Skyward AI: Advancing Astronomy with Intelligent Machines

by

Spencer Bialek
B.Sc., University of Victoria, 2017
M.Sc., University of Victoria, 2019

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

DOCTOR OF PHILOSOPHY

in the Department of Physics and Astronomy

Skyward AI: Advancing Astronomy with Intelligent Machines

by

Spencer Bialek
B.Sc., University of Victoria, 2017
M.Sc., University of Victoria, 2019

Supervisory Committee

Dr. Kim Venn, Co-Supervisor
(Department of Physics & Astronomy)

Dr. Sébastien Fabbro, Co-Supervisor
(Department of Physics & Astronomy)

Dr. J. Scott McIndoe, Outside Member
(Department of Chemistry)

Abstract

This dissertation represents the work I did in integrating advanced machine learning techniques into three important challenges that the field of astronomy currently faces.

Firstly, I tackled the emerging concern of contamination from low earth orbit satellites in the upcoming massive spectroscopic sky surveys. With the imminent launch of several hundred thousand satellites, there's a potential for significant contamination in wide field, multi-fiber spectrographs. I employed a multi-staged approach to gauge the feasibility and constraints of pinpointing and reducing the impact of such contamination in a WEAVE-like stellar spectral survey. By crafting a series of convolutional-network based architectures, I managed to identify and separate stellar spectra that were artificially tainted with satellite (solar-like) spectra. My findings revealed a promising capability to flag a majority of contaminated sources and reconstruct the clean spectra with minimal error. This work offers a suite of machine learning strategies that can be harnessed to enhance stellar parameters for contaminated spectra in the WEAVE stellar spectroscopic survey and similar endeavours.

In my second project, I introduced a novel solution to the well-studied problem of atmospheric turbulence compromising the clarity of astronomical images. By training a U-Net on simulated observations, I demonstrated how a sequence of short-exposure observations of a stellar field can be transformed into a turbulence- and noise-free image. This approach significantly boosts angular resolution over arbitrarily wide fields while preserving flux to a lower signal-to-noise than an averaged stack, without compromising the astrometric stability in the resultant image. It is technically simple as well, keeping costs of implementing and maintaining such a system low.

Lastly, I explored the potential of self-supervised learning in extracting meaningful representations of galaxies from millions of unlabelled sources. Recognizing the power of self-supervised methods, particularly SimCLR, I aimed to validate their utility for the UNIONS Survey. My efforts were geared towards automating the clustering and classification of galaxy types, refining photometric redshift estimations, and leveraging these techniques to unearth rare astronomical phenomena such as ultra-faint dwarf galaxies, gravitational lenses, and merging galaxies. The initial results show that, by using a query galaxy image, the fully trained SimCLR model can successfully find similar types of galaxies using a self-similarity search in a database of millions of galaxies.

Throughout these projects, I have combined machine learning and astronomical research, presenting innovative solutions to pressing challenges in the field. Each endeavour reflects my dedication to leveraging the capabilities of machine learning to propel astronomical dis-

coveries forward, offering fresh perspectives and tools to address longstanding and emerging issues in the discipline.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank:

Terry Bialek, Susan Bialek, and Corey Bialek, for helping me get through all of this with boundless love. I am eternally grateful for the support of a wonderful family. I love you all!

Katelyn Bunn, for being my Number One, my Little Babe, my Barnacle Babe, my Katelawn, and for being an incredibly supportive, Goofy, hilarious, loving partner. Every day with you is an absolute joy.

Kim Venn, for your endless support, encouragement, and patience. It was not easy starting a Ph.D. as a global pandemic set in, but your capacity for honest and compassionate communication, your willingness to self-reflect to grow and strengthen relationships, and your enthusiasm for supporting people and research, made it a lot easier. You have been an amazing friend and mentor throughout my undergrad, M.Sc., and Ph.D. (it's been a long time, eh?), and I look forward to many more years of friendship!

