosshairs correspond to the $p(x)$ inspection criteria used in this work for CFIS and DECaLS images, respectively.

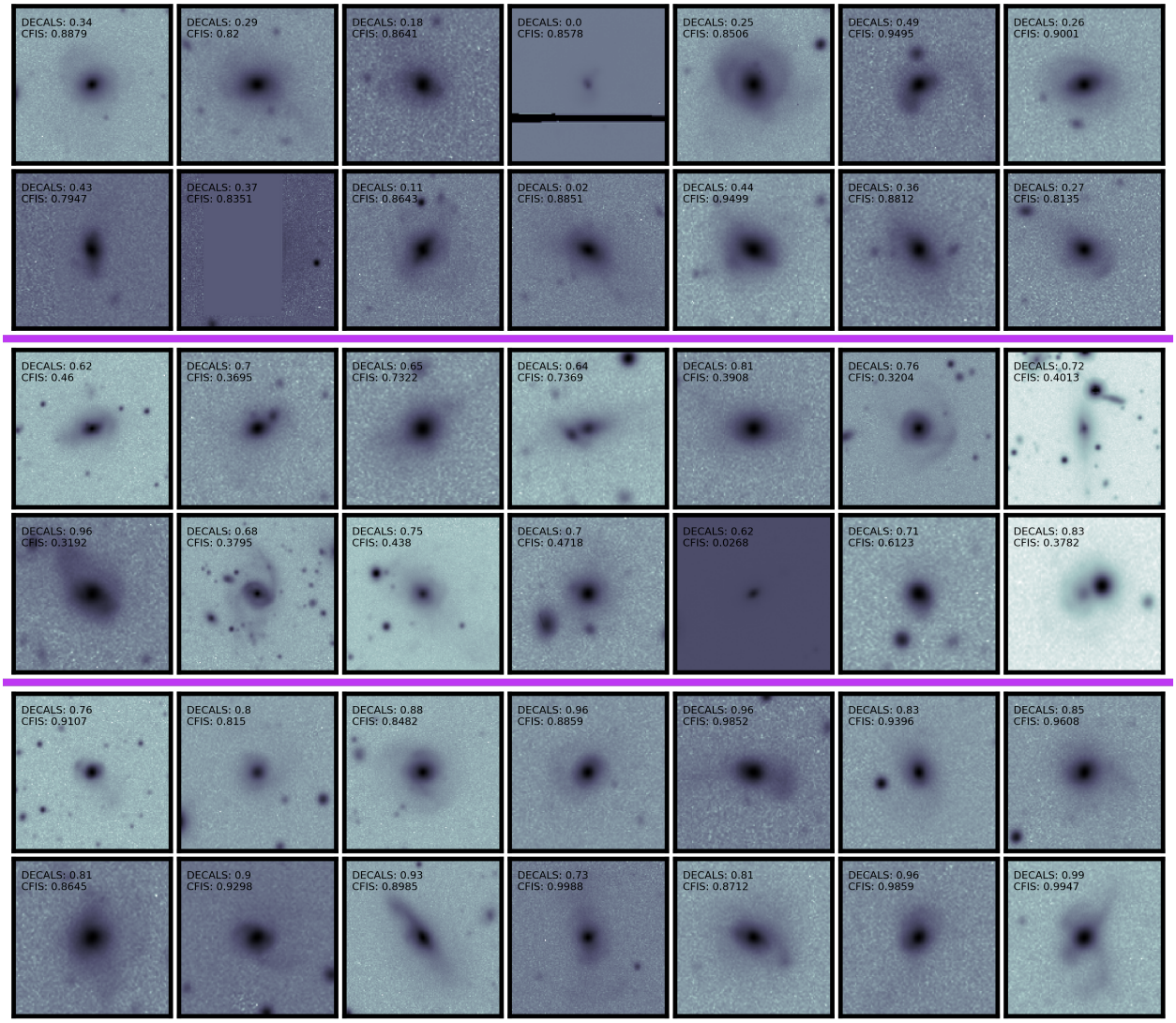

Figure 3.11: DECaLS images of galaxies that serve as case studies in cross-survey CNN classification. Top two rows: galaxies that belong to the visually confirmed CFIS post-merger sample, but which received CNN $p(x)$ predictions < 0.5, or confidently negative. All fourteen galaxies meeting these criteria are shown. In many cases, merger features can still be seen, but in two (fourth column in the first row, and second column in the second row) an imaging artifact in the DECaLS imaging is presumably responsible for the loss. Middle two rows: galaxies that meet the visual inspection criteria in DECaLS with CNN $p(x) > 0.6$, but were not flagged for inspection during the CFIS visual classification experiment, i.e., with CFIS $p(x) < 0.75$. CFIS artefacts may be to blame, and in such cases the recovery of new, genuine post-mergers is possible. Bottom two rows: galaxies belonging to the CFIS visually confirmed post-merger sample that were also recovered with CNN $p(x) > 0.6$ in DECaLS. All panels are annotated with the specific CNN classifications in both CFIS and DECaLS.

gies that have allowed the CNN to identify post-mergers in CFIS are frequently still visible and detectable by the CNN in DECaLS.

Figure 3.12 shows the same galaxies as Figure 3.11, but in CFIS imaging instead of DECaLS. The two mosaics are created with identical scaling, but the improved depth of CFIS relative to DECaLS results in better contrast between the galaxies (especially low-surface brightness structure) and the background. The images in the top two rows of Figure 3.12 demonstrate that CFIS quality imaging is beneficial to the recovery of identifiable merger morphology in many cases. In the middle two rows (galaxies not inspected during the CFIS post-merger identification effort, but with $p(x) > 0.6$ in DECaLS), some galaxies (e.g., third row, first and second panels) do have imaging artefacts in CFIS. In other cases (e.g., third row, sixth panel) the reason for a galaxy's low $p(x)$ prediction in CFIS is not obvious. The relatively clear and visually dramatic post-merger morphologies seen in the final two rows of Figure 3.12 have apparently ensured their identification in shallower DECaLS imaging. Still, a number of comparably obvious post-mergers have changed status between the two classification efforts (see again the first four panels of Figures 3.11 and 3.12).

The above experiment demonstrates two relevant effects. First, merger recovery is somewhat stochastic and not simply a function of image depth (stochastic recovery is explored in detail in Ferreira et al. 2024, in which multiple networks are tested in a jury system). Second, I expect that a statistically compelling sample of mergers can be identified from the DECaLS imaging using the CNN I already have in hand. I emphasize that completeness is relatively unimportant for this work, but purity is. I aim for a pure post-merger sample because my goal is to characterize the impact of coalescence in a bona-fide sample of mergers, rather than collect a complete sample of mergers. It is also important that the mergers identified in DECaLS are reasonably representative of the post-merger class, even though they have been identified in shallower imaging than the CFIS post-merger sample. It is encouraging that the final DECaLS post-merger sample (discussed later in Section 3.2.2) has similar $M_\star$ and z statistics to the CFIS sample (see Figure 3.2), but potential biases are discussed in Section 3.2.2 and a methodological study of biases as a function of imaging quality is presented later in Chapter 7.

Having demonstrated that the already-trained CNN can be used to recover a post-merger sample in DECaLS imaging, I next produce 100-kpc-square DECaLS-r cutouts of the $\sim$ 265000 galaxies in the sample co-included in SDSS DR7, DECaLS, and eROSITA, resize them to 138×138 pixels, normalize them in linear fashion between $0-1$, and classify them with the CNN. As with CFIS, some images contain artefacts (e.g., saturated foreground stars, missing pixels) that partially or completely mask the morphology of the galaxy. I classify these objects anyway, since the CNN's training set includes images with artefacts, and there will be an opportunity to remove

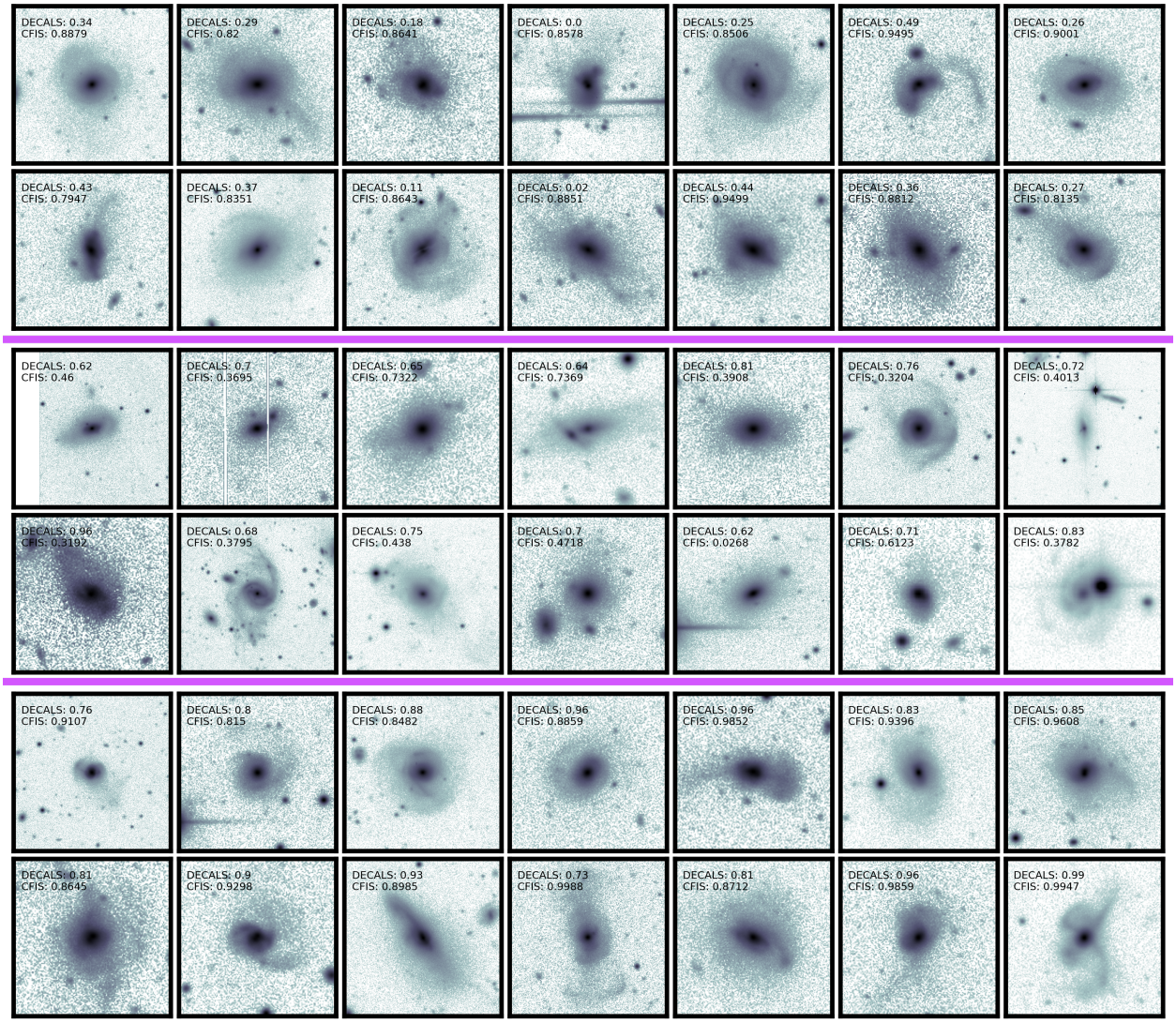

Figure 3.12: CFIS images of the exact same galaxies, and with identical scaling, as in Figure 3.11. The top two rows are visually confirmed CFIS post-mergers with CNN $p(x) < 0.5$ in DECaLS, the middle two rows are galaxies with CNN $p(x) > 0.6$ in DECaLS but $p(x) < 0.75$ in CFIS, and the final two rows are agreed upon (visually confirmed CFIS post-mergers with DECaLS CNN $p(x) > 0.6$).

any objects with prominent artefacts masquerading as post-mergers in the visual classification stage of the merger search (in practice, such objects are rare).

Finally, I conduct a visual classification effort in order to remove false positives from the sample. 4093 galaxies receive CNN post-merger predictions $p(x) > 0.6$, and they are in turn inspected by me. A consultation for select ambiguous cases is completed (again with DRP and SLE), and the conclusions of the committee are taken into account during a final check of the sample by me. Galaxies are removed from the final DECaLS post-merger sample if they:

- lack visually convincing post-merger morphology, or
- are obscured by one or more imaging artefacts such that the morphology can not be discerned, or
- have a nearby companion in DECaLS imaging that might have been interacting, or
- have two or more nuclei, suggesting pre-merger rather than post-merger status.

3.2.2 The DECaLS post-merger sample

Of the 4093 galaxies inspected, 923 are confirmed as having the unambiguous characteristics of recent coalescence; 100 random examples are shown in Figure 3.13. As with the CFIS post-merger sample, the galaxies in the final DECaLS post-merger sample all have the unambiguous characteristics of recent coalescence. The impact of the visual classification exercise is shown in Figure 3.14, which shows the fraction of galaxies visually confirmed as post-mergers as a function of $p(x)$. As with CFIS in Figure 3.6, I find that the fraction increases as a function of $p(x)$, suggesting that the CNN's certainty of post-merger status for a given galaxy is linked to the galaxy's actual likelihood of being a post-merger. In other words, the model is still well calibrated when applied to DECaLS imaging.

I note that the visual confirmation fraction (23 per cent) from the DECaLS post-merger search is lower than that in CFIS (36 per cent). I attribute the lower confirmation rate to the fact that the signatures of a recent merger are visually weaker (and therefore, less convincing; see Figure 3.9) in DECaLS compared to CFIS due to the $\sim$ 1 mag difference in sensitivity. As a result, a greater fraction of predicted post-mergers had somewhat ambiguous merger features in DECaLS, and are rejected during the visual classification effort. The sensitivity difference between the two surveys likely means that the post-merger remnants selected from DECaLS have somewhat more dramatic morphologies (following a merger with a higher mass ratio or more disruptive orbital parameters, or with less time elapsed since coalescence). Nonetheless, at the end of this process

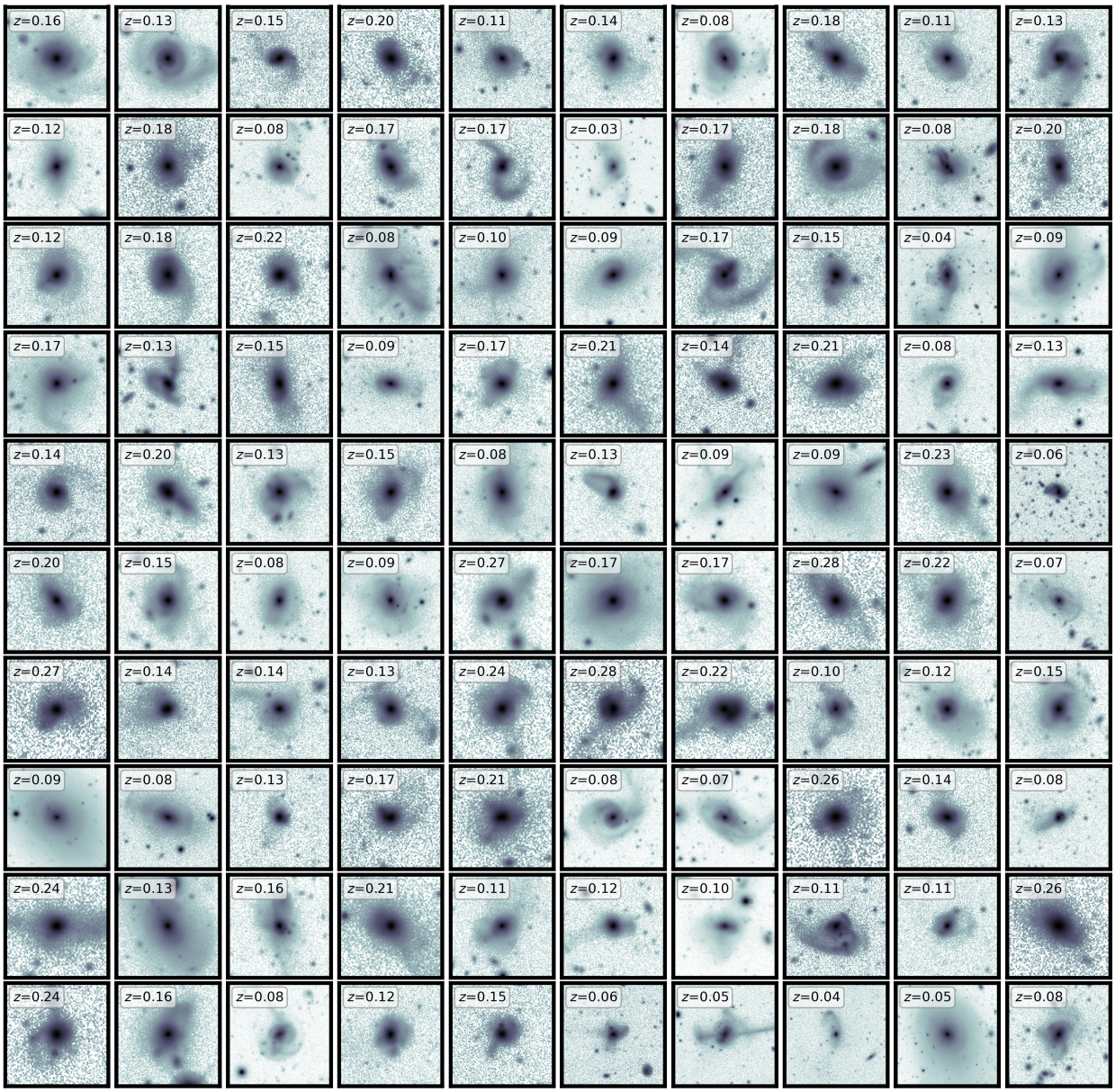

Figure 3.13: Mosaic of 100 *r*-band DECaLS images of post-merger galaxies selected at random from the visually confirmed sample of 923 identified for this work, annotated with the SDSS DR7 spectroscopic *z*. Cutouts are 100 kpc on a side. Images are pre-processed and converted to log scale in order to highlight the low-surface-brightness features that earned them post-merger classifications.

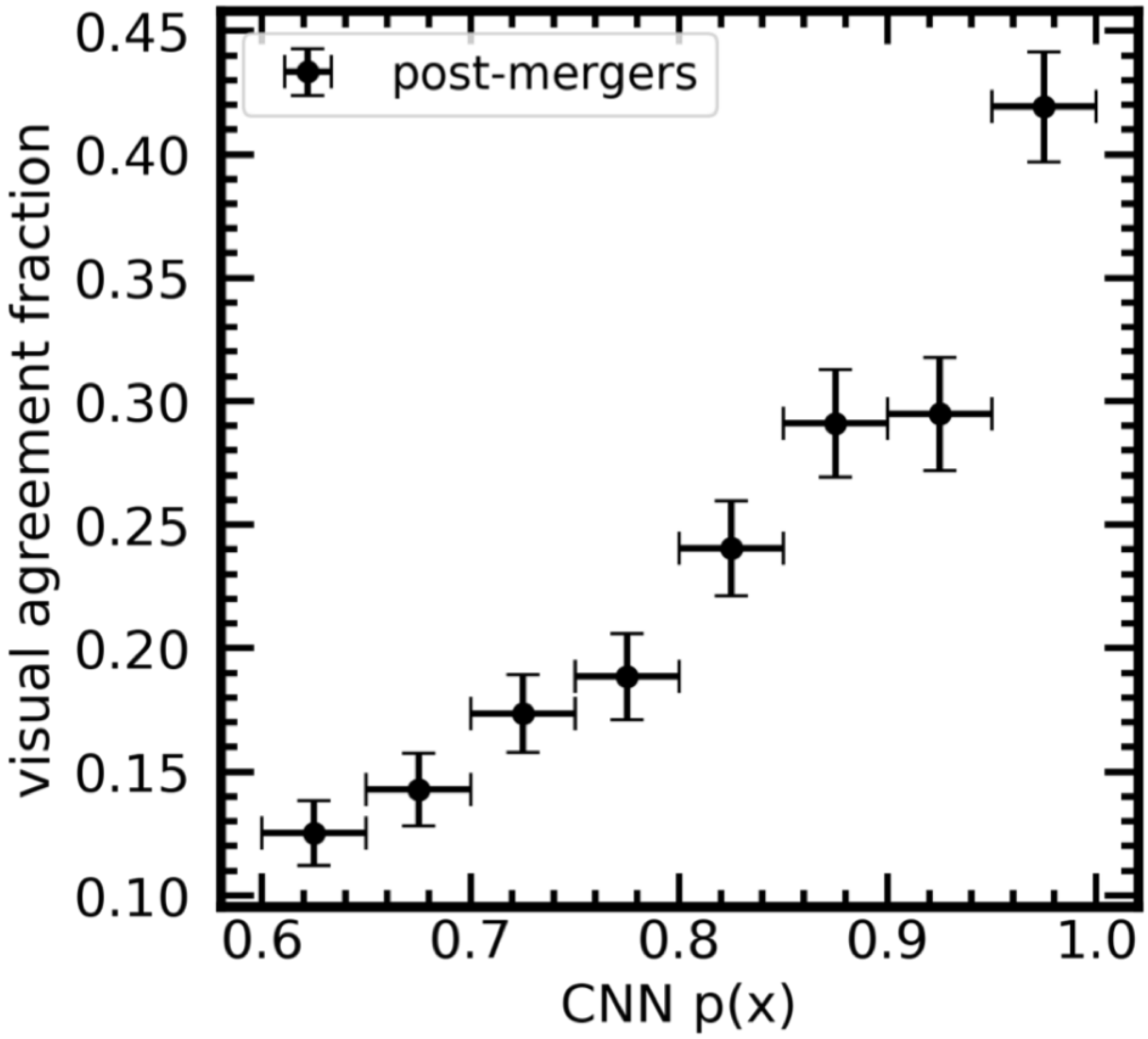

Figure 3.14: The fraction of CNN-predicted DECaLS galaxies confirmed to be post-mergers during visual inspection as a function of CNN $p(x)$. As with CFIS, the fraction increases with $p(x)$, suggesting that the model's predictions are truly indicative of a galaxy's merger status.

I have a second large, pure sample of post-mergers in DECaLS. Additional details of the DECaLS post-merger sample are presented in depth in Chapter 6 in the context of my multi-wavelength AGN study, and the sample is available in full as a digital data release with the forthcoming MNRAS edition of Bickley et al. (2024). The first ten galaxies in the DECaLS visually confirmed post-merger sample are shown in Table 3.2.

Having identified post-merger samples in both CFIS and DECaLS imaging, I can compare the outcomes of the two visual classification exercises. In Figure 3.15, I plot the z (left column) and $M_\star$ distributions of the CNN-predicted post-merger samples for both surveys in the top row. In the middle row, I plot the visual agreement fraction (as in Figures 3.6 and 3.14) against the redshifts and stellar masses of the galaxies inspected, i.e., those with $p(x) > 0.75$ in CFIS or $p(x) > 0.6$ in DECaLS. In the bottom row, I plot the final z and $M_\star$ distributions for the visually confirmed post-merger samples from each survey. The plot is constructed such that the distributions in the top row are convolved by the selection functions in the middle row to produce the final sample statistics in the bottom row. Figure 3.15 reveals that while the two samples are ultimately qualitatively similar, there were minor differences in the way they were selected. Post-mergers are unlikely to be found in either survey at the very lowest redshifts (with $z < 0.02$). In CFIS, the likelihood of selection increases as a function of redshift until $z \sim 0.2$, while in DECaLS, the curve remains comparatively flat with z (bottom left panel). In both surveys, galaxies are most likely to be selected as post-mergers near the middle of the mass training regime of the CNN, with $\log(M_\star/M_\odot) \sim 11$, and there are discrepancies in the visual selection functions at low-$M_\star$. For both mass and redshift, however, the regions of parameter space where the agreement fractions are discrepant – with $\log(M_\star/M_\odot) < 10.5$, $z < 0.1$, or $z > 0.2$ – there are relatively few galaxies flagged by the CNN for inspection to start with. The demographics of both merger samples are therefore fairly consistent with one another.

3.3 Discussion and Conclusions

The hybrid CNN and visual classification approach to merger identification has proven extremely useful in the context of this work, and the merger samples identified also have science potential beyond what is explored here. The main drawbacks of visual classification are efficiency and reproducibility, both of which are alleviated when the CNN is used as an automated preliminary filter.

I argue that the drawbacks to the method used here are inherent to the problem of binary merger identification (i.e., sorting galaxies into one merger class and one non-merger class). The act of assigning merger status on a binary basis is a useful simplification, but in doing so, I have

SDSS DR7 Obj. ID	RA	DEC	z	$M_\star$, $\log(M_\odot)$	SFR, $\log(M_\odot/\mathrm{yr})$	m_r	CNN $p(x)$
587722981746802823	195.333539	−1.192719	0.159501	11.32	−0.18	16.67	0.654
587722981755060409	214.199001	−1.070562	0.125132	11.19	0.30	16.13	0.802
587722982288589146	206.633066	−0.724307	0.148447	10.33	0.66	17.58	0.819
587722982289506481	208.662005	−0.811589	0.198613	10.82	−0.02	17.88	0.850
587722982290030864	209.951318	−0.766162	0.106379	10.89	−0.10	16.57	0.908
587722982292783382	216.223982	−0.692976	0.136630	11.03	−0.33	17.06	0.948
587722982818971825	191.804522	−0.349249	0.082776	11.12	0.53	15.68	0.936
587722982824083639	203.461371	−0.271583	0.176416	11.13	0.17	17.16	0.813
587722982826639582	209.246390	−0.342550	0.105212	10.72	−0.18	16.62	0.845
587722982829785301	216.479687	−0.271547	0.129187	10.90	1.06	16.94	0.858

Table 3.2: The first 10 objects in the hybrid visual-CNN DECaLS post-merger catalogue. The entire post-merger catalogue is available as digital supplementary material with the MNRAS edition of Bickley et al. (2024).

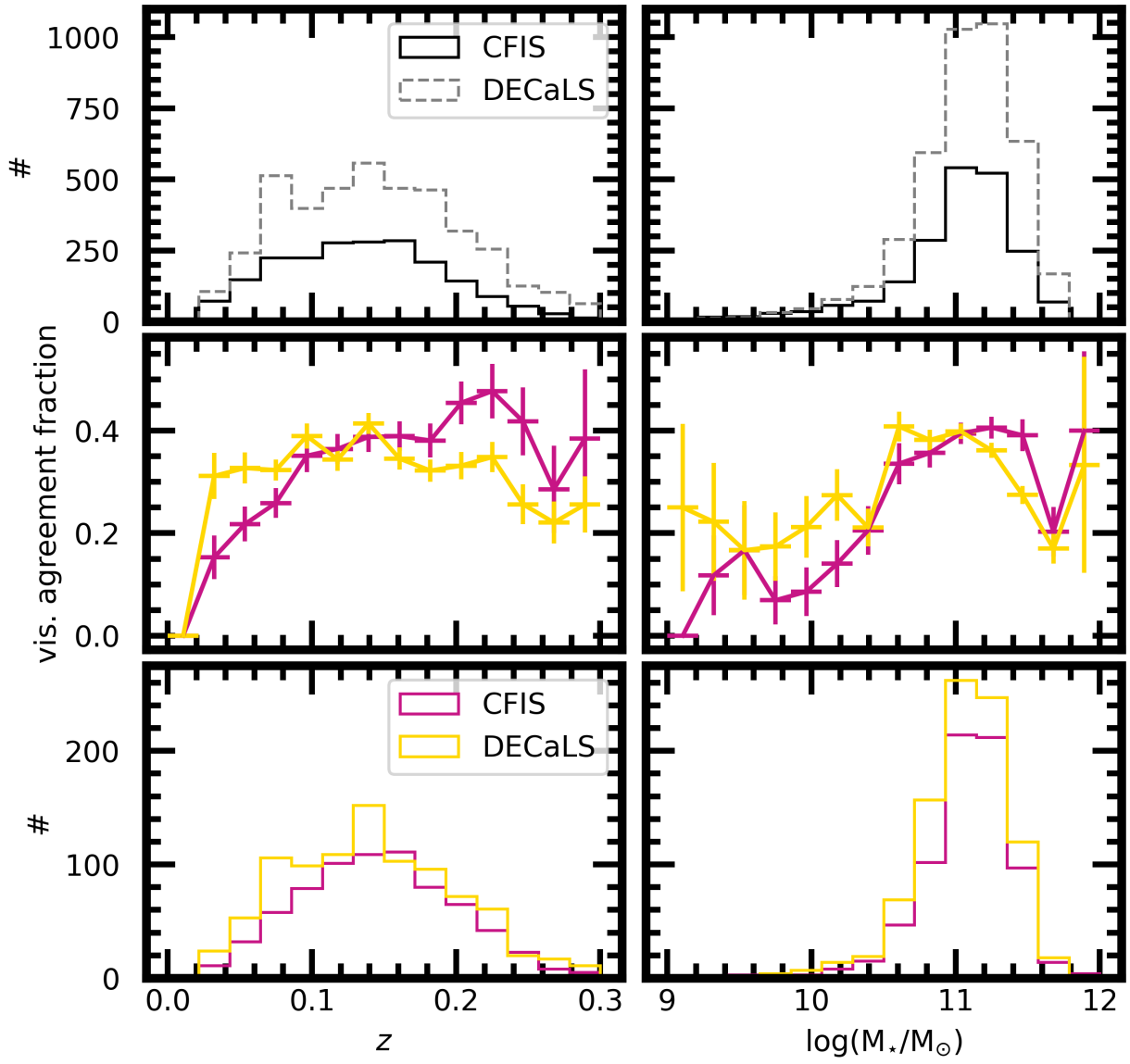

Figure 3.15: The same as for Figure 3.14, but with the agreement fractions plotted against z (middle row, left column) and stellar mass (middle row, right column) for both the CFIS merger identification effort (magenta) and for DECaLS (yellow). The trends suggest that the characteristics of CFIS and DECaLS imaging give rise to post-merger samples with similar, but not identical, demographics. For context, the z and $M_\star$ distributions for the CNN-predicted merger samples are shown in the top row of panels, and both visually confirmed post-merger samples are plotted in the bottom row of panels. The histograms in the top row are convolved by the curve in the middle row to produce the distributions seen in the bottom row.

had to choose stellar mass, temporal, mass ratio, and environmental criteria to select visually unambiguous samples of mergers and control galaxies for training. These choices bear on the final completeness of the merger samples I identify, especially near the "edges" of the criteria choosen for training (e.g., for merger remnants with lower stellar masses as shown in Figure 3.6, or for mergers with stellar mass ratios near the lower limit of 1:10). Future efforts will improve on this work by marginalizing more precisely over parameters like mass ratio and $T_{Postmerger}$, though it is important that some experimental validation of the predictions being made can be carried out. More complex questions about a galaxy's merger history (e.g., predicting a galaxy's merger tree from its current appearance) are increasingly difficult to verify by eye. The underpinning assumptions of SBI for merger characterization must therefore be examined carefully on an ongoing basis.

3.3.1 Current and future work in merger identification

The simulation-based post-merger searches carried out in both CFIS and DECaLS were the main methodological novelty in this work at the time of beginning (ca. 2019). Owing to the pace of progress in machine vision techniques in recent years, as well as the pace of adoption of these methods by the astronomy community, there are already a number of promising extensions of this method at the time of writing. Ferreira et al. (2024) have introduced the multi-model merger identifier (Mummi) machine vision suite, which currently makes three-class merger stage predictions (non-merger, pre-merger/pair, and post-merger) with high accuracy, and elevated completeness compared to the single-model approach used in this work. Mummi is trained on mock CFIS images prepared using the same method detailed in this work. Its elevated accuracy is mainly owed to the size of its training set (meaningfully larger than used here, with 20 mock redshift realizations created for each camera angle for each galaxy), and the use of a jury system, wherein twenty models (10 CNNs and 10 vision transformers, a machine vision method popularized after the identification of the original CFIS post-merger sample detailed in this Chapter) are each trained on a unique subset of the data. Predictions from each of the twenty models therefore reflect different experiences, and the each model has the potential to mitigate the errors made by other models. Subsequent Mummi developments include temporal predictions of merger stage as well, and the feasibility of machine vision methods for such predictions has already been assessed in the literature (see Pearson et al. 2024). At scale, temporal predictions will allow for better characterization of how galaxies change over time leading up to and following the moment of coalescence.

Forthcoming advancements in the quality of astronomical imaging data also have the potential to improve the quality of CNN-identified post-merger samples. Later in Chapter 7, I will

investigate in detail the potential of current and next-generation astronomical surveys for identifying mergers in the framework presented in this Chapter.

Chapter 4

Studying star formation enhancement in the merger sequence

In galaxies with gas reservoirs, simulations predict that merger events will lead to elevated SFRs leading up to and after coalescence (e.g., Mihos & Hernquist 1996; Lotz et al. 2008; Rodríguez Montero et al. 2019; Hani et al. 2020), and multiple observational studies of galaxy pairs (Larson & Tinsley 1978; Knapen et al. 2015; Zuo et al. 2018) have indicated that pre-coalescence star formation enhancements are common in galaxies experiencing a merger. In post-mergers, the observational literature is somewhat more sparse and tense, with some statistical studies finding evidence for approximately doubled SFRs in post-mergers compared to non-interacting galaxies (notably, Ellison et al. 2013; Guo et al. 2016, with 97 and 70 post-mergers studied, respectively). Others, meanwhile, report no enhancement in star formation rate amongst post-mergers (e.g., in a sample of CNN-identified post-mergers; Pearson et al. 2019b). In the previous Chapter, I used a simulation-based approach to identify large samples of post-merger galaxies in CFIS and DECaLS with a CNN, and performed visual inspections to remove false positives from both samples. With pure merger samples in hand, it is now possible to characterize the impact of merger events on galaxy evolution with newfound statistical power. In this Chapter, I will describe an experiment designed to measure the strength of merger-induced star formation enhancement using the post-merger sample from CFIS.

4.1 Data and Methods

To compare the star formation rates of galaxies in the merger sequence to those of non-interacting galaxies, I require samples of both pre- and post-coalescence galaxy mergers, and a control pool of non-interacting galaxies for comparison. Similar to Section 3.1.2, the parent sample for the study described in this chapter from which the merger and control samples are derived is the sample of galaxies appearing in CFIS DR2 and SDSS DR7 with $z < 0.5$. However, in the study described

in this Chapter, I also require that galaxies have stellar masses and star formation rates available from the MPA-JHU[1] catalogue of spectroscopic measurements for SDSS DR7 galaxies (Brinchmann et al. 2004). From the parent sample, which contains 168597 galaxies, I select samples of pre-merger galaxy pairs, post-mergers, and non-merger controls. The details of each sample selection, as well as the methods I use for matching controls and comparing SFRs of mergers to non-mergers, are described in this Section.

4.1.1 Star-forming galaxy selection and SFRs

The purpose of this study is to investigate the amplitude of star formation enhancement brought on by mergers. As such, it is ideal to compare the SFRs of merging (or merged) star-forming galaxies (SFGs) to those of non-interacting SFGs that presumably represent the progenitors of the mergers. As such, I apply criteria to select star-forming galaxies with adequate spectroscopic signal to take reliable measurements of their SFRs (i.e., to remove quiescent galaxies / galaxies with weak emission lines from the sample). I also apply criteria to ensure that galaxies whose emission lines are significantly contaminated by an AGN (e.g., from a NLR) are removed as well. To achieve these aims, I use the BPT diagram cut for star-forming galaxies described in Kauffmann et al. (2003), which limits the search to galaxies whose emission lines are dominated by star formation. I require a minimum signal to noise ratio of 3 for the lines used in BPT classification (Hα, Hβ, [NII]λ6584, [OIII]λ5007). Since the visual classification exercise suggested that the CNN's classifications are unreliable outside of its mass training regime (10^{10} – 10^{12}M$_\odot$, see Figure 3.6), I also limit the analysis presented in this Chapter to this range. There are 19113 galaxies in the final pool eligible for this study, and their $p(x)$ predictions are shown in the teal histogram in Figure 4.1.

In the resulting population of star-forming galaxies, I use total SFRs from the MPA-JHU catalogue of spectroscopic measurements (Brinchmann et al. 2004). For galaxies with strong (S/N>3) emission lines, the Brinchmann et al. (2004) SFRs are computed using an emission line-fitting model that accounts for variable metallicity, ionization, dust attenuation, and the relative quantities of dust and metals. When emission lines are weak, SFRs are computed based on the strength of the spectral break at 4000Å (hereafter, D_{4000}). The D_{4000} SFRs are calibrated based on the correlation between D_{4000} and SFR (estimated from the emission lines) for galaxies that have both. SFR estimates made using galaxy spectra in SDSS must be aperture-corrected, since the SDSS fibre generally does not enclose the entire target galaxy. In order to estimate the additional star formation taking place outside of the fibre, Brinchmann et al. (2004) construct a grid of in-fibre

[1] wwwmpa.mpa-garching.mpg.de/SDSS/DR7/

measured SFRs arranged by their photometric colours (specifically, plotting $g - r$ colour against $r - i$ colour). The empirical correction is subsequently made based on the brightness and colour of a galaxy's image outside of the fibre.

4.1.2 Post-merger selection

In this Chapter, I combine the CNN's predictions and visual classifications to select two post-merger samples. In order to study the characteristics of a merger sample identified using the CNN alone, I apply a $p(x)$ cut at 0.8. I note that $p(x) > 0.8$ is more restrictive than what is used in Chapter 3 to flag galaxies for visual inspection. I use a different $p(x)$ cut for the CNN-only post-merger sample since it does not benefit from subsequent visual inspections. Together the high $p(x)$ cut and lack of visual quality control are used to simulate a scenario in which no manual quality control is enforced. While I expect the CNN-only post-merger sample to be significantly contaminated, it allows me to characterize how contamination bears on the statistical conclusions of this work. I also use the visually confirmed post-merger sample, which I expect to be far less contaminated than the CNN-identified sample following rigorous visual classifications described in Chapter 3. 94 of the galaxies from this pool have received CNN classifications with $p(x) > 0.8$, and 45 of them belong to the visually confirmed post-merger sample (i.e., they received CNN classifications > 0.75 and were confirmed during the visual classification exercise described in Chapter 3). Figure 4.1 (constructed in the same style as Figure 3.2 in the last Chapter) shows the mass, redshift, and CNN $p(x)$ statistics for the parent sample of galaxies appearing in CFIS DR2 and SDSS DR7 (greyscale 2D histogram and black histograms), the visually confirmed CFIS post-merger sample (magenta) and galaxies eligible for the star formation enhancement study in this chapter (i.e., meeting the criteria just described in Section 4.1.1; teal).

Ellison et al. (2013) is the most direct literature predecessor to the work presented in this Chapter, since it also also includes separate samples of galaxy pairs and post-mergers. For comparison against an established benchmark, I use the Ellison et al. (2013) post-merger sample in this work alongside the two post-merger samples already mentioned. In the case of Ellison et al. (2013), post-mergers were selected by eye from SDSS imaging from the merger sample presented in Darg et al. 2010. Since the post-mergers in Ellison et al. (2013) were selected from shallower imaging, it is likely that they are the remnants of more disruptive or recent mergers with greater mass ratios on average compared to the new post-merger sample identified in CFIS for this work. Only 26 objects from the Ellison et al. (2013) sample meet my criteria for star-forming galaxies, but I present their characteristics as well in Section 4.2 for comparison.

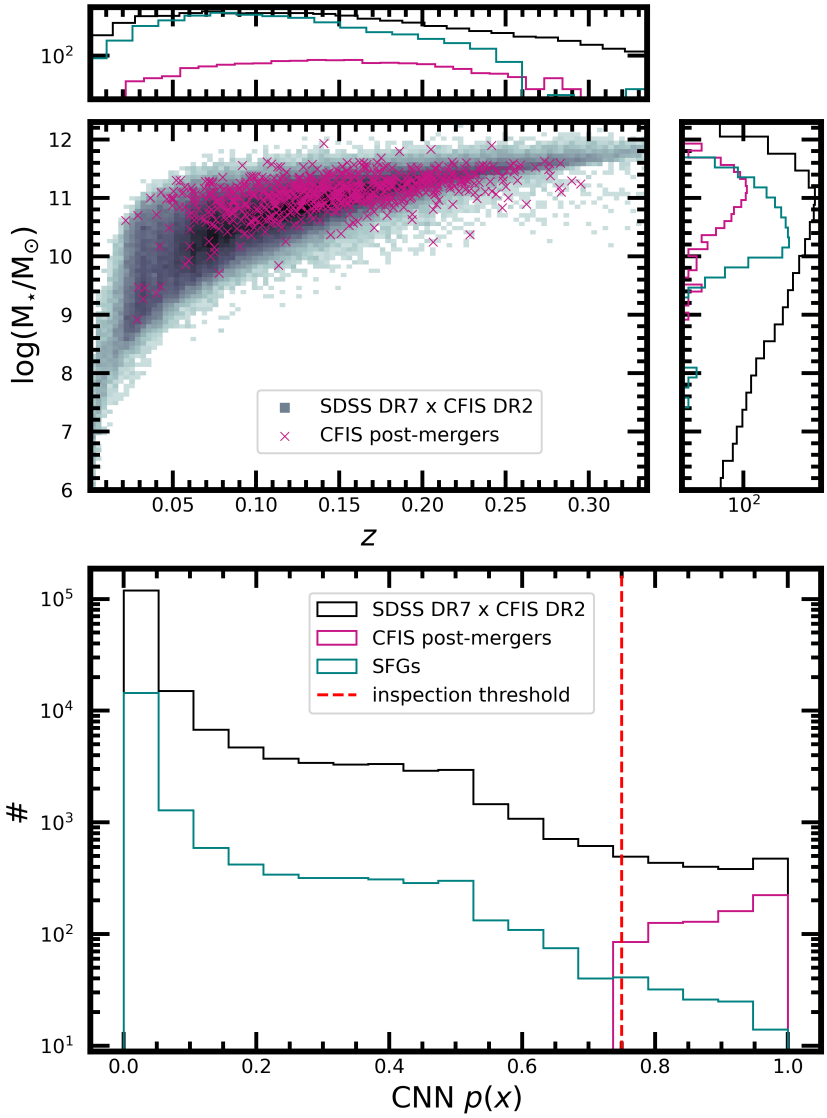

Figure 4.1: An overview of the galaxy sample studied in this Chapter. Top panel: stellar mass and redshift statistics for the parent sample of galaxies appearing in CFIS DR2 and SDSS DR7 (greyscale 2D histogram and black histograms in the auxiliary panels), the visually confirmed CFIS post-merger sample (magenta histograms and crosses), and the subset of galaxies in the parent sample meeting the star formation rate and spectroscopic signal to noise criteria for my star formation enhancement experiment (teal). Bottom panel: the CNN $p(x)$ predictions on the parent sample (black histogram), CFIS visually confirmed post-mergers (magenta) and galaxies meeting the star formation rate and signal to noise criteria (labelled SFGs) used in this Chapter. For reference, I also plot the $p(x)$ threshold (red dashed line) above which the galaxies were visually inspected.

4.1.3 Galaxy pair selection

The changes (e.g., to star formation rate) experienced by galaxies in the post-merger epoch are best presented in the context of the merger sequence, beginning with pair-phase galaxy interactions, and ending after coalescence. To add this context, I use the spectroscopic galaxy pairs catalogue from SDSS DR7 presented in Patton et al. (2016), and limit my sample to pairs whose spectroscopic characteristics suggest they will eventually merge – with on-sky projected separations (r_p) of at most 80 kpc, and velocity separations (Δv) of at most 300 km/s. My galaxy pair selection in this Chapter targets pairs with stellar mass ratios between 0.25 – 4. I make this choice to enforce methodological consistency with Ellison et al. (2013). I re-evaluate pair selection criteria again in Chapters 5 and 6 in order to choose the most relevant and inclusive sample of galaxy pairs for each study. There are 3463 galaxies meeting the pair criteria for this study. The pair galaxies are further arranged into 10 kpc bins based on their r_p, and in Section 4.2.2 I measure the median star formation enhancements in each bin in order to approximate the progression of star formation as a function of merger stage leading up to coalescence.

A larger Δv requirement, e.g., $\leq$1000 km s^{-1} could capture a larger sample of galaxies that could conceivably be experiencing interactions, but the risk of contamination by non-interacting galaxies also increases. In order to restrict my analysis to galaxies that are very likely to be pre-mergers, I continue to follow Patton et al. (2016) and use $\Delta v \leq 300$ km s^{-1}. Regardless of the choice of Δv, I expect a small amount of contamination from interlopers to persist.

4.1.4 Non-merger control pool

The merger samples studied in this chapter are compared to the properties of non-interacting control galaxies. In this Chapter, I use different criteria to identify the control pools for each merger sample in the interest of maintaining consistency with the literature. Later (e.g., in Chapter 6), I adjust this choice to tailor the control method to the experiment at hand by identifying a single pool of non-interacting control galaxies for all mergers.

In this Chapter, the CNN-only $p(x) > 0.8$ post-mergers and visually confirmed post-mergers are compared to non-interacting control galaxies from the pool of star-forming galaxies that have received CNN $p(x)$ predictions < 0.1. While 0.1 is somewhat arbitrarily chosen, 81 per cent of galaxies given negative $p(x) < 0.5$ classifications fall below $p(x) = 0.1$, and the specific choice of non-post-merger control pool does not affect the calculated star formation enhancements. In Section 4.2.1, I assess the utility of various CNN $p(x)$ cuts for selecting both mergers and controls, but $p(x) < 0.1$ is used to make my final estimate of post-merger star formation enhancement in Section 4.2.2. In an effort to maintain methodological consistency with Ellison et al. (2013), the

Ellison et al. (2013) post-mergers are compared to the pool of star-forming galaxies that have received Galaxy Zoo (Darg et al. 2010) merger vote fractions of zero. The sample of star-forming galaxy pairs described in Section 4.1.3 are compared to still another pool of star-forming control galaxies, which must have projected separations of > 80 kpc and Darg et al. 2010 merger vote fractions of zero. I emphasize that all mergers and controls in this Chapter are star-forming, i.e., meeting the criteria described above in Section 4.1.1 in addition to the further criteria by which their merger statuses are estimated.

4.1.5 Control matching and star formation enhancements

In this chapter, I aim to characterize star formation enhancements for merging and merged galaxies relative to non-merging controls. Whenever I calculate a star formation enhancement for one of the merger samples (either the CNN-identified, visually confirmed, or Ellison et al. 2013 post-merger samples, or the sample of galaxy pairs) I use the same control matching methodology to compute ΔSFR. For each merger, I identify any galaxies in the control pool with $M_\star$ within 0.1 dex and z within 0.01. If there are five or more eligible controls meeting the initial $M_\star$ and z criteria, a ΔSFR is calculated for the merger by subtracting the median control log-scale SFR from that of the merger. If fewer than 5 controls are found using the default search settings, both the $M_\star$ (in dex) and z tolerances are iteratively increased by a factor of 1.5 until at least five are identified. No prior upper limits are set on the number of tolerance growths, but due to the abundance of control galaxies, 99 per cent of the post-mergers find five controls without increasing their tolerances, and the remaining galaxies find their controls within 3 growths (i.e., within 0.34 dex or a factor of 2.2 in stellar mass).

In many studies (e.g., Ellison et al. 2013; Satyapal et al. 2014; Patton et al. 2016 and in Chapter 2), methodology is introduced to control for the influence of local environment on star formation rate in addition to $M_\star$ and z, since dense environments are often associated with a suppression of star formation (e.g., Shi et al. 2024). Since environmental statistics are not available from Patton et al. (2016) for all post-mergers, I do not match controls on environmental factors (neighbour distances or neighbourhood density) in order to include as many post-mergers as possible. When environmental factors are included, there is a minor suppression ($\leq$0.05 dex) in the calculated ΔSFRs for mergers, but all trends remain the same.

4.2 Results

Having identified samples of pre- and post-merger galaxies, controls, and star-forming galaxies, as well as a robust method for computing merger-induced star formation enhancements, I next

present the main results for this Chapter: the typical star formation enhancements in galaxy pairs and post-mergers compared to non-interacting controls.

4.2.1 Modulating the decision threshold

From the work presented in the previous Chapters (see e.g., Figure 3.6), I know that the CNN's prediction $p(x)$, while imperfect, is strongly linked to post-merger status in both simulated (from IllustrisTNG100-1) galaxies, and real galaxies that appear in CFIS. Further, I presented the results of a simulated ΔSFR experiment in a mock survey of simulated galaxies in Figure 2.19, and found that a high decision threshold, i.e., the chosen cut in $p(x)$ for a galaxy to be counted as a post-merger, could be effectively used to enforce high post-merger purity in the predicted sample. With a high-purity sample in hand, it is possible to closely estimate the true ΔSFR in recent post-mergers. I take my cue from the mock survey results and investigate ΔSFR with varying $p(x)$ cuts in order to approximate the true enhancement in CFIS post-mergers, albeit at the expense of completeness. I begin with a set of post-merger decision thresholds, and calculate a median ΔSFR for the post-mergers above each.

Figure 4.2 shows the results of my star formation enhancement study on predicted post-merger galaxies in CFIS. The green dashed line, included for context, is the true ΔSFR of galaxies in their first post-merger simulation snapshot (< 150 Myr since a major merger) in IllustrisTNG100-1 as shown in Section 2.2.3. The purple curve from the same work shows how the median predicted enhancement evolves as one modulates the decision threshold in a mock survey of simulation galaxies.

The colour-coded points with the grey error region in Figure 4.2 show the prediction of ΔSFR for CFIS post-mergers made by the CNN alone. As in the simulation, I expect based on Figure 3.6 that the CNN identifies increasingly pure CFIS post-merger samples as I enforce increasingly high decision thresholds. As a result, the median ΔSFR of the predicted sample steadily increases between $0.5 < p(x) < 0.9$, and makes a more dramatic jump in ΔSFR for the highest decision thresholds. The number of galaxies included decreases with increasing $p(x)$ cuts as well. In Table 4.1 I give selected example statistics for several $p(x)$ cuts, tabulating the number of post-mergers and the measured ΔSFR. Although few galaxies are included in the post-merger sample above $p(x) = 0.9$, the IllustrisTNG100-1 results suggest that these data points indicate most accurately the true behaviour of galaxies that have undergone a major merger very recently, with enhancements around 0.3 dex (see the purple curve approaching the green line, Figure 4.2).