Sébastien Fabbro, for your limitless creativity in pushing the boundaries of knowledge. You are somehow always able to help me get through blockages in my research, and I honestly don't know if I could have gotten to this point without your friendship, guidance, and support. I am also very grateful for the food you've made... I'm still dreaming of that *tarte au flan*. See you on the ocean, Seb!

My dear friends, of which there are far too many to name. I think one of the most important things in life is to embed yourself within a healthy community, and I have somehow ended up in several. I am so grateful for the love and support from all of you. If you made the effort to find my dissertation and read the acknowledgements, you're very likely a person who has made a positive impact in my life, so you – yes YOU – are awesome and I appreciate the heck out of you. Hey, maybe I'll even have time to hang out with you now that I'm done my Ph.D.!

Things are only impossible until they are not.
Jean-Luc Picard

Chapter 1

Introduction

The astronomy projects I worked on in my Ph.D., detailed in this dissertation, rely heavily on machine learning algorithms and tools developed over the last several decades. From mitigating communications satellite contamination, to removing atmospheric turbulence in images, and finally to automating the classification of galaxies, each project utilizes these methods to solve problems in a novel way, achieving results that, I believe, would have been unattainable without machine learning. I truly stand on the shoulders of computer science giants, building useful tools with the goal of expediting the discovery process by both enhancing the quality of and fully exploiting the data we astronomers collect. As such, this dissertation begins with a comprehensive yet concise overview of machine learning – along with the methods I adopted and tips & tricks I learned throughout the process – and leads into a brief history of the impact other astronomers have had with similar methods.

1.1 What is Machine Learning?

Machine learning is a branch of artificial intelligence that enables machines to learn and improve from *experience* without being explicitly programmed. It relies on the use of statistical methods and algorithms to learn patterns and insights from the data itself, and then make predictions or decisions based on what it learned.

A common framework for *supervised* machine learning is the use of a function f($\mathbf{x}$) that maps an input vector $\mathbf{x}$ to a known output vector $\mathbf{y}$. The function f($\mathbf{x}$) is parameterized by a set of weights or coefficients $\mathbf{w}$, and the goal is to learn the optimal values of $\mathbf{w}$, which I will call $\mathbf{w}^*$, that minimize the error between the predicted output and the true output. This can be formulated as an optimization problem:

$$\mathbf{w}^* = argmin_{\mathbf{w}} \left[\mathcal{L}(f(\mathbf{x}; \mathbf{w}), \mathbf{y}) \right]$$

where $\mathcal{L}$ is a cost function, i.e. a sum of loss functions on individual examples, that, when combined with an optional regularizing term, measures the empirical risk between the predicted model output f($\mathbf{x}$;$\mathbf{w}$) and the true output $\mathbf{y}$. This learning process can be likened to guiding a child in recognizing objects, helping them learn through feedback – receiving praise for correct identifications and gentle corrections for mistakes. In the child's context, positive reinforcement might be used, such as "Well done Katelyn, that is indeed a duck!" However, in machine learning, the feedback is typically in the form of negative reinforcement: a larger output of the loss function indicates a greater error in prediction, prompting the model to adjust its parameters to minimize this "punishment" or loss. Different tasks and model architectures necessitate distinct loss functions:

(Some) Losses for Regression

- **Mean Squared Error (MSE)**: One of the mostly commonly used losses, MSE computes the average squared difference between predicted and actual values:

$$\text{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

- **Mean Absolute Error (MAE)**: This loss calculates the average of the absolute differences between the predicted and actual values. It is less sensitive to outliers compared to MSE:

$$\text{MAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

- **Huber Loss (Huber, 1964)**: Combining properties of MSE and MAE Loss, it's less sensitive to outliers, making it beneficial for regression tasks with potential outliers:

$$\text{Huber Loss}(\mathbf{y}, \hat{\mathbf{y}}) = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & \text{for } |y_i - \hat{y}_i| \leq \delta \\ \delta|y_i - \hat{y}_i| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$

  where δ is a threshold value that balances the contributions from the MSE and MAE portions of the loss.