The orange series in Figure 4.2 shows the result of the same experiment for the visually confirmed post-merger sample. Without any additional cuts other than the initial $p(x) > 0.75$ used to

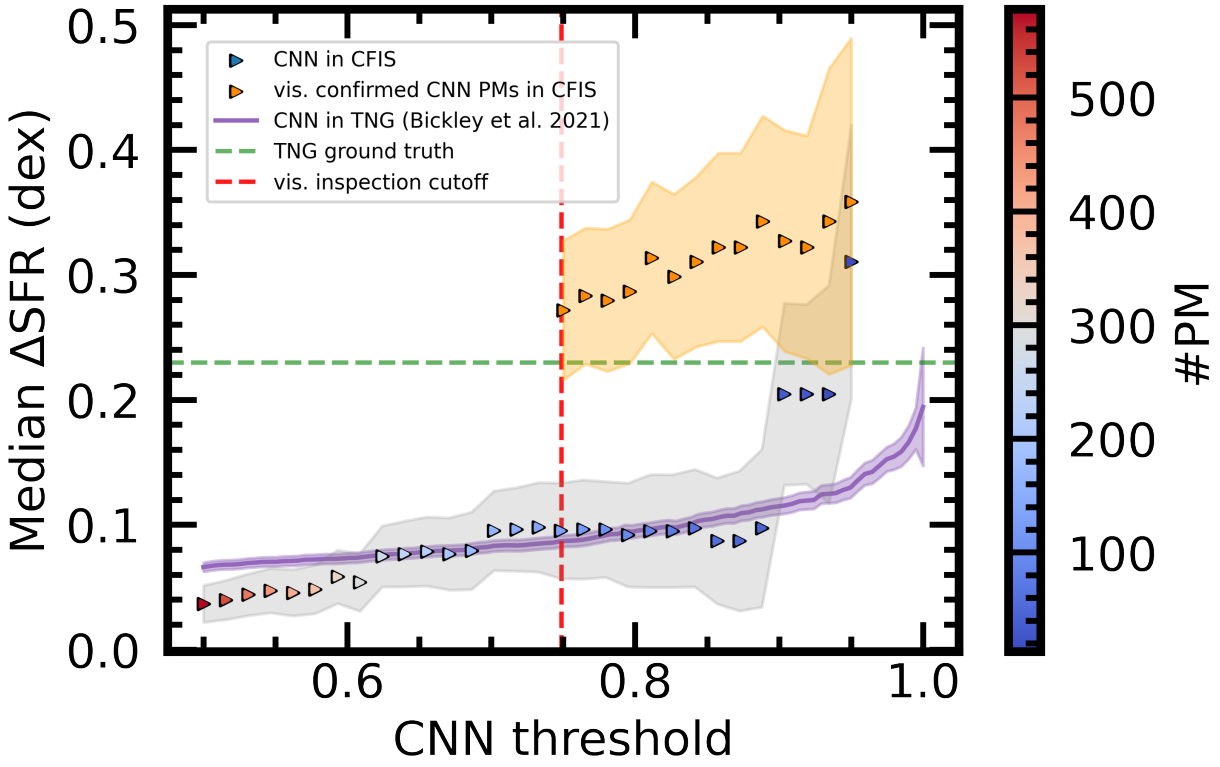

Figure 4.2: Median star formation enhancements found using a range of CNN thresholds. For each data point, all of the galaxies with $p(x)$ higher than the specified value are included in the sample. Error regions represent the statistical error on the median. I compare the trend of increasing star formation with decision threshold (and by extension, sample purity) using 1) only CNN labels (colour-coded points), and 2) the visual labels (orange points) to the trend found by applying the same methodology to galaxies from IllustrisTNG100-1 as in Section 2.2.3 (purple curve), as well as the underlying true value for recent post-mergers from IllustrisTNG100-1 (green dashed line). The red dashed line marks the CNN threshold below which galaxies were not inspected visually. Using the CNN alone, I calculate meager ΔSFR for $p(x)$ cuts below 0.6. As in the mock survey, the median ΔSFR increases when more strict $p(x)$ cuts are imposed, up to a maximum of $\sim$ 0.3 dex (or a factor of two) for $p(x) > 0.95$. The visually confirmed sample is already highly enhanced ($\sim$ 0.27 dex) before any additional $p(x)$ cut is applied. As $p(x)$ increases, the median enhancement of the visually confirmed sample increases up to $\sim$ 0.36 dex, though in a less dramatic fashion than the CNN-only sample. The relative stability of the orange series suggests that the visually confirmed sample is already highly pure.

CNN $p(x)$	#PM, full	ΔSFR, full (dex)	#PM, vis.	ΔSFR, vis. (dex)
0.5	578	0.04	45	0.27
0.8	94	0.08	27	0.31
0.9	33	0.10	24	0.32
0.95	13	0.31	7	0.36

Table 4.1: The number of CNN-predicted and visually confirmed post-mergers with CNN predictions above selected $p(x)$ cuts, with median ΔSFR values for each subset. See also Figure 4.2.

select galaxies for visual inspection, the visually confirmed galaxies are already highly enhanced in star formation relative to controls, producing nearly twice as much stellar material on average than their non-post-merger counterparts. Applying a progressive decision threshold to the visual sample does not produce a dramatic spike in enhancement for high $p(x)$ as it does in the CNN-selected samples, suggesting that the visual classification exercise has already produced a highly pure sample, but it does further enhance the median predicted ΔSFR up to ~ 0.36 dex.

Figure 4.3 shows the distribution of ΔSFR for CNN-identified post-mergers at selected $p(x)$ decision thresholds: $p(x) > 0.5$, the natural threshold for a well-calibrated model trained on balanced data, $p(x) > 0.8$, $p(x) > 0.9$, and $p(x) > 0.95$. The normalized (top) version of the figure shows the change in the shape of the distribution of post-merger ΔSFRs, while the non-normalized logscale version of the figure (bottom) shows how the successively smaller (and presumably more pure) subgroups of predicted post-mergers move towards a higher median ΔSFR as shown in Figure 4.2. It is noteworthy that for a sample with $p(x) > 0.5$, I find a negligible median enhancement in the SFR of post-mergers thus selected. An effect analogous to the null result found when using a cut at $p(x) = 0.5$ is likely responsible for the under-prediction of merger induced star formation enhancement in the literature when low $p(x)$ thresholds are used and/or post-merger samples are not visually confirmed (e.g., Pearson et al. 2019b).

Since the strength of merger features is plausibly linked to $p(x)$, the trends uncovered with the CNN alone in Figures 4.2 and 4.3 could be explained by an intuitive physical connection between merger intensity, the time since a galaxy underwent a merger, the prominence of visible merger-induced morphologies, and the strength of merger-induced star formation enhancement. As a result, the fact that ΔSFR increases with the CNN $p(x)$ used to select post-mergers is likely the combined result of two effects. First, the higher $p(x)$ cuts are improving the purity of the post-merger sample, by exchanging purity for completeness. Other than the loss of statistical power, this exchange is nominally favourable, since I aim to characterize the impact of mergers on star formation as accurately as possible. Second, higher cuts in $p(x)$ appear to select a higher proportion of mergers that are more visually dramatic. Since visual merger severity, the strength of gas inflows, and star formation enhancement are interconnected, this behaviour by the CNN is somewhat less preferred, since the measured ΔSFR above very high CNN $p(x)$ is no longer representative of the visually diverse population of post-mergers. As such, Figure 4.3 indicates that cuts in CNN $p(x)$ should be used sparingly in order to preserve a representative sample of post-mergers. In this context, quality control using visual classifications is proven to be especially important. Still, since I have used a CNN and manual hybrid approach to select my post-merger sample, I can report star formation enhancements with some confidence: ~ 0.27 dex for an inclusive, visually confirmed sample of post-mergers.

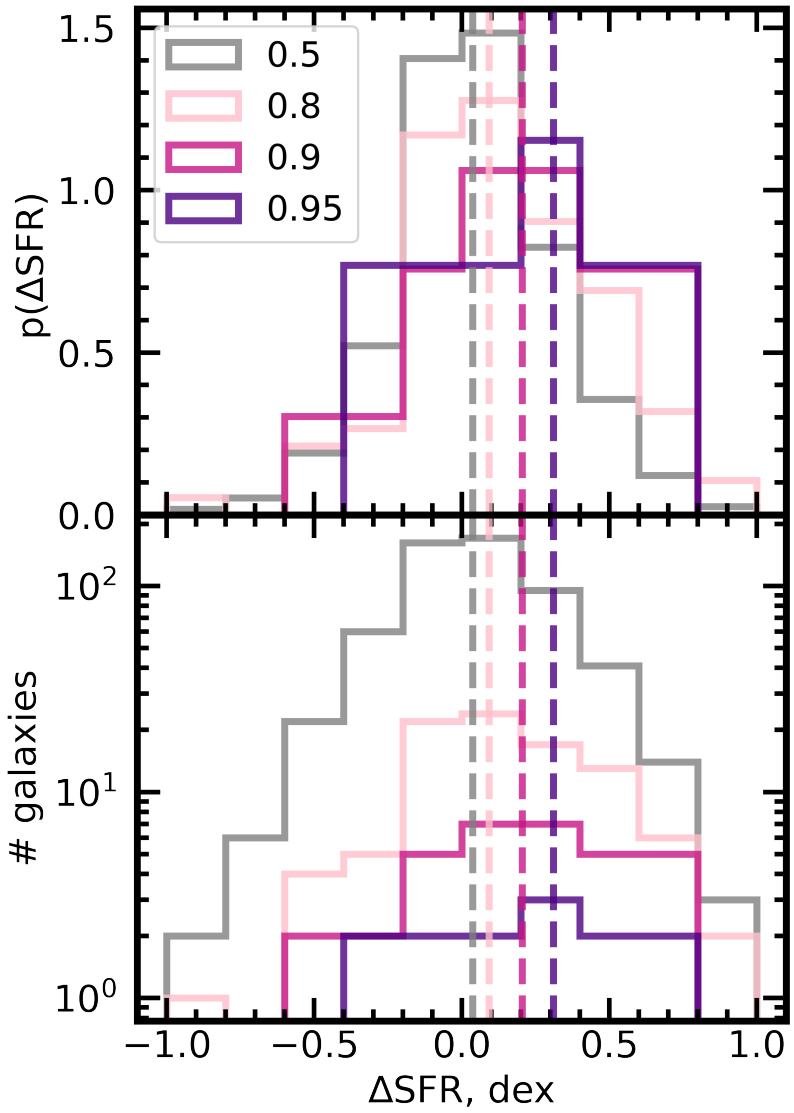

Figure 4.3: The normalized (top panel) and log-scale histogram (bottom panel) distributions of the star formation enhancements calculated for selected CNN thresholds, with vertical dashed lines representing the median enhancements for each group. As the threshold grows more extreme, from $p(x) > 0.5$ (grey curve) to $p(x) > 0.8$ (pink curve), $p(x) > 0.9$ (violet) and $p(x) > 0.95$ (dark purple), the sub-samples of galaxies are increasingly enhanced in SFR on average.

4.2.2 Star formation in the merger sequence

The results for post-mergers are best considered in the context of the merger sequence, including the sample of pre-merger galaxy pairs described in Section 4.1.3. In order to add the context of the merger sequence leading up to coalescence to my study of star formation enhancement in post-mergers, I compute the median ΔSFR for galaxy pairs in 10 kpc bins of r_p spanning 0 – 80 kpc following the method described in Section 4.1.5. In order to highlight the role of the post-merger selection method in determining the measured star formation enhancements for post-mergers, I also compare the ΔSFR for the visually confirmed CFIS post-merger sample to the CNN-only $p(x) > 0.8$ predicted post-merger sample, as well as the Ellison et al. (2013) post-merger sample.

Figure 4.4 shows that I uncover a trend consistent with Ellison et al. (2013) in the pair phase, with a peak ΔSFR of $\sim$ 0.15 dex for pairs with projected separations $<$ 10 kpc. Without the help of visual inspection, even with a relatively high threshold of $p(x) > 0.8$, the CNN predicts a median ΔSFR of 0.08 dex for the post-mergers, which is lower by $\sim$ 0.07 dex than for close pairs with projected separations below 10 kpc. However, the median enhancement for the entire population of visually confirmed post-mergers meeting the criteria for the study is consistent with the Ellison et al. (2013) result at $\sim$ 0.27 dex, confirming the link between the post-merger epoch and strongly enhanced star formation in a sample that is nearly twice as large. Further, this result is consistent with recent theoretical studies of IllustrisTNG by Patton et al. (2020) and Hani et al. (2020), suggesting that IllustrisTNG served as a suitable training ground for the CNN.

4.3 Discussion and Conclusions

I have applied a CNN trained to identify simulated, realism-added, recently-coalesced post-mergers from the IllustrisTNG simulations to r-band galaxy images from CFIS. The results of this application and the subsequent study of the out-falling predicted post-merger galaxies suggest that the simulation-based inference framework is operating in the way it was intended. Moreover, the approach has revealed the impact of coalescence on galaxy star formation with unprecedented statistical significance.

The global statistics of the merger status predictions are consistent with my expectations given the rarity of post-mergers in the Universe (Figure 4.1), and considering the results of a proof-of-concept "mock survey" detailed in Chapter 2. When I compute star formation enhancements for post-mergers identified by the CNN, as well as those confirmed by eye, I find that the behaviour of galaxies on the CNN $p(x)$, ΔSFR, and merger purity axes are very similar to the simulated results in Chapter 2: I find that increasing the decision threshold appears to enhance the purity of the CNN-predicted post-merger sample in CFIS. As the decision threshold and sam-

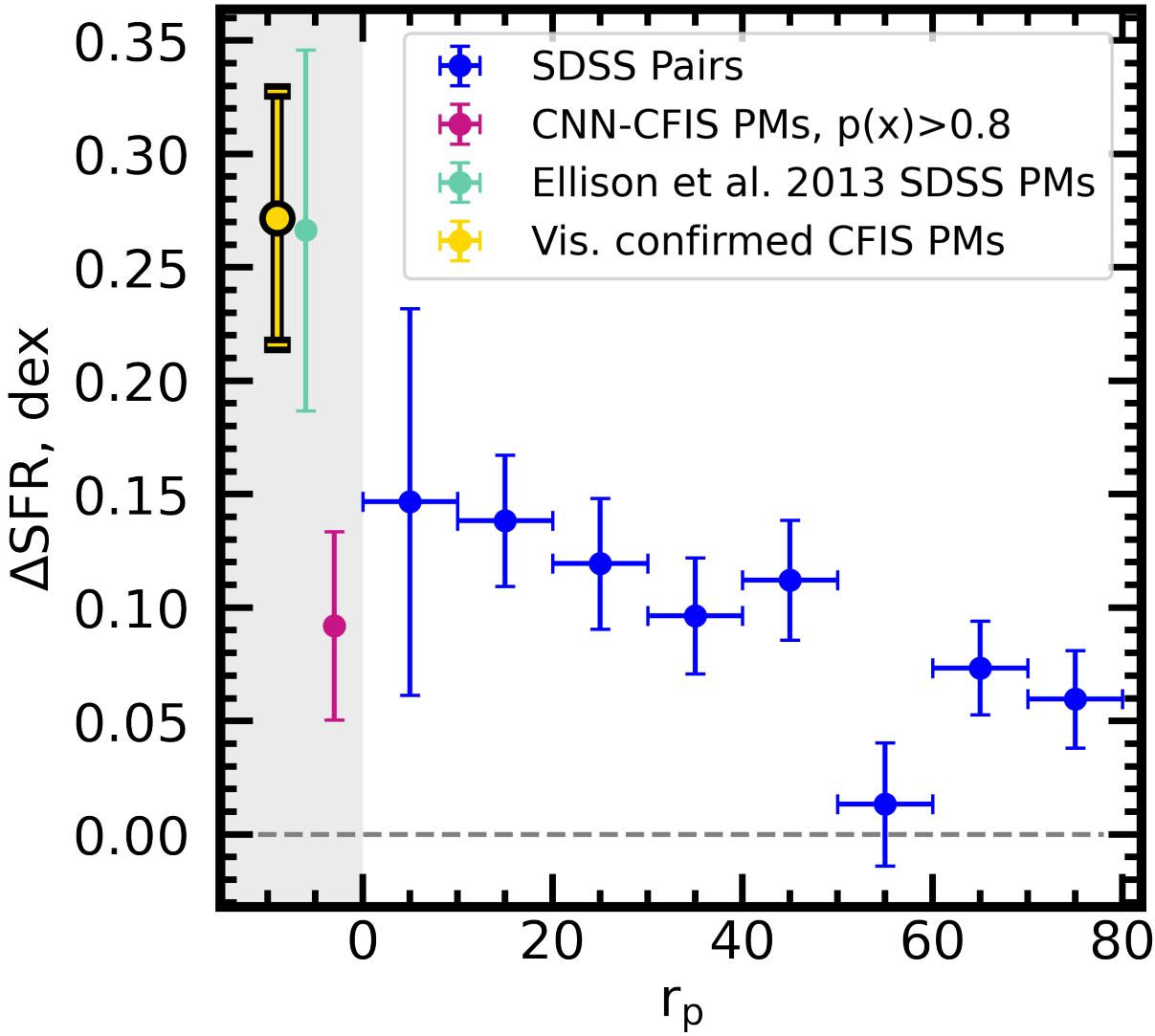

Figure 4.4: The trend between projected galaxy pair separation in SDSS-DR7 and star formation enhancement (blue series) in the style of Ellison et al. (2013) using matching, star formation criteria, and signal to noise criteria consistent with this work. I uncover the same qualitative trend; that ΔSFR increases as a function of projected separation as pre-coalescence interactions funnel gas reserves towards the centres of the participant galaxies. Of the $\sim$ 100 SDSS post-mergers in Ellison et al. (2013), 26 survive my star-forming galaxy selection criteria (turquoise point). They are even more highly enhanced in star formation than the closest pairs. Without the assistance of visual labels, the CNN uncovers increasingly high median enhancements as a function of decision threshold (see Figure 4.2) in increasingly small post-merger samples. Of the visually confirmed sample, 45 galaxies survive my selection criteria. They are as strongly enhanced on average as the post-mergers from Ellison et al. (2013).

ple purity increase, I find that the median ΔSFR increases accordingly. For the highest decision thresholds $p(x) > 0.95$, ΔSFR peaks at $\sim$ 0.3 dex (Figures 4.2 and 4.3), consistent with the trend from the mock survey in Chapter 2.

It is important to question the conviction with which I report my science results in the context of the study – i.e., that the merger status of galaxies observed in CFIS cannot be known to observers with total certainty. In IllustrisTNG, the CNN's $p(x)$ trends positively with merger status, and I see an analogous trend in Figure 3.6 when I plot the frequency of agreement between the CNN and visual classifiers as a function of $p(x)$. To first order, it seems that the CNN's classifications can be trusted to the extent that the training dataset of IllustrisTNG mock observations resemble real galaxies in CFIS. Since no star formation criteria were used to select post-mergers or non-post-mergers for the training set, there is no explicit prior on the training set that could give rise to the results presented in this Chapter. As a result, the qualitative connection reported here between coalescence and elevated SFRs amongst star-forming galaxies is likely to be a physical one.

I find that galaxies in the hybrid-identified CFIS post-merger sample are enhanced in star formation relative to controls by $\sim$ 0.27 dex. Ellison et al. (2013) report a post-merger ΔSFR of $\sim$ 0.23 dex, but when the same post-mergers are processed with my SFG selection and control matching methodology, the two are brought into near-exact agreement (see Figure 4.4). Comparing samples of far-IR (FIR) selected SFGs, Guo et al. (2016) report star formation enhancements of $\sim$ 0.33 dex within the SDSS fibre in a sample of advanced mergers. The relatively large enhancement found in Guo et al. (2016) compared to this work is consistent with the idea that fibre SFRs preferentially capture ongoing star formation in the centres of galaxies, where merger-induced star formation is expected to occur. Thorp et al. (2019) mainly focuses on the star formation enhancement profiles of galaxies, rather than their total enhancements. Still, they find that post-mergers are on average enhanced in star formation by $\sim$ 0.4 dex near the centre and $\sim$ 0.1 dex in the outskirts. Intuitively, the typical ΔSFR for post-mergers presented in this Chapter ought to (and does) fall between 0.1 – 0.4 dex, since it is (in principle) a luminosity-weighted average of the quantities reported in Thorp et al. (2019). Pan et al. (2019) perform this weighted combination for late-stage mergers observed with integral field spectroscopy (IFS) and compute a global ΔSFR for post-coalescence galaxies, reporting a typical enhancement of 0.37 dex – somewhat higher than the measurement taken in this work, but with overlapping error regions. The higher enhancement found by Pan et al. (2019) could be a result of their merger sample selection, which was done by eye without the aid of CNNs or other machine vision tools, and may therefore include a greater proportion of more visually disturbed systems. For a broad sample of late-stage mergers with masses $> 10^{10} M_{\odot}$, Osborne et al. (2020) find a typical enhancement of $\sim$ 0.3 dex,

in good agreement with this work. When collated, the literature suggests that the final stages of major, gas-rich galaxy mergers are responsible for star formation enhancements by a factor of ~ 2. Null results for star formation enhancement after coalescence (e.g., Pearson et al. 2019b) likely originate from contaminated post-merger samples.

The star formation enhancement for post-mergers reported in this Chapter is sensitive to higher-order effects, even given the understanding that I am defining post-mergers as the remnants of major mergers with mass ratios between $0.25 < \mu < 4$ and which have merged in the last ~ 150 Myr. The main driver of this sensitivity is the fact that sample purity and the false negative rate (number of "lost" post-mergers) are at odds with each other in the domains of both the CNN and visual classifications. In an effort to study an uncontaminated sample of post-mergers by increasing the $p(x)$ criterion, or by performing rigorous, purity-motivated visual classifications, I remove a significant number of true post-mergers are removed from the sample. Typically, I expect that false negatives rejected in $p(x)$ space or during visual inspection are the least visually dramatic, having experienced comparatively minor mergers somewhat less recently. Conversely, the post-mergers that make it to the final sample are more heavily disturbed. The physical hypothesis of this Chapter – that more disruptive mergers would drive stronger inflows, and lead to stronger star formation enhancement – would suggest that the false negatives would be less enhanced in star formation compared to the true positives. Including galaxies with lower $p(x)$ or visually less disturbed morphologies would (and does, see the maroon data point on Figure 4.4) lower the predicted average ΔSFR for post-mergers. There is currently a degeneracy in determining to what extent the decrease is owed to greater impurity (a larger number of false positives making it to the final sample), and to what extent it reflects the true post-coalescence ΔSFRs of less-disturbed post-mergers. The degeneracy between sample impurity and the intrinsic amplitude of star formation will be somewhat alleviated in the future by automated predictions of the time elapsed since the merger, and the merger mass ratio.

Chapter 5

Narrow-line optical, mid-infrared, and low-excitation radio galaxy AGN in CFIS post-mergers

In the last Chapter, I demonstrated that a sample of post-merger galaxies selected using a hybrid (CNN and visual) approach can be used to study how star formation changes in the ∼10s of millions of years after a merger has elapsed. The new visually confirmed post-merger sample from CFIS can also be used to characterize the impact of coalescence beyond star formation. Observations and theory both predict that post-mergers ought to be significantly enhanced in AGN activity as well, with the same gas inflows driving both central star formation and enhanced accretion onto central supermassive black holes (SMBHs). In this Chapter, I will study the statistical connections between merger status as predicted by the CNN and AGN activity, using multi-wavelength data from SDSS, the Wide-field Infrared Survey Explorer (WISE) space telescope, and the catalogue of radio-loud galaxies from Best & Heckman (2012) to characterize the post-mergers from CFIS.

5.1 Sample selection

Similar to Chapter 4, the merger, control, and AGN samples studied in this chapter are derived from the population of galaxies appearing in both SDSS DR7 and CFIS DR2, with $z < 0.5$. The parent sample is therefore the same, with 168597 galaxies. However, since the scientific aims of this chapter are different, the merger and control criteria are not identical, and the overlaps of the merger and control pools with the samples of interest (AGN categories, to be described later in Section 5.2) have changed as well.

5.1.1 Post-mergers

In this Chapter, I study the AGN demographics of the 699-galaxy visually confirmed CFIS post-merger sample identified in Chapter 3. I also compare the characteristics of the visually confirmed CFIS post-merger sample to those of Ellison et al. (2013). Though lesser in number (97 in the sample before any experiment-specific cuts are applied), the Ellison et al. (2013) merger sample provides useful context: it represents a highly pure sample as well, albeit one that is likely biased towards mergers of greater visual strength due to their selection from shallower SDSS imaging. No additional criteria are applied to the post-merger samples, since the aim of this Chapter is to holistically study AGN incidence in the post-merger sample. I no longer include a CNN-only (i.e., purely machine-identified) post-merger sample as I do in Chapter 4, since the results in Chapter 4 suggested that post-merger samples identified by the CNN alone were likely contaminated, and not representative of the true characteristics of post-mergers.

5.1.2 Galaxy pairs

As in my study of star formation, a sample of pre-coalescence galaxy pairs is required to contextualize my results for post-mergers within the larger context of the merger sequence. I again use the galaxy pair catalogue compiled by Patton et al. (2016), and select objects for comparison emulating the method employed in Ellison et al. (2013). Whenever they are shown for context in this Chapter, galaxy pairs are required to have projected separations $r_p \leq 80$ kpc, line of sight velocity differences $\Delta v \leq 300$ km s^{-1} (in order to minimize accidental projections of a false pair), and stellar mass ratios of $0.1 < \mu < 10$. As with Chapter 4, the galaxy pairs are subsequently arranged into eight bins, each spanning 10 kpc between 0 – 80 kpc.

In contrast to Chapter 4, I am now applying symmetrical mass ratio criteria to my pair sample and the simulated post-merger sample used to train the CNN (1:10 or more). At the time of completing this work, I implemented this change in order to select a sample of galaxy pairs that represent better pre-merger analogues to the post-mergers. Although merger mass ratio may scale with the degree of morphological disturbance of merger remnants, I do not find that the inclusion or exclusion of minor galaxy pairs (e.g., changing the mass ratio criterion to $0.25 \leq \mu \leq 4$) has a significant influence on the results. I also expect the median progenitor mass ratio of the CFIS post-merger sample to increase following the visual inspection effort detailed in Chapter 3, since merger candidates of unconvincing strength were removed. I therefore present the physical characteristics of merger remnants alongside galaxy pairs with the caveat that the pairs studied may not be precise pre-merger analogues to the post-mergers.

5.1.3 Control pools

Wherever post-mergers are studied in this Chapter, I will compare their physical characteristics to those of control galaxies that are matched in mass and redshift, but which are not themselves post-mergers. To this end, I again use CNN classifications. The vast majority (131168 of 168597) of galaxies in the CFIS sample are assigned CNN $p(x) < 0.1$ by the network, nominally indicating non-post-merger status. Even with an accurate classifier, Bayesian statistics suggest that a small number of genuine post-mergers will fall below this threshold. Thanks to the initial rarity of post-mergers, however, one can reasonably assume that the extreme dilution of post-mergers below $p(x) < 0.1$ will effectively eliminate their influence on my results.

Wherever galaxies from the Ellison et al. (2013) sample is shown, their physical characteristics are compared to those of their own mass- and z-matched corresponding control sample with a Galaxy Zoo (Darg et al. 2010) merger vote fraction of zero. As with controls for the visually confirmed post-merger sample, I expect that the signal from misclassified post-mergers contaminating this control sample should be effectively zero.

The controls for galaxy pairs must have projected separations of $r_p > 80$ kpc in the Patton et al. (2016) catalogue, and Galaxy Zoo merger vote fractions of zero in order to ensure that they do not belong to an interacting pair, and have not merged recently.

5.1.4 Control matching methodology

The results presented later in Section 5.3 use two different control matching methods. The first is used to calculate an AGN excess – the ratio of the AGN fractions in matched samples of post-mergers (or pairs) and controls. When controls are properly matched and statistics are of adequate quality, any excess > 1 indicates that the differentiating characteristic between the two samples (i.e., post-merger or pair status) is responsible for (or correlated with a factor that is responsible for) the increased AGN frequency. Excesses < 1 suggest that mergers may be responsible for a suppression of AGN in the target sample.

In calculating each excess, I match controls to the post-mergers or pairs (referred to collectively as "target galaxies") on stellar mass and redshift. Controlling for environment using additional parameters (halo mass, environment density) does not qualitatively change my results, suggesting that the influence of the most relevant interacting companion (in the case of galaxy pairs) or coalesced companion (for the post-mergers) is largely responsible for the signal uncovered in this Chapter. In order to ensure I have the same number of controls per target galaxy, I identify the closest non-merger match (without replacement) in 2D parameter space for each target galaxy, assigning equal weight to each parameter. After finding one control for each galaxy,

a Kolmogorov-Smirnov (K-S; Smirnov 1948) statistic is calculated for the $M_\star$ and z distributions of the target galaxies and controls. If the K-S p value is below 0.9, the control pool is finalized and I calculate the AGN fractions and excess for the relevant AGN type (described in Section 5.2). If the K-S p value remains above 0.9, I attempt to match additional equal-sized batches of controls, checking the K-S p value every time before adding an additional batch. I cap the number of controls per post-merger at 10, as the counting statistics do not improve significantly past 10. I use the same statistical control matching methodology for each target sample (visually confirmed CFIS post-mergers, Ellison et al. 2013 post-mergers, and 10-kpc bins of galaxy pairs), allowing for the control pool and number of controls per target galaxy to vary (between 1 – 10 depending on the continued success of K-S tests) between each subset of galaxy pairs (binned by projected separation) and post-merger sample I study. A similar control-matching method is used again in Chapter 6, with small improvements discussed therein. In practice, during control matching for the excess calculations in this Chapter, I reached the cap of 10 controls per target in both post-merger samples, and in each bin of projected separation, thanks to the abundance of controls.

The second control matching method is used later in Section 5.3.2, in which I measure a luminosity enhancement in the [OIII]λ5007 emission line (which approximates the SMBH accretion rate) in merger-AGN compared to non-merger-AGN. Since I am comparing measurements of a continuous variable (rather than a binary AGN status, as in the excess calculation), I use a control matching method similar to that used in Chapter 4 to compute star formation enhancements. When computing the luminosity enhancement, I require an ensemble of at least 5 non-merger-AGN controls for each merger-AGN. Rather than finding the nearest controls in parameter space, I set initial tolerances of ± 0.1 dex in $M_\star$ and ± 0.05 in z. In practice, all of the target galaxies (pairs and post-mergers alike) find at least 5 controls without any growths in parameter space. The [OIII]λ5007 luminosity enhancement, $\Delta\log(L_{[OIII]})$, is calculated as the difference between the logged [OIII]λ5007 luminosity (in units of erg s^{-1}) of the target galaxy and the median logged luminosity of the control ensemble, and hence captures the typical accretion rate difference between AGN triggered by mergers and secular AGN.

5.2 AGN criteria

In Section 1.1.2, I described the various physical mechanisms (accretion rate, viewing angle, degrees of obscuration by gas and dust) that are responsible for the broad, multi-wavelength observational diversity of AGN. In this Chapter, I account for heterogeneity within the AGN category by using three sets of AGN criteria: narrow-line AGN (NLAGN), identified using galaxies' optical spectra and the BPT diagram, dust-obscured mid-IR AGN, identified based on mid-IR colour, and

low-excitation radio galaxy (LERG) AGN, selected by Best & Heckman (2012) using a combination of radio detections and optical emission line criteria.

5.2.1 Narrow-line optical AGN

Since spectroscopic redshifts are required for appropriate treatment of the CFIS galaxy images, all post-mergers identified in CFIS *r*-band imaging by the CNN have available SDSS DR7 optical spectra as well. As in Chapter 4, I can use emission line measurements from the same optical spectra to search for the luminous echoes of AGN in the galaxies' gaseous narrow-line regions (NLRs). Galaxies exhibiting the NLR optical emission line characteristics of the Seyfert II class will hereafter be referred to as NLAGN.

In order to quantify the incidence of NLAGN in post-mergers, as well as estimate the accretion rates of their black holes, I use the "maximum starburst line" on the BPT diagram (Baldwin et al. 1981) computed by Kewley et al. (2001). The line represents the theoretical boundary separating AGN from starbursts, determined using grids of models with varied metallicities and ionization parameters, and is more stringent than the criterion of Kauffmann et al. (2003) which is often used (e.g., in Chapter 4) to separate purely star-forming galaxies from those with potential contributions from AGN. I again use Milky Way and host extinction-corrected optical emission line measurements (from the SDSS MPA-JHU catalogue and Scudder et al. 2012) to place the merger and control galaxies on the BPT diagram. To avoid contamination from shocks, which are expected in gas-rich galaxy mergers, I do not count objects on the AGN side of the Kewley et al. (2001) diagram as AGN if they fall below either the [SII]λ6717 or [OI]λ6300 BPT diagram low-ionization nuclear emission region (LINER) criteria described in Kewley et al. (2006b).

When quantifying AGN excesses (the ratio of the AGN fractions in mergers and controls) in Section 5.3.1, I require S/N $\geq$ 5 on all four emission lines used in the BPT diagram. Consequently, the nebular emission lines of galaxies that I count in my excess calculations are dominated by the high-energy ionization associated with the AGN, and the excess calculations capture the frequency with which mergers induce strong, unambiguous optical evidence of an AGN.

Later, in Section 5.3.2 when I approximate the SMBH accretion rate with the luminosity of the [OIII] emission line in Section 5.3.2, I require an ensemble of five or more controls (which must also be NLAGN) for each merger in order to calculate a robust accretion rate enhancement. The method used here is analogous to that used to characterize star formation enhancement in Section 4.1. The [OIII] emission line has been used many times to approximate the accretion rates of optically identified AGN, (e.g., Kauffmann et al. 2003; Brinchmann et al. 2004; Chen et al. 2009; Liu et al. 2012), but [OIII] flux can also be contaminated by star formation. The link between

[OIII] luminosity and star formation rate is complex, and its dependence on several parameters is explored in Kewley et al. (2001). Nonetheless, the Kewley et al. (2001) criterion I use is intended to be a maximal starburst division, such that galaxies above it can only be produced (in their models) by AGN contributions. I choose galaxies for this experiment whose emission is unambiguously AGN-dominated according to the same Kewley et al. (2001) criterion used in my excess calculations, and compare the AGN found in mergers to those found in non-mergers. The unprecedented size (699 galaxies) of the visually confirmed post-merger sample allows me to use the same S/N $\geq$ 5 cut as I do when measuring NLAGN excesses, and additionally remove LIERs / LINERs using the [SII]λ6717 and [OI]λ6300 Kewley et al. (2006b) criteria in order to ensure a robust connection between observed [OIII]λ5007 luminosity and SMBH accretion.

5.2.2 Infrared AGN

Though they are useful for identifying a particular subset of AGN, optical emission lines do not represent a complete census of SMBH activity in low-redshift galaxies (Hickox & Alexander 2018). A number of studies have demonstrated the utility of mid-IR observations to identify dust-obscured AGN (e.g., Lacy et al. 2004; Stern et al. 2005; Donley et al. 2007; Hickox et al. 2007; Donley et al. 2008; Eckart et al. 2009; Stern et al. 2012; Mateos et al. 2013). Many galaxies with ongoing SMBH accretion are obscured such that their optical emission line strengths cannot be reliably measured, while at the same time reddening their mid-IR colours (e.g., Assef et al. 2013). In fact, mergers may preferentially produce dust-obscured AGN (Satyapal et al. 2014, Weston et al. 2017, Blecha et al. 2018). Models predict that when one or both companions involved in the interaction is dusty, the kinematic disturbances induced by the final stages of a merger may distribute the dust in the centre, obscuring optical emission that might be observable in a dynamically settled galaxy (Yutani et al. 2022).

In order to identify dust-obscured AGN, I turn to legacy all-sky mid-IR observations from the Wide-Field Infrared Survey Explorer (WISE) space telescope, and follow Satyapal et al. (2014) and Ellison et al. (2013) in using a WISE $W1 - W2 > 0.5$ colour cut. Because I am interested in studying the proportion of galaxies with WISE photometry whose $W1 - W2$ colours exceed 0.5, I also require that all galaxies (post-mergers as identified by any method, galaxy pairs, and controls) included in my WISE-related results have been detected by WISE. WISE sources considered in my analysis have S/N of at least 2 for both $W1$ and $W2$, but a more stringent S/N criterion does not change the results since > 99 per cent of my sources have S/N > 10 for both $W1$ and $W2$. Mid-IR observations can be affected by star formation, but even template spectra for galaxies with extreme star formation have been shown to fall below $W1 - W2 = 0.5$ at low z (Assef et al.

2013; Satyapal et al. 2014).

5.2.3 LERGs

Optical and mid-IR AGN detections together provide a reasonably complete census of the observational phenomena associated with radiatively efficient, rapidly accreting SMBHs with a luminous accretion disk. Indeed, these AGN are most closely linked with the archetypal understanding of the role of the merger sequence in galaxy evolution, in which gas inflows simultaneously produce upticks in star formation and SMBH accretion (e.g., Hopkins et al. 2006). While useful, this narrative largely excludes mergers between gas-poor systems. Although low-excitation AGN states are most often associated with relatively isotropic accretion of hot gas from galactic haloes (e.g., Best et al. 2005, 2006; Allen et al. 2006; Hardcastle et al. 2007; Gaspari et al. 2013), there has been some evidence of a role for mergers in triggering radiatively inefficient accretion (e.g., Sabater et al. 2013; Garofalo 2019).

In order to investigate the role of mergers in triggering radiatively inefficient AGN, I also quantify the proportion of LERGs in the post-merger and pair samples relative to controls selected from the same SDSS sample as in Section 5.2.1. These low-excitation radio-selected AGN are taken from the compilation of Best & Heckman (2012), who match SDSS to a pair of radio catalogues: the National Radio Astronomy Observatory (NRAO) Very Large Array (VLA) Sky Survey (NVSS) and Faint Images of the Radio Sky at Twenty-centimeters (FIRST). After classifying their detections as either star-forming or AGN via a combination of 4000Å break strengths, radio-to-emission-line luminosities, and BPT diagnostics, they use optical emission lines and a second decision tree to distinguish the latter into high-excitation radio galaxies (HERGs) and LERGs. Because the high-excitation AGN state is well described by the optical and mid-IR observations, I consider only the LERGs in this work.

5.2.4 Overlap of AGN types

The visually confirmed post-merger sample, the Ellison et al. (2013) post-merger sample, and the SDSS galaxy pair sample contain 699, 96, and 17566 galaxies, respectively. The visually confirmed post-merger sample, which is the main new sample of interest developed for this work, contains the following:

- 32 NLAGN meeting the Kewley et al. (2001) and Kewley et al. (2006b) criteria (excluding star-forming galaxies, composite galaxies, and LINERs) with S/N> 5 for the emission lines used in BPT placement

- 66 mid-IR AGN with $W1 - W2 > 0.5$
- 14 LERGs
- 17 galaxies that host AGN identifiable using both optical and mid-IR criteria
- No overlap between either the optical or mid-IR AGN and LERGs

On average, the NLAGN have a WISE colour of ~ 0.35, and the LERGs have a typical WISE colour of ~ 0.12. My choices of criteria for optical and mid-IR AGN do not by definition preclude the possibility of a galaxy being a LERG, and such dual-status objects do appear in the SDSS DR7 and WISE catalogues. They are rare, however, and as a result none appear in the visually confirmed post-merger sample. The mass, redshift, and CNN $p(x)$ statistics for the sample of SDSS DR7 galaxies appearing in CFIS DR2 (from which the post-merger sample is derived), the visually confirmed post-merger sample, and the three AGN types are shown in Figure 5.1.

5.3 Results

Having outlined the new galaxy selection criteria used in my AGN study, I next present the main results of this Chapter – the relative frequencies of three types of AGN in mergers relative to controls, and the relative luminosities of narrow-line AGN in mergers compared to those found in controls.

5.3.1 AGN in post-mergers

In the target samples and the matched control samples, the NLAGN fraction is the number of BPT AGN above the Kewley et al. (2001) and Kewley et al. (2006b) lines with S/N ≥ 5 divided by the total number of target or control galaxies. The WISE AGN fraction is the number of galaxies with $W1 - W2 > 0.5$ divided by the total number of target or control galaxies. For the WISE mid-IR excess, the target samples and control pools are restricted to galaxies with WISE detections, even though the matching parameters are derived from SDSS spectra. The LERG fraction is the number of LERGs in each sample divided by the total number of target or control galaxies. For the LERG excess I require that the target and control pool galaxies are classified in the Best & Heckman (2012) catalogue as either star-forming galaxies, HERGs, or LERGs. All excesses are the ratio of the AGN fraction in the target sample and the AGN fraction in the relevant matched control sample.

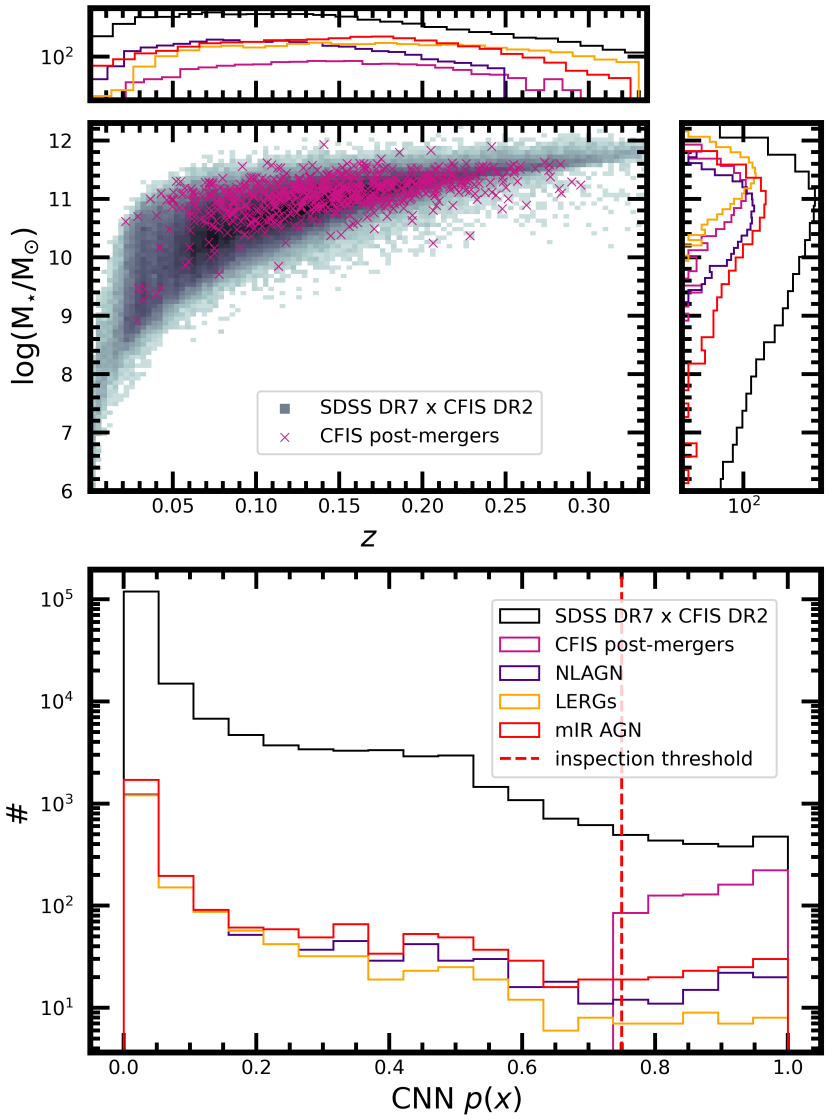

Figure 5.1: The mass, redshift, and CNN $p(x)$ statistics for galaxies relevant to this Chapter: NLAGN (indigo), mid-IR AGN (red), and LERGs (orange). The statistics for the AGN samples are shown alongside the sample of SDSS DR7 galaxies appearing in CFIS DR2 (greyscale background histogram in the top main panel and black auxiliary histograms) and the visually confirmed post-merger sample (magenta crosses in the top main panel and magenta histograms) for reference. The CNN $p(x)$ statistics for the parent, merger, and AGN samples are shown with the same colour code in the bottom panel.

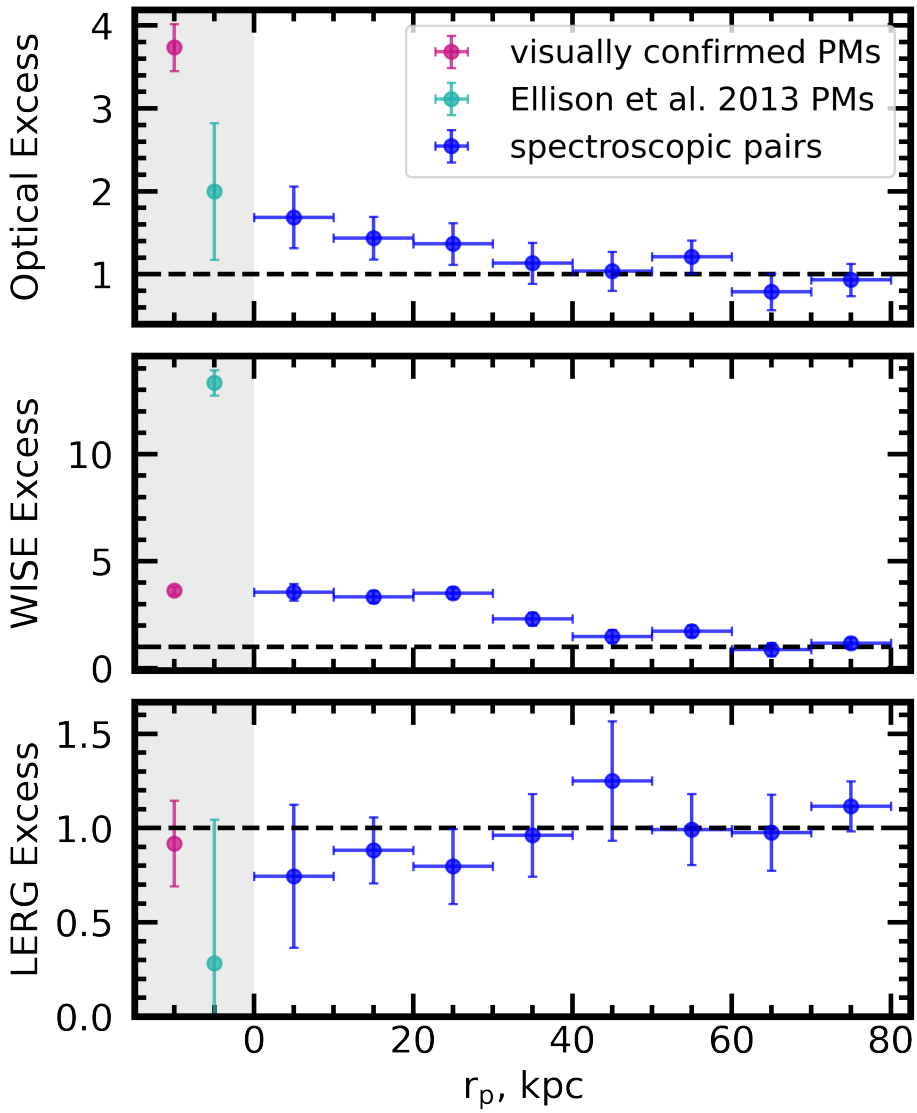

Figure 5.2: Narrow-line (top), mid-IR (centre), and LERG (bottom) AGN excesses in SDSS spectroscopic pairs (blue), the Ellison et al. (2013) post-merger sample (teal), and the visually confirmed post-merger sample (magenta). Vertical errors are calculated by adding the inverses of the binomial errors on each fraction $\sqrt{f(1-f)/N}$, where f is the AGN fraction in the target or control sample, and N is the size of that sample. Horizontal errors are the bin widths. The visually confirmed, more inclusive post-merger sample finds a stronger NLAGN excess, and a weaker mid-IR AGN excess. The excess discrepancy is consistent with the hypothesis that more minor mergers, which are present in the visually confirmed sample and largely absent from the Ellison et al. (2013) sample, are less likely to be dust-obscured after the merger. I find LERG excesses for post-mergers and galaxy pairs consistent with unity (i.e., a null result). Even with a relatively simple control-matching scheme, the merger sequence does not appear to be strongly connected to LERG status.

NLAGN excess

The top panel of Figure 5.2 shows the NLAGN excesses for spectroscopic pairs (blue) in 8 bins of projected separation between 0 – 80 kpc, as well as for Ellison et al. (2013) post-mergers (teal) and the visually confirmed post-merger sample (magenta). I uncover a decreasing trend of NLAGN excess with increasing r_p, peaking with an excess of $\sim$ 1.7 for galaxy pairs separated by $\leq$ 10 kpc. Using a more tolerant NLAGN criterion (Stasińska et al. 2006, which allows for the inclusion of composite galaxies with significant contributions from both star formation and AGN), and a different control matching algorithm (allowing for different numbers of controls to be selected for galaxies belonging to the same target sample), Ellison et al. (2011) still reported quantitatively consistent pair-phase excesses, from $\sim$ 0–2.5 for galaxy pairs spanning separations from 80 down to 0 kpc. I recalculate the post-merger NLAGN excess in the Ellison et al. (2013) sample in order to compare using my updated control-matching method, and find $\sim$ 2 times as many NLAGN compared to controls. I find that NLAGN are even more common in the visually confirmed post-merger sample, with an excess of 3.7.

Mid-IR AGN excess

There are 6106 SDSS pairs, 364 visually confirmed post-mergers, and 78 Ellison et al. (2013) post-mergers for which I can calculate a WISE mid-IR colour using the legacy WISE all-sky catalogue. The middle panel of Figure 5.2 shows the excess of mid-IR AGN with $W1 - W2 > 0.5$ in the same three main galaxy samples as in Section 5.3.1 compared to matched control samples, which are themselves required to have WISE detections so that their WISE $W1 - W2$ colours are calculable. Much like the NLAGN excess, I find a decreasing excess of mid-IR AGN in bins of r_p with a peak value of $\sim$ 3.6 for the closest spectroscopic pairs. The excesses are statistically consistent with the pair phase $W1 - W2 > 0.5$ results reported in Satyapal et al. (2014), even though a different control matching methodology was used. The new visually confirmed post-merger sample shows a mid-IR AGN excess consistent with the closest pairs, but significantly below what is reported in the Ellison et al. (2013) sample, which contains 13.3 times as many WISE AGN per capita compared to the control sample. The pair phase and visually confirmed post-merger sample excesses for both optical AGN and mid-IR AGN are in reasonable concordance (spanning $\sim$ 0 – 4 in the pair phase, and 3 – 4 in the post-mergers). The agreement supports the view that the optical and mid-IR AGN diagnostics are identifying different observational phenomena associated with the same SMBH engines.

In order to investigate the discrepancies in optical and mid-IR AGN excesses between the visually confirmed post-mergers and the Ellison et al. (2013) post-mergers, I consider the dif-

ferences in the selection methods used for their identification. Because the CNN, Galaxy Zoo classifications, and expert visual classifications all rely on morphology, I posit that a morphological bias may be responsible – specifically, that the strength of morphological disturbance bears on a galaxy's likelihood to be observed as a dust-obscured AGN. In order to investigate differences in sample morphology, I use shape asymmetry quantities computed for the sample in SDSS imaging by Scott Wilkinson. While CFIS imaging could have been used (by definition) for the CFIS post-merger sample, SDSS is used because CFIS imaging is not available for every galaxy in the Ellison et al. (2013) post-merger sample. Shape asymmetry is a non-parametric morphological measurement that takes the asymmetry of a binary mask that denotes a boundary between the galaxy and the background, detailed in Pawlik et al. (2016). The binary mask used to measure shape asymmetry is generated following the 8-connected structure detection method described in Pawlik et al. (2016), in which a galaxy image is smoothed using a 3×3 running average filter, and pixels above a limiting surface brightness of approximately 24.7 mag arcsec^{-2}, equivalent to one standard deviation above the typical sky noise level in SDSS imaging, are ascribed to the galaxy. Shape asymmetry was developed for the purpose of automated merger identification because of its emphasis on the particular asymmetry of low-surface brightness features whose importance would be overlooked by a traditional asymmetry measurement. The merit of shape asymmetry for merger identification is explored in detail in Wilkinson et al. (2024). By measuring shape asymmetry on galaxies whose post-merger status is already confirmed, shape asymmetry instead measures the morphological strength of the post-merger. Tidally disturbed post-mergers that have experienced dramatic, major mergers are more likely to have more extended morphologies, and higher shape asymmetries compared to those whose mergers have been relatively tame. Shape asymmetry has also been shown to fade in the several hundred millions of years that follow coalescence (Pawlik et al. 2016), but so too has it been shown that the $W1 - W2$ colour decreases after coalescence (Blecha et al. 2018). Shape asymmetry therefore also includes information about the recency of the merger, along with the initial intensity of the morphological disruption.

Figure 5.3 shows the shape asymmetry demographics of the two post-merger samples. The top panel shows the normalized distributions of shape asymmetry derived from SDSS imaging for the entire SDSS DR7 galaxy population with spectroscopic redshifts < 0.35, and masses > $10^{8.5}$M$_\odot$ (grey, representing the area of M$_\star$–z parameter space encompassing both the Ellison et al. 2013 and visually-confirmed post-merger samples, see Figure 3.8), the Ellison et al. (2013) post-merger sample (teal), and the new visually confirmed sample (magenta). It is important to note that shape asymmetry does not trend strongly with either stellar mass or redshift, and that the qualitative results of the shape asymmetry study do not change when I compare the merger samples to their matched control galaxies from Sections 5.3.1 and 5.3.1 instead of the SDSS parent sample used

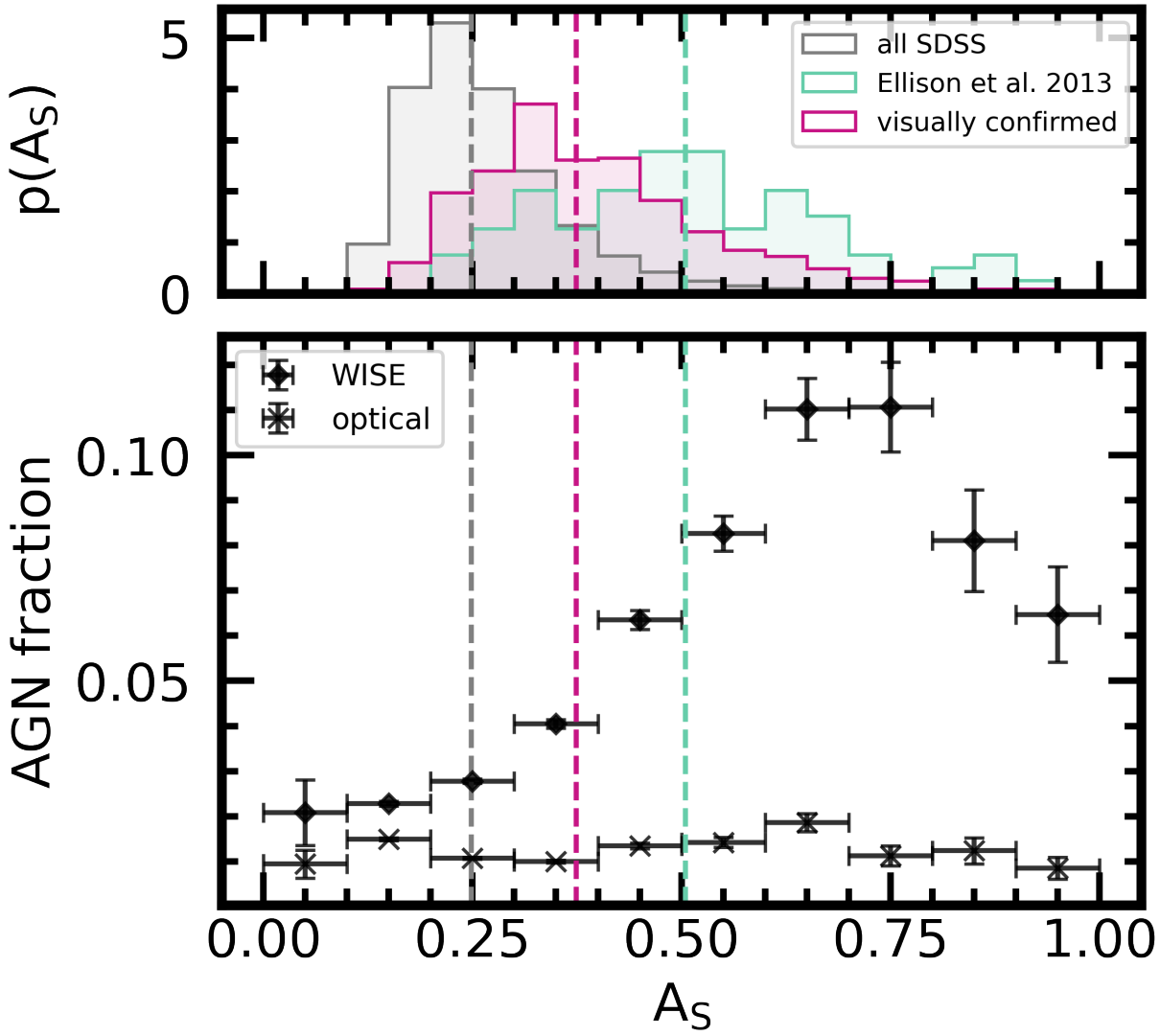

Figure 5.3: Shape asymmetry statistics derived from SDSS imaging for SDSS galaxies with $z < 0.35$ and $M_\star > 10^{8.5} M_\odot$ (grey), the Ellison et al. (2013) post-mergers (teal), and the visually confirmed post-mergers (magenta, top panel). The vertical bars spanning both panels represent the median of each distribution. While both more disturbed than the typical SDSS galaxy, the visually confirmed post-merger samples are less disturbed on average than the Ellison et al. (2013) mergers, suggesting that a greater number of minor mergers are included. The difference in disturbance would be a natural consequence of the CNN's inclusive simulated training set. The bottom panel shows the AGN fractions in the SDSS sample as identified by NLAGN (X-markers) and mid-IR (diamonds) criteria, see Section 5.1 for details. The vertical errors are the binomial errors on each fraction $\sqrt{f(1-f)/N}$, and the horizontal errors are the bin widths. Since the hybrid method typically selects galaxies with smaller shape asymmetries, it also selects a proportionally lower fraction of mid-IR AGN. The effect presented here accounts for the discrepancy between the two post-merger data points in Figure 5.2.

here. The median SDSS-derived shape asymmetry of each sample is plotted over both panels as a dashed line of the same colour. Shape asymmetries derived from CFIS r-band imaging are available for the visually confirmed post-mergers, but I present only SDSS shape asymmetries in order to allow for direct comparison to the Ellison et al. (2013) sample, which does not appear in full in CFIS. Note that the shapes of the visually confirmed post-merger sample, while of course more asymmetrical than SDSS in general, are significantly less disturbed than the Ellison et al. (2013) post-mergers, with an average difference of $\Delta A_S \sim 0.13$. I posit that the typical difference in SDSS-derived shape asymmetry between the two post-merger samples is the result of the fact that the Ellison et al. (2013) post-mergers were identified by eye in shallow SDSS imaging. Conversely, the visually confirmed post-mergers were first identified by a CNN trained on CFIS-depth simulated imaging of post-mergers with mass ratios as small as 1:10. It is therefore plausible that a number of relatively minor post-mergers were preserved by the CNN and confirmed during visual classification.

The bottom panel of Figure 5.3 shows the local NLAGN and WISE AGN fractions of the SDSS parent sample (with $z < 0.35$ and $M_\star > 10^{8.5} M_\odot$) in ten bins of A_S between 0 and 1. I find that the NLAGN fraction is generally low and consistent with increasing shape asymmetry. While the WISE AGN fraction does not trend monotonically with A_S, the data show that more morphologically disturbed galaxies in SDSS are indeed more likely to host an AGN that is identifiable by its mid-IR colour. These results indicate that the degree of disturbance is unlikely to have a strong impact on the NLAGN fraction, but the same is not true for mid-IR AGN, since more disturbed galaxies typically have higher mid-IR AGN fractions. Consequently, the Ellison et al. (2013) sample is more likely to contain highly disturbed post-mergers, which host proportionally more mid-IR AGN and a consistent number of NLAGN, while the visually confirmed sample is more inclusive of less-disturbed mergers, which are less likely to host mid-IR AGN.

If the degree of morphological disturbance is responsible for the difference in mid-IR AGN excess between the two post-merger samples, a subset of the visually confirmed post-mergers with the same SDSS shape asymmetry demographics as the Ellison et al. (2013) post-mergers ought to exhibit an excess that is in better agreement. In order to test this hypothesis, I match exactly one galaxy (without replacement) from the visually confirmed post-merger sample to each Ellison et al. (2013) post-merger on shape asymmetry. Where multiple visually confirmed post-mergers have shape asymmetries within ± 5 per cent of an Ellison et al. (2013) post-merger, I select the single best match. Where there are no matches, I grow the tolerance from 5 per cent until a single match can be found. Of the 85 galaxies in the Ellison et al. (2013) sample with shape asymmetries available, 82 have a match within 5 per cent of their shape asymmetry in the visually confirmed sample. The remaining 3 galaxies require 2, 3, and 5 growths, respectively; they are included for

completeness but their exclusion does not affect the results. The shape asymmetry-matched visually confirmed post-mergers have a NLAGN excess consistent with the visually confirmed sample taken as a whole, but their mid-IR AGN excess is increased from 3.6 up to 6.5. While still not in perfect agreement with the Ellison et al. (2013) sample, this experiment confirms that the degree of morphological disturbance ($\Delta A_S \sim 0.13$) and sample selection are linked to an increased likelihood of a mid-IR AGN detection. The change in the excess result is consistent with the physical narrative presented in Yutani et al. (2022), wherein rapidly accreting AGN in extremely recent (within $\sim$ 4Myr) post-coalescence systems are more likely to be observed as dust-obscured galaxies (DOGs) on account of central and/or galaxy-scale dispersion of dust from the progenitor galaxies. I posit that the longevity and intensity of the dust obscuration may scale with the dynamic intensity of the merger. As a result, the specific method used to identify mergers based on their morphology has a significant impact on the quantitative excesses calculated.

LERGs in post-mergers

The bottom panel of Figure 5.2 investigates the role of mergers in triggering LERGs. In my mass- and redshift-matched study, I find that LERGs are no more likely to exist in the post-merger or pair samples than in controls. The lack of a LERG signal is qualitatively discrepant with the Pace & Salim (2014) finding that galaxies hosting radio AGN have a 50 per cent excess in the number of satellites, and the link between tidal forces associated with pair phase interactions and LERG incidence suggested by Sabater et al. (2013). Ellison et al. (2015) find a modest pair phase LERG excess of 3.8 ± 0.4, and a small excess of $\sim 4\pm2$ (nearly consistent with unity as well) in the post-merger phase when they match controls on stellar mass and redshift. The lack of an elevated LERG incidence rate in post-mergers or galaxy pairs in this work is therefore in mild tension with the literature.

5.3.2 Optical AGN accretion rate enhancements

In addition to the initial triggering of AGN, I can use the post-merger sample to determine typical merger-induced accretion rate enhancements in NLAGN, using [OIII]λ5007 luminosity as a proxy for accretion rate (see also Kauffmann et al. 2003; Brinchmann et al. 2004; Chen et al. 2009; Liu et al. 2012). As stated in Section 5.2.1, I again use a S/N criterion of at least 5 for the four BPT emission lines, and explicitly disallow LIERs and/or LINERs using the [SII]λ6717 and [OI]λ6300 criteria of Kewley et al. (2006b). I ultimately include 31 post-mergers from the visually confirmed sample, 3 from the Ellison et al. (2013) sample, and 263 SDSS pairs in bins of separation between 0 – 80 kpc with mass ratios $0.1 \leq \mu \leq 10$ selected from the Patton et al. (2016) catalogue with

non-LINER NLAGN and S/N of at least 5 on all emission lines used for placement on the BPT diagram.

Figure 5.4, which shows $L_{[OIII]}$ (top panel) and $\Delta\log(L_{[OIII]})$ (bottom panel) for NLAGN in spectroscopic pairs, the Ellison et al. (2013) post-mergers, and the visually confirmed post-merger sample, suggests that accretion rates in NLAGN hosted by galaxies with a close companion are consistent with isolated NLAGN. Conversely, both post-merger samples are significantly enhanced compared to the matched non-merger NLAGN control ensembles. The visually confirmed post-merger sample is in fact ~ 2 times as luminous in [OIII]λ5007 on average compared to non-post-merger controls. The accretion rate enhancements for NLAGN ushered in by coalescence therefore appear to be significant, while those produced by pair-phase interactions may just as likely be produced by secular or ambient processes. A higher positive enhancement of ~ 1.4 dex is found for the three post-mergers that remain after applying my quality control cuts to the Ellison et al. (2013) post-mergers. While the sample size does not invite extensive interpretation, it is possible that the increased morphological disturbance that is typical of the Ellison et al. (2013) sample is linked to more rapid gas infall and elevated accretion rates in these systems.

The relationship between star formation rate and $L_{[OIII]}$ is approximately proportional (with significant scatter) in BPT star-forming galaxies beneath the Kauffmann et al. (2003) criterion with S/N>5. Therefore, in order to achieve similar $\Delta\log(L_{[OIII]})$, star-forming galaxies would need to have approximately doubled star formation rates (SFRs). Still, since I select galaxies whose nebular emission could not plausibly be produced by star formation alone, the [OIII]λ5007 luminosity enhancements uncovered in this Section are primarily indicative of SMBH accretion-driven ionization.

The result does not appear to be an effect of the SDSS fibre aperture on the Ellison et al. (2013) sample, which lies at low-z relative to the visually confirmed sample (see Figure 3.8) as I find no correlation of $\Delta\log(L_{[OIII]})$ with z for individual galaxies in the visually confirmed sample. In the sample of Kewley et al. (2001) AGN with S/N$\geq$ 5 for the four BPT emission lines and stellar masses between $10^{10-12}M_{\odot}$, the median measured [OIII]λ5007 luminosity actually increases from $10^{39.2}$ to $10^{41.6}$ erg s^{-1} between $0 \leq z \leq 0.25$. The fibre aperture effect therefore gives rise to higher [OIII]λ5007 luminosities at higher z, and moreover, my control-matching methodology accounts for systematic changes in [OIII]λ5007 luminosity with stellar mass and redshift. Aperture effects (or more broadly, any redshift or stellar mass effects) are not responsible for the difference in $\Delta\log(L_{[OIII]})$ calculated between the Ellison et al. (2013) and visually confirmed post-mergers.

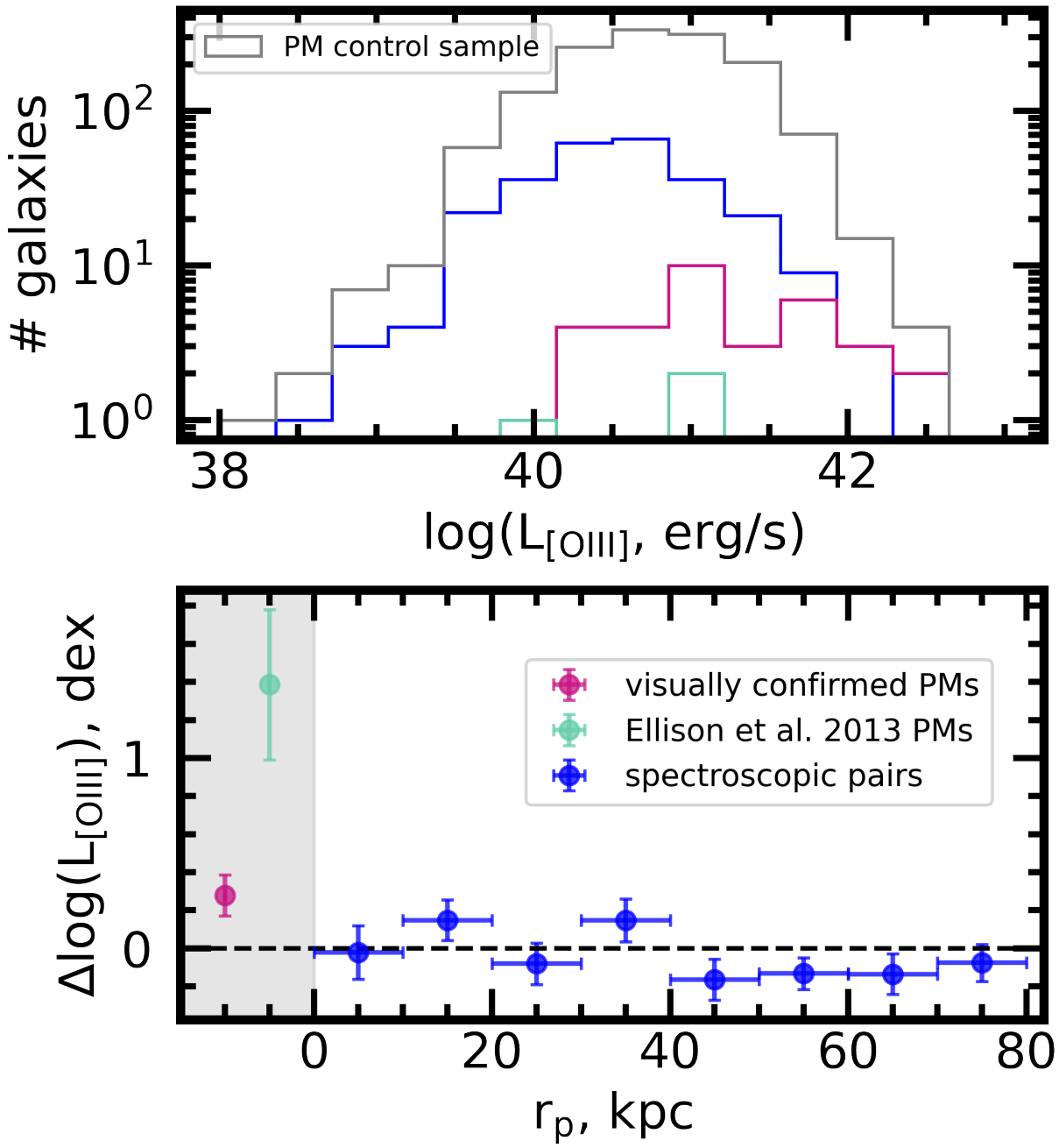

Figure 5.4: [OIII]λ5007 luminosities and luminosity enhancements in post-mergers and pairs. The top panel shows log-scale [OIII]λ5007 luminosity histograms for NLAGN in the the galaxy pair sample described in Section 5.1.2 (blue), the visually confirmed post-merger sample (magenta), the Ellison et al. (2013) post-merger sample (teal), and the NLAGN control pool for the visually confirmed post-merger sample (grey). The bottom panel shows $\Delta\log(L_{[OIII]})$ for the same three target samples (galaxy pairs, visually confirmed post-mergers, and Ellison et al. 2013 post-mergers). Vertical error bars are the statistical error on the median, $1.253\sigma/\sqrt{N}$, and horizontal error bars are the bin widths. I find enhancements approximately consistent with zero in the pair phase, with small local suppressions past 40kpc. In both post-merger samples, I find significant positive excesses. Post-mergers in the visually confirmed sample are $\sim$ 2 times as luminous in [OIII]λ5007 as their non-post-merger controls.

5.4 Discussion and conclusions

In this Chapter, I have used the CNN-identified and visually confirmed post-merger sample introduced in Chapter 3 to study the triggering and accretion of supermassive black holes in post-merger galaxies. I also offer pair-phase results in order to contextualize the post-merger results within the merger sequence. I match control (non-post-merger or non-pair) galaxies on $M_\star$ and z to the target (either post-merger or pair) galaxies in order to study the AGN excess – that is, the ratio of the AGN fractions in the target sample and the control sample – using optical narrow-line region (via SDSS, the BPT diagram, and the Kewley et al. 2001 NLAGN selection), mid-IR (via WISE and the colour criterion $W1 - W2 > 0.5$ used by Satyapal et al. 2014 to select dust-obscured AGN), and LERG classifications (capturing the low-excitation, isotropic SMBH accretion mode, as determined by Best & Heckman 2012).

The excess results presented in Section 5.3.1 suggest a connection between available gas, merger events, AGN triggering, and AGN observability, which is in qualitative agreement with the literature. Using different BPT diagram criteria to select AGN, Ellison et al. (2013) compute an excess of 3.75, in good agreement with the excess of 3.7 result of this work. Satyapal et al. (2014) use the same $W1 - W2$ colour criterion as I do to select mid-IR AGN, and report a mid-IR AGN excess of $\sim$ 11 in the Ellison et al. (2013) post-merger sample. While the analogous computation in this work is quantitatively discrepant (with an excess $\sim$ 13 for the same post-merger sample), both results are consistent with a striking statistical connection between coalescence following a visually dramatic major merger, the triggering of AGN, and dust obscuration. The cause of the lower mid-IR AGN excess I calculate for the visually confirmed CFIS post-merger sample is discussed earlier in Section 5.3.1. For a sample of galaxy mergers (including systems with multiple nuclei, which are nominally excluded here), Weston et al. (2017) report an excess range of 5 – 17, which comfortably encompasses my results for the Ellison et al. (2013) sample, but lies above my results for CFIS post-mergers. The discrepancy between Weston et al. (2017) and this work could also be attributed to elevated levels of morphological disturbance in the Weston et al. (2017) sample, which was selected by eye from SDSS imaging. Goulding et al. (2018) report a mid-IR excess of 2 – 7 in mergers compared to non-interacting mergers, which is in agreement with my findings, while Gao et al. (2020) report lower (but still > 1σ significant) AGN excesses in merger samples identified using the Pearson et al. (2019a) CNN. Using visual classifications alone to identify their merger sample, Li et al. (2023b) find AGN excesses ranging from 2 – 4 depending on the specific AGN diagnostic, again in good agreement with my results for optical and mid-IR AGN in the CFIS post-merger sample. Comerford et al. (2024) report a multi-wavelength AGN excess in major mergers (identified using non-parametric morphological statistics combined via

LDA) of 1.8. The fact that an AGN excess is recovered by so many publications using different merger and AGN selection techniques suggests that the underlying connection between late-stage galaxy mergers and AGN is strong. Still, the specific AGN excess measurements reported vary widely. I suggest that for a given AGN diagnostic, the variation is driven by the differences between merger identification techniques, each of which bears a unique combination of accuracy, precision, and biases. The sensitivity of the measured AGN excess to the method of AGN identification is explored in greater detail in Chapter 6.

The elevated incidence rate of mid-IR AGN in mergers with more dramatically disturbed morphologies (see Figure 5.3) is particularly noteworthy, since it suggests that merger events modulate both the underlying AGN fraction, as well as the frequency with which an AGN can be detected at a given wavelength or using a particular AGN diagnostic. The connection between the merger sequence and AGN observability is explored to a greater extent in the next Chapter, in which I present the detectability of AGN with multiple criteria as a function of merger stage.

In addition to my census of AGN in post-coalescence galaxies, I can also revisit the merit of a hybrid (CNN and visual) post-merger identification framework, which has allowed me to improve on the statistics of literature studies of post-mergers, and propose revisions to other results whose quantities were more heavily influenced by the selection functions of their merger identification method.

Chapter 6

AGN triggering and obscuration in DECaLS post-mergers with eROSITA

In the previous Chapter, I reported a number of trends that are qualitatively consistent with the current literature on the role of galaxy mergers in fueling the growth of SMBHs in the centres of galaxies: namely, that AGN are triggered more frequently as galaxies move through the merger sequence, with AGN being identified most often near the time of coalescence. I also found new observational evidence for a connection between merger stage, the strength of the merger, and the likelihood of an AGN being detected in the mid-IR. In this chapter, I will use new observations from the eROSITA X-ray space mission to characterize the incidence and multi-wavelength characteristics of AGN in galaxy mergers, and dissect the observability of AGN in specific wavelength bands as a function of merger stage. The main results of this Chapter are presented in Bickley et al. (2024), and here I also detail additional experiments for context.

6.1 Data and methods

The work presented in this Chapter is mainly an X-ray focused study of ongoing and completed galaxy mergers, supplemented by additional multi-wavelength observations. In this Section, I describe the galaxy sample studied, as well as the selection methods for post-mergers, galaxy pairs, control pool galaxies, and multi-wavelength AGN.

6.1.1 Galaxy samples

The results presented later in Section 6.2 depend on homogeneous availability of multi-wavelength data products from SDSS, WISE, and eROSITA in order to search for signs of AGN. First, I will identify the parent sample of galaxies for which the data products are available. I will next describe my selection of three main phases of the merger sequence relevant to this study – post-

mergers, galaxy pairs, and non-interacting control galaxies. Since the overlap between available data from CFIS and eROSITA is limited, I use DECaLS imaging to identify the post-mergers studied in this Chapter; the merger identification effort is described earlier in Section 3.2.

6.1.2 Parent sample and a new stellar mass estimate

The parent sample of galaxies in this study (from which all subsets in this Chapter are derived) is the same galaxy sample described in Section 3.2.1, since the availability of merger status predictions is essential to this study. The results presented in this Chapter also require redshift (which is available for the parent catalogue via SDSS DR7) and stellar mass measurements in order to match mergers to controls. Before now, a uniform stellar mass estimate was not available for the entire parent catalogue, since the MPA-JHU (Brinchmann et al. 2004) stellar mass catalogue is constructed mainly for galaxies with narrow emission lines. As a result, mass estimates are generally not available from MPA-JHU for galaxies with broad emission lines (see the MPA-JHU inclusion criteria below in Section 6.1.8). Upon inspecting the optical spectra of X-ray AGN from the first data release of the eROSITA all-sky survey (eRASS1; Predehl et al. 2021; Merloni et al. 2024), I found that many of them have broad emission lines. As I will show in Section 6.1.10, there is a significant overlap such that using only the MPA-JHU mass catalogue would lead to an incomplete census of AGN in mergers. Instead, I use new stellar mass estimates based only on SDSS *griz*-band[1] photometry throughout this Chapter so that galaxies with broad emission lines can be included.

The new stellar masses are computed by Samir Salim following the method of Salim et al. (2016). Salim et al. (2016) used the SED fitting code, Code Investigating GAlaxy Emission (CIGALE) (Noll et al. 2009; Boquien et al. 2016) to fit grids of synthetic stellar continuum templates from Bruzual & Charlot (2003) and star formation emission line templates to produce probability density functions (PDFs) for stellar mass ($M_\star$), and a similar procedure is used here.

One potential complication of using photometry for stellar mass estimates is that an unobscured accretion disk or BLR may contribute significantly to the UV and blue optical photometry of a galaxy, biasing the photometric estimates of $M_\star$ from the SED fitting. Salim et al. (2016) allowed for additional photometry (SDSS *u*-band, UV from GALEX, and mid-IR from WISE) to be considered when performing SED fitting for stellar mass and SFR estimates. However, in this work SDSS *u* and UV photometry are excluded on account of potential contamination by emission from the AGN accretion disk, and mid-IR photometry is also excluded from the stellar mass fits in order to avoid biases from AGN-heated dust. To investigate the influence of UV and *u*-

[1] In order of increasing wavelength, SDSS photometric products are available for the *u* (354 nanometers), *g* (475 nm), *r* (622 nm), *i* (763 nm), and *z* (905 nm) photometric bands.

band photometry on stellar mass predictions from CIGALE, Samir Salim made estimates of $M_\star$ with and without UV from the Galaxy Evolution Explorer (GALEX, a UV space telescope; Martin et al. 2005) and *u*-band (from SDSS) photometry. For galaxies with broad-line AGN, he finds that the predictions made including the UV and optical photometry (*FNugriz* bands[2]) are higher by as much as ~ 0.9 dex (on average only ~ 0.01 dex, since the statistics are dominated by broad-line AGN with a weaker total contribution by the BLR to the SED) from those made with *F*, *N* and *u* excluded, whereas the masses of non-broad-line AGN stay consistent to 0.05 dex. Further removal of the *g*-band does not significantly change the masses of broad line AGN (BLAGN), so I conclude that excluding UV+*u* photometry appears to correct for the AGN contamination.

Figure 6.1 illustrates this point. On the horizontal axis, I plot $\log(\chi^2)$, which represents the quality of the fit obtained through SED modelling that includes all of the bands (*F*, *N*, *ugriz*). A high value of $\log(\chi^2)$ indicates that the model developed for the stellar SED is not optimally suited to the observed SED of a galaxy, presumably because of AGN contamination. On the vertical axis, I plot the log difference between the stellar mass estimate obtained using *FNugriz* and *griz* photometry, such that positive values indicate an overestimate of the stellar mass when *FNu* are included compared to using *griz* only. The background 2D histogram is for a subset of parent catalogue galaxies appearing in GALEX, and which are neither NLAGN nor BLAGN according to the criteria set in Sections 6.1.8 and 6.1.7. The average trend and 1σ error regions for non-AGN and NLAGN are shown in orange and indigo, respectively. BLAGN are marked by the black crosses on Figure 6.1, and the average trend and 1σ error region for BLAGN are shown in cyan. Amongst NLAGN, the mass offset is not sensitive to the quality of the fit (i.e., the indigo series is effectively flat), although there is a small zero-point offset between the mass estimates made with and without *FNu*-band photometry. The offset is to be expected, since contribution to the total stellar mass by young, UV-bright stars is underestimated without UV fluxes included. Still, the lack of a trend amongst NLAGN indicates that the quality of the fit does not lead to an over- or under-estimate of $M_\star$ when galaxies fall within the parameters of the modelling assumptions.

For BLAGN, however, the trend turns upward at high $\log(\chi^2)$, corresponding to a population of problematic cases (note the scatter points with high χ^2 on the horizontal axis). In these cases, the large fitting errors indicate that the model (which does not include BLAGN or accretion disk templates) is having difficulty reproducing the observed SED using stellar and nebular emission templates alone. In these cases, the *FNugriz* masses are severely overestimated due to UV contamination by the unobscured SMBH accretion disk. Taken together, the trends shown in Figure 6.1 indicate that *griz* masses are preferable for my study, since they are not as affected by contamination from the accretion disk when it is exposed.

[2]The near-UV *N*-band spans 175 – 280 nm, and the far-UV *F*-band spans 135 – 175 nm.

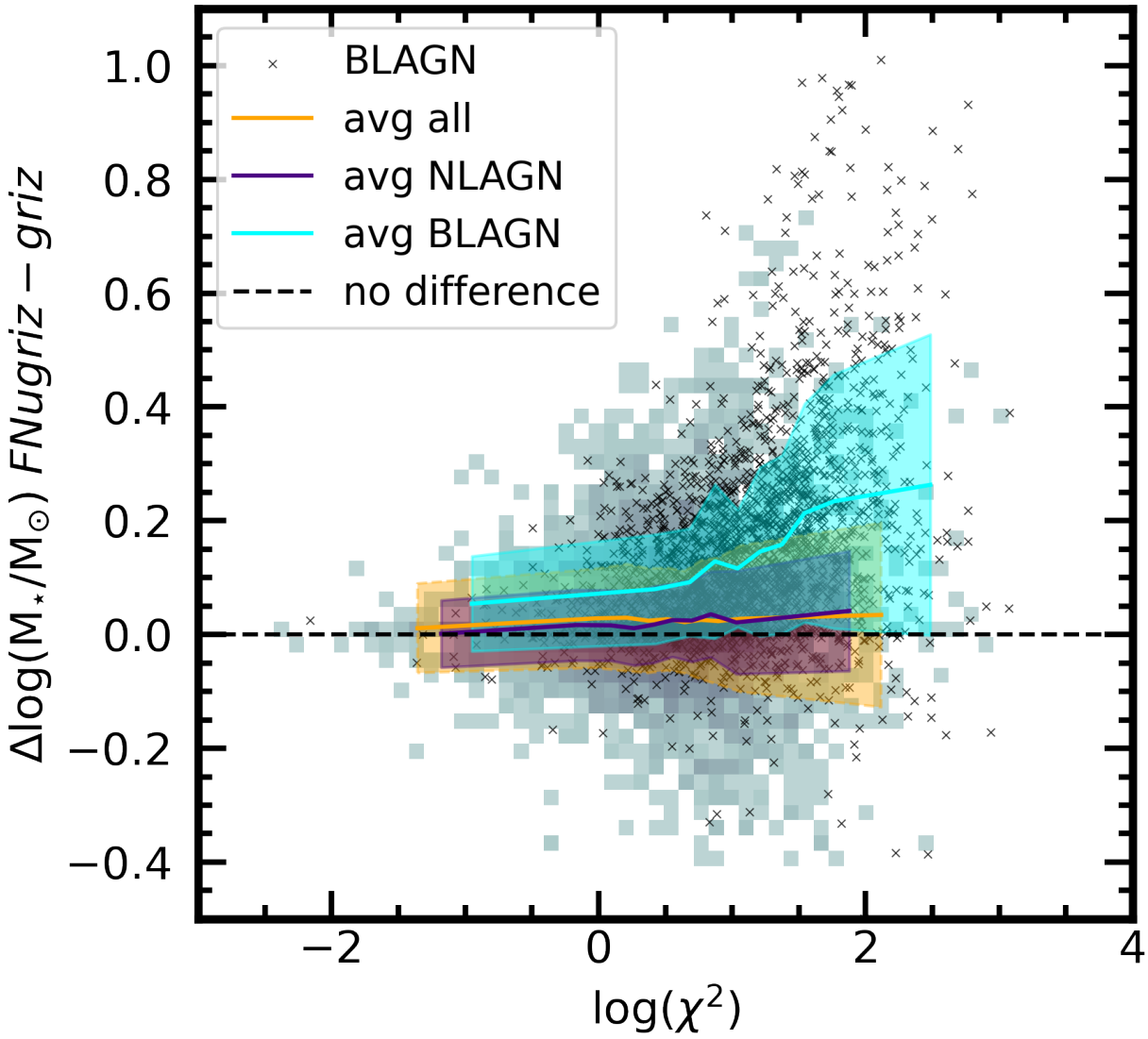

Figure 6.1: A comparison of the stellar mass fitting results using *griz* photometry from SDSS only, and GALEX *FN* plus SDSS *ugriz*. I plot the log-scale difference between the *FNugriz* and *griz* stellar masses against $\log(\chi^2)$, representing the goodness-of-fit for the mass estimate. The 2D histogram in the background shows the extent of the data for a subset of parent catalogue galaxies appearing in GALEX that meet neither my NLAGN nor BLAGN criteria. The mean trend and 1σ region for the non-AGN set are shown in orange. The indigo error region shows the trend for NLAGN, and the black scatter points and cyan data series show the entire BLAGN sample appearing in GALEX, and the mean trend, respectively.

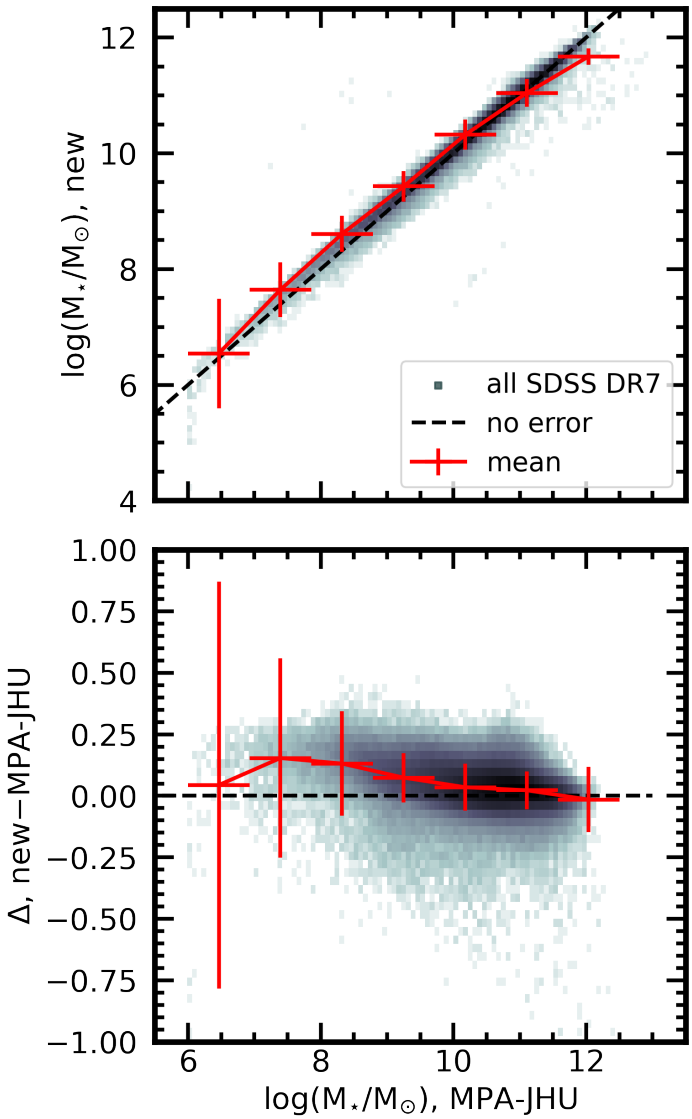

Figure 6.2: A comparison between the stellar mass estimates for galaxies in SDSS DR7 from the MPA-JHU catalogue and the new mass estimates used in this Chapter. MPA-JHU does not report stellar mass estimates for the majority of galaxies with broad emission lines, and broad emission lines frequently appear in the spectra of the X-ray AGN studied in this Chapter. As such, a new photometric $M_\star$ estimate is derived. Top panel: The new mass estimate plotted against the mass from MPA-JHU for galaxies that have both (2-D histogram), the unity line representing no difference between the two (black dashed line), and the median trend (red). The error bars show the standard deviation in the bin. Bottom panel: the difference between the two mass estimates plotted against the MPA-JHU estimate. The two estimates are in 1σ agreement.

I inspect the distribution of $M_{\star,\mathrm{new}}$–$M_{\star,\mathrm{MPA-JHU}}$ and the results of the comparison are shown in Figure 6.2. I find that $\sim$ 88 per cent of galaxies listed in both MPA-JHU and the new $M_\star$ catalogue agree within 0.1 dex. The median difference is $\sim$ 0.03 dex (with the new mass estimate being slightly higher on average), and the standard deviation of the mass difference distribution is $\sim$ 0.08 dex. Figure 6.2 reveals that galaxies with $\log(M_\star/M_\odot) < 9$ in MPA-JHU often have somewhat higher masses in the new photometric mass estimate. Still, since relatively few galaxies in the parent sample lie below $10^9 M_\odot$, the impact of the discrepancy on the results of this study is small, particularly since most galaxies in the post-merger, pair, and X-ray AGN samples have higher stellar masses (see Figure 6.3 below). The broad agreement between the new stellar mass estimates and the MPA-JHU catalogue suggests that the new $M_\star$ catalogue offers an important improvement in the availability of stellar mass estimates without sacrificing precision.

Photometry flags are raised by CIGALE when the SED fitting cannot account for an unusual colour for an object in the catalogue, typically due to imaging artefacts in SDSS. When I remove the sources with photometry flags, the sample only shrinks by 0.05 per cent, and there is no impact on my results. To be included in the parent catalogue, I require that galaxies have $\log(M_\star/M_\odot) > 8$ and no photometry flags raised during stellar mass fitting. The mass cut is chosen to be inclusive of the low-mass wings of the eRASS1-detected sample. There are 264820 galaxies in the parent catalogue with spectroscopic redshifts in the range $0.005 < z < 0.3$ and *griz*-determined stellar masses in the range $8 < \log(M_\star/M_\odot) < 12.75$. Applying an explicit upper limit in $M_\star$, e.g., at $\log(M_\star/M_\odot) < 12$, also has no impact on the statistical conclusions of this work.

Chapters 3-5 use MPA-JHU stellar mass estimates for the analysis, since BLAGN are not studied. The new *griz* masses are used for the analysis only in this Chapter. However, since *griz* stellar mass estimates have since been made for all galaxies with a spectroscopic z in SDSS DR7, I also use them for data visualization purposes the introductory Figures in each Chapter (e.g., Figure 5.1 in Chapter 5) to allow for at-a-glance comparison of the $M_\star$, z, and CNN $p(x)$ statistics of the galaxies studied.

6.1.3 Post-mergers

The post-merger sample used in this Chapter is the same one detailed in Section 3.2.2. As with every galaxy in the parent sample, the data products relevant to AGN characterization used throughout this Chapter are homogeneously available for all 923 of the post-mergers.

6.1.4 Pair sample

For this chapter, I identify a bespoke sample of galaxy pairs in the parent sample (with shared overlap between SDSS DR7, DECaLS-*r*, and eRASS1) using the on-sky coordinates of the galaxies in the parent catalogue, their spectroscopic redshifts from SDSS, and the cosmological parameters assumed in this work (see Section 1.4). I follow generally the method of Patton et al. (2016) with small modifications (described below) in order to identify a more complete catalogue meeting specific criteria relevant to this study. The main departure from the method used in Patton et al. (2016) is that I adopt the new *griz* masses.

For each galaxy in the parent catalogue, I find the closest companion (with the smallest projected separation r_p) whose stellar mass is within a factor of 10 and whose $\Delta v < 300$ km/s. As in Chapter 5, the mass ratio criterion is motivated by the fact that simulated post-mergers with $0.1 < \mu < 10$ were selected to train the CNN used for post-merger classifications. The pairs thus selected plausibly represent pre-coalescence analogues to the post-merger sample. Any galaxy with such an interacting companion closer than 120 kpc in r_p is labelled as a member of a pair for this Chapter. A projected separation of 120 kpc is (at least) twice the separation where previous studies (e.g., Ellison et al. 2013; Bickley et al. 2023, the results of which are presented in Chapter 5) have found an AGN excess and therefore gives a good buffer. Searching for an AGN signal out to wider separations allows for the possibility that X-ray AGN as detected by eROSITA may be triggered under different conditions or at earlier times in the merger sequence. Finally, I require that the galaxy pairs not be included in the post-merger sample described in Section 6.1.3. By these criteria, there are 15485 galaxies with a companion with $r_p < 120$ kpc.

In addition to the sample of closest companions described above, I separately define a sample of "mutual galaxy pairs" (MGPs). MGPs are sets of exactly two galaxies defined by mutual companionship (i.e., galaxy *A*'s tabulated companion in the catalogue is galaxy *B*, and galaxy *B*'s companion is *A*). Mutual pairs are distinct from galaxies belonging to more complex spatial families on the sky (e.g., where galaxy *B*'s nearest neighbour is a third galaxy, *C*, instead of *A*). The sample of MGPs is required in my analysis of AGN in eRASS1, since the resolution of that survey is insufficient to resolve individual galaxies in a close pair. X-ray incidence in galaxy pairs is therefore computed on a pairwise basis (see Section 6.1.5). From the sample of close companions described above, I identify 4565 MGPs (i.e., 9130 galaxies) with projected separations < 120 kpc.

6.1.5 Control pool

Post-merger and pair galaxies are compared to matched controls throughout this work. Control galaxies are selected from those in the parent sample that have no close companion within 1000

km/s in Δv, $0.1 < \mu < 10$ in mass ratio, and $r_p < 120$ kpc. I use these criteria because galaxies with $\Delta v > 1000$ km/s are likely not destined to merge, or at different z entirely. I select control galaxies that are separated from their companions by at least 120 kpc in order to compare the merger samples to controls that are evolving in effective isolation compared to the sample of galaxy pairs. I use the same mass ratio criteria as for galaxy pairs in order to avoid a mass ratio asymmetry in the analysis. Control galaxies must also have CNN $p(x) < 0.1$ predictions from DECaLS imaging. In contrast to Chapters 4 and 5, the galaxy pair and post-merger samples are compared to control galaxies taken from the same control pool. Using the same control pool improves the significance of results wherein the properties of galaxy pairs are compared to post-mergers – since both samples are now compared to control galaxies taken from the same pool, differences between the measured properties of galaxy pairs and post-mergers are a result of the merger criteria only, rather than the merger and non-merger criteria. Informal visual inspection of several hundred control pool galaxies by the author reveals that they are generally non-interacting, with undisturbed elliptical or spiral morphologies.

To construct a control sample for each galaxy in the pair and post-merger samples, I begin by identifying the best simultaneous match in z and $M_\star$ in the control pool. I create normalized $\log(M_\star)$ and z statistics for the control pool by subtracting their mean value and dividing the result by their standard deviation. I then use the normalized $\log(M_\star)$ and z to create a KDTree[3] an algorithm which can be used to efficiently identify a galaxy's nearest neighbours in parameter space. I apply the same normalization to the merger sample, and identify the best control galaxy (without replacement) for each merger. I repeat the process, allowing for as many as 10 unique controls to be matched to each merger. After each batch of controls is matched, I apply a two-sample K-S test to the $M_\star$ and z statistics of the merger and control samples. The K-S test is designed to estimate the likelihood that two samples were drawn from the same parent distribution. In this context, high K-S p-values indicate that the control matching is working well to select control samples that are statistically indistinct from the mergers in $M_\star$ and z. If the K-S p-value falls below 0.9 for either galaxy statistic, I reject the most recent batch of controls and terminate control matching. Typically, 10 batches of controls are successfully matched to the merger (pair, MGP, and post-merger) samples. The most poorly matched control galaxy in this Chapter is offset in $\log(M_\star)$ by 0.7 dex, but more than 99 per cent of galaxies are matched to controls within 0.02 dex in $\log(M_\star)$. The most poorly matched control galaxy in z is offset by 0.01, and more than 99 per cent of galaxies are matched to controls within 0.002 in z.

As in Chapter 5, I find that matching controls on environmental parameters (environment density, neighbour distances) in addition to $M_\star$ and z does not qualitatively change my results,

[3] docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.KDTree.html

even though the methods of selection for post-mergers and galaxy pairs means that they reside in somewhat sparse (for the post-mergers) and dense (for the pairs) environments relative to the average control galaxy. Moreover, the inclusion of additional control-matching parameters worsen the quality of the matched controls in $M_\star$ and z, and typically results in smaller control samples and weaker statistics, since the K-S tests often fail earlier when additional parameters are included. Patton et al. (2013) found evidence for enhanced star formation rates in galaxies experiencing pair-phase interactions at separations as wide as 150 kpc, but there is little evidence for merger-induced AGN triggering at $r_p \sim 150$ kpc. Alternative control methods with a wider r_p exclusion criterion for the control sample are found to have little effect on the results presented, since the additional exclusion criteria for Δv, μ, and CNN $p(x)$ further diminish the likelihood of any pre-merger galaxy pairs with $120 < r_p < 150$ appearing in the control sample.

6.1.6 X-ray AGN sample

While X-ray AGN identified in eROSITA are the main data novelty in this Chapter, I also present an analysis of AGN in the optical and mid-IR, as assessed in the previous Chapter, for comparison. I also include a new analysis of a sample of BLAGN identified homogeneously from the SDSS DR7 sample in order to investigate the connection between galaxy mergers and unobscured AGN. In a departure from Chapter 5, I do not study LERGs any further, since results up to this point have suggested at most a weak connection between LERGs and galaxy mergers (see also Ellison et al. 2015).

The X-ray fluxes used in this work are taken from eRASS1. eRASS1 surveys the entire sky down to a flux limit of 5×10^{-14}erg/s/cm^{-2} in the 0.2–5 keV band (Merloni et al. 2024), corresponding to limiting luminosities of $L_X = 10^{42}$, $10^{42.7}$, and $10^{43.1}$ erg/s at $z = 0.1$, 0.2, and 0.3, respectively. The typical eROSITA point spread function (PSF) full width at half maximum (FWHM) is ~ 30 arcsec. X-ray detections considered in this work have S/N> 5, and spurious detections and extended sources have been removed in advance.

eRASS1's ability to detect AGN is sensitive to obscuration, and I anticipate that Compton-thick AGN (typical column densities $N_H > 10^{24}$/cm^2 e.g., in Corral et al. 2019) and some Compton-thin obscured AGN (where the obscuration is still sufficient to suppress soft X-rays, $N_H \gtrsim 10^{22}$/cm^2) will not be detected (Liu et al. 2022). More gently obscured AGN may still be detected in eROSITA, but with attenuated fluxes. Since highly luminous AGN are rare, a relatively large proportion of the sample is likely intrinsically faint, and thereby close to the eRASS1 detection limit (i.e., observed in eRASS1 with only a few counts). Consequently, the most common effect of flux suppression is to remove sources from the sample. I therefore expect that heavily obscured AGN

are more often absent from the X-ray sample than they are detected through the obscuration. Still, the relative contributions of intrinsic luminosity and obscuration are not observationally discernible in the eRASS1 sample.

Optical counterparts for the X-ray sources are identified probabilistically from multi-band Legacy Survey Data Release 10 (LS10) imaging, based on the Bayesian method outlined in Salvato et al. (2018). The counterparts used in this work are identified by Mara Salvato, and presented in Salvato et al. (in prep). Candidate counterparts within the positional uncertainty of the X-ray source are each assigned a probability p_i of being the correct counterpart of the X-ray detection, and are included in this work if the source with the maximum p_i for a given X-ray detection appears in the parent catalogue described in Section 6.1.2. Maximum likelihood fluxes from the eRASS1 catalogue are converted to luminosities based on the SDSS spectroscopic redshifts of the parent sample galaxies. eRASS1 is able to detect AGN with observed luminosities $> 10^{43.1}$ erg/s for any galaxy in this study (with $z < 0.3$). For a typical redshift (z=0.1) in my sample, eRASS1 is more than an order of magnitude more sensitive.

AGN are not the only source of X-ray emission in galaxies, with some binary star systems also producing significant emission (Ranalli et al. 2003; Lehmer et al. 2019). However, at the sensitivity of eRASS1, the X-ray sources detected and associated with galaxies in the parent catalogue are highly luminous, and therefore unlikely to be the product of X-ray binaries in young stellar populations in the host galaxies. For example, only six of the 3718 (0.2 per cent) eRASS1-detected galaxies have MPA-JHU SFRs in excess of the Ranalli et al. (2003) L_X criterion adapted by Agostino & Salim (2019), who parameterize the correlation between the 0.5 – 10 keV X-ray luminosity produced by star formation (in erg s^{-1}) and the SFR (in $M_\odot$ yr^{-1}) as:

$$\mathrm{SFR} = 0.66 \times 10^{-40} L_X. \tag{6.1}$$

Only one of the six cases is a post-merger, and it is an AGN by both my mid-IR (see Section 6.1.9) and narrow-line (Section 6.1.8) AGN criteria. I therefore count any X-ray source in the parent sample as an X-ray AGN, and note that removing the small number of galaxies with SFRs in excess of the criterion given in Equation 6.1 from the X-ray AGN sample does not impact the results. There are 3718 galaxies with X-ray detections in the parent catalogue. Of the X-ray AGN detected galaxies, 58 are post-mergers, 232 are in an interacting pair, and 49 are in an MGP.

Figure 6.3 shows the stellar mass (left panels) and redshift (right panels) distributions of the parent catalogue (grey series), the X-ray AGN sample (magenta), post-merger sample (yellow), pair galaxy sample (blue), and MGP sample (blue dashed). I also show the statistics for the sample of post-mergers, pairs, and MGPs that host X-ray detections in the filled histograms appearing

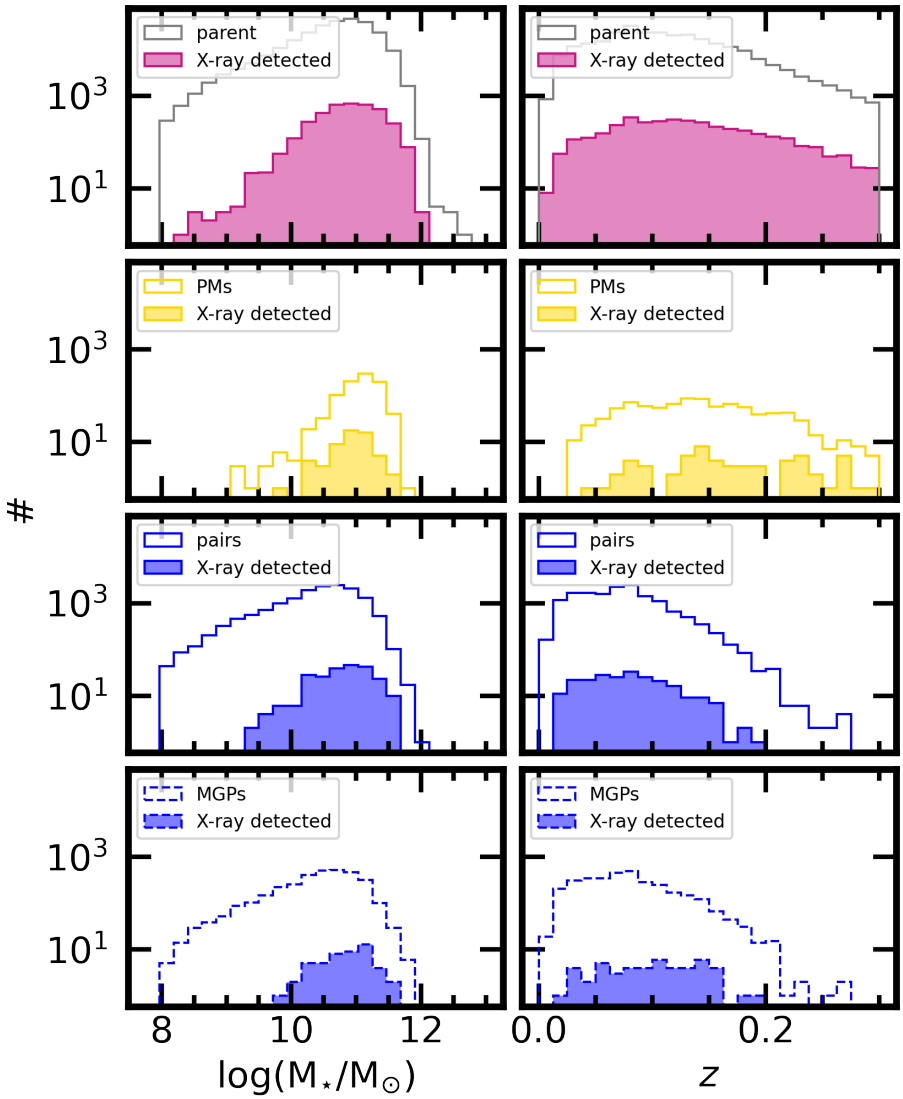

Figure 6.3: Stellar masses (left column) and spectroscopic redshifts (right column) for galaxies in the parent sample (top row, grey histograms), further subdivided into X-ray AGN from eRASS1 (top row, magenta histograms), post-mergers (second row, yellow), galaxies belonging to pairs meeting my criteria for this Chapter (third row, blue), and galaxies in mutual galaxy pairs (fourth row, blue dashed). For post-mergers, pairs, and MGPs, I also show the statistics for the X-ray detected subsets in the same panels (filled histogram series). Post-mergers and X-ray AGN both lie at preferentially high stellar masses and redshifts due to the volume-limited nature of SDSS at high luminosity. Pairs and MGPs, meanwhile, are typically found at lower-z since fibre collision is less common, and galaxies in close pairs can be more reliably distinguished from one another.

in each panel. Relative to the parent sample, galaxy pairs tend to appear more often at lower z. The z bias for galaxy pairs is due to the fibre collision effect of spectroscopic targets in SDSS DR7: galaxy pairs at close physical separations are more widely separated on the sky at low-z, and the participant galaxies are more likely to be targeted individually for spectroscopy since the SDSS DR7 fibres are separated by a minimum of 55 arcsec (York et al. 2000; Patton & Atfield 2008; Patton et al. 2016). The stellar mass distribution of pair galaxies traces the parent catalogue well. Meanwhile, the eRASS1 and post-merger samples lie at relatively high z and $M_\star$ compared to the parent sample. In both cases, the bias has to do with the volume-limited nature of the relevant survey at high luminosities. The morphological features associated with a major merger event are likely to be brighter on the sky when the galaxy is more intrinsically luminous, galaxies with high stellar masses are more likely to be identified as post-mergers, and one needs to look out to higher z to find them (the same effect explored in Section 3.1.3). An analogous selection bias results in the preferential appearance of highly X-ray luminous sources in eRASS1 at higher z.

Figure 6.4 presents $M_\star$ and z statistics for the parent catalogue (background histogram) and for eRASS1 sources (data points colour-coded by L_X). The relative positions of the parent catalogue and eRASS1 sample on these axes highlights the difference between the SDSS DR7 and eRASS1 selection functions: at high redshifts, proportionally more X-ray AGN are found in galaxies with intermediate masses, $\log(M_\star/M_\odot) \sim 10-11$. That the populations do not trace one another perfectly suggests that X-ray AGN in eRASS1 are preferentially hosted by galaxies in this mass domain. Still, in Section 6.2, all results account for such biases by performing rigorous matching between merger and non-interacting control samples on $M_\star$ and z.

Pairwise treatment of X-ray detections in galaxy pairs

When assessing X-ray AGN incidence in galaxy pairs, I expect that close pairs are likely to appear as crowded within the $\sim$ 30 arcsec eROSITA PSF (which corresponds to a projected physical distance of $\sim$ 57 kpc at $z = 0.103$, the median redshift of the parent catalogue studied in this Chapter). I must therefore count X-ray AGN status on a pairwise basis, rather than considering the properties of individual galaxies. I study the sample of MGPs whenever X-ray incidence is considered in the context of galaxy pairs. If either or both MGP galaxies have been assigned to an X-ray source in the eRASS1 catalogue, I count the MGP collectively as an AGN. If neither has been assigned to an X-ray source, I count it as a non-AGN. The same approach is used for the controls; although separate controls are matched to each member of the MGP, I count the AGN pairwise in their controls. In this way, I am fairly counting X-ray detections between the MGPs and their controls.

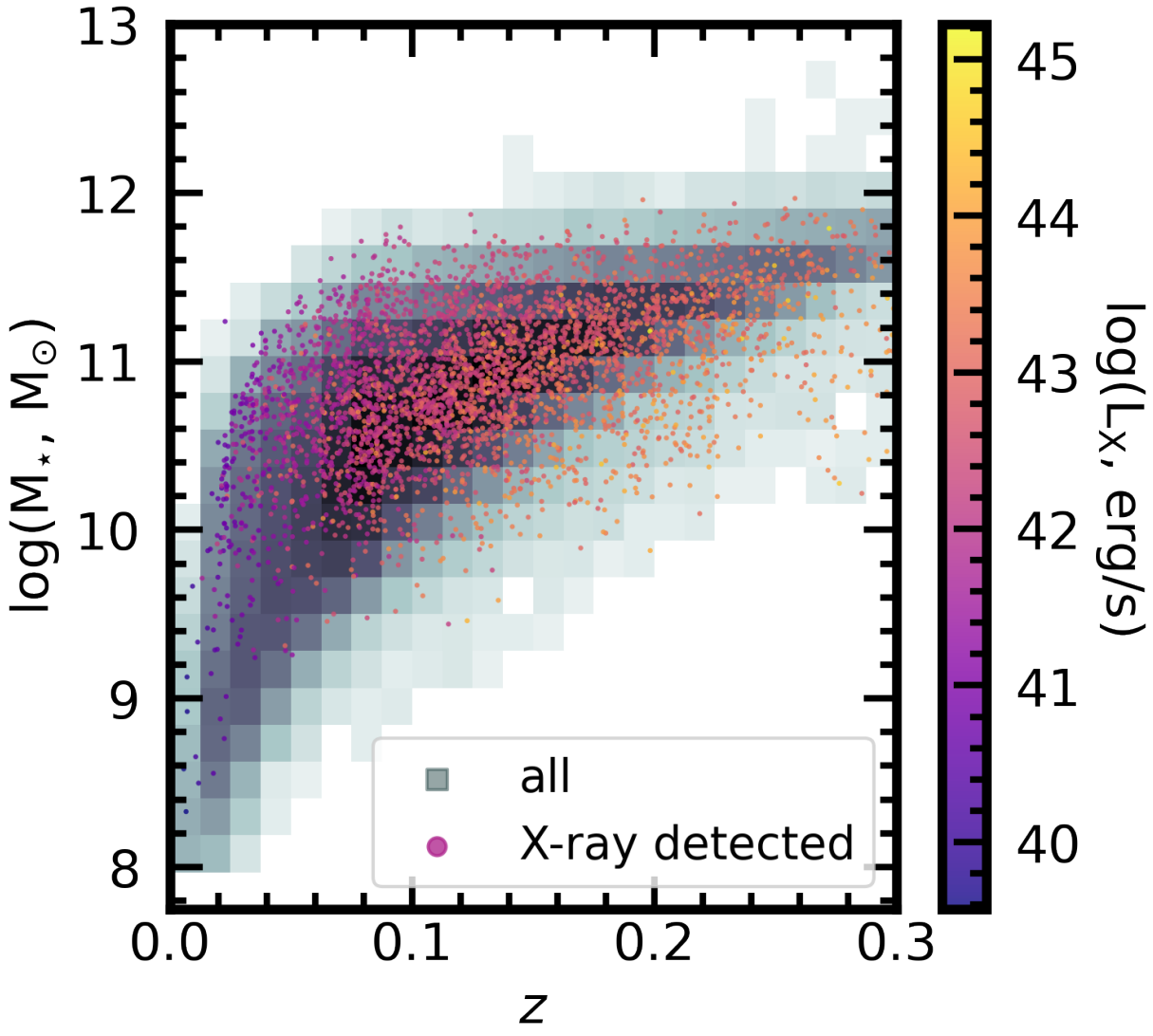

Figure 6.4: The stellar mass-redshift distribution of the parent sample for this study (2D histogram), with eRASS1 detections superimposed. The colour scale represents X-ray luminosity in eRASS1.

Given my pairwise approach to AGN statistics, it is necessary to derive a statistical correction for the influence of X-ray PSF blending on the sample of MGPs. The measured AGN fraction amongst MGPs in a given bin of projected separation (f_m) is a function of the true AGN fraction (f_A) and the fraction of galaxies crowded below the effective resolution of the survey (f_c). I can write f_A in terms of p_0, the chance of a galaxy pair hosting zero AGN, p_s, the chance of a pair with a single AGN, and p_d, the chance of a pair with a double AGN. The coefficient for p_d is one, since an MGP with two AGN has an AGN fraction of 1, and the coefficient for p_s is $\frac{1}{2}$, since an MGP with only a single AGN has an AGN fraction of one half. Since a given pair must have either zero, one, or two AGN, p_0, p_s and p_d sum to one, allowing a simplification:

$$f_A = \frac{p_d + \frac{1}{2}p_s}{p_0 + p_d + p_s} = p_d + \frac{1}{2}p_s. \tag{6.2}$$

When crowding occurs at a rate f_c, the measured AGN fraction f_m is no longer the same as the actual AGN fraction f_A. In cases where double-AGN pairs are blended within the PSF, p_d is multiplied by a crowding factor $(1-\frac{1}{2}f_c)$, since only half of the AGN are counted in a crowded MGP with a double AGN. Meanwhile, there is no effect on the number of AGN detected in single-AGN pairs, so the coefficient for p_s remains the same at $\frac{1}{2}$:

$$f_m = (1 - \frac{1}{2}f_c)p_d + \frac{1}{2}p_s. \tag{6.3}$$

In order to correct the measured AGN fraction up to the true AGN fraction, I am interested in f_m/f_A, which can be written as the ratio of Equations 6.2 and 6.3:

$$\frac{f_m}{f_A} = \frac{(1 - \frac{1}{2}f_c)p_d + \frac{1}{2}p_s}{p_d + \frac{1}{2}p_s}. \tag{6.4}$$

Conveniently, p_d, p_0, and p_s can all be written in terms of f_A. The likelihood of an MGP with two AGN is set by combining the independent AGN likelihoods of two galaxies:

$$p_d = f_A^2. \tag{6.5}$$

The likelihood of an AGN-free MGP is also the combination of the independent chances that two galaxies do not host an AGN:

$$p_0 = (1 - f_A)^2. \tag{6.6}$$

Any MGPs that are neither double AGN nor AGN-free must be single AGN, represented by p_s. I can therefore subtract p_d and p_0 from the total to obtain p_s:

$$p_s = 1 - f_A^2 - (1 - f_A)^2 = 2(f_A - f_A^2). \tag{6.7}$$

Substituting the last three definitions into Equation 6.4, I can rewrite and simplify (with some intermediate steps omitted for brevity):

$$\frac{f_m}{f_A} = \frac{-1}{2} f_A f_c + 1. \tag{6.8}$$

Finally, I derive the expression for the true AGN fraction:

$$f_A = \frac{1 - \sqrt{1 - 2 f_c f_m}}{f_c}. \tag{6.9}$$

In combining the AGN statuses of MGPs and control pairs, I force a crowding fraction of unity, f_c=1. This avoids the difficulty of attempting to predict which galaxy pairs in the sample will be crowded. With f_c methodologically forced to 1, the true AGN fraction is $1 - \sqrt{1 - 2f_m}$. This expression yields in the two extreme cases: when there are no AGN ($f_m = 0$ and $f_A = 0$), and when all galaxies in a sample host an AGN ($f_m = 0.5$ and $f_A = 1$, since both single- and double-AGN pairs would be detected as single AGN). The pairwise treatment of galaxy pairs and analytical correction are applied to both the MGPs and matched control samples when I study X-ray AGN incidence in galaxy pairs.

6.1.7 BLAGN sample

Galaxies in this Chapter are counted as BLAGN if they are included in the sample derived homogeneously from SDSS DR7 spectroscopic targets by Liu et al. (2019). Liu et al. (2019) use a sophisticated spectral fitting technique that allows for multiple components to the Balmer and forbidden lines appearing in the spectra, and include galaxies in their final BLAGN sample if the quality of the fit is improved by a broad Hα component, and if that component is detected with S/N> 5. There is no lower limit on the velocity of the broad Hα component, so long as it is broader than the narrow component. In practice, BLAGN in the Liu et al. (2019) catalogue have Hα line velocities of at least 428 km/s, and a median velocity of 2763 km/s. In my parent catalogue, 2626 galaxies are identified as BLR hosts by Liu et al. (2019). For additional detail, I refer the reader to Tables 1 and 2 of Liu et al. (2019).

6.1.8 NLAGN sample

In order to identify galaxies with narrow-line AGN (NLAGN) emission lines, I refer to the widely-used MPA-JHU catalogue of spectroscopic measurements for SDSS DR7 (Brinchmann et al. 2004), the same catalogue used in Chapters 4 and 5. While the emission line fitting method used by MPA-JHU considers only one component each for Balmer and forbidden emission lines, it is well suited to characterize emission line fluxes in the absence of a dominant BLR contribution to the optical spectrum. MPA-JHU also derive their measurements from the SDSS DR7 spectroscopic sample, and so the entire parent catalogue is technically eligible to be classified by the MPA-JHU emission line measurements as a NLAGN (even though emission line measurements for every source in the parent catalogue are not tabulated). MPA-JHU apply the following criteria to SDSS DR7 sources before reporting emission lines used in this work to select NLAGN:

- Fluxes are always reported when the spectroscopic class and the target type are "galaxy", the redshift is less than 0.7, and the spectrum has a median pixel-wise S/N> 0.

- Fluxes are also reported when the spectroscopic class is "galaxy" (even when the target type is not "galaxy"), when the redshift is less than 0.7, and the pixel-wise S/N> 2.

- Fluxes are reported for objects with spectroscopic class "quasar" when the object was targeted as a galaxy, the redshift is less than 0.7, the S/N is > 2, the χ^2 < 2 for the spectral fit to the continuum, the (single-component) Balmer line width is < 500 km/s, and the forbidden lines are no less than 85 per cent as wide as the Balmer lines.

Given the velocity thresholds in the MPA-JHU catalogue, and the typical velocities in the Liu et al. (2019) BLAGN sample, the two catalogues are generally mutually exclusive. In the 469 (28 per cent of BLAGN) cases where a galaxy exists in both the MPA-JHU and Liu et al. (2019) catalogues, I prioritize the Liu et al. (2019) classification as a BLAGN. For the remaining (narrow emission line) galaxies in the MPA-JHU catalogue, I assess the presence of an AGN using the Kewley et al. (2001) criterion on the BPT diagram, requiring also that each of the four lines used for BPT classification are detected with S/N> 5, and that the Hα equivalent width is greater than 6Å. The first criterion is designed to select galaxies with narrow line nebular emission unambiguously dominated by an AGN contribution, while the second is deployed to remove galaxies dominated by low-ionization emission regions (LIERs, e.g., Stasińska et al. 2006; Cid Fernandes et al. 2011; Belfiore et al. 2016; typically produced by shocks, weak AGN, or old / hot stellar populations) from the sample. Removal of LIERs is important in merger studies, since hydrodynamic shocks are an intuitive (and studied; see Monreal-Ibero et al. 2006; Farage et al. 2010; Rich et al. 2014,

2015) consequence of galaxy mergers. While the choice of BPT diagram criteria (e.g., a lower S/N criterion, including composite galaxies between the Kewley et al. 2001 and Kauffmann et al. 2003 criteria on the diagram, or using a different criterion to remove LIERs / LINERs as I did in Chapter 5) does bear on the NLAGN fraction in the sample and some of my results, it does not affect the trends presented in this Chapter. There are 2723 galaxies in the parent catalogue that meet my criteria for classification as a NLAGN.

6.1.9 Mid-IR AGN sample

Characterization of dust-obscured AGN is central to this work, and I use photometry from the $W1$ and $W2$ bands of the WISE space telescope to select galaxies whose circum-nuclear or galaxy-scale dust has been heated by UV emission from the accretion disk. Depending on the geometry of the dust, mid-IR observations can be used to identify galaxies whose X-ray emission, BLR, and/or NLR are not observationally detected. Since my sample lies at low-z, I again follow several literature efforts (e.g., Satyapal et al. 2014; Ellison et al. 2013) and use a single colour criterion, $W1 - W2 > 0.5$, to select galaxies with a dust-obscured AGN. In order to ensure that mid-IR data products are homogeneously available (and available with high S/N) for the entire SDSS DR7 parent catalogue described in Section 6.1.2, I use $W1$ and $W2$ magnitudes derived from unWISE[4] forced photometry[5] at the positions of SDSS galaxies (Lang et al. 2016). As a result, no sources from the parent catalogue are missing $W1$ or $W2$ photometry, and I can characterize the hot dust in the entire galaxy sample. In the parent catalogue, 6391 galaxies are classified as mid-IR AGN using these criteria. Earlier in Chapter 5, I used the same mid-IR colour criterion as in this Section, but applied to the WISE all-sky data release. The unWISE data products are of better quality than those in the WISE all-sky release, and they are uniformly available for the sample. The work presented in this Chapter was completed after that in Chapter 5, and my adoption of the improved dataset occurred between the two.

6.1.10 Multi-wavelength AGN

Having laid out the criteria used to identify AGN in the X-ray, optical (narrow- and broad-line diagnostics) and mid-IR, I can investigate the global connections between the different AGN classes independent of the merger sequence.

Figure 6.5 offers a surface-level orientation to the characteristics of the AGN samples studied in this Chapter, with the $M_\star$, z, and CNN $p(x)$ statistics shown for each AGN class as well as

[4]https://unwise.me/

[5]Forced photometry refers to de-blended photometry from WISE observations at the locations of galaxies in SDSS.

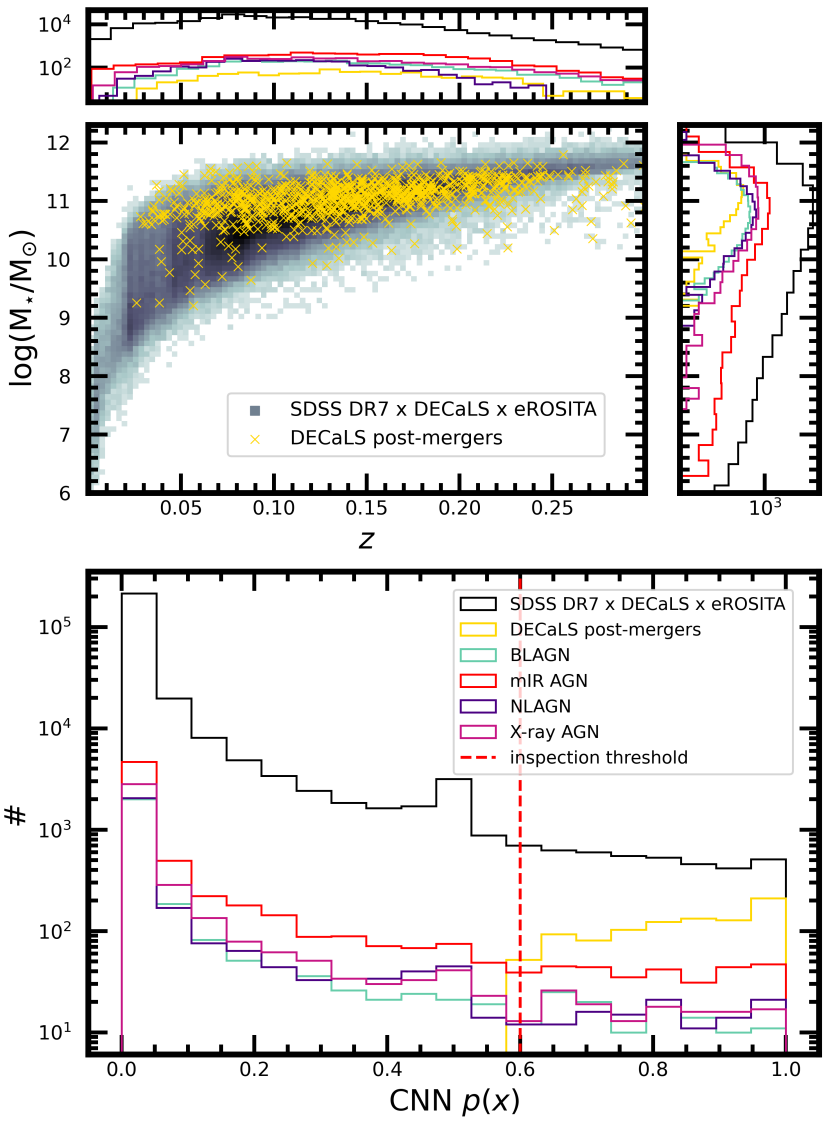

Figure 6.5: The mass, redshift, and CNN $p(x)$ statistics for galaxies relevant to this Chapter: NLAGN (indigo), mid-IR AGN (red), X-ray AGN (magenta), and BLAGN (teal). The statistics for the AGN samples are shown alongside the sample of SDSS DR7 galaxies appearing in CFIS DR2 (greyscale background histogram in the left main panel and black histograms) and the visually confirmed DECaLS post-merger sample (yellow crosses in the left main panel and yellow histograms) for reference.

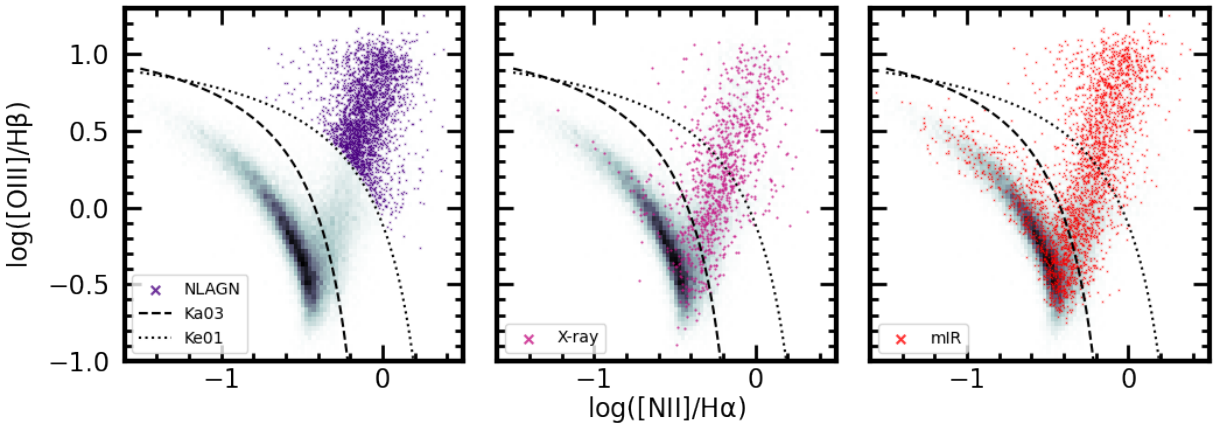

Figure 6.6: BPT AGN positions of the entire sample (2D histograms in each panel), galaxies meeting my NLAGN criteria (indigo points, left panel), galaxies with X-ray detections eligible for placement on the BPT diagram (middle) and galaxies with unWISE $W1 - W2$ colours suggesting a dust-obscured AGN (right) eligible for placement on the BPT diagram.

for the post-merger sample. Since the DECaLS post-merger sample was identified specifically for this project, the statistics are the same as those presented in Section 3.2.2. More interesting connections between the AGN classes can be found in the demographics of dual-detected or multiple-detected AGN (i.e., where a single galaxy belongs to more than one of the AGN groups), so I present those next.

Figure 6.6 shows the BPT diagram positions of the parent catalogue (background histogram in each panel), NLAGN (left panel, indigo), X-ray AGN (middle panel), and mid-IR AGN (right panel). Notably, galaxies meeting my other AGN criteria do not always lie in the AGN wing of the BPT diagram; the positions of the X-ray AGN echo the results from Agostino & Salim (2019) (and other works investigating the incompleteness of NLAGN selection, e.g., Trump et al. 2015; Jones et al. 2016). Mid-IR AGN are even more widely distributed throughout the BPT diagram, with 39 per cent of BPT-eligible (with S/N> 5 for each of the four BPT diagram optical emission lines) mid-IR AGN falling in the star-forming locus of the diagram defined by the Kauffmann et al. (2003) criterion (dashed lines, Figure 6.6). Taken as a whole, the narrow-line characteristics of the various AGN classes highlight the necessity of multiple AGN criteria for a complete study. I do not show the positions of BLAGN in Figure 6.6 because they are de facto absent from the MPA-JHU catalogue.

Figure 6.7 shows the X-ray luminosities (left panel) and mid-IR colours (right panel) of galaxies meeting the various AGN criteria in this Chapter. Generally, galaxies with NLAGN classifications lie at lower X-ray luminosities than those with either BLAGN or mid-IR AGN detections.

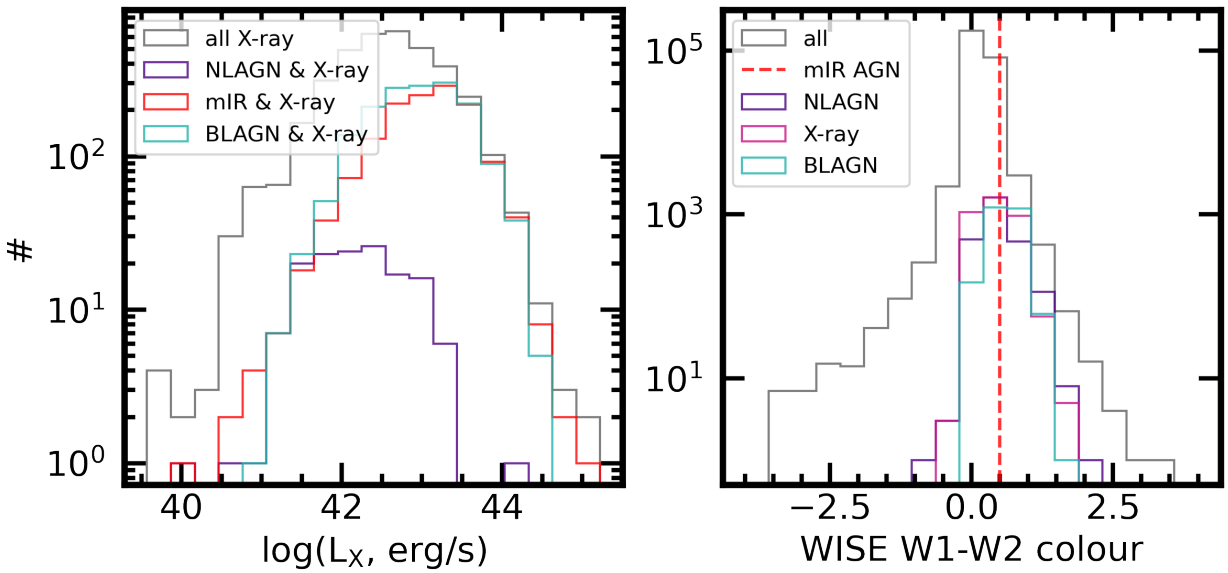

Figure 6.7: X-ray luminosities from eRASS1 (left) and $W1 - W2$ colour from unWISE (right) for the galaxies eligible for this study (grey series). The galaxies with X-ray detections (left) that also meet my narrow-line (indigo), broad-line (teal), or mid-IR (red) AGN criteria are also plotted. The $W1 - W2$ colours from unWISE (right) are also shown for the subsets of galaxies meeting my other AGN criteria (same colour code as left, and with eRASS1 sources shown in magenta).

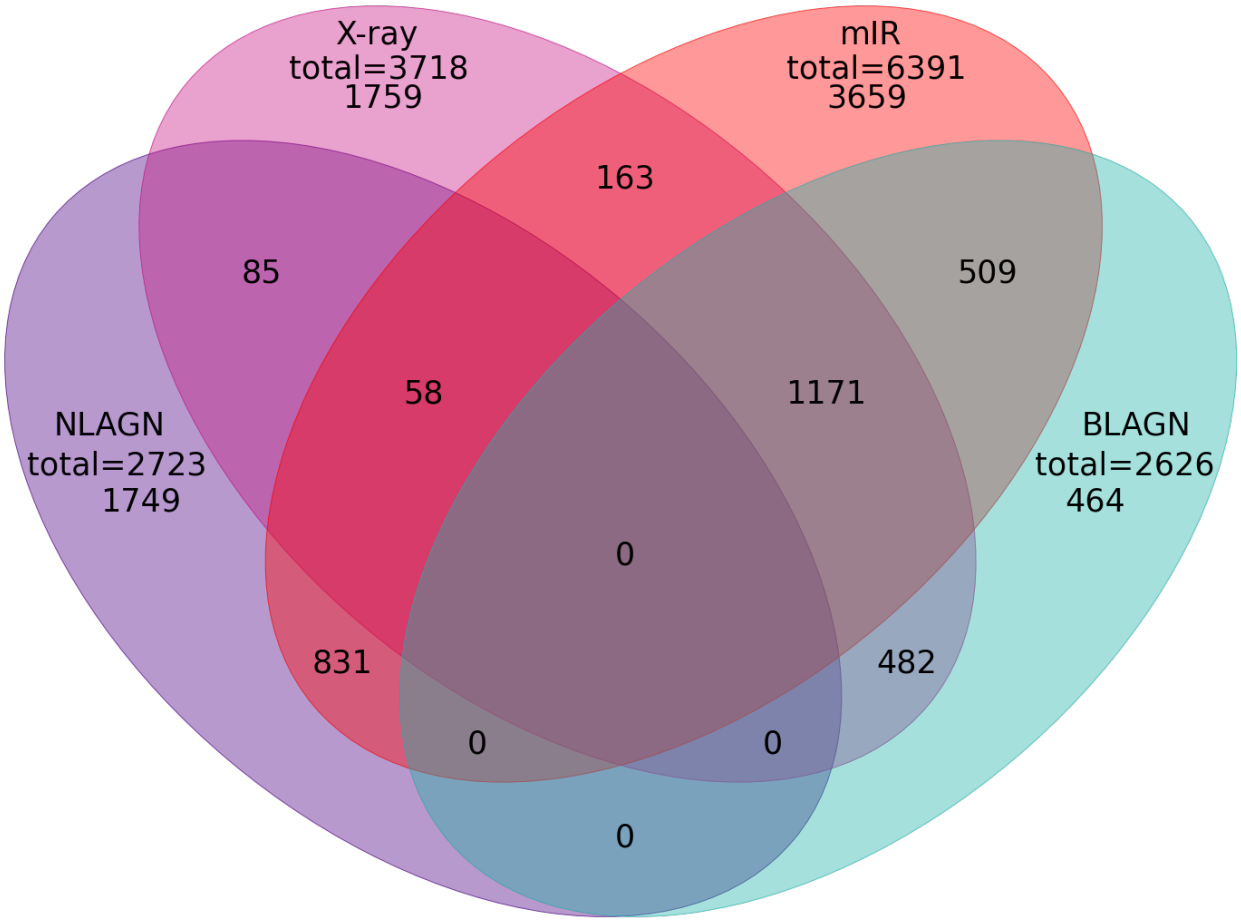

Figure 6.8: Four-class Venn diagram describing the overlap between the AGN criteria in the sample. Overlap exists between all classes, except where it is expressly forbidden (the BPT AGN criteria, indigo, and BLAGN criteria, teal, are explicitly mutually exclusive).

The dominance of BLAGN and mid-IR AGN at high L_X hints at the accretion rates corresponding to each AGN type, although the effect is partly caused by BLAGN "forcing out" NLAGN from the high L_X domain, since BLAGN classifications supersede NLAGN classifications in this Chapter. At lower X-ray luminosity, however, it is likely that the NLAGN are preferentially X-ray attenuated along the line of sight compared to BLAGN. Additional evidence for lower rates of X-ray obscuration amongst BLAGN is presented later in Section 6.2.5. The WISE colours of the galaxies in the sample illustrate that galaxies with other AGN classifications do not always meet my mid-IR AGN criterion (to the right of the red dashed line), a result expected in cases where the AGN is not bolometrically dominant.

Figure 6.8 summarizes the multi-wavelength AGN selection. Overlap exists between all classes in the sample, except where it is forbidden by construction (BLR detection forbids the inclusion of

a galaxy in the NLAGN sample). While this Chapter prioritizes the novel statistical results made possible by the eRASS1 X-ray catalogue, the intersection of the X-ray AGN sample with other AGN types offers essential context. It is worth noting that my BLAGN (Section 6.1.7), NLAGN (Section 6.1.8), and mid-IR (Section 6.1.9) samples are selected homogeneously from the SDSS DR7 catalogue, thus ensuring that any object in the parent catalogue could be classified as an AGN by any of my multi-wavelength AGN criteria. The most striking connection on the Venn diagram is the intersection between the X-ray, mid-IR, and BLAGN samples (overlap of the magenta, red, and teal ovals). In particular, the overlap between the X-ray and BLAGN ovals emphasizes the importance of including BLAGN in my analysis; without them, $\sim$ 44 per cent of the X-ray AGN sample would be eliminated. Figure 6.8 also illustrates the utility of X-ray selection in multi-wavelength AGN studies: nearly half (47 per cent) of the X-ray AGN (the magenta oval) are not identified as AGN according to any other criteria.

Except in cases where obscuration is significant ($N_H \gtrsim 10^{22}/cm^2$), X-ray luminosity is a comparatively reliable indicator of the state of the SMBH, since the BLR is more easily obscured by dust, and the observability of NLAGN and mid-IR AGN is also conditional. Figure 6.9 characterizes the X-ray luminosities most often associated with each of the other AGN criteria. The top panel (repeated for context from the left panel of Figure 6.7) shows the luminosity distributions of the entire X-ray AGN sample (magenta dashed histogram), as well as the subsets of the sample meeting the BLAGN (teal), mid-IR (red), and NLAGN (indigo) criteria. The bottom panel shows the fractions of X-ray AGN co-detected as BLAGN, mid-IR AGN, and NLAGN as a function of L_X. X-ray AGN are most often co-detected as NLAGN at the low end of luminosities in the sample, between $\sim 10^{41-42}$ erg/s (shown also in Figure 6.7).

The fraction(s) of X-ray galaxies co-detected as BLAGN and/or mid-IR AGN grows steadily with increasing L_X, and there is nearly total overlap between the X-ray, BLAGN, and mid-IR samples for $L_X > 10^{43.5}$ erg/s. The multi-wavelength demographics of the X-ray sample suggest that hot circum-nuclear or galaxy-scale dust is more often observed in galaxies with higher AGN luminosities. Higher X-ray luminosities are also frequently associated with the detection of an exposed BLR, consistent with either a higher likelihood of removal of obscuring material (dust and/or gas; Ueda et al. 2003; Maiolino et al. 2007; Treister et al. 2008; Burlon et al. 2011; Ricci et al. 2017) or degrees of attenuation of the X-ray flux by the obscuring material.

Implicitly absent from Figure 6.9 are low-luminosity or obscured AGN without an X-ray detection (fainter than the eRASS1 flux limit at 5×10^{-14}erg/s/cm^{-2} in the 0.2–5 keV band). Still, the general AGN demographics of the sample suggest an AGN "observability sequence" with L_X, in which AGN with low X-ray luminosities, heavy obscuration, or degrees of both are often seen as NLR hosts. Progressively higher L_X is consistent with being concurrent with removal of the NLR,

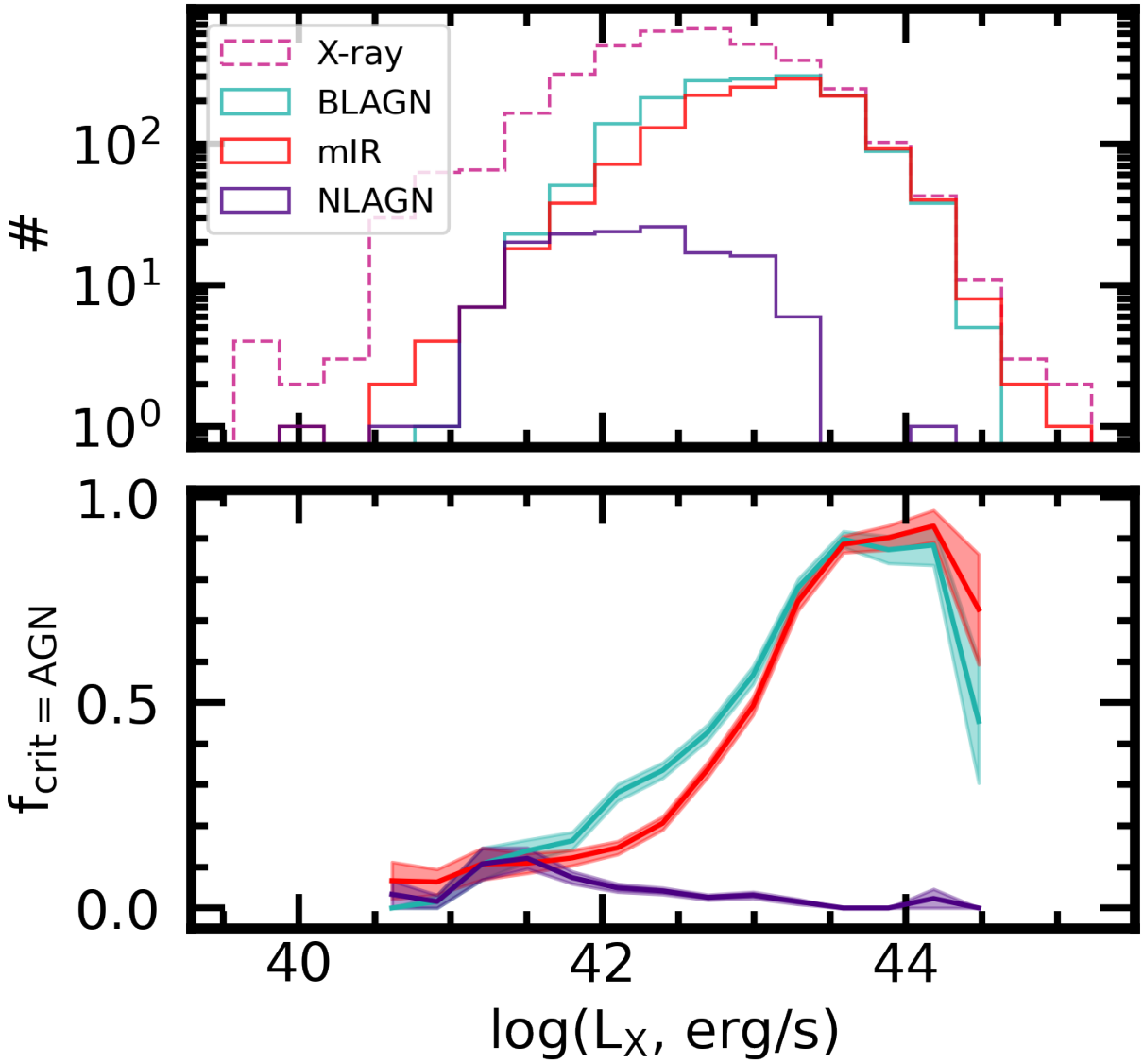

Figure 6.9: The number (top panel) and fraction (bottom panel) of eRASS1 galaxies in a given bin of X-ray luminosity meeting one of the other AGN criteria (NLAGN, mid-IR, or BLAGN). Fractions are only plotted when there are more than 10 X-ray galaxies in a given bin of L_X. The shaded error regions are defined by the binomial error on the fraction, $\sqrt{f(1-f)/N}$ where f is the measured fraction and N is the number of X-ray galaxies in the bin.

exposure of the BLR, and more prominent nuclear hot dust emission relative to the contribution of starlight to the mid-IR (Ricci et al. 2017; Blecha et al. 2018). Understanding these trends and the interplay between multi-wavelength diagnostics will be vital in my interpretation of the results for mergers, presented in the next section.

6.2 Results

After assembling a catalogue with homogeneous multi-wavelength data, I commence my study of the influence of pre- and post-coalescence galaxy mergers on the X-ray and multi-wavelength AGN in the sample.

6.2.1 X-ray AGN incidence in mergers

Numerous statistical studies (e.g., Ellison et al. 2011, 2013; Satyapal et al. 2014; Weston et al. 2016; Bickley et al. 2023) have used optical and mid-IR diagnostics to uncover an increasing AGN fraction in mergers at progressively closer separations, with a peak in the post-merger regime. Although some previous studies have assessed X-ray AGN frequency in low-z galaxy mergers, statistics have either been poor (e.g., Secrest et al. 2020) and/or limited in the merger stage that they assess (e.g., Li et al. 2023a). Thanks to the eRASS1 sample, I am now able to make the same assessment for X-ray selected AGN in a large statistical sample of both galaxy pairs and post-mergers for the first time.

As mentioned earlier, the blending of pairs of galaxies within the eROSITA PSF means that I treat pre-coalescence galaxies in a pairwise fashion. Hence, for this analysis, I will use the MGP sample. MGPs and their matched controls are sorted into six 20 kpc bins (from 0 – 120 kpc). Their X-ray AGN fractions are assessed and corrected using the method described in Section 6.1.6. For post-mergers, I match individual galaxies to individual controls, since post-mergers do not belong to crowded galaxy pairs. The mergers in each subset all succeed in finding 10 matched isolated control galaxies each, with no significant discrepancy in $M_\star$ or z statistics. In the pair, post-merger, and control samples, I count AGN detections and fractions, and then divide the AGN fractions in each merger sample by the AGN fractions in the relevant control sample to compute an excess.

The results of this experiment are shown in Figure 6.10, which shows the numbers of MGPs in each bin of r_p and number of post-mergers (blue and yellow histograms, respectively) with the number of X-ray detections associated with each bin in the top panel. The middle panel shows the recovered AGN fractions in the merger samples and matched control samples. Controls have fairly consistent f_A of $\sim 2 - 2.5$ per cent, although the control sample for post-mergers is slightly

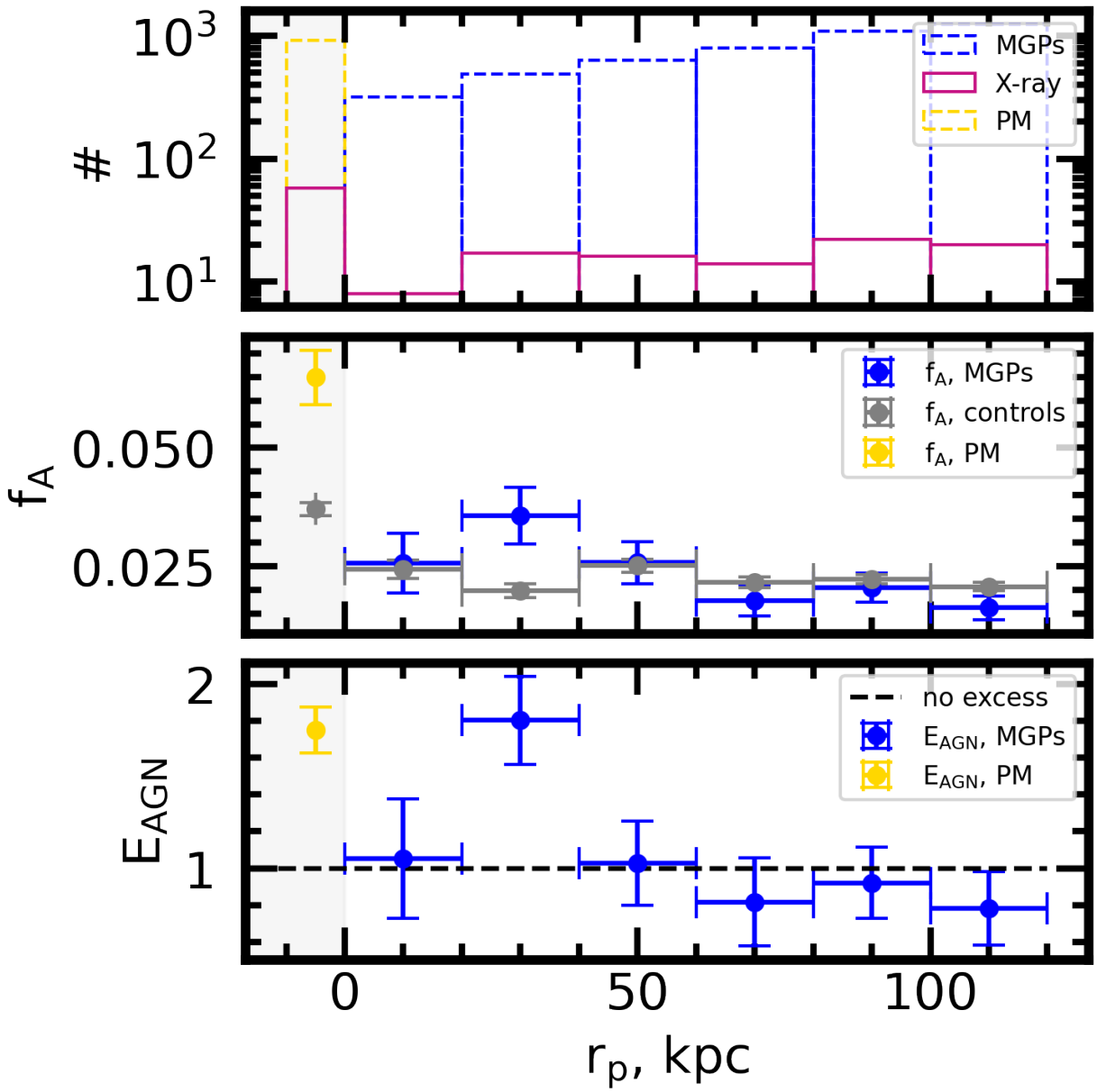

Figure 6.10: The number (top), fraction (middle), and excess over matched controls (bottom) of X-ray detections in the sample as a function of projected separation, using my pairwise treatment for galaxy pairs. Error bars on the AGN fractions are the binomial errors, and the errors on the excess are calculated as $\sigma_{f_{Mer}}/f_{Mer} + \sigma_{f_{Ctrl}}/f_{Ctrl}$ where f_{Mer} is the AGN fraction in the merger sample and f_{Ctrl} is the AGN fraction in the controls.

higher ($\sim$ 3.5 per cent) because the post-mergers and their controls are offset towards higher masses (see Figure 6.3). The bottom panel shows the excesses computed by taking the ratio of f_A in mergers and controls. In line with previous studies of optical and mid-IR selected AGN in galaxy pairs (e.g., Ellison et al. 2013; Satyapal et al. 2014), pairs at wide separations ($r_p > 40$) have X-ray AGN fractions that are statistically similar to non-interacting galaxies (although, the statistical significance, strength, and physical extent in r_p of the excess varies with sample size and methodology), suggesting that the influence of early pair-phase interactions are ineffective at triggering AGN with luminosities detectable in eRASS1 at these separations. Pairs with r_p between 20 – 40 kpc do appear to host X-ray AGN in excess of the isolated control sample by a factor of $\sim$ 1.8. The excess indicates that the merger sequence has begun to drive gas inflows strong enough to trigger eRASS1-detectable AGN at these separations. Optical and mid-IR AGN fractions in Bickley et al. (2023), the results of which are shown in Chapter 5, Figure 5.2, first become statistically distinguishable from controls near $\sim$ 40 kpc, even though the method is not identical.

Perhaps unexpectedly, Figure 6.10 shows that the closest galaxy pairs with $r_p < 20$ kpc in my sample do not appear to host X-ray AGN in excess of their matched controls. The lack of an AGN excess in close galaxy pairs is in contrast to my own findings in Chapter 5 and at other wavelengths in this Chapter (see Section 6.2.4). It is unlikely that X-ray AGN triggered by mergers have actually "switched off" at these separations. Rather, I hypothesize that merger-induced obscuration is effectively masking the underlying AGN excess in this bin of r_p, and the AGN luminosities triggered by pair-phase interactions are not sufficient to remove the obscuring material via radiation pressure. I offer observational evidence for this hypothesis later in Section 6.2.4.

The X-ray AGN excess recovered for post-mergers in Figure 6.10 is once again in qualitative agreement with my results at other wavelengths (e.g., in Chapter 5): I find that X-ray AGN appear to be triggered $\sim 1.8 \pm 0.1$ times as frequently in post-mergers compared to isolated controls, within the error bars of Secrest et al. (2020) and Li et al. (2023b); I return to a more detailed comparison with the literature in Section 6.3. The positive and significant excess of X-ray AGN in post-mergers verifies the role of coalescence in triggering luminous AGN.

6.2.2 Post-mergers in X-ray AGN

In Figure 6.10, I demonstrated that there is an excess of X-ray AGN in galaxy pairs with separations of 20 – 40 kpc and in post-mergers, indicating that, statistically, mergers can trigger AGN. I now turn to the complementary experiment (e.g., Hewlett et al. 2017; Ellison et al. 2019) and assess what fraction of AGN are in mergers. The merger excess experiment aims to assess the preva-

lence of mergers as the AGN triggering mechanism. I use the same control matching approach to match the X-ray detected eRASS1 galaxy sample to a control pool of X-ray-undetected galaxies (i.e., non-AGN). Because the X-ray-detected galaxies are significantly offset towards higher masses, control matching is truncated after 3 batches when the $M_\star$ K-S p-value falls just below the quality control criterion to 0.89. I then compute the fraction of X-ray AGN and non-AGN that have a post-merger classification. Of the 3718 X-ray AGN studied, 1.6 ± 0.2 per cent have a post-merger classification, and 0.8 ± 0.08 per cent of the 11154 controls have a post-merger classification. I report a post-merger excess of 2.0 ± 0.24 in X-ray detected galaxies over X-ray undetected controls. The post-merger fractions of the X-ray detected and undetected control samples (1.6 per cent and 0.8 per cent, respectively) are lower limits, since my post-merger selection prioritizes purity over completeness. However, the quantitative post-merger excess remains accurate; since the post-merger sample was identified independently of the X-ray AGN sample, and the fractional post-merger completeness is in principle identical in both the X-ray and non-X-ray samples. Combining this evidence with the X-ray AGN excess result for post-mergers, I conclude that mergers and X-ray AGN are unambiguously linked at low-z.

6.2.3 X-ray luminosity in mergers

Having established that some pair-phase galaxy interactions and coalescence events both appear to trigger luminous X-ray AGN, I next narrow my focus and compare the luminosities of AGN in mergers to AGN in non-merging isolated galaxies. The comparison of AGN in mergers and AGN in non-mergers again requires control matching. I use the same galaxy-wise control method described in Section 6.1.5 to identify at least 5, and as many as 10 controls with an eRASS1 detection for each merger-AGN (in either the pair or post-merger sample, also requiring an eRASS1 detection). I use an ensemble of at least 5 controls per merger because the goal of this experiment is to determine whether the merger's X-ray luminosity is atypical for galaxies with similar $M_\star$ and z. Every merger subset (post-mergers and bins of galaxy pairs) finds 10 controls per merger except for the $60 < r_p < 80$ kpc bin, which finds 5 batches before the K-S p-value for z falls to 0.899. I finally calculate a log-scale X-ray luminosity offset $\Delta\log(L_X)$ for each merger by subtracting the mean log X-ray luminosity of the control ensemble from that of the merger.

Figure 6.11 shows the results of this test. The top panel histogram shows the numbers of pair galaxies in each bin of r_p (blue) and post-mergers (yellow), as well as the number of isolated controls matched to them (grey). Note that all post-mergers, pairs, and controls included in this experiment must have an eRASS1 detection associated with them, so that $\Delta\log(L_X)$ can be computed. The middle panel violin plots show the L_X distributions of the merger samples (in either

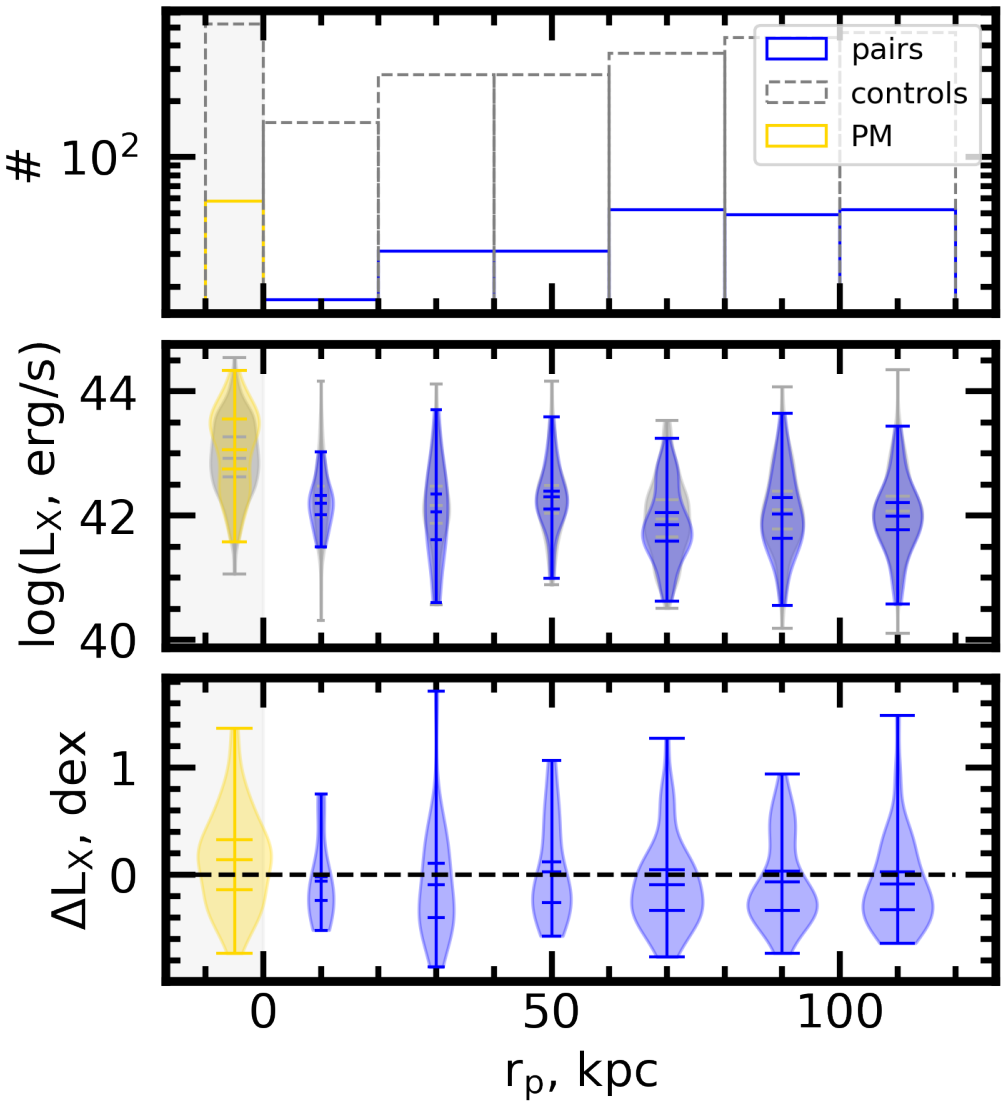

Figure 6.11: X-ray luminosity offset plot for galaxy pairs (blue) and post-mergers (yellow) in the sample relative to X-ray AGN in isolated controls (grey). The top panel shows the numbers of post-merger and pair AGN, and the number of control AGN matched for each merger sample. The middle panel violin plot shows the luminosity distributions for the merger and control samples, with tick marks at the extrema, means, and 1σ positions. The bottom panel violin series shows the distribution of luminosity enhancements calculated for the mergers compared to their individual control ensembles. X-ray AGN in pairs are shown in the blue data series, and post-mergers are shown in yellow. Galaxy pairs and post-mergers are both found to have X-ray luminosities in eROSITA that are statistically consistent with their non-interacting controls. The violin series in the third panel highlights that galaxies across the merger sequence exhibit a range of X-ray luminosities.

blue or yellow for pairs and post-mergers, respectively) and the control galaxies matched to them (grey). The tick marks on the violins show the maxima and minima of each $\log(L_X)$ distribution, the means, and the 1σ region. Finally, the bottom panel shows the actual result of my experiment, in which the median $\log(L_X)$ of the control ensemble for each merger is subtracted from the merger's $\log(L_X)$ to compute a luminosity enhancement. The distribution of $\Delta\log(L_X)$ values for each merger sample is used to create the violin series.

The distributions of $\log(L_X)$ for mergers and their matched controls are in good agreement (reasonable overlap between the blue/yellow and grey violins in the second panel). The statistical comparison in the bottom panel shows that once the offsets are calculated, the luminosities for both post-mergers and pair galaxies are consistent with those of their control ensembles (i.e., no merger sample is offset from its control sample by more than 1σ, even when the signal from all bins of r_p are combined). There is a diversity of $\Delta\log(L_X)$ within the post-merger and pair samples, the extent of which is illustrated by the violin series in the third panel. Possible explanations for the moderate luminosities seen in the galaxy pair and post-merger AGN populations are explored further in Section 6.3. Later in Section 6.2.5, I present evidence for the probable attenuation of X-ray fluxes in late stage mergers, and previous studies (e.g., Ricci et al. 2021; Yamada et al. 2021) have shown evidence for increasing obscuration with decreasing galaxy separation. In this context, the lack of any luminosity offsets > 1σ suggests that the luminosities of galaxy pairs with $r_p < 20$ kpc and post-mergers are likely intrinsically elevated, and attenuated by the obscuring material. I note that luminosity offsets are sometimes presented with error regions representing the statistical error on the average ($\frac{\sigma}{\sqrt{N}}$, e.g., in Ellison et al. 2013 and for [OIII]λ5007 luminosity enhancements in Chapter 5), but in Figure 6.11 I elect to show the full distributions and unadjusted 1σ standard deviations (i.e., the violin series is composed of histograms and the error bars enclose 68 per cent of the data). I adjust the methodology in this Chapter in order to better highlight the true range of X-ray luminosities in the galaxies.

6.2.4 Multi-wavelength AGN excesses

My X-ray results derived from eROSITA observations suggest that galaxy mergers are capable of triggering (or enhancing) and fueling AGN. Even considering the effects of obscuration (and the connection between late-stage mergers and obscuration), X-rays from the engine in the centre of the AGN are less easily obscured than the BLR, and less conditional than NLAGN or mid-IR AGN detection (see Menzel et al. 2016). Still, a better physical understanding can be reached by supplementing the X-ray data with observations at other wavelengths.

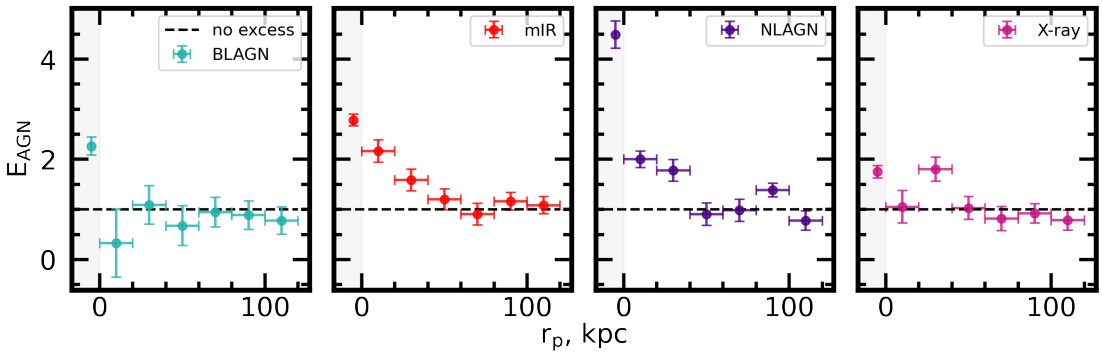

Figure 6.12: Multi-wavelength excesses for BLAGN, mid-IR, NLAGN, and X-ray AGN plotted side by side for comparison. BLAGN, mid-IR, and NLAGN statistics are not computed pairwise, since I have unblended measurements for each individual galaxy. The eRASS1 X-ray AGN excess is the same as in Figure 6.10.

Individual AGN criterion excesses

Figure 6.12 offers an introduction to the multi-wavelength characteristics of the galaxy pairs and mergers in the sample. The first, second, and third panels show the AGN excess for interacting galaxy pairs and post-mergers in the style of Figure 6.10 for galaxies with a BLR detection, mid-IR AGN detection, and NLR detection, respectively. The pair-phase results in these three panels use the entire sample of pair galaxies, rather than MGPs, since individual galaxies in a pair are always resolved in SDSS fibre spectroscopy and unWISE photometry. The full pairs sample is preferable where it can be reliably studied since it improves the statistics, but the qualitative results are in agreement when the MGPs are used instead. The eRASS1 excess series shown in the fourth panel is identical to that plotted in Figure 6.10.

In the AGN unification model (Antonucci 1993; Urry & Padovani 1995; and see Hickox & Alexander 2018 for a recent review), the observability of the BLR is considered a function of viewing angle, and when the BLR is not visible it is due to obstruction by a dusty torus. Given the assumption that post-merger and pair status do not bear on the viewing angle of the AGN, the BLAGN series in the first panel indicates that the observability of the BLR in galaxy pairs remains governed by obscuration effects. The lack of a strong positive or negative signal suggests that either 1) pair-phase interactions do not preferentially trigger AGN with accretion rates strong enough to change the covering fraction or depth of obscuring material in the nucleus (Fabian et al. 2008; Ricci et al. 2017), or 2) pair-phase interactions are triggering and obscuring BLAGN with similar frequencies, giving the appearance of no effect. After coalescence, however, BLRs are observed in excess of what is seen in isolated controls, indicating that the conditions brought on by coalescence facilitate observability of the BLR. One interpretation of the emergence of a

BLAGN excess in post-mergers is that a subset of SMBHs in post-mergers reach high enough accretion states to remove obscuring material.

The excess of AGN in the mid-IR series in the second panel of Figure 6.12 increases steadily from right to left, from wide pairs with $40 < r_p < 60$ kpc down to the post-merger epoch. The increasing trend of mid-IR AGN incidence with r_p is consistent with the underlying hypothesis of this work: that AGN are triggered with increasing frequency as merger events progress. The signal is also affected by the increasing likelihood of obscuration on either galactic or nuclear scales over the course of the merger sequence (Blecha et al. 2018). It is therefore unsurprising that this panel shows the smoothest evolution as a function of r_p.

The NLAGN trend (third panel of Figure 6.12) also suggests an increasing likelihood of AGN triggering with decreasing r_p. The conditions for the excitation and detection of the NLR seem to be particularly common in post-mergers, resulting in the highest excess for any plot in this work at more than a factor of four. The particular NLAGN signal recovered is sensitive to the choices of S/N and BPT diagram criteria used, but the trend is not. The excess signal is slightly weaker when galaxies above the Kauffmann et al. (2003) criterion on the BPT diagram are included along with those exceeding the Kewley et al. (2001) maximum starburst criterion, suggesting that major mergers are preferentially associated with galaxies whose emission lines are unambiguously NLR-dominated. Lowering the S/N criterion for the BPT diagram emission lines (e.g., to S/N> 3) has no statistically meaningful influence on the signal shown in the NLAGN panel of Figure 6.12.

Taken together, the multi-wavelength AGN excesses shown in Figure 6.12 suggest more frequent AGN triggering and obscuration by gas clouds and dust as mergers progress during the pair phase, with circum-nuclear obscuration often clearing via radiation pressure (e.g., Fabian et al. 2008) in the post-merger epoch. Within the merger sequence, AGN are most likely to be triggered in the post-merger epoch (as suggested by the mid-IR and NLAGN panels, which are less impeded by the effects of obscuring material). Notably, post-mergers appear to be unique, either in their ability to trigger AGN with accretion rates strong enough to remove obscuring material from the nucleus, or in that enough time has passed in a high accretion state that obscuring material has been gradually removed.

Any AGN criterion

None of my AGN selections can be truly complete, with NLAGN and dust-obscured AGN requiring the presence and detection of the excited / heated obscuring material, BLAGN detection requiring an exposed nucleus, and X-ray detection requiring a relatively high accretion rate (due to the relatively shallow threshold of $L_X \sim 10^{40}$ erg/s) and lack of heavy obscuration. Together, a

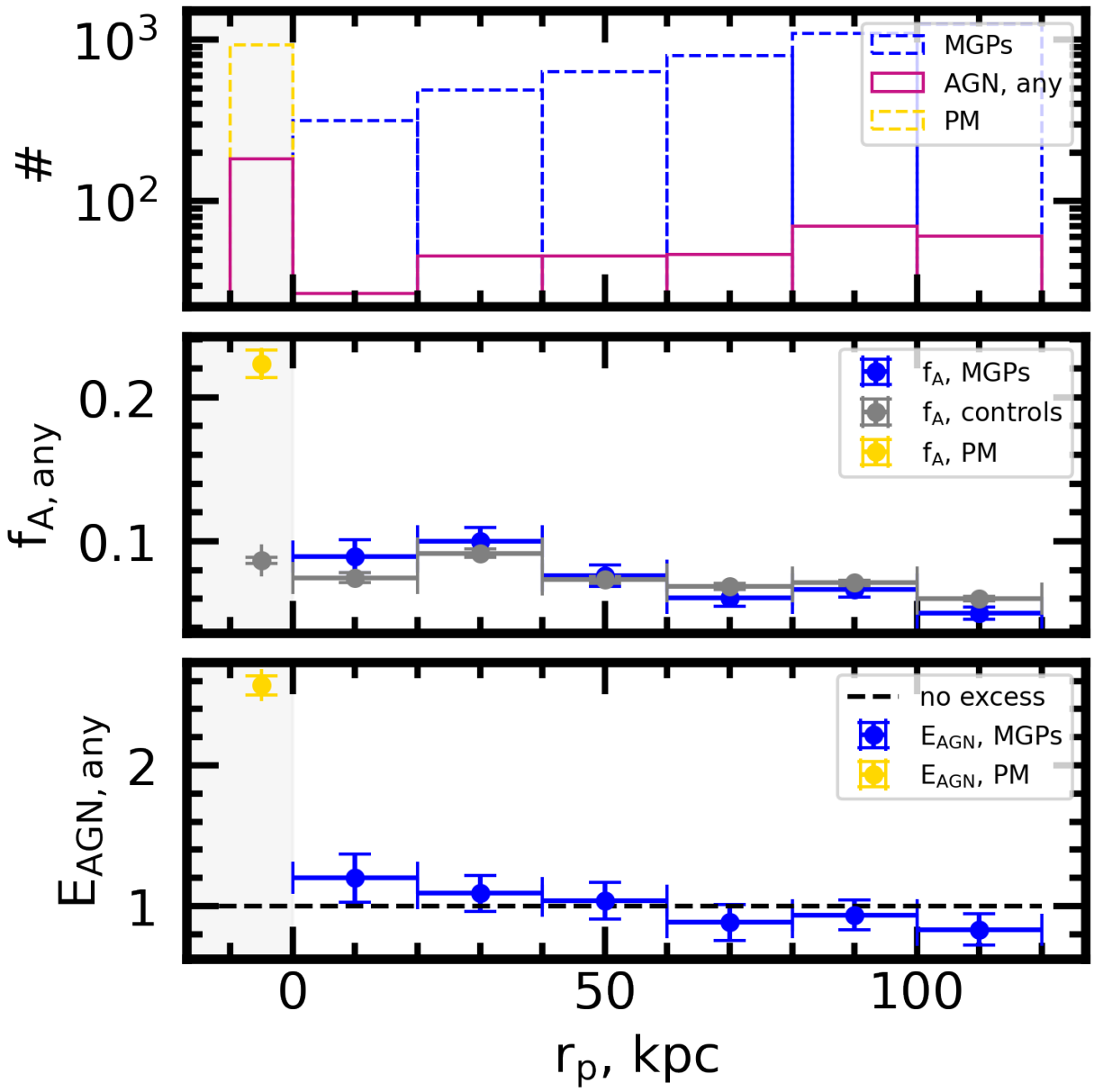

Figure 6.13: "Any" AGN excess for the same pair and merger samples as Figure 6.10. I find an excess of 1.2 in close galaxy pairs, and a dramatic 2.6× excess in post-mergers when I allow for any AGN criterion to be counted as a detection.

more complete AGN sample can be assembled, although there are certainly gaps in the selection (e.g., where narrow-line sources probably host AGN but are counted as "composite" sources between the Kewley et al. 2001 and Kauffmann et al. 2003 criteria on the BPT diagram, where a BLR is detected weakly with S/N< 5 in the Liu et al. 2019 catalogue, or when an AGN is not luminous enough in X-rays to reach the eRASS1 flux limit). Still, I can approach completeness by counting galaxies meeting any of my criteria as an AGN, and compute an excess.

I plot the results of an any-AGN excess study in Figure 6.13. For this experiment, I again must consider MGPs instead of individual pair galaxies, since X-ray detections are counted. If either pair galaxy is detected as a BLAGN, mid-IR, NLAGN, or X-ray AGN, I count the MGP as an AGN. I use the same criteria to classify pairs of matched control galaxies (the same matched controls as in Figure 6.10) and apply the correction derived in Section 6.1.6. Post-mergers are considered individually and matched to batches of single controls, so no correction is required.

The AGN fractions in this experiment and in Figure 6.13 are likely the most complete presented in this work: 6 – 10 per cent for isolated galaxies depending on $M_\star$ and z, 8 – 10 per cent for close galaxy pairs, and as much as 22 per cent for post-mergers. These fractions are accurate insofar as the pair and post-merger samples selected are representative of their respective classes, which in turn depends on the selection described in Sections 6.1.4, 6.1.3, and 6.1.5. I note that the post-merger multi-wavelength AGN fraction is substantially lower than that reported in Li et al. (2023b) (77 per cent). The difference could be attributed to the fact that Li et al. (2023b) use a more inclusive S/N criterion on the BPT diagram for their optical selection, and their X-ray AGN sample is dominated by galaxies with lower L_X, with the X-ray AGN threshold set at a 2 – 10 keV luminosity of $\sim 10^{40.5}$ erg/s.

I find small and statistically insignificant excesses of 1.04 and 1.09 for galaxy pairs between $40 < r_p < 60$ kpc and $20 < r_p < 40$ kpc, respectively. For galaxy pairs, the any-AGN approach favours control galaxies over the pairs, with many controls that were counted as non-AGN in my single-criterion results (in the denominator of f_A in the excess panels of Figure 6.12) receiving an AGN classification when I include more AGN types. The reclassification of some control galaxies acts to diminish the AGN excess seen in pairs compared with Figure 6.12. The loss of a significant excess in galaxy pairs with $r_p > 20$ kpc is a mathematical consequence of the higher baseline AGN fraction amongst galaxies in general when any AGN criterion is allowed (note the higher control AGN fractions in this figure compared to Figure 6.10). The higher global AGN fraction proportionally shrinks the difference in AGN fraction between the pair and control samples, which constitutes the excess signal. Nonetheless, for close galaxy pairs with $r_p < 20$ kpc, I recover a positive and $> 1\sigma$ excess of 1.2. Post-mergers with any AGN classification remain highly in excess of isolated controls with the same, by a factor of 2.6.

Obscuration excess

One of the main caveats of my X-ray results is that unconstrained degrees of obscuration are affecting the AGN fractions in eRASS1 for late-stage galaxy pairs and post-mergers. Obscuration effects have been implied by trends studied up to this point in Figures 6.10 and 6.12, and strengthened by the assumption that many heavily obscured and dusty AGN should be recoverable as power-law dust-obscured galaxies in the mid-IR. To further extend my investigation of obscuration, I next study particularly the incidence of galaxies that are detected as AGN in the mid-IR but not in the X-ray.

Figure 6.14 has identical construction to Figure 6.13, except I count galaxies in the numerator of $f_{mIR\neg X}$ when MGPs, post-mergers, or controls have $W1 - W2 > 0.5$ in the mid-IR, and are not detected in eRASS1. For MGPs and their control pairs, I require eRASS1 non-detections for both members, and $W1 - W2 > 0.5$ for at least one member. Post-mergers are again counted individually. It is worth noting that galaxies meeting the mIR¬X criterion are not necessarily Compton-thick AGN: some X-ray AGN may have only thin obscuration, and are too faint to be detected in eRASS1. The middle panel, $f_{mIR\neg X}$, should therefore be interpreted not as a Compton-thick fraction, but as a diverse subset that includes both heavily obscured galaxies and intrinsically X-ray faint galaxies in some proportion. My control-matching methodology should remove any signal to do with luminosity distance, $M_\star$, or secular processes. The bottom panel therefore represents the excess of AGN that are observable in the mid-IR and unobservable in the X-ray specifically due to a merger.

The lower panel of Figure 6.14 shows that there is no excess of mid-IR detected, X-ray undetected AGN down to $r_p = 20$ kpc. Below 20 kpc, however, mergers are associated with mIR¬X AGN 1.31 ± 0.27 times as often as isolated galaxies. Since the excess is designed to trace the influence of the merger, it is reasonable to assume that these are heavily obscured (or Compton thick) cases owed to a merger-induced influx of dusty gas (Blecha et al. 2018). Such cases are likely the cause of the unity X-ray AGN excess measurement for MGPs with $r_p < 20$ kpc in Figure 6.10.

Somewhat surprisingly, even though post-mergers exhibit a significant excess of X-ray AGN compared to controls in Figure 6.10, I find here that they also host (presumably heavily obscured or Compton-thick) mIR¬X AGN 3.1 times more often than matched controls. The large mIR¬X excess in post-mergers suggests a diversity of observability in post-mergers, where heavy columns of obscuring material are sometimes removed by radiation pressure from the AGN, and sometimes continue to enshroud the nucleus. If I assume that the proportion of $f_{mIR\neg X}$ AGN in post-mergers in excess of $f_{mIR\neg X}$ in controls are indeed heavily obscured cases, it would suggest that the f_A measurement for X-ray luminous post-mergers is underestimated by ~ 8 per cent. In

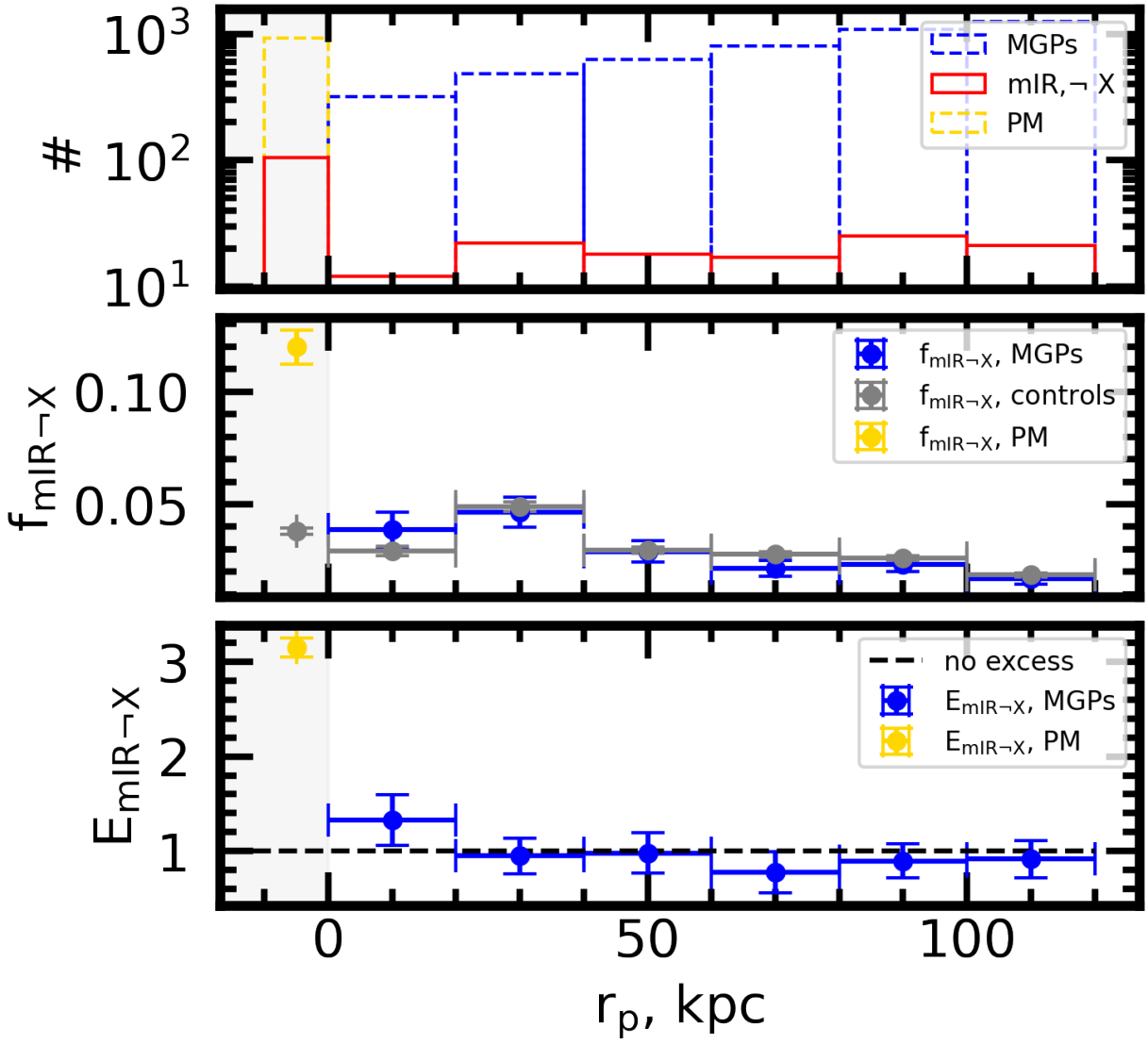

Figure 6.14: Investigating the heavily obscured AGN class in close pairs and post-mergers. The top panel shows number of pairs where neither member has an X-ray detection, and one or both members have $W1 - W2 < 0.5$. For post-mergers, the mid-IR and X-ray measurements for the target galaxy are assessed individually. The middle shows the fraction meeting the obscured AGN criterion, no X-ray plus $W1 - W2 > 0.5$, in the merger and control samples. The bottom panel shows the excess. I note that AGN in the numerator of the fraction plotted in the second panel likely includes galaxies that are only thinly obscured but intrinsically faint in the X-ray, in addition to heavily obscured AGN. By matching controls on $M_\star$ and z, I control for secular factors and luminosity distance so that the excess represents mid-IR AGN whose X-ray non-detection is a consequence of the merger event.

turn, I can compute an "obscuration-corrected" X-ray AGN excess:

$$E_{X,corr} = \frac{f_{X,PM} + (f_{mIR\neg X,PM} - f_{mIR\neg X,Ctrl})}{f_{X,Ctrl}} \tag{6.10}$$

Applying this correction to post-mergers, I calculate a new excess of $\sim$ 4.1, bringing the measurement into agreement with my NLAGN result in Figure 6.12. Even though NLAGN and X-ray AGN only overlap partially, and NLAGN do not trace well the high-L_X tail of the X-ray AGN sequence, the degree of overlap between NLAGN detections and X-ray AGN detections appears to remain relatively consistent as a function of merger stage. As such, AGN selected using the NLAGN criteria used in this work mirror the trends (but not the total fraction) of X-ray AGN incidence in galaxy mergers.

6.2.5 Multi-wavelength observability of X-ray AGN in mergers

There are several interesting multi-wavelength intersections to explore in the context of the merger sequence, but the small numbers of galaxies co-detected as X-ray AGN plus another AGN criterion (see Figure 6.8) do not invite rigorous statistical characterization. Still, I can inspect the overlap between different AGN diagnostics as a function of merger stage.

Figure 6.15 shows the multi-wavelength characteristics of post-mergers, three subsets of pair galaxies, and isolated galaxies. The trends of the dual-detected (X-ray plus another criterion) samples with merger stage reflect and verify the trends presented earlier (e.g., in Figure 6.12). The lowest two r_p bins in Figure 6.15 span the same r_p windows as those in the Figures (e.g., Figure 6.10) above, and a third bin is added to characterize generally any wider pairs with $r_p > 40$ kpc, for which I do not uncover a strong merger-AGN connection in this work. These five merger sequence categories are plotted in columns. In rows, I plot the overlap between the eRASS1 X-ray AGN sample and the AGN sample recovered using my NLAGN criterion (first row), mid-IR (second row), and BLAGN (third row) using Venn diagrams. The width of each Venn diagram is normalized, so that the horizontal overlap between adjacent diagrams can be compared directly. The percentage of X-ray AGN that are co-detected as an AGN using the second diagnostic are also listed below the Venn diagrams in each panel.

The result for NLAGN mirrors what I find in Section 6.2.4 – the overlap between narrow-line and X-ray detected AGN is small, but comparatively stable as a function of merger stage. NLR observability is therefore less influenced by the turmoil of an ongoing or completed galaxy merger, even though the fraction of the X-ray AGN sample recovered using the narrow-line criteria is relatively low.

The mid-IR and BLAGN results tell the same story from two sides. Amongst isolated galaxies,

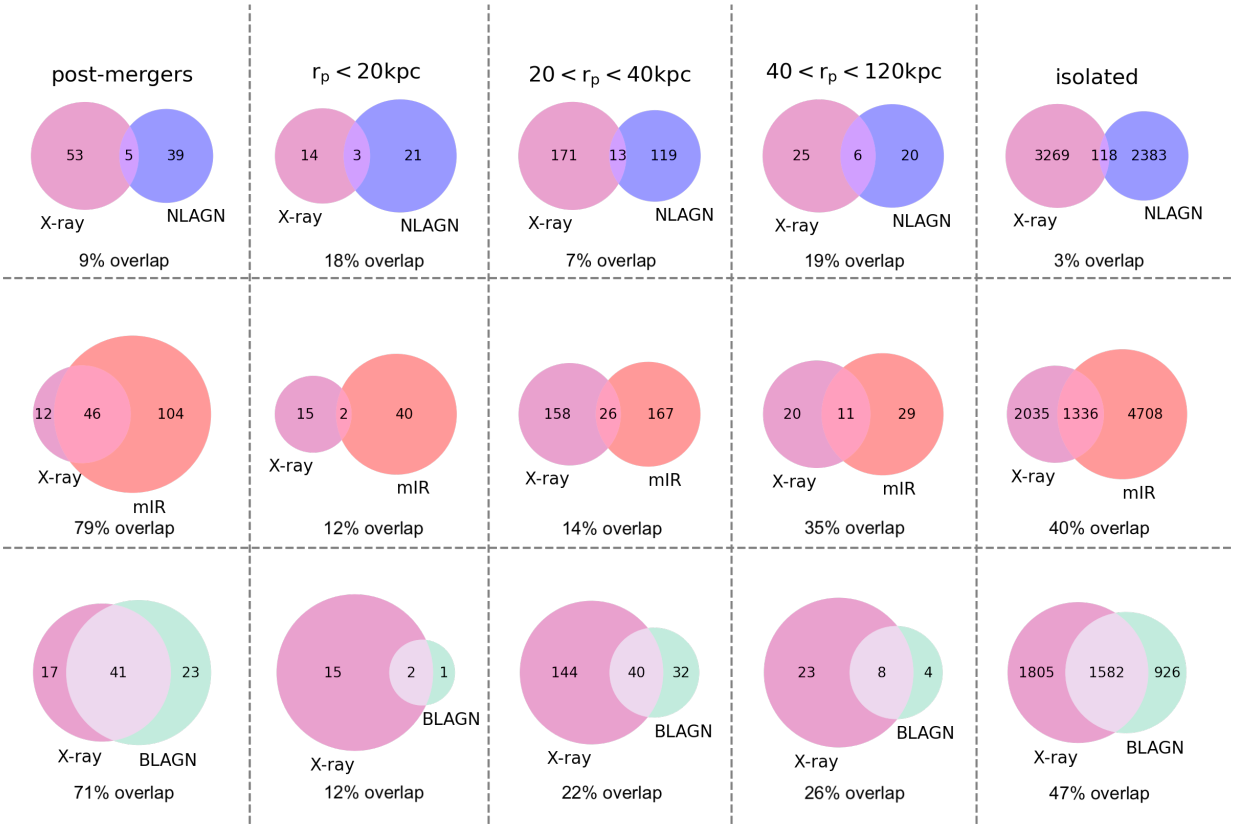

Figure 6.15: The multi-wavelength observability of X-ray AGN in coarse bins of projected separation. I have taken the isolated (control pool) galaxies, galaxy pairs where I find little to no signal on my excess plot (40 – 120kpc), and the two bins of interacting galaxy pairs where I uncover significant merger-AGN connections in this work (< 20, and 20 – 40 kpc), as well as post mergers. The percentage of X-ray AGN co-detected as an AGN via the second diagnostic is stated in each panel as well.

47 per cent X-ray AGN are co-detected as BLAGN, and mid-IR AGN detection coincides with X-ray AGN detection in 40 per cent of isolated X-ray cases. The BLAGN - X-ray and mid-IR - X-ray overlap fractions in isolated galaxies are likely decided by the orientation of the torus and accretion rate (see Figure 6.9). As mergers progress, X-ray and mid-IR AGN increasingly separate from one another as dust-obscured AGN are more often heavily obscured. At the same time, galaxies with a detected BLR become increasingly rare on account of the same physics.

In post-mergers, I find much stronger overlap between the eRASS1 X-ray, mid-IR, and BLAGN samples, since the accretion rates brought on by many mergers are sufficient to blow out circumnuclear dust and expose the BLR. A meaningful subset of the mid-IR post-mergers that are not seen in the X-ray, meanwhile, are presumably either hopelessly obscured or lacking the strength to push the obscuring material away as explored in Figure 6.14.

The physical scenario suggested by Figures 6.14 and 6.15 is explored further in Figure 6.16, which investigates the ratio of fluxes in the X-ray and mid-IR (unWISE $W1$) as a function of the $W1$ flux. Flux in $W1$ is converted to erg/s/cm^2/Å from the values reported in the unWISE data release (Lang et al. 2016), and multiplied by the effective width of the $W1$ band, which spans ~2.8 – 3.8 μm. For eROSITA, I use the total maximum likelihood X-ray flux from eRASS1. Since galaxies included in Figure 6.16 must have an X-ray flux, they are all counted as AGN in this study. Galaxies with unobscured AGN are expected to have fairly consistent X-ray to mid-IR flux ratios (the upper "cloud" population in Figure 6.16), especially in hard X-rays (e.g., Fiore et al. 2009; Ichikawa et al. 2012; Toba et al. 2019). Cases where the flux ratio is lower are likely the result of degrees of attenuation, an effect that is particularly relevant due to the relatively soft X-ray regime to which eROSITA is sensitive (lower "sequence", Figure 6.16). The Unlike Figure 6.14, which investigates the frequency with which obscuration results in the loss of X-ray AGN from the sample, Figure 6.16 shows how often X-rays are attenuated (but still detected) as a function of merger stage. To verify the connection between the amount of obscuration and the position of AGN on the diagram in Figure 6.16, I colour-code the data points for post-mergers and pairs based on whether they were (teal data series) or were not (yellow series) detected as BLAGN. Unsurprisingly, BLAGN are commonly found in the region of the diagram associated with little obscuration, while other merger-AGN are generally found in the lower sequence of X-ray attenuated AGN. The flux ratio corresponding to the minimum sensitivity of eROSITA for a given $W1$ flux is marked with the red dashed line. AGN whose X-rays are obscured to the point of non-detection by eROSITA would populate the vacant region below and to the left of the red dashed line. I also draw an empirical separation between the unobscured and attenuated populations at $\log(f_X/f_{W1}) = -1.575$ (black line, Figure 6.16). The distribution of $\log(f_X/f_{W1})$ has two peaks (corresponding to the mode values for the unattenuated and attenuated populations),

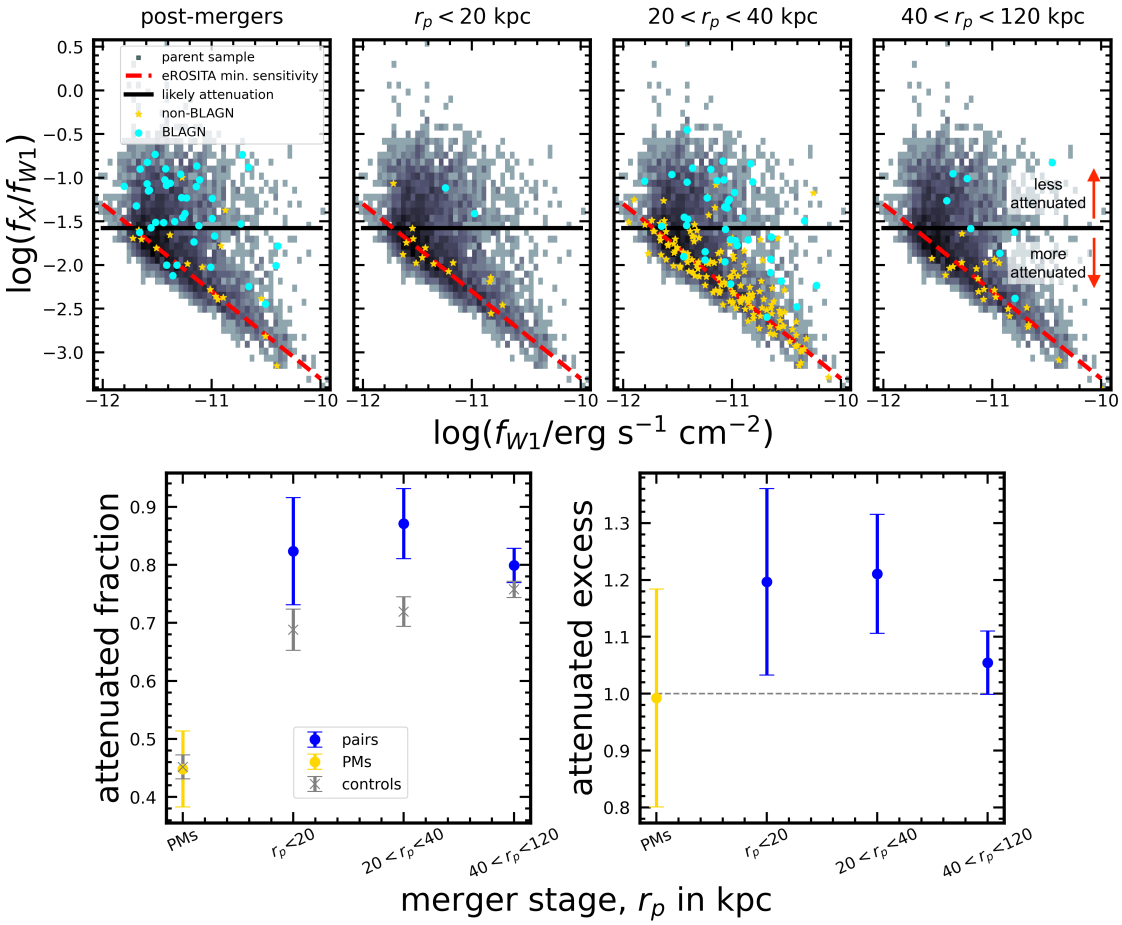

Figure 6.16: Another view of the evolution of AGN observability as a function of merger stage, using the same four bins of galaxy mergers as in Figure 6.15. For the top four panels, post-mergers are shown in the leftmost panel, galaxy pairs with $r_p < 20$ kpc are shown in the second, pairs with $20 < r_p < 40$ kpc are shown in the third, and pairs with $40 < r_p < 120$ kpc are shown in the fourth. All panels show the ratio of observed fluxes in the X-ray (as observed by eROSITA) and in the W1 mid-IR band (from unWISE) as a function of the W1 flux. The figure shows the number of mergers belonging to each category that have the multi-wavelength appearance of being partially obscured or unobscured as a function of merger stage. In each panel, I plot the flux ratio corresponding to the minimum observable flux for eROSITA using the red dashed line. In the bottom left panel, I plot the fraction of merger-AGN appearing below the black line (in the region associated with X-ray attenuation by obscuring material) in each of the four panels above, with errors computed as the binomial error on the fraction, $\sqrt{f(1-f)/N}$. Post-mergers are shown in yellow, and galaxy pairs belonging to each of the main categories from the panels above are shown in blue. I also plot the fraction of attenuated galaxies in isolated control samples matched to the mergers on $M_\star$ and z. In the bottom right panel, I plot the attenuation excess for the merger samples over the controls.

and −1.575 lies between them. In the bottom panels of Figure 6.16, I also compute the fraction of mergers appearing below the empirical attenuation criterion in each of the four merger stage bins studied earlier in Figure 6.15. I also make a statistical comparison between the mergers and matched samples of non-interacting control galaxies, following the same methodology described in Section 6.1.5. The attenuated fractions for the matched controls are shown in the grey data series in the bottom left panel of Figure 6.16. In the bottom right panel, I compute the attenuation excess by taking the ratio of the attenuated fractions in the merger and matched control samples.

My earlier results have shown that galaxy pairs with projected separations $40 < r_p < 120$ kpc do not host AGN in excess of controls (see Figure 6.12), leading to only a few galaxy pairs plotted in teal (for BLAGN) and red (for non-BLAGN) in the rightmost panel of Figure 6.16. Even at wide separations, I detect an elevated fraction of sources with X-ray attenuation, suggesting that inflows may already be common in early-stage galaxy pairs. More advanced mergers with $20 < r_p < 40$ kpc do host elevated incidence rates of NLAGN, mid-IR AGN, and X-ray AGN (Figure 6.12), and the corresponding panel in Figure 6.16 is thus more populated. The majority of galaxy pairs in the third panel of Figure 6.16 are X-ray attenuated and non-BLAGN, suggesting that merger-induced gas inflows are both triggering and obscuring AGN in galaxy pairs that have begun to experience strong tidal torques. The second panel in Figure 6.16 suggests that X-ray detected galaxy pairs with $r_p < 20$ kpc are very often attenuated, and rarely host BLAGN. Meanwhile, Figure 6.14 indicates that many AGN become so obscured at close separations that they are no longer detected in eROSITA. The loss of X-ray AGN likely accounts for the rarity of galaxy pairs in the second panel of Figure 6.16 relative to the adjacent panels, as well as the lack of an X-ray AGN excess for pairs with $r_p < 20$ kpc. The leftmost panel of Figure 6.16 shows that a large proportion of post-mergers appear in the unobscured "cloud", consistent with my earlier finding in Figure 6.12 that they frequently host visible BLAGN. I also find that circum-nuclear blowout is not ubiquitous after coalescence, since a number of post-mergers remain attenuated, and some lack a BLR detection. A diversity of intrinsic accretion rates and degrees of obscuration in post-mergers are likely responsible for this result, as well as the distribution of observed X-ray luminosities in post-mergers presented earlier in Figure 6.11. A temporal hypothesis could also explain the diversity of attenuation seen in post-mergers, since more recent post-mergers would have had less time to clear obscuring material. Since the CNN-induced mass bias in post-mergers gives rise to a z bias (see Figure 3.4), and galaxies at higher z are less likely to be intrinsically attenuated, the precise roles of merger status and z in setting the quantities in the bottom panel of Figure 6.16 are not easily disentangled. In other words, it is unclear whether coalescence is responsible for the lack of a post-merger attenuation deficit in the bottom panel of Figure 6.16, or whether the z distribution of the post-merger sample has resulted in a biased attenuation fraction.

6.3 Discussion and conclusions

The results of this Chapter are best collated chronologically, starting with the first infall of a galaxy that will eventually merge, and ending after coalescence when the merging galaxies' nuclei become indistinguishable from one another. In general, my findings suggest that it is essential to keep track of two AGN phenomena: SMBH triggering / accretion, and obscuration.

Even without the influence of a merger, isolated galaxies studied in this work appear to span the entire dynamic range of AGN properties. In general, they have relatively low AGN fractions of about 6 – 10 per cent (between 25 – 50 per cent as many as in late-stage galaxy mergers, see Figure 6.13). When AGN are present, they appear to have (on average, and perhaps by definition) moderate luminosities governed by the strength of the secular accretion events that have triggered them. Their observability depends mainly on this luminosity, and is modulated by the orientation of the BLR, dusty torus, and NLR relative to the line of sight (see Figures 6.9 and 6.15).

After galaxies have experienced their first pericentric passage in a major merger (here approximated as $0.1 < \mu < 10$, since my selection targets such galaxies) the tidal torques begin to funnel gas towards the galaxies' centres and onto their SMBHs. I find that this effect is visible for pairs with $r_p < 40$ kpc. While small projected separations do not guarantee post-pericentre status, simulations suggest that pairs at these separations are advanced in the merger sequence: Patton et al. (2024) recently reconstructed the orbits of pre-merger galaxy pairs from IllustrisTNG 100-1, and $\sim$ 91 per cent of galaxies in the Patton et al. (2024) sample with physical separations between $20 < r < 40$ kpc, $\Delta v < 300$ km/s, and $0.1 < \mu < 10$ have experienced at least one pericentric passage with their interacting companion. Inflows ushered in by close pair-phase interactions begin triggering AGN with higher frequency than in isolated galaxies, with an excess of $\sim$ 1.5 – 2 depending on the AGN flavour studied. While AGN are more common, they are observed to have statistically typical uncorrected X-ray luminosities relative to AGN in isolated galaxies (Figure 6.11), and partial attenuation by obscuring gas likely plays a role in determining the observed X-ray luminosities of many well-separated galaxy pairs (third panel, Figure 6.16). In well-separated galaxy pairs, I do not find evidence for elevated incidence of heavy obscuration leading to removal from the eROSITA catalogue ($20 < r_p < 40$ kpc, Figure 6.14). While pair-phase interactions are therefore better at triggering detectable AGN than secular processes, the relative strengths of secular and merger-induced accretion are not well constrained in eROSITA due to obscuration effects.

AGN are still being triggered and fueled at the closest separations (Figures 6.12 and 6.13), and their uncorrected X-ray luminosities are still observed to be typical relative to AGN in isolated galaxies. In the closest galaxy pairs with $r_p < 20$, as with Figure 6.10, it is likely that degrees of

obscuration (but not always total obscuration, as the galaxies studied in Figures 6.11 have eRASS1 detections and are not Compton-thick) of the X-ray source also contribute to the appearance of a relatively sparse (when obscuration leads to an eROSITA non-detection) and average-luminosity (when obscuration attenuates the detected flux) X-ray AGN population. I find that an increasing number of AGN are obscured by dense (potentially Compton-thick, although I do not have measurements of N_H from eROSITA) dust columns in late stage pairs (lowest r_p bin, Figure 6.14). This worsens the reliability of AGN identification methods relying on observations from the SMBH engine, and likely attenuates the X-ray luminosities of galaxies that are detected (Figure 6.16). The precise degree of attenuation is not explored in this work, but future observations allowing characterization of the dust column in heavily dust-obscured AGN will allow this connection to be explored with greater nuance.

When coalescence occurs, there is a noticeable increase in the frequency of AGN. Depending on the AGN selection criteria, AGN are $\sim 2-4$ times more common in post-mergers compared to isolated controls (Figure 6.12). Due to the effects of heavy obscuration in late stage mergers, I argue that these quantitative results do not fully represent the intrinsic accretion states of the SMBHs. In a subset of cases, I expect based on models (e.g., Fabian et al. 2008) and observations (e.g., Ricci et al. 2017) that coalescence-epoch gas inflows can fuel AGN that remain heavily obscured by dense obscuring gas columns (Blecha et al. 2018). Even more than in close pairs, however, obscuration bears on the observability of AGN in post-mergers (Figure 6.14), so the true AGN excess is likely ~ 4.1. For the same reasons as in close galaxy pairs, I expect that the true unobscured luminosity enhancements of the post-mergers are somewhat higher than what is computed in Figure 6.11 (see the left panel of Figure 6.16), but more detailed characterization of the obscuring column would require X-ray spectroscopy (e.g., an analysis similar to that in Pfeifle et al. 2023). Since post-mergers best represent the cumulative effect of a merger event on galaxy evolution, deep X-ray spectroscopy of post-merger samples in the future would help to develop a more nuanced understanding. The ubiquity, demographics, and observability of dual SMBHs in late stage galaxy mergers are lately being explored in multi-wavelength studies (e.g., Ellison et al. 2017; Tubín et al. 2021; Koss et al. 2023), and future case studies on the topic will lend additional context to the statistical results presented in this work. Broadly, the results of this work serve as priors on AGN observability in mergers and isolated galaxies that should be carefully accounted for in subsequent studies of AGN incidence or characteristics.

In comparing my results with the literature, I restrict myself to a discussion of the low-redshift Universe. The X-ray AGN excess in post-mergers I compute is in statistical agreement with what is found in Secrest et al. (2020) and Li et al. (2023b), the two closest literature analogues to this study. In a sample of 43 post-mergers from the Ellison et al. (2013) sample and a matched sample

of 430 non-interacting control galaxies, Secrest et al. (2020) check for 3σ hard (2 – 10 keV) X-ray detections from XMM Newton and find an excess of $2.22^{+4.44}_{-2.22}$. I posit that the smaller post-merger sample and lower statistical power in the Secrest et al. (2020) experiment is primarily responsible for the difference in their results compared to mine, since the excess quantities computed are in good agreement. Additional differences may follow from the X-ray AGN selection used in Secrest et al. (2020), since eROSITA is sensitive to softer X-rays than XMM Newton. Still, the error bars on the AGN fractions and excess in Secrest et al. (2020) wholly encompass the equivalent measurements in this work, so any discrepancies owed to AGN selection methodology are secondary to the influence of sample size.

The Li et al. (2023b) post-merger sample is selected by eye from SDSS and Legacy Survey (which includes DECaLS) imaging, and the AGN in Li et al. (2023b) are identified using 2 – 10 keV fluxes from both Chandra and XMM (for the X-ray), BPT diagram criteria similar to this work (for optical AGN; though, with a minimum emission line S/N of > 3, while I use S/N> 5) and the same WISE $W1 - W2 > 0.5$ colour criterion as this work (in the mid-IR). Compared to matched controls, Li et al. (2023b) find post-merger AGN excesses of 2.4 – 2.6, 3.4 – 3.8, and 6.7 for X-ray, NLAGN, and mid-IR AGN, respectively. Li et al. (2023b) also use radio observations from the VLA FIRST survey to check for an excess of radio cores in post-mergers, but do not find a significant excess. While the radio AGN criteria used by Li et al. (2023b) are different than the ones I used in Chapter 5, the lack of a statistically significant result in the radio by Li et al. (2023b) is consistent with my findings. The multi-wavelength analysis of Li et al. (2023b) is therefore in broad agreement with my own, with both studies indicating a multi-wavelength AGN excess of $\sim$ 2 – 4 depending on the criteria used. The measurements agree even though the multi-wavelength AGN fraction reported in Li et al. (2023b) is much higher than my own (77 per cent), owing to differences in multi-wavelength AGN selection (particularly for NLAGN and X-ray AGN). I also combine my study of heavily obscured AGN candidates (in Figure 6.14) with my X-ray AGN results (Figure 6.10) and attempt a correction for the effects of obscuration. Implementing the correction, I estimate that the intrinsic (or obscuration-corrected) X-ray AGN excess in post-mergers is $\sim$ 4.1, also in statistical agreement with the result presented for AGN with $L_X > 10^{42}$ or $L_X > 10^{43}$ in Figure 11 of Li et al. (2023b).

My suggestion of a high obscured AGN fraction in close galaxy pairs is in tension with Hou et al. (2023), who find that Compton-thick obscuration is not responsible for the low incidence rate of AGN in a sample of 7 galaxy pairs (14 galaxies) interacting within $r_p < 5$ kpc, instead positing that in late-stage galaxy pairs, AGN feedback has already begun to expel circum-nuclear gas, leading in turn to lower accretion rates. These two effects are not easily distinguished using the observations in this work. I therefore allow for the possibility that the mIR¬X AGN popu-

lation studied in Figure 6.14 could contain both AGN that are obscured by dusty gas brought in by merger events and AGN whose accretion rates are diminished by feedback. The late-stage pair feedback scenario is not obviously compatible the multi-wavelength analysis presented in Section 6.2.4, since both NLAGN and mid-IR AGN are still being triggered in late galaxy pairs, but the timescales for dust cooling (for the mid-IR case; Ichikawa & Tazaki 2017) and narrow-line recombination (for the NLAGN; Peterson et al. 2013) could allow for sustained detection of narrow-line and mid-IR AGN even after AGN feedback has diminished the instantaneous accretion rate. Conversely, the detection of a merger excess amongst luminous and obscured AGN in Koss et al. (2018) suggests that obscuration and accretion can coincide in late stage mergers. Blecha et al. (2018) uncover an analogous result in simulations, reporting that mergers achieve simultaneous maxima in both N_H and L_{AGN} at the time of coalescence (Figure 1 of Blecha et al. 2018). It is therefore likely that some proportion of my mIR¬X AGN are in fact heavily obscured. Other studies tracing AGN via hard X-rays in close galaxy pairs (e.g., Koss et al. 2012, 2016) have found a merger-AGN connection; such results are consistent with the hypothesis that the relatively soft X-ray response of eROSITA may be responsible for a number of X-ray undetected AGN in this work. Since BLAGN are typically only observable through low column densities ($N_H < 10^{22}/cm^2$), the co-detection of many X-ray AGN as BLAGN in this work indicates that my X-ray AGN sample is generally not heavily obscured (Oh et al. 2022). My suggestion that obscuration can often clear post-coalescence is in degrees of tension with the findings of Matzko et al. (2022), who find no statistical evidence for an elevated incidence rate or luminosity of ionized outflows in post-mergers. The tension can potentially be reconciled by the fact that Matzko et al. (2022) require MPA-JHU stellar masses for their work, and BLAGN are nominally excluded from the MPA-JHU sample. In this work, meanwhile, an abundance of BLAGN in post-mergers are the primary piece of evidence for preferential clearing of obscuration in the post-merger phase. Evidence for outflows could potentially be found using the method of Matzko et al. (2022) in galaxies that are excluded from the Matzko et al. (2022) post-merger sample, but a different stellar mass estimate and an adjusted spectral fitting procedure accounting for a BLR component would need to be implemented.

Throughout this work, post-mergers are shown to be unique by several criteria: they more often host AGN than secular or pair-phase galaxies, and they are a major channel for the production of both heavily obscured AGN and type 1 AGN with visible BLRs and accretion disks after the removal via radiation pressure of dusty gas in the nucleus.

Chapter 7

The potential of CNNs for merger identification in current and next-generation imaging surveys

In Chapters 4, 5, and 6, I used the post-merger sample identified in Chapter 3 to measure the typical impact of coalescence after a merger on galaxy properties. The hybrid CNN and visual inspection framework used in this work is a straightforward way to identify a highly pure sample of likely post-merger candidates for subsequent study. However, the completeness and representativeness – the extent to which the merger sample identified contains post-mergers that are like the underlying sample of true recent post-mergers in the Universe – of the samples identified in both CFIS (Section 3.1.3) and DECaLS (Section 3.2.2) depends on a number of factors, including:

1. The morphological similarity of galaxies in IllustrisTNG100-1 to galaxies in the Universe,
2. the quality of observational realism applied to the simulated galaxies used for CNN training,
3. the extent to which the signatures of recent mergers are still distinguishable above the noise after an observation,
4. the ability of a CNN to correctly identify mergers with distinguishable merger features, and
5. the frequency with which mergers successfully identified by the CNN are confirmed during visual inspection.

Item (1) is already explored in some depth in the literature, notably in Rodriguez-Gomez et al. (2019) and Eisert et al. (2024), who demonstrate 1σ statistical agreement in the non-parametric morphological parameters (for the former) and 70 per cent overlap in the parameter space uncovered via contrastive learning (for the latter) between galaxies in the Universe as imaged by two different surveys and galaxies from IllustrisTNG. Ferreira et al. (2024) also demonstrate that Uniform Manifold Approximation and Projection (UMAP) representations of the parameter spaces

occupied by CFIS and IllustrisTNG galaxies with CFIS realism are in good qualitative agreement. The similarity between real galaxies and those simulated in IllustrisTNG is also verified to some extent by the efficacy of the CNN used in Chapter 3. While the realistic appearances of TNG galaxies therefore place limitations on the maximum potential of my approach, they are not so much a hinderance as to prevent the successful identification of mergers in the low-z Universe in this work.

Item (2) is the main topic of Bottrell et al. (2019b) and is discussed in additional detail in Bottrell (2020), wherein it is concluded that observational realism sets another limit on the completeness of the merger samples identified using machine vision. It is shown in Bottrell et al. (2019b) that one of the main simplifications used in Chapter 2 (foregoing radiative transfer and instead using two-dimensional stellar mass maps as the starting point for mock observations) can be done with acceptable levels of loss (a multi-objective optimization score of 90.2 per cent, compared to 96.0 per cent when the CNN in that study is both trained and evaluated on images where radiative transfer is used). I also note that adopting an approach using radiative transfer for simulated galaxy images would require a new set of assumptions (e.g., that the stellar populations, chemical evolution, and dust characteristics of simulated galaxies are realistic) that would need to be independently verified.

Items (4) and (5) are already explored in Sections 2.2.1 and 2.2.5 of Chapter 2, demonstrating that in CFIS imaging, the CNN can successfully identify genuine mergers $\sim$ 88 per cent of the time. Other studies in the literature have been successful in applying CNNs towards the goal of merger identification (see Table 1.1). I also characterized statistically my expectation that visual classifications by people are highly sensitive to the amount and type of experience of the classifier, and therefore difficult to reproduce. One of the principal benefits of using a CNN first is that the classifications are reproducible by construction, and the role of visual classification is partially limited. Taken together, the results of Chapter 2 suggest that $>$ 10 per cent of post-mergers in the Universe may be lost by the CNN in CFIS imaging, and that an additional proportion will be lost during visual inspection. Some of the loss during visual classifications is necessary for the identification of a sample that is rhetorically defensible, since a number of genuine post-mergers in the Universe must have ambiguous morphologies (the most obvious examples of ambiguous cases are genuine post-mergers that belong to a galaxy pair as well).

Item (3) has remained largely unexplored in my work up to this point, since merger identification work is downstream of the availability of large galaxy imaging catalogues. Still, some efforts have been made in the literature begin to address this question: for example, Martin et al. (2022) investigate the potential observability of low-surface brightness features in the Legacy Survey of Space and Time (LSST) taken at the Vera Rubin observatory, and Domínguez Sánchez et al.

(2023) perform a similar assessment using CNNs to identify tidal features in Hyper Suprime-Cam imaging. The most direct literature precursor to this study, Wilkinson et al. (2024), contains a systematic analysis of merger observability on a grid of depth and resolution, using individual and machine-combined non-parametric morphological statistics. It has been the assumption of this work that better (i.e., deeper, higher resolution) imaging data is beneficial, and that quality can sometimes be sacrificed for quantity (i.e., extent of on-sky coverage) in pursuit of large merger samples and scientific results with good statistical significance. As large, high-quality imaging data is more often becoming widely available, survey quality has become a factor that astronomers can control. Greater degrees of control will be available in coming years, with data from next-generation observatories like the Vera Rubin Observatory becoming available. In the context thus outlined, item (3) becomes more interesting. In this Chapter, I will explore the extent to which the success of low-z post-merger searches is limited by imaging, using five real surveys – SDSS, DECaLS, CFIS, the Hyper Suprime-Cam Subaru Strategic Program Wide field (hereafter HSC-W), and the projected 10-year depth of LSST (hereafter referred to simply as LSST) – to sample the technical parameter space broadly referred to as "imaging quality".

7.1 Data and methods

Several methodological improvements and insights have been developed since producing the results presented in Chapter 2, so my approach to this comparative study is described here in detail, even though the experiments are philosophically similar (involving training CNNs on mock images of post-merger and control galaxies from IllustrisTNG100-1). In this Section, I will explain how post-merger and non-post-merger control galaxies are selected from the simulation, how mock observations of the galaxies are performed, and how the CNN is trained. I will also detail how predictions of merger status are made.

7.1.1 IllustrisTNG 100-1 galaxy selection

In this Chapter, I use matched samples of post-merger galaxies and control galaxies to explore the potential of CNNs as a function of image quality. The selection for both classes is updated from that used in Chapter 2, and is designed to select a somewhat smaller number of galaxies that are more unambiguously post-mergers or non-post-mergers.

Post-mergers

The post-mergers and control galaxy selection for this work is conducted by Leonardo Ferreira for use in Ferreira et al. (2024), but is included here for completeness in narrative. Ferreira et al.

(2024) itself uses the Byrne-Mamahit et al. (2024) approach of searching the IllustrisTNG 100-1 merger trees. As I did in Chapter 2, Leonardo Ferreira uses galaxy metadata from IllustrisTNG to select galaxies from snapshots between simulation redshifts $0 < z < 1$. As argued in Ferreira et al. (2024), galaxy morphologies outside of this redshift range are expected to be statistically and qualitatively different from those found at low-z, e.g., with $z < 0.3$ where the post-merger samples from Chapter 3 lie.

Simulation quality control cuts are also enforced; SubhaloFlag=True ensures that galaxies have dark matter and total mass properties consistent with cosmological origin, and $\log(M_\star/M_\odot) > 10$ ensures that each galaxy contains at least ~ 7000 star particles. Departing from the selection used in Chapter 2, there is also an upper mass cut at $\log(M_\star/M_\odot) < 11$ – since a disproportionate number of galaxies with high stellar masses are experiencing a merger, it is difficult to match a representative sample of non-merging control galaxies to them. From this subset Ferreira et al. (2024) select a general merger sample with stellar mass ratios $\mu > 0.1$, and identify a broad post-merger class including all galaxies that have coalesced after a merger that took place in the last 1.7 Gyr. Ferreira et al. again follow Byrne-Mamahit et al. (2024) and use an updated measurement of μ, which considers the mass ratio between interacting companion galaxies at a time when they were at least 50 kpc apart. In comparing the masses of interacting galaxies while they are still well separated, my colleagues avoid the effects of "numerical stripping", a simulation problem in which galaxy masses can be under-estimated when particles are wrongly associated with a nearby neighbour. For this work, I take the subset of immediate post-mergers for which coalescence took place in the simulation snapshot prior to imaging. The final sample of post-mergers for this chapter contains 1627 galaxies. The main differences between the post-merger sample used in this chapter and that used in Chapter 2 are the upper limit on $M_\star$, the updated μ estimate for the simulation, and the enforcement of the SubhaloFlag.

Controls

An equal number (1627) of control galaxies are matched by Leonardo Ferreira to the post-mergers just described in Section 7.1.1. The control pool includes two types of non-merger galaxies: "true" and "potential", the former of which has not experienced a merger in the last 1.7 Gyr, and the latter of which has not experienced a merger in the last 1.7 Gyr, and is unlikely to experience a merger event in the near future based on a nearest neighbour separation criterion ($r_1 > 50$ kpc). The potential non-mergers are distinct from true non-mergers because the simulation runtime came to an end before their merger trees can be fully inspected for potential future merger events.

Individual control galaxies are matched to the post-mergers on three parameters: $M_\star$, z, and

gas fraction f_{gas}, where f_{gas} is defined as the gas mass divided by the sum of the stellar and gas masses. The $M_\star$ matching tolerance is 0.05 dex, and the f_{gas} tolerance is 0.05. Control galaxies must also be taken from the same simulation snapshot (i.e., with identical lookback time) as their post-merger counterparts. The inclusion of f_{gas} as a matching parameter is different from what is used previously (e.g., in Chapter 2). The change is motivated by the scientific goals of Ferreira et al. (2024), which include an eventual interest in fairly characterizing increases in star formation efficiency as a function of merger stage. Furthermore, gas content and the availability of gas for forming new stars is known to affect the observed morphologies of post-mergers; see Wilkinson et al. (2024), Lotz et al. (2010a), and Bell et al. (2006). The nearest neighbour in the $M_\star - f_{gas}$ plane is taken as the best control for each merger. Controls are matched without replacement so that the final samples of mergers and controls contain the same number of galaxies. Figure 7.1 shows that the merger identification and control matching procedures have been effective in identifying statistically comparable galaxy samples whose primary difference is merger status.

7.1.2 Mock observations

The mock observation approach used in this Chapter is the same as in Chapter 2, with methodological substitutions aligning with this Chapter's role as a methods comparison. First, redshifts are no longer chosen at random from a real population of galaxy redshifts. Instead, I identify five redshift bins between $0 < z < 0.3$ that contain equal numbers of galaxies in SDSS DR7. The redshift values in the centres of each bin are $z = 0.036, 0.101, 0.150, 0.198, 0.256$. At each of the five redshifts, I perform one mock observation for each of four camera angles at the vertices of an equilateral tetrahedron with the galaxy at its centre. I therefore make a total of 20 (five redshifts times four camera angles) mock observations of each galaxy. Since real skies are not available from all of the surveys I plan to study (10-year imaging co-adds from LSST are not yet available at the time of writing), I use synthetic image backgrounds based on the reported observational parameters for each survey. I repeat the mock observation procedure for each of the five surveys, so that the complete image dataset used in this Chapter includes 20 images for each of 1627 post-mergers and 1627 controls for each of five surveys. The observational parameters for each of the surveys are outlined in Table 7.1, including the CCD pixel scale (on-sky angle subtended by a single pixel), the PSF (expected or known instrumental PSF for the survey, dominated by the atmosphere in all cases), σ_{sky} (the standard deviation of the Gaussian sky noise in magnitudes per square arcsecond used to generate the mock observation, derived by Scott Wilkinson for this work), and the 5σ point-source depth (a convenient quantity reported as an at-a-glance metric of the limiting depth for many surveys). The essential statistics used to determine σ_{sky} for each

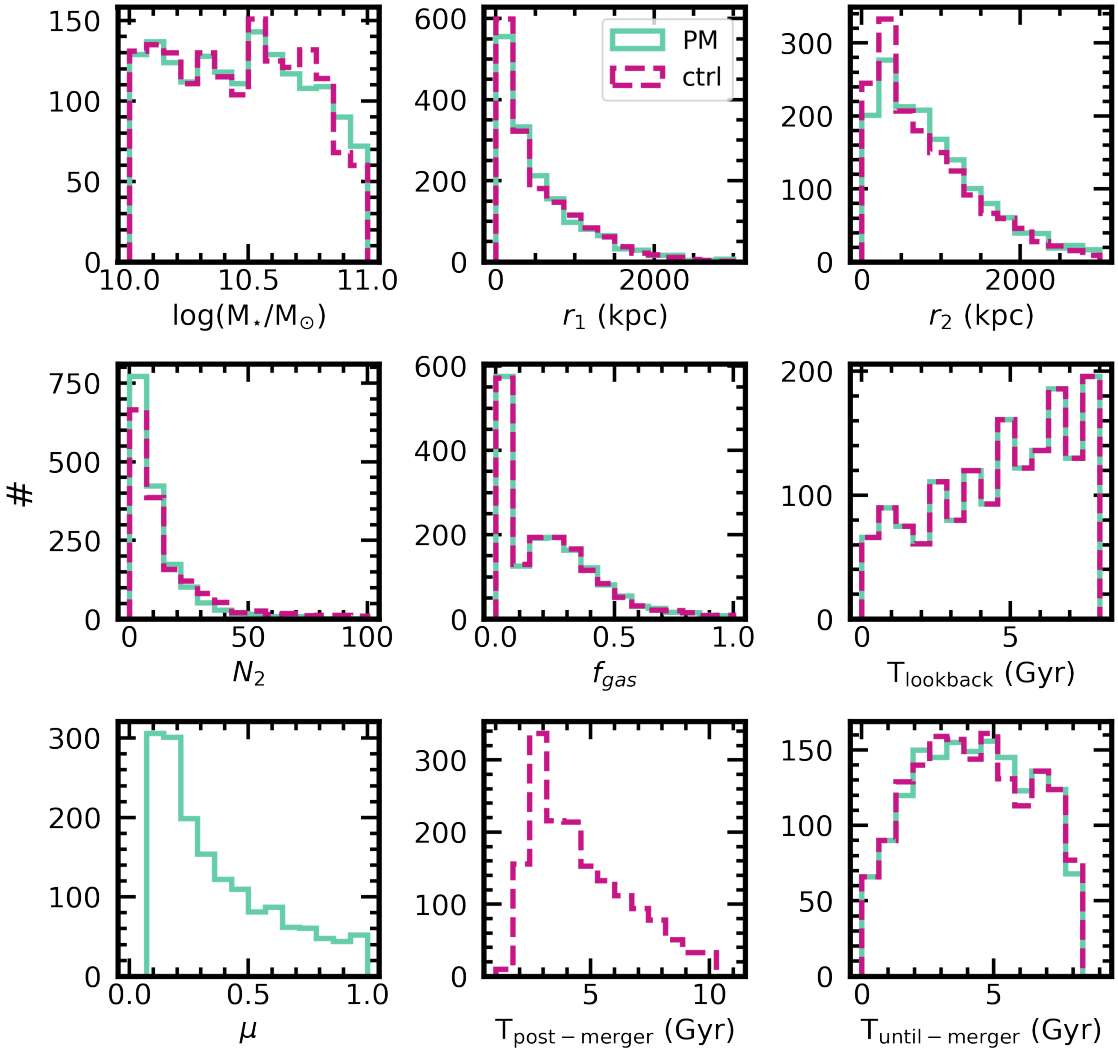

Figure 7.1: The stellar mass, lookback time, gas fraction (f_{gas}), mass ratio μ, and environment (r_1, r_2, N_2) statistics for the post-mergers (teal) and non-merger controls (magenta) used to train the CNNs in this work. I also include $T_{Postmerger}$ for the controls (but not for the post-mergers because they all have $T_{Postmerger} = 0$) and $T_{Untilmerger}$ for both classes.

Survey (units)	CCD pixel scale (arcsec)	PSF (arcsec)	σ_{sky} (mag)	5σ point-source depth (mag)
SDSS	0.396	1.4	24.33	22.7
DECaLS	0.262	1.18	24.97	23.54
CFIS DR5	0.186	0.69	25.6	25.0
HSC-W	0.17	0.75	26.95	26.5
LSST 10y	0.2	0.7	28.12	27.5

Table 7.1: Essential parameters describing the image quality of *r*-band data from each of the surveys included in this comparison. The 5σ point-source depths and PSFs are plotted against one another later in Figure 7.10.

survey were found at the following sources: SDSS[1], DECaLS (Dey et al. 2019), CFIS DR5[2], HSC-W (Aihara et al. 2022), and the LSST 10-year co-adds (Ivezić et al. 2019).

A new approach for generating synthetic sky backgrounds was developed for this effort by Scott Wilkinson. After the light from the synthetic galaxy is re-binned to the required CCD pixel scale and blurred with a Gaussian approximation of the survey PSF (see Section 2.1.3), the parameter σ_{sky} is needed to specify the amplitude of the noise that will be added to approximate the sky. Computing σ_{sky} analytically is non-trivial, since σ_{sky} is sensitive to the CCD pixel scale, the PSF, the survey's true sensitivity (approximated here by the 5σ point source depth), and the aperture within which the signal (e.g., from a galaxy image or a point source) is compared to the noise. In order to determine σ_{sky} in a homogeneous way for each of the five surveys, Scott Wilkinson used the following procedure:

- Create a point-source with a brightness equal to the reported limiting 5σ point-source depth.
- Convolve the source with the PSF of the relevant survey.
- Add the source to a test image containing an arbitrarily high level of Gaussian noise at the pixel scale of the survey.
- Using SourceExtractor (Bertin & Arnouts 1996) as implemented in Python by Barbary (2016), measure the flux and flux error within some circular aperture.

[1]https://classic.sdss.org/dr7/
[2]https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/community/unions/MegaPipe_CFIS_DR3.html

- Repeat all previous steps with incrementally less noise in steps of 0.025 mag arcsec^{-2} until the flux divided by the flux error is at least 5 (i.e., 5σ).

- Repeat all previous steps 500 times and take median to ensure there is no dependence on noise.

- If the aperture used to estimate the 5σ point source depth is not given, repeat all previous steps with varying circular apertures until the flux (at the point of a 5σ detection) is equal to the reported 5σ limiting point source depth of the survey in magnitudes.

- Report the standard deviation used to generate the final sky background as $\sigma_{\rm sky}$.

Since the modified version of RealSim (Bottrell et al. 2019b) used to generate mock CFIS images in Chapter 2 accepts a false sky σ, CCD pixel scale, and PSF as input, the availability of σ_{sky} means that image quality is now fully parameterized. Other than the modifications to the redshift selection and false sky generation methods, the mock observation and image normalization procedures are the same as outlined in Section 2.1.3.

Figures 7.2 and 7.3 show a useful cross-section of the data used to train and evaluate the CNNs in this Chapter. In each row, both Figures show the five z realizations created for a single camera angle of one post-merger (for Figure 7.2) and one control (for Figure 7.3) in a single survey. Subsequent rows show the corresponding images from each of the five surveys considered. In SDSS, the visibility of the prominent tidal tail in the post-merger galaxy is hidden in the image noise by $z = 0.2$, and the post-merger and control look indistinguishable from one another in SDSS in my highest-z realization. The other surveys retain the low-surface brightness features associated with the merger across the redshift range, but the deepest images from the LSST camera (fourth row in each mosaic) allow for obvious distinction between the two classes even at the highest z studied here.

7.1.3 CNN architecture and training strategy

In this chapter, I train CNNs on image data prepared according to the realism parameters of each survey. Images for training are partitioned into training, test, and validation sets, which represent 80, 10, and 10 per cent of the data respectively. All 20 images associated with each post-merger galaxy, and all 20 images associated with the matched control for a given post-merger galaxy, only ever appear within one partition. The aim of this strategy is to prevent any amount of cross-contamination between the partitions, which could give rise to artificially

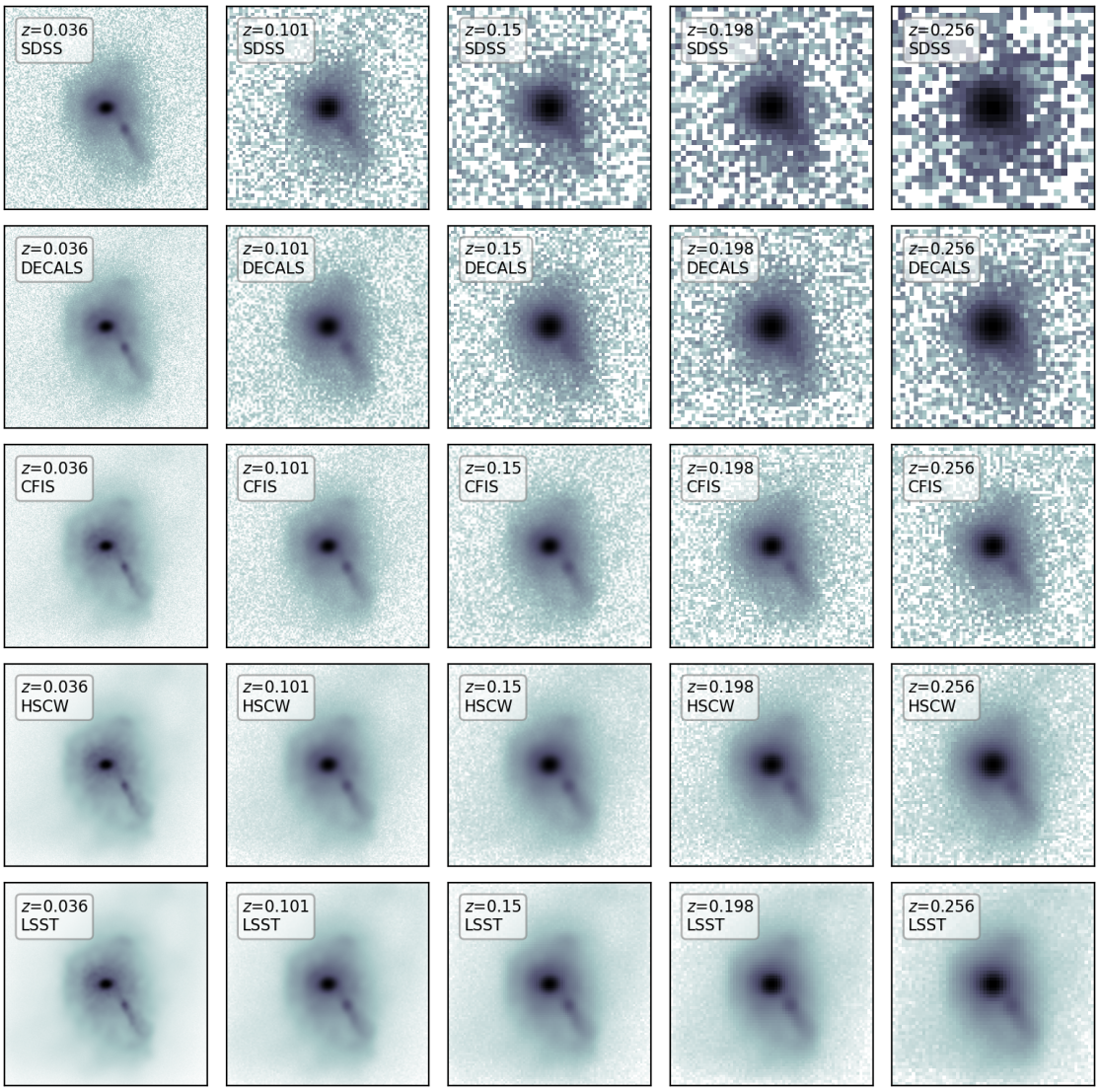

Figure 7.2: A mosaic showing all of the survey and redshift realizations for one camera angle for one galaxy from the post-merger sample. Redshift increases from left to right, and particularly in the shallower surveys studied in this Chapter (e.g., SDSS, DECaLS) the decrease in visibility of the merger-induce tidal tail with increasing mock observation z can be noted. The differences between rows highlight the importance of PSF, CCD scale resolution, and limiting depth in preserving the morphological signatures of a recent merger event. The galaxy images are shown with 1σ log-normalized scaling to better highlight the low-surface brightness features.

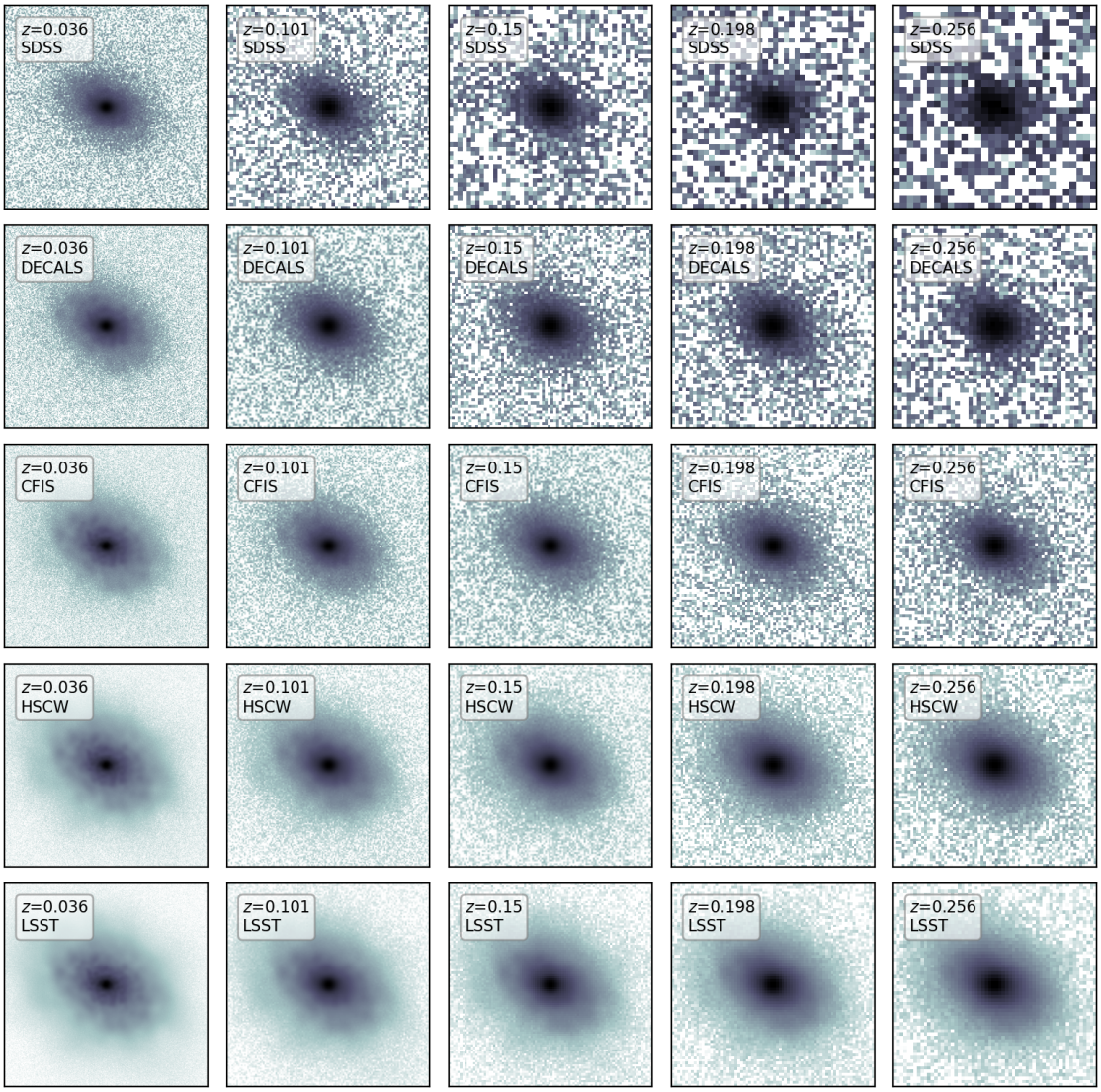

Figure 7.3: The same as Figure 7.2, but for one galaxy from the control sample. Even for an undisturbed galaxy, substructure detail becomes increasingly visible for surveys with deeper limiting magnitudes or finer resolution.

high performance metrics. Moreover, the images corresponding to each galaxy appear within the same partition for all five surveys. In this way, each of the five CNNs is trained on nearly identical data, with the only difference being image quality. Controlling for as many nuisance parameters as possible (e.g., information leaking between data partitions, small fluctuations in performances due to differently-shuffled data) maximizes the significance of the comparison between the CNNs conducted later in this Chapter.

The CNN architecture used in this Chapter is identical to that used throughout this work (first described in Chapter 2) except for the dimensions of the input layer. Since a 138×138-pixel input image was chosen in Chapter 2 based on the z demographics of galaxies in CFIS, it is not necessarily fair to enforce the same input image size for all five surveys, whose selection functions include galaxies with different z distributions than CFIS. Images in this chapter are resized instead to 128×128 pixels using the anti-aliasing resize algorithm from SKIMAGE[3] (van der Walt et al. 2014), and the final images are large enough to capture the essential detail in the galaxy images. The same on-the-fly augmentation algorithm as in Chapter 2 is used again. All models are allowed to train for an arbitrarily long time as long as their performance is improving – I monitor the value of the loss function (binary cross-entropy, which characterizes cross contamination between the post-merger and control classes), and if the loss does not improve for 50 training epochs[4], training is terminated and the model weights from the epoch with the best validation performance are restored. Training is accelerated using multiple GPUs and the "mirrored strategy" approach currently being implemented in TENSORFLOW[5], in which the responsibilities of training are parallelized across multiple graphics processing units (GPUs). Importantly, other than a factor decrease in training time proportional to the number of cores used, the models behave the same whether or not the mirrored strategy is used.

The model architecture from Chapter 2 was not the only one considered for this work, since a number of updated machine vision models have been developed and popularized in the years between Chapter 2 and this Chapter. I tested two other architectures on the deepest and shallowest image datasets (LSST and SDSS, respectively) in order to assess whether the performance achieved in this Chapter was limited by the data (this is the goal, as I am interested in the influence of data on the accuracy of the classifiers) or by the model. The first model tested was an implementation of AlexNet (Krizhevsky et al. 2012), on which my CNN architecture was originally based. The second was the stock KERAS implementation of an EfficientNetB0[6] architecture.

[3] scikit-image.org

[4] One epoch represents a complete study of the training dataset, followed by a performance check on the validation set.

[5] https://www.tensorflow.org/api_docs/python/tf/distribute/MirroredStrategy

[6] keras.io/api/applications/efficientnet/

AlexNet and the EfficientNetB0 models were chosen because my CNN architecture sits between them in sophistication and modernity. Both of the new models were also adjusted to accept 128×128-pixel monochromatic images as input. Ultimately, both AlexNet and EfficientNetB0 find performance maxima on the SDSS and LSST datasets that are similar (within ~ 2 per cent in class-wise accuracy) to what is reported in Section 7.2, but the training histories (the path of improvement over the course of training epochs) of the two stock models are not smooth. The difference in behaviour during training is likely owed to the large training-time dropout fractions implemented in my CNN, which in Chapter 2 were necessary to encourage the model to generalize, and appear to serve the additional purpose of avoiding sometimes-problematic "tangents" in parameter space wherein models spend several epochs considering fruitless alternative solutions before returning to the global minimum. It is nevertheless encouraging that the minima found by all three models are in good agreement with one another. The models' similar global performance suggests that the results of this study are data-limited, rather than model-limited. As with Chapter 2, I note that with enough iteration, it is possible that a more sophisticated architecture could have yielded improvements in certain regions of the image quality parameter space. Still, the primary aim of this Chapter is to conduct a comparative study of CNN performance as a function of imaging quality, and my tests using the EfficientnetB0 architecture suggest that the conclusions presented below in Section 7.2 are not highly sensitive to the particular CNN chosen.

7.1.4 Inference

After the training phase of this project, I have produced five CNNs, one each for the five surveys studied here. Each CNN is evaluated on an unseen test dataset (the 10 per cent of reserved galaxies) in order to study the best-case scenario, in which a model is applied to data with composition similar to its training data. However, additional utility may be extracted from trained machine vision models by applying them outside of their training regimes. To quantify this value, I also perform an exhaustive cross-survey inference effort in which each model is evaluated on the other four training datasets (e.g., the SDSS model is evaluated on the SDSS, DECaLS, CFIS, HSC-W, and LSST test datasets). The model classifications taken from each of 25 inference sets (five models times five test sets) constitute the new material used to characterize the sensitivity of merger classifications to imaging quality in the Results Section below.

7.2 Results

Having trained all five models and used them to make merger predictions on the test set galaxies, I now present the results, which have been separated into two main categories: same-survey results

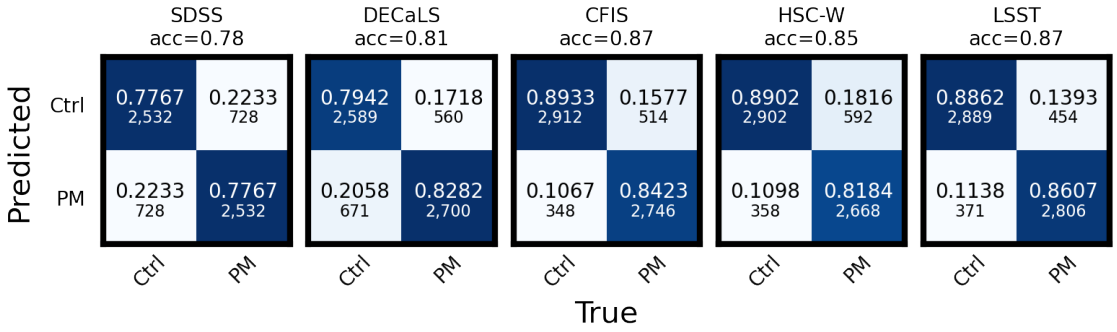

Figure 7.4: Confusion matrices for the five trained CNN models, evaluated on test data with the same survey realism as their training data. The accuracy on post-mergers (bottom right corner of each confusion matrix) and on non-merger control galaxies (top left corner of each matrix) are clearly sensitive to depth, but resolution plays a role in determining whether models are more successful at identifying post-mergers or controls. It is likely that very deep imaging reveals additional diffuse or faint structure in normal galaxies that might falsely appear to have been the result of a merger event.

(Section 7.2.1, in which models are evaluated on data with the same construction as their training sets) and cross-survey results (Section 7.2.3, in which the potential for models to succeed outside of their training regime is explored). I will also pay particular attention to the trends of CNN performance with observational parameters that could influence the apparent strength of faint tidal morphologies, including redshift, depth, spatial resolution (pixel scale and PSF), as well as physical galaxy parameters (merger mass ratio). In all cases, "completeness" and "accuracy" refer to the fraction of galaxies belonging to a given class (post-merger or control) that are classified correctly by the model, while "purity" and "precision" refer to the fraction of galaxies predicted by the model to be a given class that truly belong to that class.

7.2.1 Same-survey results

The overall success of the CNNs in distinguishing post-mergers from controls can be summarized by conventional machine learning figures of merit for classification – class-wise accuracy (i.e., the fractions of post-mergers and controls that are classified correctly), ROC curves, and purity-completeness curves. The CNNs' performances can best be interpreted in the fundamental context of automated classification: that each model is attempting to generalize over the intrinsic and observational diversity of the merger and control samples in order to correctly classify as many galaxies as possible. When models are less successful in classifying a certain subset of galaxies (e.g., one of the classes, or galaxies belonging to a certain mass range), it is because the model's overall performance was improved even as the subset in question was comprimised.

Figure 7.4 shows the confusion matrices for models trained and tested on data with realism based on the same survey. The surveys are arranged in order of increasing 5σ r-band limiting point-source depth from left to right. SDSS has the lowest total accuracy, as well as the lowest class-wise accuracies, for both post-mergers and controls. Based on the visual quality of the images, this is unsurprising – galaxies at the highest z studied cannot typically be separated by eye in SDSS (see Figure 7.2), and visual separability is often a reasonable predictor of the potential of machine vision methods for a given problem. The exact balance between the post-merger and control classes in SDSS suggests that the visual degeneracy is equally confusing in both classes. In trying to adjust itself to account for the lack of merger features in many galaxies that are labeled as mergers, the model ultimately suffers a decrease in overall accuracy.

In DECaLS quality imaging, the accuracy is improved somewhat for both classes, but especially for the post-mergers, presumably since a larger number of galaxies with post-merger labels actually exhibit merger-like morphologies in the training data. Accuracies for both classes are again improved in CFIS imaging, since the morphological differences between mergers and controls are more visually distinct with better depth.

Referring to Figures 7.2 and 7.3, one can begin to interpret the changes in performance for surveys with better 5σ limiting point-source depths than CFIS. At CFIS depth and better, the tidal tail feature in the merger image is visible across the entire redshift range, suggesting that CNNs will nominally be able to distinguish between the classes unless there is a spurious effect (e.g., a chance viewing angle from which a galaxy's tidal features are obstructed, or a merger that is intrinsically unusually faint). Critically, the non-merger control galaxies in HSC-W and LSST images have very different appearances. Particularly at low-z (either $z = 0.036$ or $z = 0.101$), increasing amounts of substructure are visible in the non-mergers. In some cases, the substructure could be a source of confusion when it is visible.

For the HSC-W and LSST model accuracies in Figure 7.4, I posit that a unique combination of depth and spatial resolution effects give rise to the metrics presented. Neither the HSC-W nor LSST models exceed the control class accuracy achieved by the CFIS model, suggesting that a reduced ability to distinguish between substructure and merger features in deep imaging is persistent. The benefits of deep imaging mainly appear to benefit the LSST model's ability to identify post-mergers. The LSST model has the highest completeness score for the post-merger class. In interpreting the global performance results, I emphasize that all five models have converged (without improving on the validation data set for 50 epochs). Rolling any of the models back to epochs where their class-wise performances were less discrepant (i.e., choosing an earlier training epoch for which the models had a smaller performance gap for post-mergers and controls) results in a significant decrease in total accuracy. Instead, I present the final accura-

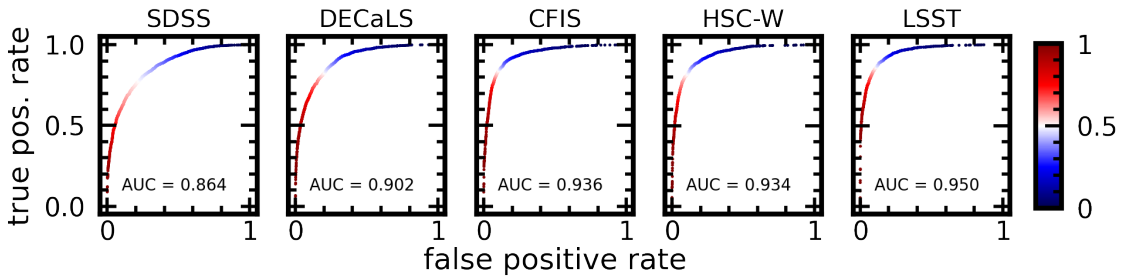

Figure 7.5: Receiver operating characteristic (ROC) curves, another merit diagram for CNN classifiers, for the five CNNs. ROC curves plot the true positive rate (fraction of post-mergers correctly identified) and the false positive rate (fraction of incorrectly classified post-mergers) as a function of model decision threshold. The area under the ROC curve is also a figure of merit, with 0.5 equivalent to random performance, and 1.0 indicating perfect separation between the classes. ROC AUC score also increases generally with depth, reflecting again the importance of imaging depth in merger identification.

cies of the models without additional tuning or adjustment, since any out-falling issues with the classification performance of a given model is a function of the imaging quality.

There is also an apparent ceiling (at least for the degrees of survey realism studied here) for CNN performance for any class-wise or global accuracy at $\sim$ 90 per cent. Since the galaxy morphologies are easily visible across the entire redshift range in the deepest data, I propose that this apparent limit is a consequence of intrinsic degeneracy between the stellar morphologies of galaxies belonging to the post-merger and control classes (e.g., some mergers have truly undisturbed appearances as a result of small μ or inconvenient viewing angles, regardless of image quality) or global difficulties in generalizing over the z domain. I do not compare the performance of the CNN on survey-realistic images to its performance on "idealized" data, since multiple z realizations are not possible in a realism-free context. As such, the CNN would need to be trained on a significantly smaller dataset, and the comparison would no longer be direct.

I next use the receiver operating characteristic (ROC) curves to collapse each model's performance down to a single figure of merit, an area under the curve (AUC) score. Figure 7.5 shows ROC curves for the five CNNs, with the AUC score shown in each panel. The curves generally enclose more area with increasing depth, particularly in SDSS (with AUC=0.864), DECaLS (AUC=0.902), CFIS (AUC=0.936), and LSST (AUC=0.950) imaging. The AUC score for HSC-W (AUC=0.934) is lower than for CFIS, reflecting the difficulties in classifying galaxies at a range of z as depth increases. In the context of this experiment, additional depth and resolution beyond what is offered by CFIS are not categorically beneficial to the models' classification accuracies or the purity of the predicted post-merger samples.

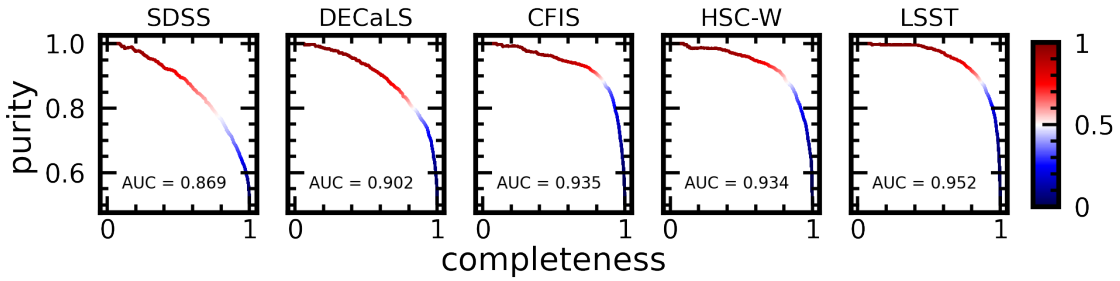

Figure 7.6: Purity-completeness (or precision-recall, in machine learning parlance) curves for all five models evaluated on like data. Each panel show the purity (or precision of the predicted merger sample) and completeness (recall of the predicted merger sample) as a function of the model's decision threshold (shown on the colour bar). The area under the curve is also a figure of merit, with an AUC of 0.5 indicating performance consistent with random, and an AUC of 1.0 indicating perfect separation between the classes by the model. The area under the curve scales generally with the depth of the survey studied, indicating that the ability of my CNN architecture to identify samples that are degrees of pure and complete depends on the observability of faint features in the images.

Survey	PM compl.	Ctrl compl.	Accuracy	Purity	ROC area	PCC area
SDSS	0.78	0.78	0.78	0.78	0.86	0.87
DECaLS	0.83	0.79	0.81	0.80	0.90	0.90
CFIS DR5	0.84	0.89	0.87	0.89	0.94	0.94
HSC-W	0.82	0.89	0.85	0.88	0.93	0.94
LSST 10y	0.86	0.88	0.87	0.88	0.95	0.95

Table 7.2: The class-wise completeness, overall accuracy, and post-merger purity statistics, as well as AUC scores for the ROC and purity-completeness curves for the CNNs trained for each survey.

I also show the purity-completeness curves (PCCs) and the corresponding AUCs for all five CNNs in Figure 7.6. Purity-completeness curves are distinct from ROC curves in that they highlight the model's ability to return a sample of post-mergers that is pure and complete as a function of CNN $p(x)$. The area under the purity-completeness curve somewhat masks the issue that the HSC-W trained model had with identifying balanced samples of post-mergers and controls. Since the HSC-W model (AUC=0.94) has exchanged some amount of completeness to identify a more pure post-merger sample, it ultimately earns a nearly identical purity-completeness AUC score to the CFIS model. The CFIS-trained model (AUC=0.94) impressively achieves strong performance for both classes at a shallower depth. The same-survey figures of merit (completenesses on post-mergers and controls, and the areas under the ROC and purity-completeness curves, total accuracy, and post-merger sample purity) for all five models are summarized in Table 7.2.

Having stated the overall figures of merit, I will present the accuracies of the classifiers as a function of several quantities that I expect would bear on a CNN's ability to distinguish between mergers and controls. Figure 7.7 shows the post-merger completeness and purity scores for the five CNNs, plotted as a function of the reported r-band limiting 5σ point-source depth. Viewing the performance metrics presented in Figure 7.4 in the context of each survey's limiting depth, it is clear that sufficiently deep imaging is helpful for reliable merger classifications, but that there is a diminishing return in taking observations deeper than $\sim$ 25 mag. As a result, the CNN is already finding mergers at low-z in CFIS imaging at a success rate similar to what is expected for HSC-W and LSST.

Figure 7.8 has the same construction as Figure 7.7, but instead plots the surveys' performance metrics against the reported PSF in arcseconds. The PSFs for all of the surveys are dominated by contributions from Earth's atmosphere. There seems to be a trend between PSF and completeness, but I argue that this is mainly because survey depth generally trends with PSF (i.e., state of the art surveys are improved in both depth and resolution compared to legacy surveys; see Figure 7.10

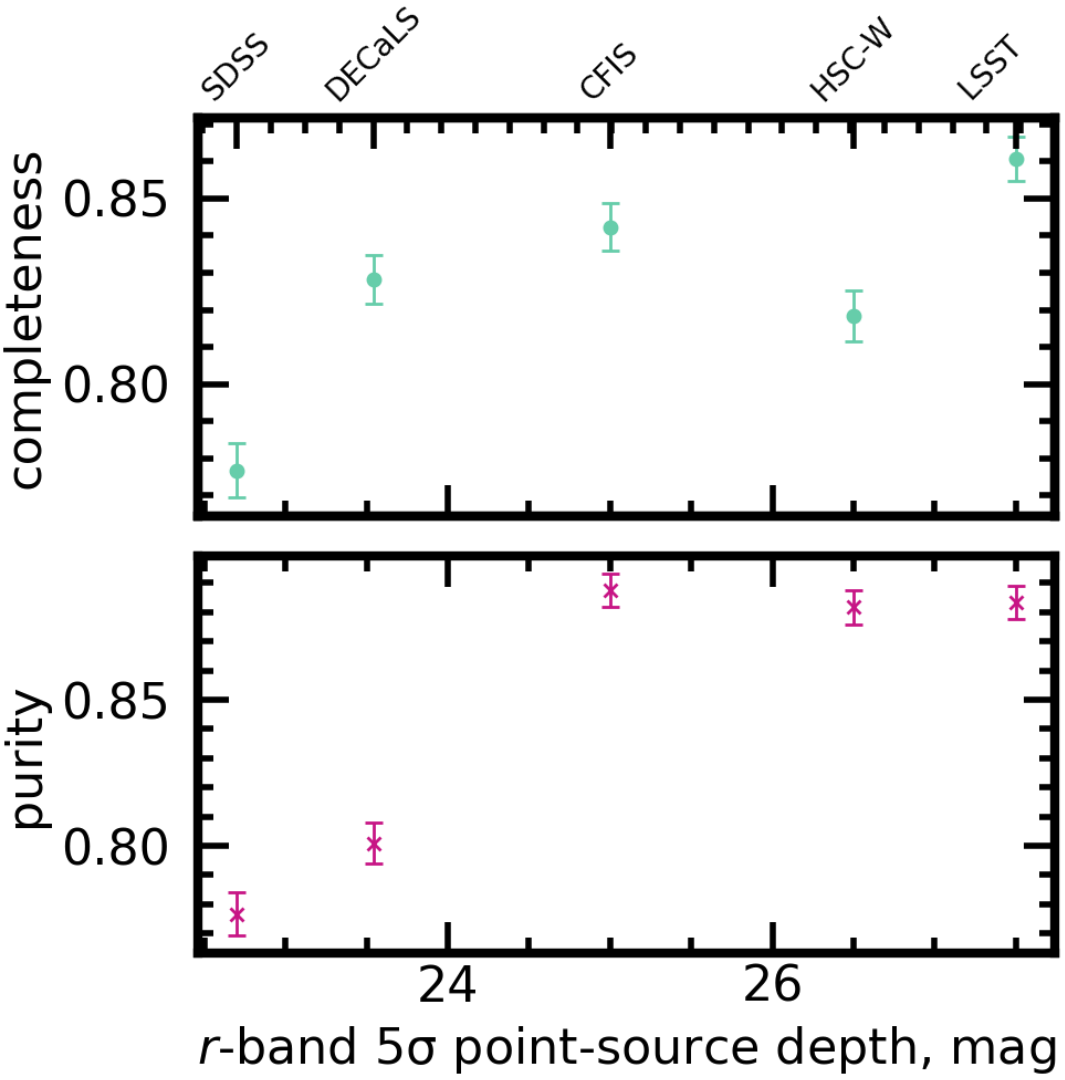

Figure 7.7: The post-merger completeness and purity scores for models trained with five different survey realism parameters as a function of the reported 5σ limiting point-source depth for each survey. Completeness on the post-merger class is shown in teal and purity of the predicted post-merger class is shown in magenta. Generally, deeper imaging is helpful to accuracy, but there is a diminishing return above the depth of CFIS.

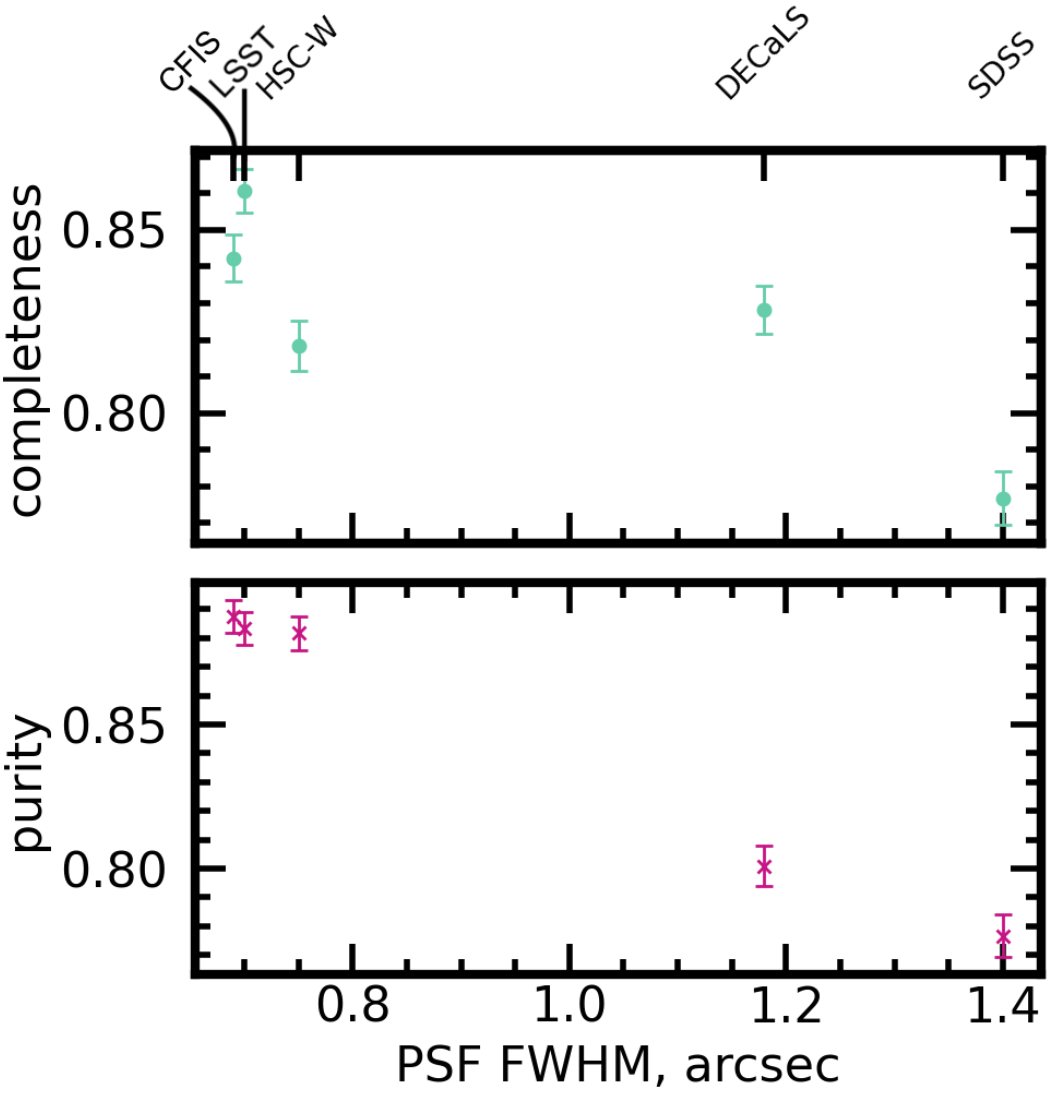

Figure 7.8: The same as Figure 7.7 but for the typical PSF for each survey, which constrains the effective angular resolution. The completeness scores for the LSST model, HSC-W model, and CFIS model are closely grouped together on these axes, suggesting that reliable merger classifications can be completed for the z range studied even when the PSF is atmosphere-dominated.

below). While it is difficult to estimate the specific contributions of depth and PSF to the final accuracy of each model without generating a more extensive grid of mock models, I refer to Wilkinson et al. (2024), who find that depth is the main determinant of merger identification accuracy when combining non-parametric morphological parameters in a random forest (RF) classifier.

The angular scale of the individual pixels of the camera CCD used to take images for each survey also bears on effective resolution and depth; Figure 7.9 plots the same performance metrics for the five models again but as a function of CCD pixel size. The trend with CCD scale is interesting for two reasons. First, the CCD scale determines the extent to which the maximum spatial resolution (set by the PSF) is preserved in the final image; in this way, it is a more direct determinant of spatial resolution than the PSF itself. Second, it plays a role in determining the effective S/N of the image, since signal and noise photons are binned on a pixel-wise basis. Similar to the results presented in Figure 7.8 for the PSF, completeness increases with decreasing pixel scale from SDSS to DECaLS and from DECaLS to LSST, but turns down again for smaller pixel scales. It is unlikely that CCD pixel scale is the main driver of the final accuracies for each model, but they reflect the two main trends suggested by the results up to this point: that increasing depth is generally good for accuracy, but that an excess of spatial detail can be a source of confusion.

I also plot the post-merger completeness scores of the models on a two-dimensional plane of depth and resolution in Figure 7.10. The completeness scores are represented by the blue-green colour scale for each marker on the plane. Viewed in two dimensions, the relative importances of the trends outlined in Figures 7.7 and 7.8 can be summarized. Since the SDSS- and DECaLS-trained models have the largest PSF and shallowest depth, it is difficult to determine the precise roles of depth and resolution in setting the final accuracy of each model. Still, the visual characteristics of the post-mergers in SDSS and DECaLS data (especially at high-z, see Figure 7.2) suggest that the surveys' comparatively shallow depth may set a practical upper limit on the visibility of tidal features.

7.2.2 Performance trends with galaxy parameters

In Section 7.2.1, I demonstrated that the efficacy of CNN models for merger and non-merger classification is sensitive to survey depth and resolution, but that the relationships are not necessarily linear or monotonic. Connections between survey parameters and accuracy can be used as a design reference for future merger identification tasks for the surveys presented here. Conversely, connections between CNN accuracy and galaxy parameters are not directly under the control of astronomers, but rather, will bear on the completeness and representativeness of samples that

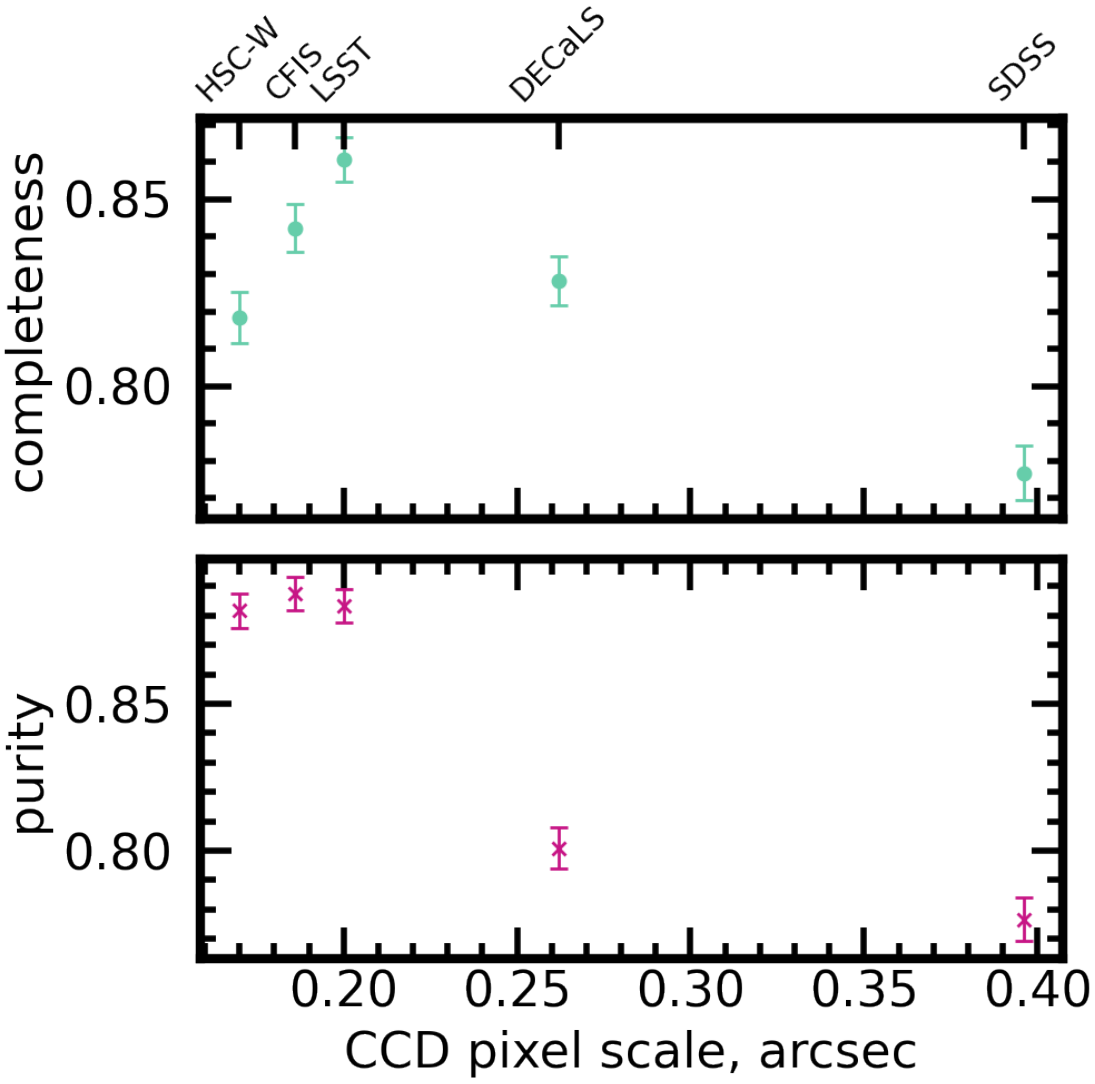

Figure 7.9: The same as Figure 7.7 but for the angular pixel scale of the charged couple device (CCD) for each survey, which bears on both effective angular resolution and the pixel-wise S/N. The completeness scores for the LSST model, HSC-W model, and CFIS model are grouped together on these axes as well.

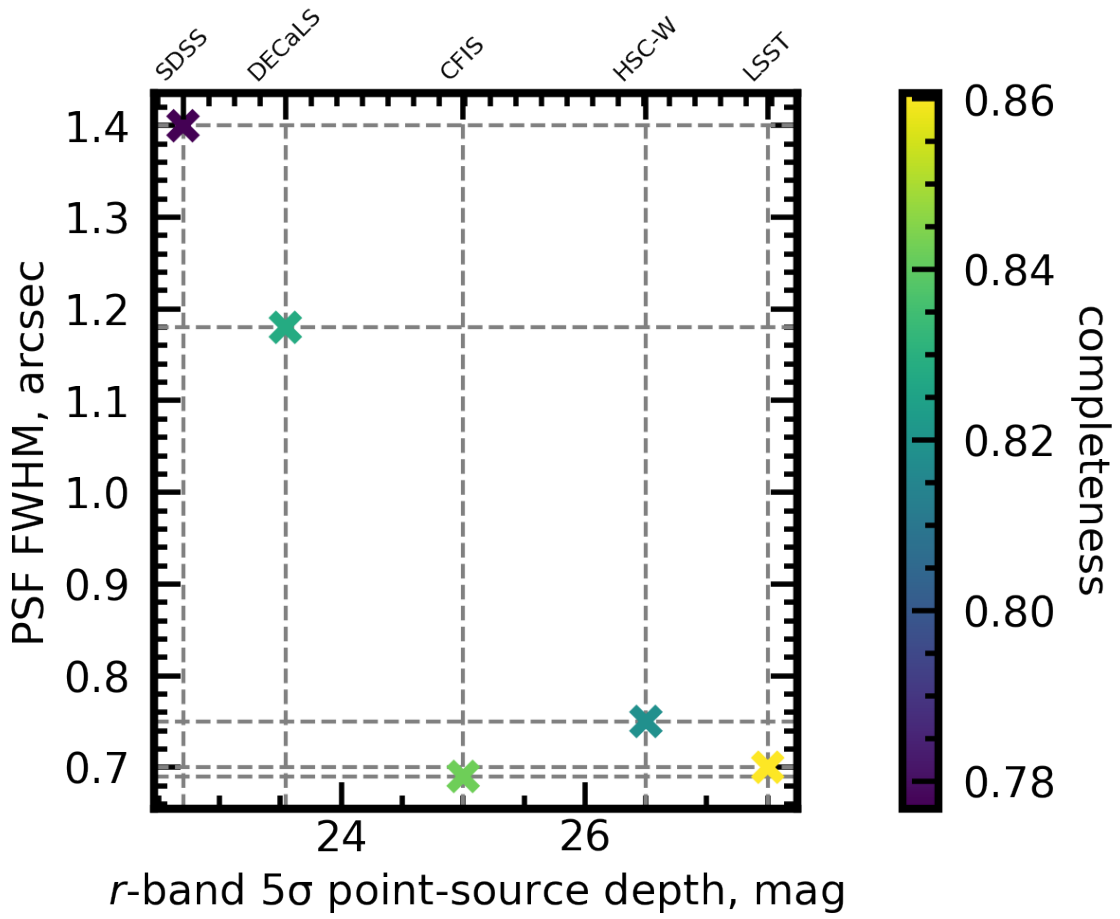

Figure 7.10: The post-merger completeness scores for each of the five models (colour scale) plotted in the depth (here approximated by limiting 5σ point-source depth) and resolution (PSF FWHM in arcseconds) plane. Performance is sensitive to both parameters, with an apparent minimum limiting depth of $\sim$ 24 – 25 mag being important to performance. Resolution also plays a role in setting the final accuracy of each model, but the shallowest two surveys (SDSS and DECaLS) also have the worst spatial resolution, making it somewhat difficult to disentangle the individual contributions of each parameter.

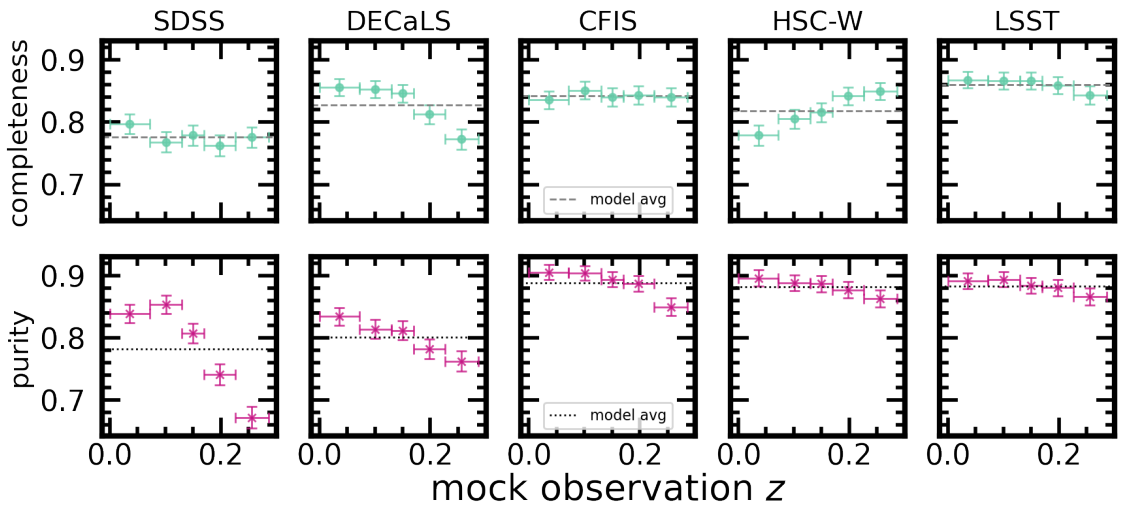

Figure 7.11: The completeness of the five CNNs for post-mergers (teal), purity of the predicted post-merger samples (magenta), and the average completeness (grey dashed line) and average purity (black dotted line) for each model binned as a function of z. Statistics are reported at each of the five discrete redshifts used for mock observations (given in Section 7.1.2). The error regions are the binomial errors on the statistics at each z. The purity of predicted post-merger samples decreases with z in all cases, but the effect is strongest in shallow imaging.

are identified in a given survey.

Low-surface brightness features fade with redshift as a result of cosmological dimming (as seen in Figure 7.2), so it is reasonable to expect that CNNs will struggle with calibration (learning to identify the range of surface brightnesses associated with tidal features) in shallow imaging and at high-z. The post-merger completeness and predicted post-merger sample purity for each of the five CNNs is shown as a function of mock observation z in Figure 7.11. The SDSS, CFIS, and LSST models retain approximately consistent completeness as a function of mock observation z. The DECaLS model identifies post-mergers more successfully at low-z, reflecting the limited visibility of merger-like morphology in the highest two redshift realizations in DECaLS imaging (see Figure 7.2). The HSC-W model has the opposite behaviour, possibly due to the resolution of the HSC CCD (see Figure 7.9, which reveals decreasing performance with finer CCD resolution between LSST, CFIS, and HSC-W).

Figure 7.11 also reveals that the purity of predicted post-merger samples decreases for five of the five CNNs with increasing z. This is to be expected, since at high-z, images in all five datasets contain larger random fluctuations due to sky noise. In some cases, I expect that these fluctuations and the muted appearance of some post-mergers' morphologies together give rise to more contaminated samples. The models' sensitivity to redshift is broadly the result of challenges

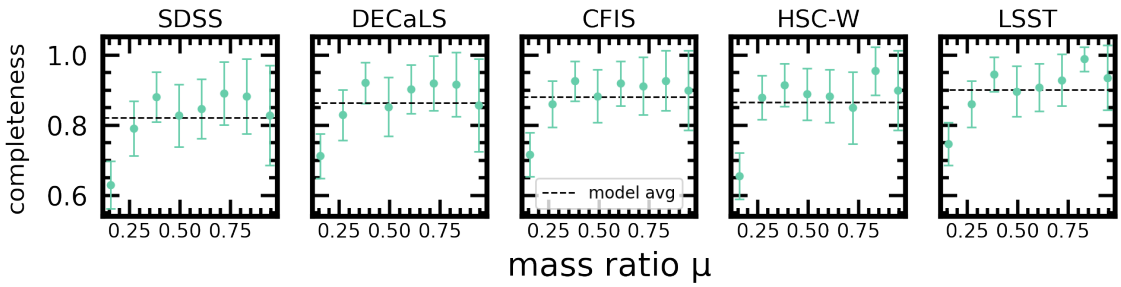

Figure 7.12: The same as Figure 7.11, but with completeness shown for galaxies arranged in eight bins of μ. I find very similar trends as a function of μ across all five surveys, even though the models have shown themselves to behave very differently in other tracts of parameter space. The similarity of the trend in all five panels (increasing and stabilizing completeness for higher mass ratios) is intuitive, and suggests that merger mass ratio is one of the primary factors affecting whether a given galaxy will be selected as a post-merger. Post-mergers with $\mu < 0.25$ are likely to be proportionally under-represented by some 20 per cent compared to the remnants of merger events with larger mass ratios. The data is only shown for post-mergers, since the mass ratios of long-past merger events for my control sample galaxies are not relevant.

inherent to the task of generalizing over a range of z.

The progenitor mass ratio μ is expected to be among the most important factors governing merger observability at a given location in $z - M_\star$ parameter space. Post-mergers whose progenitors were similar in mass are expected to be more dramatically disturbed compared to the remnants of minor mergers (e.g., with $\mu < 0.1$). Figure 7.12 emphasizes this point, and suggests that the influence of μ on merger observability is the most ubiquitous one presented in this work. All five CNNs experience very similar trends (albeit with their amplitudes governed by the depth and resolution of each survey) with significant loss of post-mergers with $\mu < 0.25$. Mergers near the lower limit of μ selection for the training set are therefore at risk of under-representation in the final merger sample predicted by the CNN, and the bias is likely worsened if visual classifications are used for quality control (e.g., as in Chapter 3) since more dramatic post-mergers are also more likely to be confirmed visually.

7.2.3 Cross-survey results

The relative performance of the five CNNs presented in Section 7.2.1 illustrate the potential of the "best-case scenario" in which models are trained and evaluated on data from the same image quality domain. But machine vision models (and deep learning models in general) cost time and substantial quantities of energy to train, both of which are valuable resources in science research (Strubell et al. 2019). The practice of transfer learning (i.e., training an already-trained model

on a new dataset for fine tuning) shows promise (Ackermann et al. 2018; Domínguez Sánchez et al. 2019), but transfer learning still mandates the creation of a new dataset and additional computing resources. If CNNs can be applied without re-training as a matter of course outside of their original domain without sacrificing scientific rigor, the efficiency of merger searches in the coming years could be improved substantially. Moreover, in Chapter 6, I already conducted a case study that indicated my CNN trained on IllustrisTNG galaxies with CFIS realism could be applied to shallower and lower-resolution imaging from DECaLS without incurring a prohibitive amount of loss or sample impurity. I will investigate the potential for merger searches using cross-survey classifications for the five trained CNNs described earlier in Section 7.1.3.

Figure 7.13 summarizes the results of the cross-survey inference experiment, in which all five CNNs are applied (without any re-training) to the test datasets for all five surveys. The classification results achieved by a single model are shown in rows, while the results for a given dataset are shown in columns. As in Figure 7.4, the true labels are shown on the horizontal axes, and the machine-predicted labels are shown on the horizontal axes of each of the confusion matrices. The exact confusion matrices from Figure 7.4 are shown along the diagonal as well, for reference.

The results of the cross-survey inference experiment can be broadly characterized as an illustration of the importance of calibration for automated merger searches. If a model learns to identify the features of a recent merger event in a certain range of brightness, it will (perhaps intuitively) misclassify control galaxies if non-mergers as imaged by a different survey appear to have diffuse or asymmetric features with similar brightness levels. Conversely, if mergers in a different survey appear to lack tidal features within the learned range of brightness, a model will misclassify those mergers as controls. The importance of calibration is verified in the class-wise accuracies of the confusion matrices above and below the diagonal in Figure 7.13. Above the diagonal, models are used to classify deeper imaging than their training regime. As a result, models typically mis-classify a large number of controls as post-mergers; for an extreme case, refer to the top-right confusion matrix, in which the SDSS-trained model misclassifies nearly half of the controls as post-mergers when applied to the LSST-like imaging. Below the diagonal, models are applied to shallower imaging than in training. As a result, a large number of post-mergers are mistakenly labeled by the models as controls; see for example the bottom-left confusion matrix, in which the LSST-trained CNN classifies the vast majority of galaxies as non-mergers regardless of their true label when applied to SDSS imaging.

Figure 7.13 also reveals some surprises in the cross-survey inference experiment that depart from the general trend just described. The HSC-W model and CFIS model perform very well when applied to deeper domains than their training data, perhaps thanks to the plateau in the trend

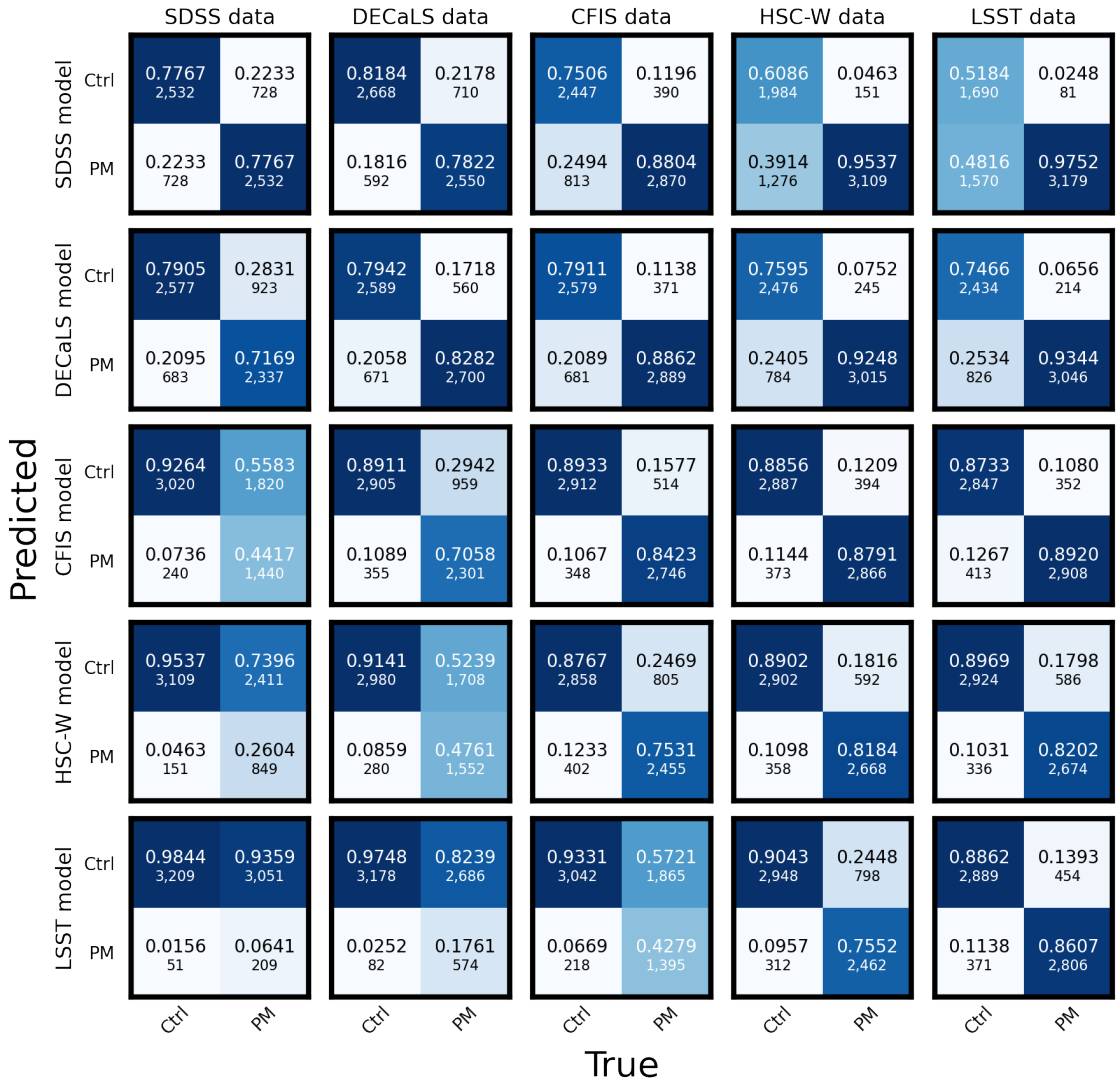

Figure 7.13: Confusion matrices for all five trained CNNs after being applied to all five test datasets. Rows of matrices show the classification results for a single model (e.g., for the SDSS-trained model in the first row) on each of the test datasets, while columns show the classification results for each of the five trained CNNs on a given test set (e.g., for the CFIS test set in the third column). The confusion matrices on the diagonal are the same as shown in Figure 7.4. Broadly, the matrices highlight the fact that calibration is essential for CNN-based merger searches. When models are applied to shallower data than their training set, they tend to under-predict mergers and over-predict controls. When models are applied to deeper data, they tend to misclassify a larger number of control galaxies as mergers.

of performance versus 5σ limiting point-source depth shown in Figure 7.7. Since the practical visibility of merger features (see again Figure 7.2) do not appear to change significantly at depths better than $\sim$ 25 mag, perhaps the appearance of merger features become standardized after images are resized and normalized via the method described in Section 2.1.3. In the context of a hybrid CNN and visual inspection merger identification framework, the CFIS and HSC-W models are therefore extremely useful, i.e., allowing for the recovery of a large proportion of post-mergers outside of their training regimes. In all cases, using visual inspection or other ensemble classification methods as a form of quality control *a posteriori* is advisable.

The particular combination of training and inference data used in Chapter 6 is shown in the third row, second column, where the CFIS-trained model is applied to DECaLS data. The model misclassifies an additional 14 per cent of post-mergers compared to its performance in CFIS data, a result consistent with the lower global visual agreement fractions for post-mergers reported in Sections 3.1.3 and 3.2.2 for the post-merger searches in CFIS and DECaLS, respectively. Still, the apparent purity of the post-merger sample from DECaLS and the results in Figure 7.13 indicate that merger searches can be conducted responsibly using cross-survey inference as long as additional quality control is enforced and any physical biases in the sample are accounted for (or at least discussed) in post.

Figure 7.14 directly addresses one of the central questions of the cross-survey inference experiment: how much post-merger completeness is lost when a CNN is applied outside of its training domain? In each cell of Figure 7.14, I subtract the model's completeness score on a given dataset from its baseline completeness, i.e., its completeness score when applied to the test set from its training domain. Values along the diagonal are therefore zero by definition. Figure 7.14 illustrates that completeness generally increases when models are applied to deeper imaging than their training domain (note the trend of positive Δ completeness above the diagonal, shown in purple) and decreases when they are applied to shallower imaging than their training domain (negative Δ completeness below the diagonal, ochre).

Figure 7.15 has the same construction as Figure 7.14, but shows the change in purity of the predicted post-merger sample when each model is applied outside of its training domain. The baseline purities for the same-survey experiment are given above in Table 7.2. Figure 7.15 demonstrates that Δ purity is usually smaller than Δ completeness, though I note that this may be somewhat misleading: since this study uses test datasets for each survey that include equal numbers of post-mergers and non-post-mergers, the purity scores presented throughout this work are artificially high compared to what could be achieved with a single CNN in observations.

Figure 7.15 indicates that purity is generally lost when models are applied outside of their training regime, but there are noteworthy exceptions to the rule. The SDSS-trained model im-

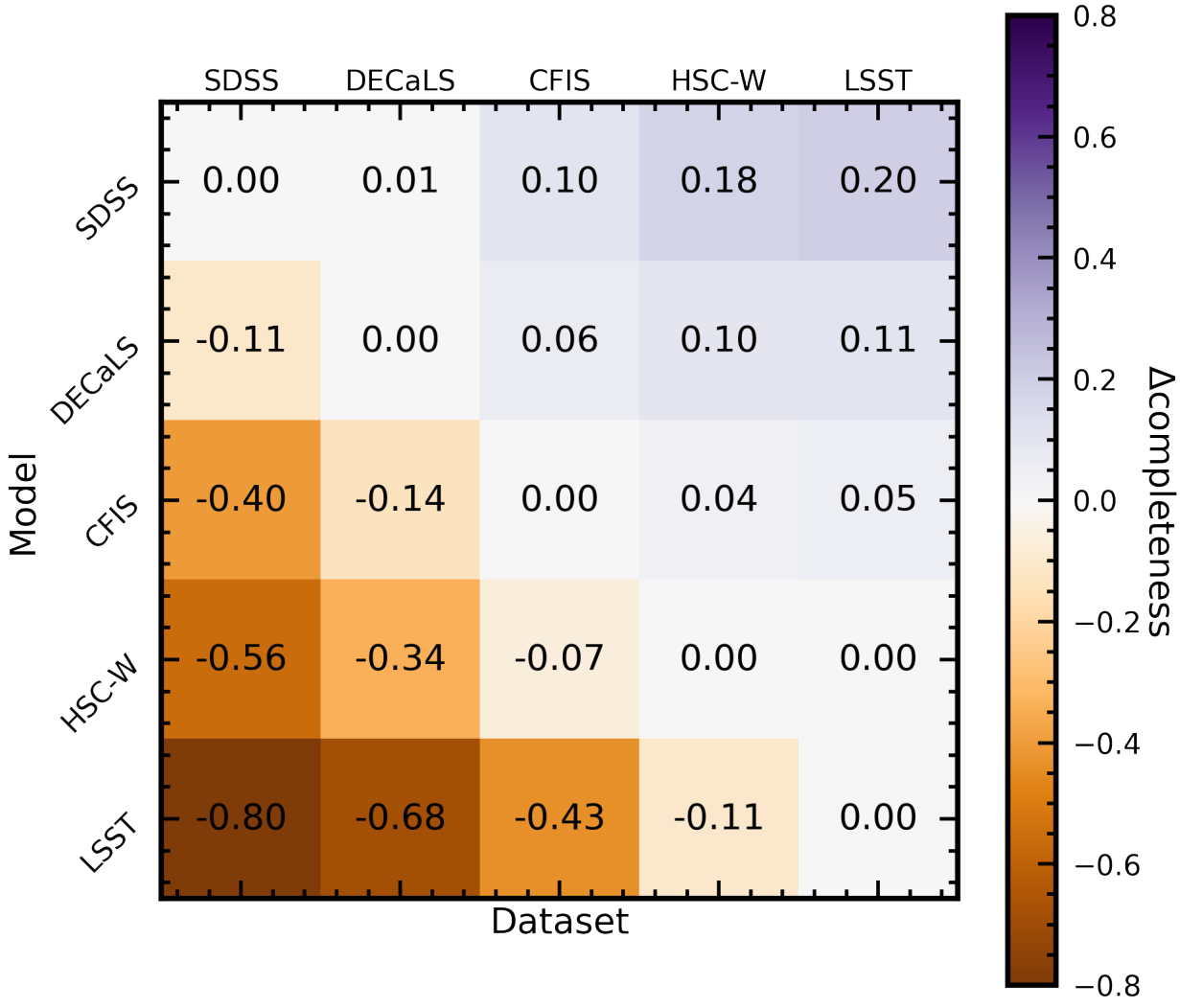

Figure 7.14: The same configuration as Figure 7.13, but reporting the change in post-merger completeness for each model when it is applied outside of its training domain compared to its same-survey completeness score. Positive Δ completeness (shown in purple) indicate that a model identifies a greater proportion of the true post-mergers in the test set, while negative Δ completeness (shown in ochre) indicate that the model identifies fewer post-mergers compared to its baseline. Values along the diagonal are zero by definition.

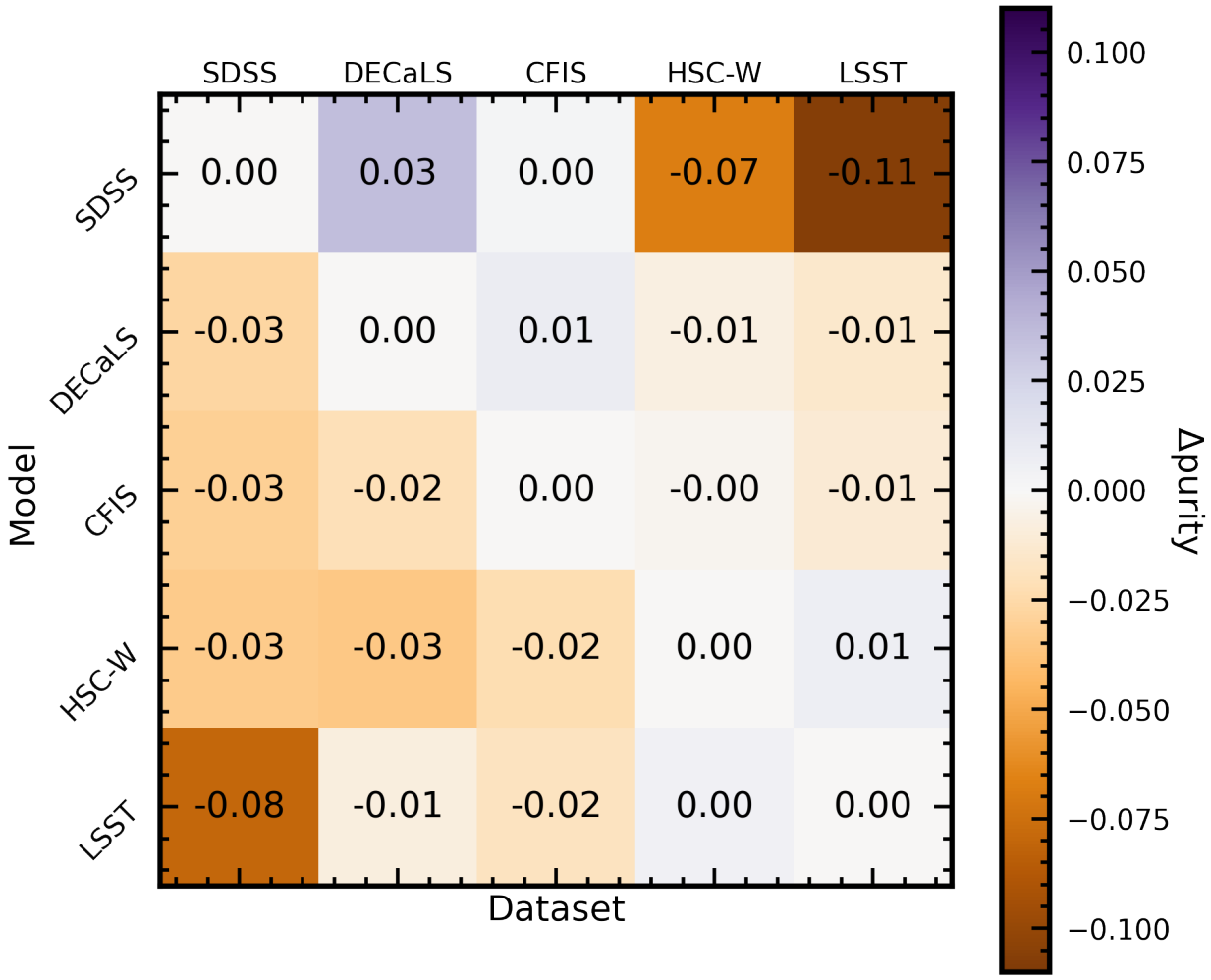

Figure 7.15: The same as Figure 7.14, but reporting the change in the purity of the predicted post-merger sample for each model when it is applied outside of its training domain compared to its same-survey purity score. Values along the diagonal are zero by definition.

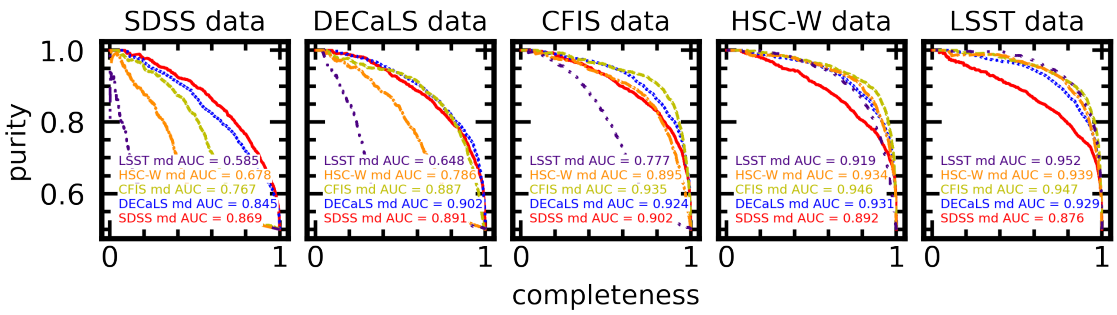

Figure 7.16: Purity-completeness (or precision-recall) curves for the cross-survey inference experiment. The series of colour-coded cuves in each panel show the purity-completeness curves for each of the five models applied to one of the datasets. The annotations show the AUC scores for each curve, illustrating each model's potential to identify samples of post-mergers that are pure and complete when applied outside of their training regimes.

proves in both purity and completeness (see Figure 7.14) when it is applied to DECaLS imaging. This is somewhat intuitive, since mergers in the DECaLS dataset at higher-z are more likely to retain their merger-like appearance. In general, the results of the cross-survey inference experiment confirm that SDSS imaging is too shallow to conduct a holistic merger search spanning the survey's redshift domain.

Figure 7.16 shows the results of the cross-survey inference experiment in another way, folding in the utility of a cut in CNN $p(x)$ for each model. The purity-completeness curves are shown only for the post-merger classes, as the primary class of interest in the context of a merger search. Mirroring the results in Figure 7.13, the curves for models applied to shallower data are typically lacking in completeness for a given $p(x)$, even though the samples they identify may be very pure. Increasing performance by each of the models in deeper data again suggests the conclusion that there is an effective standardization of the appearance of merger features in deep imaging, allowing for more successful cross-survey inference. Still, the purity completeness curves indicate that the models trained specifically to identify post-mergers in a given image survey are optimal in most cases.

Owing to the rarity of post-mergers in the low-z Universe, CNNs have been used throughout this work as a filtration tool (removing the majority of non-mergers from the predicted merger pool), rather than an implicitly trustworthy merger identification tool. One of the essential characteristics of a useful CNN in this context is the purity of the post-merger sample it is able to identify; in other words, how much time and effort is saved with the help of a CNN? In Figures 7.14 and 7.15, I exchange the confusion matrices from Figure 7.13 for individual figures of merit, the changes in completeness and purity of the out-falling post-merger sample. Recall that

purities in the context of a balanced machine learning experiment will be artificially inflated, since post-mergers in the Universe are outnumbered $\sim$ 100:1 by non-post-mergers. The purities can be adjusted using the form of Bayes rule presented earlier in Section 1.3.3. The purity of merger samples identified by the CFIS, HSC-W, and LSST-trained models in all data qualities are noticeably high, for two reasons. First, these models tend to identify a small number of visually dramatic post-mergers in shallower imaging, yielding highly pure but incomplete (small) samples of mergers. Second, the standardization of the appearance of tidal features in deep imaging (with 5σ limiting point source depths above $\sim$ 25 mag) already discussed in the context of Figure 7.13 allows these four models to classify data from deeper surveys with relative ease. The depth of the imaging used to train CNNs is therefore of principle importance to the final purity of the predicted merger samples.

7.3 Discussion and conclusions

In this Chapter, I have detailed the limits placed on low-z merger searches by image quality, using five real surveys as benchmarks, and holding all other methodological variables constant. I find the relationship between image quality and the efficacy of CNNs to be complex. Intuitively, images must be deep enough to detect faint tidal features across the entire redshift range studied in order to be successful. Perhaps unexpectedly, greater depth is not categorically beneficial: in detecting recent (within some 150 Myr) and major (with $\mu > 0.1$) mergers, I find that the depth of CFIS imaging is adequate. Additional imaging depth is met with a diminishing return in accuracy. Additional detail detected in extremely deep (e.g., LSST after 10 years of co-adds) imaging may be beneficial for more advanced merger characterization tasks that are only lately being explored (e.g., temporal estimation of merger timescales or identification of merger remnants with smaller progenitor μ values). Imaging with the depth and angular resolution of surveys like CFIS or HSC-W are already near the apparent limit of what is possible for binary classification for major merger remnants.

The results of this work are particularly useful when considered alongside the results of Wilkinson et al. (2024), who use non-parametric morphological statistics to perform a systematic analysis of merger recovery as a function of depth, resolution, viewing angle, and a variety of galaxy properties. Wilkinson et al. (2024) used different observational realism and merger identification techniques than this work, but the trends of performance with depth and resolution still offer a useful point of comparison. Figure 7.17, reproduced from Figure 10 of Wilkinson et al. (2024) (with permission of the author), shows the result of an experiment very much like the one used here. Wilkinson et al. incorporate multiple non-parametric morphological statistics (asym-

metry, outer asymmetry, shape asymmetry, Gini, and M20) in a random forest classifier, which is used to predict merger status. The approach of Wilkinson et al. (2024) is quite similar to the one used in this work, in that CNNs also combine a feature extraction component (although the features in the CNN are learned in the convolution layers, rather than prescribed) with a machine learning classification tool (the fully connected layer of the CNN).

Figure 7.17 shows two main trends that are echoed in this Chapter. First, greater depth at a given resolution is generally beneficial, but that the improvement in accuracy with increasing depth above 25 mag arcsec^{-2} is marginal compared to the improvement between 23 – 25 mag arcsec^{-2}. Wilkinson et al. (2024) also report a turnover in accuracy as a function of PSF FWHM, with the peak accuracy at 0.75 arcsec. The other results on the depth-resolution plane explored in Wilkinson et al. (2024) offer helpful context by filling in the bigger picture for this Chapter's results, since the five surveys studied here represent discrete selections from a multi-variate grid of depth, resolution, and CCD scale.

While CCD scale is held constant in Wilkinson et al. (2024), the parameter may bear significantly on the results presented here, since the CCD scale sets the value of S/N in each pixel for a given observation by binning both the signal and the noise. In practice, a larger choice of CCD scale could enable the detection of a faint tidal feature at the expense of spatial detail, while a smaller CCD scale could make the opposite exchange.

The figures of merit for the CNN trained to identify tidal features in HSC images in Domínguez Sánchez et al. (2023) (overall accuracy of 0.84, merger purity of 0.72, and merger completeness of 0.85) are in reasonable agreement with the results for HSC in this work (overall accuracy of 0.86, merger purity of 0.82, merger completeness of 0.82), though the HSC-W model presented here is able to improve on the purity of the predicted merger class via a statistical "sacrifice" of 3 per cent completeness on the merger class and an overall accuracy that is higher by 2 per cent. Domínguez Sánchez et al. (2023) also report significant loss when their model is applied to real galaxies imaged in HSC. While this work offers no direct comparison regarding the accuracy of simulation-trained models to HSC imaging, visual classification exercises have revealed (e.g., in Chapter 3) that the CFIS-trained CNN used to identify the post-merger samples studied in Chapters 4-6 performs well (qualitatively) in the observational domain. Domínguez Sánchez et al. (2023) suggest that their loss of performance when classifying real galaxies may be the result of the fact that contaminating sources and image artefacts (e.g., foreground stars, zero-flux artefacts) are not included in the training set. It is therefore likely that the proof-of-concept models trained for this Chapter would not perform well in the observational domain either – since this Chapter is mainly intended as a methodological comparison, no artefacts were included in the training data used here.

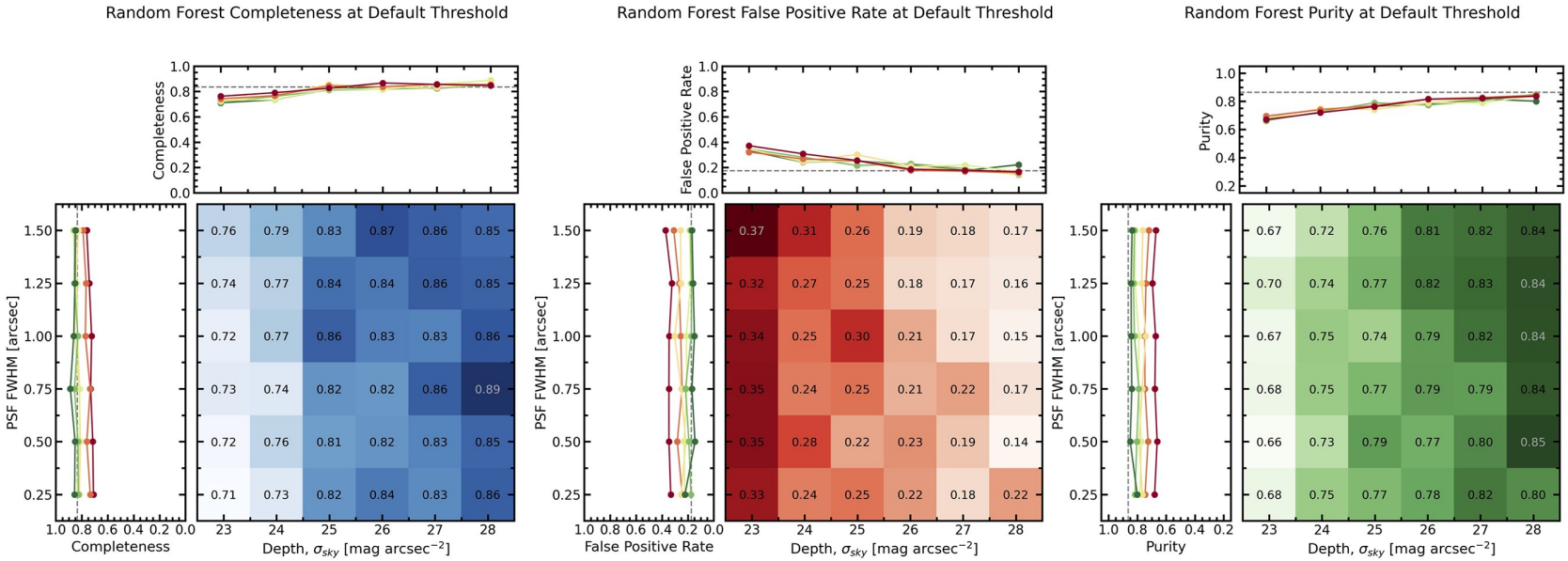

Figure 7.17: Reproduction of Figure 10 from Wilkinson et al. (2024) (with permission of the author). The completeness, false positive rate, and purity of the merger samples identified using a random forest classifier trained on non-parametric morphological statistics to distinguish between galaxy mergers and non-mergers. The colour gradient in the bottom right panel shows the trend of each of the statistics (completeness, false positive rate, and merger sample purity) as a function of imaging depth and PSF size. Cells are annotated with the specific value of each figure of merit for a given combination. The colour-coded lines in the top and side panels of each sub-figure show how performance scales as a function of each individual observational parameter (i.e., either depth or PSF).

I note that the results above lack an assessment of Euclid, the forthcoming space-based imaging survey. Euclid will also have ample depth (a limiting 5σ point source depth of 26.2 mag in the visible band) and an excess of spatial resolution compared to the surveys included in the main results of this Chapter (PSF of 0.18 arcsec, CCD scale of 0.1 arcsec; see Euclid Collaboration et al. 2022). The results for Euclid are not shown since the softening length of IllustrisTNG100-1 (at $\sim$ 1kpc) dominates over the PSF of Euclid across the entire redshift range except for the realization at z=0.256. Including results from the Euclid experiment would therefore be philosophically distinct from the rest of my results (for which the atmospheric PSF dominates the effective resolution), and therefore could be misleading. Euclid could be evaluated in the future, however, using a higher redshift analogue to the study conducted here, or using a higher-resolution simulation suite for training.

The results presented in Martin et al. (2022) indicate that 10-year depth LSST imaging taken at the Vera Rubin observatory should capture some 60 – 80 per cent of the flux from tidal features in Milky Way-mass (or greater mass) galaxies at $z \sim 0.05$, depending on the assumed final depth of the survey. To the extent that simulated galaxies processed with LSST observational realism have similar appearances to real galaxies in LSST, one would expect LSST to be an excellent opportunity for merger searches using a simulation-based approach. While LSST may be able to capture a groundbreaking amount of information about galaxies' recent assembly histories in the form of low-surface brightness detail, the results presented in this Chapter indicate that 10-year LSST-quality imaging is not required for the identification of a pure and complete sample of major mergers. I find that the change in the CNN's performance between CFIS and LSST is marginal, within $\sim$ 2 per cent for both mergers and non-mergers. I argue that merger searches at low-z with LSST will not require 10-year depth to be effective, since shallower imaging from earlier data releases will already likely be of adequate quality.

In light of the fact that ground-based surveys like CFIS and HSC are already available, I posit that images from 10-year LSST co-adds can be used to investigate more difficult and granular questions in galaxy evolution astronomy; for example, probing the impact of mergers with smaller mass ratios, at high redshift, and / or with greater temporal specificity. Space for improvement remains for major mergers at low-z (i.e., the domain studied in this work), but the potential of next-generation hydrodynamical simulations and imaging surveys for temporal, high-z, and small μ merger characterization is much greater.

Chapter 8

Conclusions

The aims of the research presented in the preceding chapters fall into two main categories. The first is to conceptually demonstrate, carry out, and validate a simulation-based deep learning approach to post-merger galaxy identification in extragalactic surveys. The second objective capitalizes on the foundation built in pursuit of the first by using the identified post-merger samples to characterize statistically the evolutionary impact of coalescence after a major merger. The main scientific conclusions of this work are therefore divided into two channels as well, with several methodological findings related to the application of deep learning methods to morphological characterization of galaxies, and a number of astronomical results addressing and/or constraining the consequences of mergers.

In Chapter 2, I detailed the development, training, and validation of a convolutional neural network (CNN) designed for efficient post-merger classifications in *r*-band imaging from the Canada-France Imaging Survey (CFIS). The CNN was trained on matched samples of post-coalescence major mergers (with mass ratios of $0.1 < \mu < 10$) and non-post-merger controls from the IllustrisTNG simulation, processed with an observational realism suite RealSim to give them the appearance of having been observed by CFIS. The trained CNN was subsequently evaluated on test data that had been produced in the same way, and its performance characterized. I found that the CNN performed well on test data, with accuracies of $\sim$ 87 and $\sim$ 89 per cent on post-mergers and controls, respectively (Figure 2.6), and a receiver operating characteristic (ROC) area under the curve (AUC) score of 0.95 (Figure 2.7). The success of the CNN on data similar to its training set indicates that the merger status of galaxies is recoverable by the CNN from monochromatic imaging when the training set is well designed and representative of the data to which the CNN will ultimately be applied for inference.

I found the CNN's performance to be highly stable in the parameter space occupied by the diverse galaxies in IllustrisTNG, with little variation in performance as a function of observed redshift, stellar mass, merger mass ratio, and gas fraction (Figures 2.8-2.13). My tests revealed that

the model misclassified a number of control (non-post-merger) galaxies that had close interacting companions, suggesting that the morphological signatures of ongoing and completed mergers may be degenerate. In Chapter 2, I argue that the degeneracy between interacting and post-merger galaxies is not especially concerning, since galaxy pairs (like post-mergers) are relatively rare, and can be identified reliably by eye. Ultimately, interacting galaxy pairs do frequently contaminate CNN-predicted post-merger samples in this work, but they are generally visually distinct.

I applied the trained CNN in a proof-of-concept "mock survey" of galaxies that contained post-mergers and non-post-mergers in representative quantities (i.e., with non-post-mergers outnumbering post-mergers more than 100:1). The model still correctly classified the majority of galaxies belonging to each class ($\sim$ 89 per cent accuracy for both non-post-mergers and post-mergers; see Figure 2.15), but the mock survey highlights the difficulty of post-merger identification in a cosmological context. Owing to the intrinsic rarity of post-mergers, even a very accurate classifier will struggle to yield a predicted post-merger sample that is pure, since the small fraction of misclassified non-post-mergers will outnumber the correctly classified post-mergers (Figure 1.6). As such, I argue in Chapter 2 that the certainty of the CNN's predictions on galaxies can be leveraged to enhance purity, and that input from an additional classifier is required. In this work, I advocate for the use of visual classifications for quality control, but more recent work by Ferreira et al. (2024) has shown that the benefit of multiple classifiers is actually the result of differing experience – whether the experience is automated or human is less important.

Projecting the performance of the CNN onto a hypothetical experiment in the observational domain, I conduct a study of merger-induced star formation enhancement using the CNN's predictions on simulated galaxies in IllustrisTNG to select the merger sample. I show that the the certainty of the CNN's predictions on a given galaxy, $p(x)$, can be used to enforce higher purity in the machine-predicted post-merger sample at the cost of completeness. Physically representative measurements of post-mergers can be taken in this way (Figure 2.19), but the small samples resulting from using an aggressive cut in CNN $p(x)$ worsens the statistical strength of the result, highlighting again the need for additional classifications from another source. Finally, I compare the CNN's performance against a number of other merger identification methods including both non-parametric morphological statistics and visual classifications. I find that visual classifications can be used to identify highly pure, but generally incomplete post-merger samples (Figure 2.22).

In Chapter 3, I apply the "hybrid approach" – filtering using the CNN before conducting manual quality control on the predicted merger sample – to two surveys. The CNN was originally designed to identify a merger sample in CFIS, and I detail the successful recovery of a highly pure post-merger sample of statistically novel size (Figure 3.5) containing 699 galaxies in the second

data release of CFIS (DR2). While the post-mergers span the full range of the input sample in mass, brightness, and redshift, I find that they are somewhat biased towards higher masses (and as a result, higher redshifts; see Figure 3.4) since the tidal features of more massive post-mergers are typically brighter on the sky relative to their less-massive peers at the same z.

I also conduct a study on the viability of applying the CNN to classify galaxy images from a survey other than CFIS, the Dark Energy Camera Legacy Survey (DECaLS). Deep learning models are ideally trained on and applied to data of similar composition, but the model performs surprisingly well on DECaLS data even though it is somewhat shallower and lower-resolution (Figure 3.10). I use a post-merger identification framework similar to that used in CFIS to identify a highly pure sample of 923 post-mergers in DECaLS. The DECaLS post-mergers are somewhat more visually dramatic than the CFIS post-mergers on account of the survey's shallower depth, but their merger statuses are nevertheless unambiguously post-coalescence.

In Chapter 4, I present the first scientific results on the influence of coalescence on galaxy evolution. Mirroring the proof-of-concept study of star formation presented in Chapter 2, I select galaxies from CFIS DR2 and the seventh data release of the Sloan Digital Sky Survey (SDSS DR7) whose optical emission lines are unambiguously dominated by signs of star formation, with BPT diagram positions beneath the empirical Kauffmann et al. (2003) criterion for purely star-forming galaxies. After selecting the science sample, I present a comparison of the star formation rates in the new post-merger sample, a previously-identified and smaller sample of SDSS post-mergers from Ellison et al. (2013), and a sample of interacting (and presumably pre-coalescence) galaxy pairs from the Patton et al. (2016) catalogue. I find that star-forming post-mergers exhibit star formation rates (SFRs) in excess of controls by 0.27 dex or a factor of $\sim$ 1.9, in good agreement with the result I find for the Ellison et al. (2013) sample. I also find that post-mergers exhibit greater star formation enhancements than their pre-coalescence peers (Figure 4.4). As with the simulated study of star formation enhancement, I demonstrate that purity can be enforced in the CNN-predicted post-merger sample by increasing the $p(x)$ threshold at the cost of completeness (Figure 4.2).

In Chapter 5, I begin my investigation of the supermassive black hole (SMBH) and active galactic nucleus (AGN) demographics of post-mergers, again using the CFIS-identified post-merger sample. I investigate whether four different AGN classes – optical narrow-line AGN (NLAGN), dusty mid-infrared AGN (mid-IR AGN), and low-excitation radio galaxies (LERGs) – preferentially appear in post-mergers (as well as pre-coalescence galaxy pairs, for comparison). I find that galaxy pairs with projected separations between 0 – 80 kpc are more likely to host mid-IR AGN and optical NLAGN with decreasing separation, consistent with the idea that successive close passages with an interacting companion are likely to funnel gas towards the centres of merging

galaxies, resulting in increased SMBH activity. The closest galaxy pairs with projected separations $r_p < 10$ kpc are found to host $\sim 2-4$ times as many NLAGN and mid-IR AGN *per capita* compared to non-interacting controls. CFIS post-mergers are found to host a similarly elevated proportion of mid-IR AGN to the closest galaxy pairs, with an excess of ~ 3.6, and even more NLAGN than the closest pairs, with an excess of ~ 3.7 (see Figure 5.2 for the details of AGN excesses in the CFIS study). Neither interacting galaxy pairs nor post-mergers are found to have an excess of LERGs in this work.

I also find that more visually disturbed post-mergers are significantly more likely to host mid-IR AGN compared to their less-disturbed peers (Figure 5.3), suggesting that the severity of the elapsed merger event bears not only on the likelihood of SMBH activity being detected but also the multi-wavelength observability of the AGN. The connection between merger stage and multi-wavelength AGN observability is investigated further in Chapter 6. I also find that NLAGN in post-mergers have elevated [OIII]λ5007 emission line luminosities compared to galaxy pairs or non-interacting galaxies, suggesting that the accretion rates in post-merger NLAGN may be uniquely high compared to other stages of the merger sequence (Figure 5.4).

In Chapter 6, which is the first to use the DECaLS-identified post-merger sample, I extend my multi-wavelength characterization into the X-ray via the first data release from the extended ROentgen Survey with an Imaging Telescope Array (eROSITA). The spectral response of eROSITA makes it sensitive to obscuration effects, so I supplement the X-ray analysis with additional multi-wavelength AGN characterization. As in Chapter 5, I identify mid-IR and NLAGN samples in DECaLS, and also adopt a catalogue of broad-line AGN (BLAGN), since I found that the spectral signatures of the broad-line region (BLR) are very often co-observed in X-ray AGN in eROSITA (Figure 6.9). I find that X-ray AGN are ~ 1.8 times as abundant in galaxy pairs with $20 < r_p < 40$ kpc compared to controls, no statistically significant excess of X-ray AGN in galaxy pairs with $r_p < 20$ kpc, and ~ 1.8 times as many X-ray AGN in post-mergers compared to controls. I posit that the AGN in galaxy pairs with $r_p < 20$ kpc have not actually "switched off", but that heavy obscuration in the final stages of mergers are frequently obscuring X-ray AGN in close galaxy pairs below the point of detection in eROSITA.

Using a multi-wavelength analysis, I demonstrate that confirmed AGN in close pairs are more often absent from the eROSITA X-ray sample (Figure 6.14) and that they are more frequently attenuated (Figure 6.16). Through the same analysis, I present new evidence that the event of coalescence results in elevated rates of optically unobscured BLAGN (Figure 6.12) with frequently unattenuated X-ray fluxes (Figure 6.16). Generally, the multi-wavelength results suggest that post-mergers are an intrinsically diverse class, with some galaxies in a state of obscuration after coalescence, and others achieving high accretion rates, thereby removing circum-nuclear obscur-

ing material via AGN feedback.

Finally, in Chapter 7, I assess the utility of CNNs for post-merger searches in past, current, and next-generation imaging surveys. I project the success of my merger identification framework for six different surveys: SDSS, DECaLS, CFIS, the Hyper Suprime-Cam Subaru Strategic Program Wide field (HSC-W, taken at the Subaru telescope), and the Legacy Survey of Space and Time (LSST, a next generation ground-based survey that will be taken at the Vera Rubin Observatory). I find that some post-mergers can be recovered in SDSS and DECaLS, but that shallow imaging (especially for SDSS at high-z) prevents models from generalizing successfully (Figure 7.2). The difficulty in generalizing over z in shallow imaging gives rise to relatively low accuracies of $\sim$ 78 per cent (for both the post-merger and control classes) for the model trained on SDSS imaging (Figure 7.4).

Completeness generally improves with increasing depth and image quality, with successively higher scores earned by the models trained to identify mergers in DECaLS (79 per cent for controls, 83 per cent for post-mergers) and CFIS (89 per cent for controls, 84 percent for post-mergers). For surveys deeper than CFIS (HSC-W and LSST), I find that performance does not increase substantially relative to what I find in CFIS, suggesting that current-generation imaging data may be adequate for the identification of major merger remnants.

I find that cross-survey inference (i.e., training a model on imaging data from one survey and applying it to a different one) can be done safely in cases where the depth and resolution of the surveys are similar. Generally, CNNs applied to imaging deeper than their training domain misclassify a larger number of controls as post-mergers, resulting in highly contaminated post-merger samples. Conversely, CNNs applied to shallower imaging than their training domain (as was done to identify the DECaLS post-merger sample in Chapter 3) tend to identify reasonably pure but incomplete post-merger samples (Figures 7.13, 7.14, 7.15, and 7.16).

Because the CNN-identified, visually confirmed CFIS and DECaLS merger samples are biased only by the training of the CNN and the decisions of the visual classification team, I believe the samples will continue to provide value in the form of subsequent cross-survey characterization in the near future. Moreover, the hybrid classification framework itself (and more broadly, ensemble classification) shows promise for future questions in astronomy surrounding rare and elusive observational phenomena.

The results of Chapter 7 offer an opportunity to discuss how new datasets and methodological improvements could improve future studies of galaxy mergers. The surprising strength of the CFIS-trained CNN compared to the models trained on nominally better imaging indicates that significant potential for morphological characterization of merging galaxies may still lie in CFIS data. CFIS imaging is already being leveraged for granular merger classifications using an

expanded version of the mock CFIS survey introduced in this work and a multi-model classification framework by Ferreira et al. (2024). In a forthcoming follow-up (Ferreira et al. in prep), the same training set is used to perform detailed temporal classifications on post-mergers, rather than grouping them into a single class (as is done throughout this work). Temporal classifications partially avoid the philosophical difficulty of choosing semi-arbitrary boundaries in time for mergers and non-mergers, and allow for a more granular characterization of how mergers evolve after coalescence.

Other boundaries (e.g., in merger mass ratio, post-merger remnant mass) and morphological degeneracies (e.g., between close galaxy pairs and post-mergers) may also be dissolved in the near future, since simulations and astronomical imaging are growing in scope and sophistication. Deep learning machine vision models themselves are continuing to improve (e.g., the vision transformers used in the Ferreira et al. 2024 ensemble) and still newer architectures combining the benefits of CNNs and vision transformers into a single architecture are growing in popularity (e.g., Wu et al. 2021).

The non-monotonic trends of performance with resolution (either PSF or pixel scale, see Figures 7.8 and 7.9) indicate that the CNN architecture is not always able to use additional spatial information to its advantage. With larger training datasets and more sophisticated deep learning methods, models will likely be able to use higher-resolution imaging to identify the signatures of more minor mergers (e.g., "mini mergers" as explored in Bottrell et al. 2024) with greater temporal detail (e.g., Ferreira et al. in prep; Pearson et al. 2024).

Throughout this work, I have avoided estimating a global post-merger fraction (or a specific post-merger fraction, e.g., for a particular z range or AGN category), since my post-merger identification framework sacrifices completeness in the interest of identifying a pure and representative post-merger sample. However, as accuracy improves, it will become possible to identify merger samples that are both pure and complete. Complete merger samples will allow astronomers to constrain the precise contribution of mergers to the galaxy evolution budget (e.g., star formation, SMBH growth, or quenching). New morphological classifications at high-redshift (e.g., with the James Webb Space Telescope in Ferreira et al. 2022b) will also likely be improved in the near future by deep learning machine vision methods, allowing astronomers to place new constraints on the merger-induced evolution of the primordial ancestors of galaxies like our own Milky Way.

Afterword

Any seeming disconnection between astronomy research and terrestrial life is strictly perceptual, and responsibility must be taken for the human and environmental costs incurred. The prices in labour, energy, and water for computing, air travel, and coffee (as a start) are essential context for all of the research described above.

Technical innovations from science are frequently appropriated and brokered by those who would use them for purposes of war, policing, genocide, and myopic extraction of natural resources, and deep learning technology is a noted non-exception. In our time, it is essential that such misuse be anticipated and prevented as part of the daily practice of research.

Time and time again, new discoveries in the natural sciences have highlighted the oddity and rarity of our brief existence – let's cherish it, innovate with sincerity, and empower more people to look forward with hope.

Bibliography

Ackermann S., Schawinski K., Zhang C., Weigel A. K., Turp M. D., 2018, MNRAS, 479, 415

Agostino C. J., Salim S., 2019, ApJ, 876, 12

Aihara H., et al. 2022, Publications of the Astronomical Society of Japan, 74, 247

Allen S. W., Dunn R. J. H., Fabian A. C., Taylor G. B., Reynolds C. S., 2006, MNRAS, 372, 21

Antonucci R., 1993, ARA&A, 31, 473

Arpit D., et al. 2017, arXiv e-prints, p. arXiv:1706.05394

Assef R. J., et al. 2013, ApJ, 772, 26

Avirett-Mackenzie M. S., et al. 2024, MNRAS, 528, 6915

Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5

Barai P., Viel M., Murante G., Gaspari M., Borgani S., 2014, Monthly Notices of the Royal Astronomical Society, 437, 1456

Barbary K., 2016, The Journal of Open Source Software, 1, 58

Barnes J. E., 2001, in von Hippel T., Simpson C., Manset N., eds, Astronomical Society of the Pacific Conference Series Vol. 245, Astrophysical Ages and Times Scales. p. 382

Barnes J. E., Hernquist L. E., 1991, The Astrophysical Journal, 370, L65

Barnes J. E., Hernquist L., 1993, in Smith G. H., Brodie J. P., eds, Astronomical Society of the Pacific Conference Series Vol. 48, The Globular Cluster-Galaxy Connection. p. 615

Baron D., Netzer H., Poznanski D., Prochaska J. X., Förster Schreiber N. M., 2017, Monthly Notices of the Royal Astronomical Society, 470, 1687

Bayes M., Price M., 1763, Philosophical Transactions of the Royal Society of London Series I, 53, 370

Belfiore F., et al. 2016, MNRAS, 461, 3111

Bell A. R., 1978, Monthly Notices of the Royal Astronomical Society, 182, 147

Bell E. F., Phleps S., Somerville R. S., Wolf C., Borch A., Meisenheimer K., 2006, ApJ, 652, 270

Bershady M. A., Jangren A., Conselice C. J., 2000, AJ, 119, 2645

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Best P. N., Heckman T. M., 2012, MNRAS, 421, 1569

Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005, MNRAS, 362, 25

Best P. N., Kaiser C. R., Heckman T. M., Kauffmann G., 2006, MNRAS, 368, L67

Bickley R. W., et al. 2021, MNRAS, 504, 372

Bickley R. W., Ellison S. L., Patton D. R., Bottrell C., Gwyn S., Hudson M. J., 2022, MNRAS, 514, 3294

Bickley R. W., Ellison S. L., Patton D. R., Wilkinson S., 2023, MNRAS, 519, 6149

Bickley R. W., et al. 2024, arXiv e-prints, p. arXiv:2401.17277

Bignone L. A., Tissera P. B., Sillero E., Pedrosa S. E., Pellizza L. J., Lambas D. G., 2017, MNRAS, 465, 1106

Bilicki M., et al. 2018, A&A, 616, A69

Blandford R., Meier D., Readhead A., 2019, Annual Review of Astronomy and Astrophysics, 57, 467

Blanton M. R., et al. 2017, AJ, 154, 28

Blecha L., Snyder G. F., Satyapal S., Ellison S. L., 2018, MNRAS, 478, 3056

Bluck A. F. L., Conselice C. J., Buitrago F., Grützbauch R., Hoyos C., Mortlock A., Bauer A. E., 2012, ApJ, 747, 34

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Boissay R., Ricci C., Paltani S., 2016, A&A, 588, A70

Boland P. J., 1989, Journal of the Royal Statistical Society. Series D (The Statistician), 38, 181

Boquien M., et al. 2016, A&A, 591, A6

Boroson T. A., 2002, ApJ, 565, 78

Bottrell C., 2020, PhD thesis, University of Victoria, Canada

Bottrell C., Torrey P., Simard L., Ellison S. L., 2017, MNRAS, 467, 1033

Bottrell C., Simard L., Mendel J. T., Ellison S. L., 2019a, MNRAS, 486, 390

Bottrell C., et al. 2019b, MNRAS, 490, 5390

Bottrell C., Hani M. H., Teimoorinia H., Patton D. R., Ellison S. L., 2022, MNRAS, 511, 100

Bottrell C., et al. 2024, MNRAS, 527, 6506

Bournaud F., 2011, in Charbonnel C., Montmerle T., eds, EAS Publications Series Vol. 51, EAS Publications Series. pp 107–131 (arXiv:1106.1793), doi:10.1051/eas/1151008

Bovy J., et al. 2011, ApJ, 729, 141

Bower R. G., Lucey J. R., Ellis R. S., 1992, Monthly Notices of the Royal Astronomical Society, 254, 601

Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151

Brown W., Patton D. R., Ellison S. L., Faria L., 2023, MNRAS, 522, 5107

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Burlon D., Ajello M., Greiner J., Comastri A., Merloni A., Gehrels N., 2011, ApJ, 728, 58

Bustamante S., Ellison S. L., Patton D. R., Sparre M., 2020, MNRAS, 494, 3469

Byrne-Mamahit S., Hani M. H., Ellison S. L., Quai S., Patton D. R., 2023, MNRAS, 519, 4966

Byrne-Mamahit S., Patton D. R., Ellison S. L., Bickley R., Ferreira L., Hani M., Quai S., Wilkinson S., 2024, MNRAS, 528, 5864

Carroll B. W., Ostlie D. A., 2017, An introduction to modern astrophysics, Second Edition

Casteels K. R. V., et al. 2014, MNRAS, 445, 1157

Chabrier G., 2003, PASP, 115, 763

Chainakun P., Watcharangkool A., Young A. J., Hancock S., 2019, MNRAS, 487, 667

Chambers K. C., Pan-STARRS Team 2016, in American Astronomical Society Meeting Abstracts #227. p. 324.07

Chambers K. C., et al. 2016, arXiv e-prints, p. arXiv:1612.05560

Chandrasekhar S., 1943, Astrophysical Journal, 97, 225

Chen Y.-M., Wang J.-M., Yan C.-S., Hu C., Zhang S., 2009, ApJ, 695, L130

Choi E., Naab T., Ostriker J. P., Johansson P. H., Moster B. P., 2014, Monthly Notices of the Royal Astronomical Society, 442, 440

Chollet F., et al. 2015, Keras, `https://keras.io`

Cid Fernandes R., Stasińska G., Mateus A., Vale Asari N., 2011, MNRAS, 413, 1687

Ćiprijanović A., Snyder G. F., Nord B., Peek J. E. G., 2020, Astronomy and Computing, 32, 100390

Ćiprijanović A., et al. 2021, arXiv e-prints, p. arXiv:2111.00961

Cisternas M., et al. 2011, ApJ, 726, 57

Collister A. A., Lahav O., 2004, PASP, 116, 345

Comerford J. M., et al. 2024, ApJ, 963, 53

Conselice C. J., 2003, VizieR Online Data Catalog, p. J/ApJS/147/1

Conselice C. J., 2006a, The Astrophysical Journal, 638, 686

Conselice C. J., 2006b, The Astrophysical Journal, 638, 686

Conselice C. J., Mundy C. J., Ferreira L., Duncan K., 2022, ApJ, 940, 168

Corral A., Georgantopoulos I., Akylas A., Ranalli P., 2019, A&A, 629, A133

Cox T. J., Dutta S. N., Di Matteo T., Hernquist L., Hopkins P. F., Robertson B., Springel V., 2006, ApJ, 650, 791

Cox T. J., Jonsson P., Somerville R. S., Primack J. R., Dekel A., 2008, MNRAS, 384, 386

Darg D. W., et al. 2010, MNRAS, 401, 1043

Davies J. J., Pontzen A., Crain R. A., 2022, MNRAS, 515, 1430

De Vaucouleurs G., 1959, Handbuch der Physik, 53, 275

Debuhr J., Quataert E., Ma C.-P., 2011, Monthly Notices of the Royal Astronomical Society, 412, 1341

Dey A., et al. 2019, AJ, 157, 168

Di Matteo P., Combes F., Melchior A. L., Semelin B., 2007, Astronomy & Astrophysics, 468, 61

Di Matteo P., Combes F., Chilingarian I., Melchior A. L., Semelin B., 2008, Astronomische Nachrichten, 329, 952

Di Matteo P., Jog C. J., Lehnert M. D., Combes F., Semelin B., 2009, A&A, 501, L9

Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441

Djorgovski S., Davis M., 1987, ApJ, 313, 59

Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661

Domínguez Sánchez H., et al. 2019, MNRAS, 484, 93

Domínguez Sánchez H., et al. 2023, MNRAS, 521, 3861

Donley J. L., Rieke G. H., Pérez-González P. G., Rigby J. R., Alonso-Herrero A., 2007, ApJ, 660, 167

Donley J. L., Rieke G. H., Pérez-González P. G., Barro G., 2008, ApJ, 687, 111

Donnari M., et al. 2021, MNRAS, 500, 4004

Dosovitskiy A., et al. 2020, arXiv e-prints, p. arXiv:2010.11929

Du M., Ho L. C., Debattista V. P., Pillepich A., Nelson D., Zhao D., Hernquist L., 2020, ApJ, 895, 139

Dullemond C. P., van Bemmel I. M., 2005, A&A, 436, 47

Duncan K., et al. 2019, ApJ, 876, 110

Dwork C., Minow M., 2022, Daedalus, 151, 309

Eckart M. E., McGreer I. D., Stern D., Harrison F. A., Helfand D. J., 2009, ApJ, 708, 584

Eisenstein D. J., et al. 2011, AJ, 142, 72

Eisert L., Bottrell C., Pillepich A., Shimakawa R., Rodriguez-Gomez V., Nelson D., Angeloudi E., Huertas-Company M., 2024, MNRAS, 528, 7411

Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, AJ, 135, 1877

Ellison S. L., Patton D. R., Mendel J. T., Scudder J. M., 2011, MNRAS, 418, 2043

Ellison S. L., Mendel J. T., Patton D. R., Scudder J. M., 2013, MNRAS, 435, 3627

Ellison S. L., Patton D. R., Hickox R. C., 2015, MNRAS, 451, L35

Ellison S. L., Secrest N. J., Mendel J. T., Satyapal S., Simard L., 2017, MNRAS, 470, L49

Ellison S. L., Viswanathan A., Patton D. R., Bottrell C., McConnachie A. W., Gwyn S., Cuillandre J.-C., 2019, MNRAS, 487, 2491

Ellison S. L., et al. 2022, MNRAS, 517, L92

Euclid Collaboration et al. 2022, A&A, 662, A112

Faber S. M., Jackson R. E., 1976, ApJ, 204, 668

Fabian A. C., 2012, Annual Reviews of Astronomy & Astrophysics, 50, 455

Fabian A. C., Iwasawa K., Reynolds C. S., Young A. J., 2000, PASP, 112, 1145

Fabian A. C., Vasudevan R. V., Gandhi P., 2008, MNRAS, 385, L43

Farage C. L., McGregor P. J., Dopita M. A., Bicknell G. V., 2010, ApJ, 724, 267

Federrath C., Roman-Duval J., Klessen R. S., Schmidt W., Mac Low M. M., 2010, A&A, 512, A81

Ferreira L., Conselice C. J., Duncan K., Cheng T.-Y., Griffiths A., Whitney A., 2020, ApJ, 895, 115

Ferreira L., Conselice C. J., Kuchner U., Tohill C.-B., 2022a, ApJ, 931, 34

Ferreira L., et al. 2022b, ApJ, 938, L2

Ferreira L., Bickley R. W., Ellison S. L., Patton D. R., Byrne-Mamahit S., Wilkinson S., Bottrell C., 2024, Galaxy Mergers in UNIONS – I: A Simulation-driven Hybrid Ensemble for Pure Galaxy Merger Classification, accepted for publication.

Fiore F., et al. 2009, ApJ, 693, 447

Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195

Fritz J., Franceschini A., Hatziminaoglou E., 2006, MNRAS, 366, 767

Gaia Collaboration et al. 2016, A&A, 595, A1

Gaia Collaboration et al. 2018, A&A, 616, A1

Gao F., et al. 2020, A&A, 637, A94

Garofalo D., 2019, MNRAS, 489, 2308

Gaspari M., Ruszkowski M., Oh S. P., 2013, MNRAS, 432, 3401

Gavrikov P., 2020, visualkeras, https://github.com/paulgavrikov/visualkeras

George I. M., Fabian A. C., 1991, MNRAS, 249, 352

Gilmore G., Wyse R. F. G., Kuijken K., 1989, Annual Review of Astronomy and Astrophysics, 27, 555

Gordon Y. A., et al. 2017, MNRAS, 465, 2671

Goubert P. H., Bluck A. F. L., Piotrowska J. M., Maiolino R., 2024, MNRAS, 528, 4891

Goulding A. D., et al. 2018, PASJ, 70, S37

Guainazzi M., Piconcelli E., Jiménez-Bailón E., Matt G., 2005, A&A, 429, L9

Guainazzi M., et al. 2021, MNRAS, 504, 393

Guo R., Hao C.-N., Xia X.-Y., Wei P., Guo X., 2016, Research in Astronomy and Astrophysics, 16, 113

Guo C., Pleiss G., Sun Y., Weinberger K. Q., 2017, CoRR, abs/1706.04599

Gwyn S. D. J., 2008, PASP, 120, 212

Gwyn S., 2019, in Teuben P. J., Pound M. W., Thomas B. A., Warner E. M., eds, Astronomical Society of the Pacific Conference Series Vol. 523, Astronomical Data Analysis Software and Systems XXVII. p. 649

Haardt F., Maraschi L., Ghisellini G., 1994, ApJ, 432, L95

Hani M. H., Gosain H., Ellison S. L., Patton D. R., Torrey P., 2020, MNRAS, 493, 3716

Hardcastle M. J., Evans D. A., Croston J. H., 2007, MNRAS, 376, 1849

Harrison C., 2014, PhD thesis, Durham University, UK

Hausen R., Robertson B., 2019, Morpheus: Library to generate morphological semantic segmentation maps of astronomical images (ascl:1906.012)

He K., Zhang X., Ren S., Sun J., 2015, arXiv e-prints, p. arXiv:1512.03385

Heckman T. M., Best P. N., 2014, ARA&A, 52, 589

Hernández-Toledo H. M., Cortes-Suárez E., Vázquez-Mata J. A., Nevin R., Ávila-Reese V., Ibarra-Medel H., Negrete C. A., 2023, MNRAS, 523, 4164

Hernquist L., 1992, ApJ, 400, 460

Hernquist L., 1993, ApJ, 409, 548

Hewlett T., Villforth C., Wild V., Mendez-Abreu J., Pawlik M., Rowlands K., 2017, MNRAS, 470, 755

Hickox R. C., Alexander D. M., 2018, ARA&A, 56, 625

Hickox R. C., et al. 2007, ApJ, 671, 1365

Hogg D. W., 1999, arXiv e-prints, pp astro–ph/9905116

Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, ApJS, 163, 1

Hopkins P. F., Hernquist L., Cox T. J., Keres D., Wuyts S., 2009, ApJ, 691, 1424

Hopkins P. F., Cox T. J., Hernquist L., Narayanan D., Hayward C. C., Murray N., 2013, Monthly Notices of the Royal Astronomical Society, 430, 1901

Hopkins P. F., Torrey P., Faucher-Giguère C.-A., Quataert E., Murray N., 2016, MNRAS, 458, 816

Hou M., Li Z., Liu X., 2020, ApJ, 900, 79

Hou M., Li Z., Liu X., Li Z., Li R., Wang R., Wang J., Ho L. C., 2023, ApJ, 943, 50

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Hubble E. P., 1936, Realm of the Nebulae

Huertas-Company M., et al. 2015, ApJS, 221, 8

Huertas-Company M., et al. 2019, MNRAS, 489, 1859

Ichikawa K., Tazaki R., 2017, ApJ, 844, 21

Ichikawa K., Ueda Y., Terashima Y., Oyabu S., Gandhi P., Matsuta K., Nakagawa T., 2012, ApJ, 754, 45

Ivezić Ž., et al. 2019, ApJ, 873, 111

Jacobs C., et al. 2019, ApJS, 243, 17

Janowiecki S., Mihos J. C., Harding P., Feldmeier J. J., Rudick C., Morrison H., 2010, ApJ, 715, 972

Jeans J. H., 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Ji I., Peirani S., Yi S. K., 2014, A&A, 566, A97

Johansson P. H., Naab T., Burkert A., 2009, ApJ, 690, 802

Jones M. L., Hickox R. C., Black C. S., Hainline K. N., DiPompeo M. A., Goulding A. D., 2016, ApJ, 826, 12

Kartaltepe J. S., et al. 2010, ApJ, 721, 98

Kartaltepe J. S., et al. 2015, ApJS, 221, 11

Kauffmann G., et al. 2003, MNRAS, 346, 1055

Keenan R. C., et al. 2014, ApJ, 795, 157

Kennicutt Robert C. J., 1998, ApJ, 498, 541

Kent S. M., 1985, ApJS, 59, 115

Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121

Kewley L. J., Geller M. J., Barton E. J., 2006a, AJ, 131, 2004

Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006b, MNRAS, 372, 961

Kewley L. J., Rupke D., Zahid H. J., Geller M. J., Barton E. J., 2010, ApJ, 721, L48

Kitzbichler M. G., White S. D. M., 2008, MNRAS, 391, 1489

Knapen J. H., Cisternas M., Querejeta M., 2015, MNRAS, 454, 1742

Kocevski D. D., et al. 2012, ApJ, 744, 148

Kormendy J., 1977, ApJ, 218, 333

Koss M., Mushotzky R., Treister E., Veilleux S., Vasudevan R., Trippe M., 2012, ApJ, 746, L22

Koss M. J., et al. 2016, ApJ, 825, 85

Koss M. J., et al. 2018, Nature, 563, 214

Koss M. J., et al. 2023, ApJ, 942, L24

Krizhevsky A., Sutskever I., Hinton G. E., 2012, 25, 1097

Lacey C., Cole S., 1993, MNRAS, 262, 627

Lackner C. N., et al. 2014a, AJ, 148, 137

Lackner C. N., et al. 2014b, The Astronomical Journal, 148, 137

Lacy M., et al. 2004, ApJS, 154, 166

Lang D., Hogg D. W., Schlegel D. J., 2016, The Astronomical Journal, 151, 36

Larson R. B., 1974, Monthly Notices of the Royal Astronomical Society, 169, 229

Larson R. B., Tinsley B. M., 1978, ApJ, 219, 46

Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692

Lehmer B. D., et al. 2019, ApJS, 243, 3

Li W., et al. 2023a, MNRAS, 523, 720

Li W., et al. 2023b, ApJ, 944, 168

Lintott C. J., et al. 2008, MNRAS, 389, 1179

Liu X., Shen Y., Strauss M. A., 2012, ApJ, 745, 94

Liu H.-Y., Liu W.-J., Dong X.-B., Zhou H., Wang T., Lu H., Yuan W., 2019, ApJS, 243, 21

Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B., 2021, arXiv e-prints, p. arXiv:2103.14030

Liu T., et al. 2022, A&A, 661, A5

Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163

Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2008, MNRAS, 391, 1137

Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010a, MNRAS, 404, 575

Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010b, MNRAS, 404, 590

Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, ApJ, 742, 103

Luo W., Yang X., Zhang Y., 2014, ApJ, 789, L16

Madau P., Dickinson M., 2014, Annual Review of Astronomy and Astrophysics, 52, 415

Maiolino R., Shemmer O., Imanishi M., Netzer H., Oliva E., Lutz D., Sturm E., 2007, A&A, 468, 979

Man A. W. S., Zirm A. W., Toft S., 2016, ApJ, 830, 89

Mancillas B., Duc P.-A., Combes F., Bournaud F., Emsellem E., Martig M., Michel-Dansac L., 2019, A&A, 632, A122

Mantha K. B., et al. 2018, MNRAS, 475, 1549

Margalef-Bentabol B., et al. 2024, arXiv e-prints, p. arXiv:2403.15118

Marinacci F., et al. 2018, MNRAS, 480, 5113

Martin D. C., et al. 2005, ApJ, 619, L1

Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Pichon C., 2018, MNRAS, 480, 2266

Martin G., et al. 2022, MNRAS, 513, 1459

Martínez-Delgado D., et al. 2010, AJ, 140, 962

Mateos S., Alonso-Herrero A., Carrera F. J., Blain A., Severgnini P., Caccianiga A., Ruiz A., 2013, MNRAS, 434, 941

Matzko W., et al. 2022, MNRAS, 514, 4828

Mayer L., Bonoli S., 2018, Reports on Progress in Physics, 82, 016901

McAlpine S., Harrison C. M., Rosario D. J., Alexander D. M., Ellison S. L., Johansson P. H., Patton D. R., 2020, MNRAS, 494, 5713

Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3

Menzel M. L., et al. 2016, MNRAS, 457, 110

Merloni A., et al. 2024, A&A, 682, A34

Mihos J. C., Hernquist L., 1996, The Astrophysical Journal, 464, 641

Monreal-Ibero A., Arribas S., Colina L., 2006, ApJ, 637, 138

Moreno J., Torrey P., Ellison S. L., Patton D. R., Bluck A. F. L., Bansal G., Hernquist L., 2015, MNRAS, 448, 1107

Moreno J., et al. 2019, MNRAS, 485, 1320

Müller A. L., Romero G. E., 2020, A&A, 636, A92

Mushotzky R. F., Done C., Pounds K. A., 1993, ARA&A, 31, 717

Naab T., Burkert A., 2003, ApJ, 597, 893

Naiman J. P., et al. 2018, MNRAS, 477, 1206

Nair P. B., Abraham R. G., 2010, ApJS, 186, 427

Nair V., Hinton G. E., 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML'10. Omnipress, Madison, WI, USA, p. 807–814

Natarajan P., Pacucci F., Ricarte A., Bogdán Á., Goulding A. D., Cappelluti N., 2024, ApJ, 960, L1

Navarro J. F., White S. D. M., 1993, Monthly Notices of the Royal Astronomical Society, 265, 271

Negroponte J., White S. D. M., 1983, MNRAS, 205, 1009

Nelson D., et al. 2018, MNRAS, 475, 624

Nelson D., et al. 2019, Computational Astrophysics and Cosmology, 6, 2

Nenkova M., Ivezić Ž., Elitzur M., 2002, ApJ, 570, L9

Nevin R., Blecha L., Comerford J., Greene J., 2019, ApJ, 872, 76

Nikolic B., Cullen H., Alexander P., 2004, MNRAS, 355, 874

Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793

Ntampaka M., et al. 2019, ApJ, 876, 82

O'Leary J. A., Moster B. P., Naab T., Somerville R. S., 2021, MNRAS, 501, 3215

Oh K., et al. 2022, ApJS, 261, 4

Omori K. C., et al. 2023, A&A, 679, A142

Osborne C., et al. 2020, ApJ, 902, 77

Osterbrock D. E., Ferland G. J., 2006, Astrophysics of gaseous nebulae and active galactic nuclei

Pace C., Salim S., 2014, ApJ, 785, 66

Padovani P., et al. 2017, A&ARv, 25, 2

Pan H.-A., et al. 2019, ApJ, 881, 119

Park J., Smith R., Yi S. K., 2017, ApJ, 845, 128

Patton D. R., Atfield J. E., 2008, ApJ, 685, 235

Patton D. R., Carlberg R. G., Marzke R. O., Pritchet C. J., da Costa L. N., Pellegrini P. S., 2000, ApJ, 536, 153

Patton D. R., Grant J. K., Simard L., Pritchet C. J., Carlberg R. G., Borne K. D., 2005, AJ, 130, 2043

Patton D. R., Torrey P., Ellison S. L., Mendel J. T., Scudder J. M., 2013, MNRAS, 433, L59

Patton D. R., Qamar F. D., Ellison S. L., Bluck A. F. L., Simard L., Mendel J. T., Moreno J., Torrey P., 2016, MNRAS, 461, 2589

Patton D. R., et al. 2020, MNRAS, 494, 4969

Patton D. R., Faria L., Hani M. H., Torrey P., Ellison S. L., Thakur S. D., Westlake R. I., 2024, MNRAS, 529, 1493

Pawlik M. M., Wild V., Walcher C. J., Johansson P. H., Villforth C., Rowlands K., Mendez-Abreu J., Hewlett T., 2016, MNRAS, 456, 3032

Pearson W. J., Wang L., Trayford J. W., Petrillo C. E., van der Tak F. F. S., 2019a, A&A, 626, A49

Pearson W. J., et al. 2019b, A&A, 631, A51

Pearson W. J., et al. 2022, A&A, 661, A52

Pearson W. J., Rodriguez-Gomez V., Kruk S., Margalef-Bentabol B., 2024, arXiv e-prints, p. arXiv:2404.11166

Pedregosa F., et al. 2011, Journal of Machine Learning Research, 12, 2825

Peterson B. M., 2006, in Alloin D., ed., , Vol. 693, Physics of Active Galactic Nuclei at all Scales. p. 77, doi:10.1007/3-540-34621-X_3

Peterson B. M., et al. 2013, ApJ, 779, 109

Petrucci P. O., Ursini F., De Rosa A., Bianchi S., Cappi M., Matt G., Dadina M., Malzac J., 2018, A&A, 611, A59

Pfeifle R. W., Weaver K., Satyapal S., Ricci C., Secrest N. J., Gliozzi M., Blecha L., Rothberg B., 2023, ApJ, 954, 116

Pier E. A., Krolik J. H., 1992, ApJ, 401, 99

Pierce J. C. S., et al. 2022, MNRAS, 510, 1163

Pillepich A., et al. 2018, MNRAS, 475, 648

Predehl P., et al. 2021, A&A, 647, A1

Quai S., Hani M. H., Ellison S. L., Patton D. R., Woo J., 2021, MNRAS, 504, 1888

Quai S., Byrne-Mamahit S., Ellison S. L., Patton D. R., Hani M. H., 2023, MNRAS, 519, 2119

Ramos Almeida C., Ricci C., 2017, Nature Astronomy, 1, 679

Ranalli P., Comastri A., Setti G., 2003, A&A, 399, 39

Reeves A. M. M., Hudson M. J., 2024, MNRAS, 527, 2037

Renaud F., Segovia Otero Á., Agertz O., 2022, MNRAS, 516, 4922

Renzini A., Peng Y.-j., 2015, The Astrophysical Journal, 801, L29

Ricci C., et al. 2017, Nature, 549, 488

Ricci C., et al. 2021, MNRAS, 506, 5935

Ricci C., et al. 2022, ApJ, 938, 67

Rich J. A., Kewley L. J., Dopita M. A., 2014, ApJ, 781, L12

Rich J. A., Kewley L. J., Dopita M. A., 2015, ApJS, 221, 28

Ridpath I., 2004, A Dictionary of Astronomy

Robotham A. S. G., et al. 2014, MNRAS, 444, 3986

Rodriguez-Gomez V., et al. 2015, MNRAS, 449, 49

Rodriguez-Gomez V., et al. 2019, MNRAS, 483, 4140

Rodríguez Montero F., Davé R., Wild V., Anglés-Alcázar D., Narayanan D., 2019, MNRAS, 490, 2139

Rosario D. J., et al. 2015, A&A, 573, A85

Sabater J., Best P. N., Argudo-Fernández M., 2013, MNRAS, 430, 638

Salim S., et al. 2016, ApJS, 227, 2

Salvato M., et al. 2018, MNRAS, 473, 4937

Sandage A., Visvanathan N., 1978, ApJ, 223, 707

Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749

Santucci G., et al. 2024, MNRAS, 528, 2326

Satyapal S., Ellison S. L., McAlpine W., Hickox R. C., Patton D. R., Mendel J. T., 2014, MNRAS, 441, 1297

Sazonova E., et al. 2021, ApJ, 919, 134

Schawinski K., Simmons B. D., Urry C. M., Treister E., Glikman E., 2012, MNRAS, 425, L61

Schawinski K., et al. 2014, Monthly Notices of the Royal Astronomical Society, 440, 889

Schaye J., et al. 2015, MNRAS, 446, 521

Schmidt M., 1959, ApJ, 129, 243

Schwarzschild K., 1916, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, pp 189–196

Schweizer F., 1982, ApJ, 252, 455

Scudder J. M., Ellison S. L., Torrey P., Patton D. R., Mendel J. T., 2012, MNRAS, 426, 549

Secrest N. J., Ellison S. L., Satyapal S., Blecha L., 2020, MNRAS, 499, 2380

Shi K., Malavasi N., Toshikawa J., Zheng X., 2024, The Astrophysical Journal, 961, 39

Simonyan K., Zisserman A., 2014, arXiv e-prints, p. arXiv:1409.1556

Sivasankaran A., et al. 2022, MNRAS, 517, 4752

Smethurst R. J., et al. 2016, Monthly Notices of the Royal Astronomical Society, 463, 2986

Smirnov N., 1948, Ann. Math. Statist., 19, 279

Smith B. J., Campbell K., Struck C., Soria R., Swartz D., Magno M., Dunn B., Giroux M. L., 2018, AJ, 155, 81

Sola E., et al. 2022, arXiv e-prints, p. arXiv:2203.03973

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15

Springel V., 2000, Monthly Notices of the Royal Astronomical Society, 312, 859

Springel V., 2010, MNRAS, 401, 791

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Springel V., et al. 2018, MNRAS, 475, 676

Stalevski M., Fritz J., Baes M., Nakos T., Popović L. Č., 2012, MNRAS, 420, 2756

Stasińska G., Cid Fernandes R., Mateus A., Sodré L., Asari N. V., 2006, MNRAS, 371, 972

Steinborn L. K., Hirschmann M., Dolag K., Shankar F., Juneau S., Krumpe M., Remus R.-S., Teklu A. F., 2018, MNRAS, 481, 341

Stern D., et al. 2005, ApJ, 631, 163

Stern D., et al. 2012, ApJ, 753, 30

Strateva I., et al. 2001, AJ, 122, 1861

Strubell E., Ganesh A., McCallum A., 2019, arXiv e-prints, p. arXiv:1906.02243

Tacchella S., et al. 2019, MNRAS, 487, 5416

Tan M., Le Q. V., 2019, arXiv e-prints, p. arXiv:1905.11946

Tanaka M., et al. 2023, PASJ, 75, 986

Tang S., et al. 2023, MNRAS, 521, 5272

Taranu D. S., Dubinski J., Yee H. K. C., 2013, ApJ, 778, 61

Thorp M. D., Ellison S. L., Simard L., Sánchez S. F., Antonio B., 2019, MNRAS, 482, L55

Thorp M. D., Ellison S. L., Pan H.-A., Lin L., Patton D. R., Bluck A. F. L., Walters D., Scudder J. M., 2022, MNRAS, 516, 1462

Tinsley B. M., 1980, Fundamentals of Cosmic Physics, 5, 287

Toba Y., Ueda Y., Matsuoka K., Shidatsu M., Nagao T., Terashima Y., Wang W.-H., Chang Y.-Y., 2019, MNRAS, 484, 196

Toomre A., 1978, in Longair M. S., Einasto J., eds, Vol. 79, Large Scale Structures in the Universe. p. 109

Toomre A., Toomre J., 1972, ApJ, 178, 623

Treister E., Krolik J. H., Dullemond C., 2008, ApJ, 679, 140

Tristram K. R. W., et al. 2007, A&A, 474, 837

Trump J. R., et al. 2015, ApJ, 811, 26

Tubín D., et al. 2021, ApJ, 911, 100

Tully R. B., Fisher J. R., 1977, A&A, 54, 661

Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, ApJ, 598, 886

Urry C. M., Padovani P., 1995, PASP, 107, 803

Ventou E., et al. 2017, A&A, 608, A9

Villaescusa-Navarro F., et al. 2022, ApJ, 929, 132

Villforth C., 2023, The Open Journal of Astrophysics, 6, 34

Villforth C., et al. 2014, MNRAS, 439, 3342

Villforth C., et al. 2017, MNRAS, 466, 812

Villforth C., Herbst H., Hamann F., Hamilton T., Bertemes C., Efthymiadou A., Hewlett T., 2019, MNRAS, 483, 2441

Voronoi G., 1908, Journal für die reine und angewandte Mathematik (Crelles Journal), 1908, 97

Waddell S. G. H., et al. 2023, arXiv e-prints, p. arXiv:2306.00961

Walmsley M., Ferguson A. M. N., Mann R. G., Lintott C. J., 2019, MNRAS, 483, 2968

Walmsley M., et al. 2022, MNRAS, 509, 3966

Wang L., Pearson W. J., Rodriguez-Gomez V., 2020, A&A, 644, A87

Weidner C., Vink J. S., 2010, Astronomy & Astrophysics, 524, A98

Weigel A. K., et al. 2017, ApJ, 845, 145

Weinberger R., Ehlert K., Pfrommer C., Pakmor R., Springel V., 2017, MNRAS, 470, 4530

Weston M. E., McIntosh D. H., Brodwin M., Mann J., Cooper A., McConnell A., Nielsen J. L., 2016, MNRAS, 464, 3882

Weston M. E., McIntosh D. H., Brodwin M., Mann J., Cooper A., McConnell A., Nielsen J. L., 2017, MNRAS, 464, 3882

White S. D. M., 1984, ApJ, 286, 38

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Wien W., 1978, Eine neue Beziehung der Strahlung schwarzer Körper zum zweiten Hauptsatz der Wärmetheorie. Vieweg+Teubner Verlag, Wiesbaden, pp 156–165, doi:10.1007/978-3-663-13885-3_12, https://doi.org/10.1007/978-3-663-13885-3_12

Wilkins D. R., Cackett E. M., Fabian A. C., Reynolds C. S., 2016, MNRAS, 458, 200

Wilkinson S., Ellison S. L., Bottrell C., Bickley R. W., Gwyn S., Cuillandre J.-C., Wild V., 2022, MNRAS, 516, 4354

Wilkinson S., Ellison S. L., Bottrell C., Bickley R. W., Byrne-Mamahit S., Ferreira L., Patton D. R., 2024, MNRAS, 528, 5558

Woo J.-H., Yoon Y., Park S., Park D., Kim S. C., 2015, ApJ, 801, 38

Wu H., Xiao B., Codella N., Liu M., Dai X., Yuan L., Zhang L., 2021, arXiv e-prints, p. arXiv:2103.15808

Wuyts S., Cox T. J., Hayward C. C., Franx M., Hernquist L., Hopkins P. F., Jonsson P., van Dokkum P. G., 2010, ApJ, 722, 1666

Yamada S., Ueda Y., Tanimoto A., Imanishi M., Toba Y., Ricci C., Privon G. C., 2021, ApJS, 257, 61

York D. G., et al. 2000, AJ, 120, 1579

Yutani N., Toba Y., Baba S., Wada K., 2022, arXiv e-prints, p. arXiv:2205.00567

Zanisi L., et al. 2021, MNRAS, 501, 4359

Zeiler M. D., 2012, arXiv e-prints, p. arXiv:1212.5701

Zheng Y., Dave R., Wild V., Montero F. R., 2022, MNRAS, 513, 27

Zuo P., Xu C. K., Yun M. S., Lisenfeld U., Li D., Cao C., 2018, ApJS, 237, 2

van der Walt S., et al. 2014, PeerJ, 2, e453