- **Structural Similarity Index (SSIM, Wang et al., 2004)**: This is a type of perceptual loss typically used in image quality assessment . It is designed to provide a more comprehensive assessment of perceived image quality by considering changes in

structural information, luminance, and texture. The SSIM index between two images x and y is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where:

- μ_x and μ_y are the averages of x and y respectively.
- σ_x^2 and σ_y^2 are the variances of x and y respectively.
- σ_{xy} is the covariance between x and y.
- c_1 and c_2 are constants introduced to avoid instability when the denominators are close to zero.

The SSIM value lies in the range $[-1, 1]$, where a value of 1 indicates that the two images being compared are identical in terms of structural information, luminance, and texture.

The work in this dissertation was primarily concerned with regression tasks, and so I grew quite familiar with testing the different regression loss functions mentioned above. It is always advisable to try various loss functions and use an independent metric to test for the quality of predictions; although MSE is very common, it does suffer from known limitations in *some* cases such as creating "blotchy" patches in flat regions of images, "smoothing" faint features, and poor quality when judged by a human (Zhang et al., 2012). See Zhao et al. (2016) for a fuller discussion on this topic.

Optimization of the loss function in machine learning is a vast field, encompassing a variety of techniques tailored to different types of models and data (e.g., De Jong, 1988; Geem, Kim, and Loganathan, 2001). The essence of optimization is to adjust the model parameters iteratively, searching the parameter space to find the values that yield the lowest loss. Some algorithms perform a global search, considering a wide range of parameter values, while others use local search strategies, refining the parameters in small steps. The choice of optimization algorithm and its settings, such as learning rate and convergence criteria, can significantly influence the model's performance and training time (Zaheer and Shaziya, 2019; Soydaner, 2020; Haji and Abdulazeez, 2021). More on optimization later, but first let's discuss model specification.

Machine learning has experienced a series of evolutionary steps in its model development over the past several decades. One of the earliest attempts at creating machine learning

models was inspired by the human brain, leading to the development of the perceptron (Rosenblatt, 1958), a foundational model for neural networks (see next section). Despite the initial enthusiasm, limitations in the perceptron model (e.g., constrained to binary classification and only linearly-separable data) and the challenges in training more complex neural networks (e.g., limited data set sizes, computational resources, and the propagation of gradients) led to a period known as the "AI winter," where funding and interest in neural network research waned. In its post-neural network phase, simple yet effective algorithms like linear and logistic regression were the mainstay for prediction and classification tasks (Fu, 1968; Austin et al., 1988). The subsequent decades introduced more sophisticated models: decision trees (Quinlan, 1986) offered interpretable decision-making processes, while their ensemble counterparts, such as random forests and boosting, enhanced accuracy and robustness (Breiman, 2001; Freund and Schapire, 1995). Machine learning models further saw the rise of Support Vector Machines (SVMs), which excelled in finding optimal boundaries for classification (Cortes and Vapnik, 1995; Burges and Schölkopf, 1996).

As data grew both in volume and complexity, the need for more intricate models became evident. Along with technological progress (e.g., more advanced Graphical Processing Units (GPUs) for faster training and inference) and better training strategies (e.g. the backpropagation algorithm for propagating gradients, Rumelhart, Hinton, and Williams (1986)), this led to the resurgence of neural networks, setting the stage for the current *deep learning* era. While some of the earlier models are still used today, such as random forests, neural networks are at the forefront of machine learning research.

1.1.1 Neural Networks

A neural network is a type of machine learning algorithm that is loosely modeled after the structure and function of the human brain. It is composed of layers of interconnected nodes called neurons that "fire" when presented with patterns it was trained to recognize. One of the first types of neural network was the feed-forward neural network, specifically the "fully connected" feed-forward neural network. It consists of an input layer, one or more hidden layers, and an output layer. In this architecture, each neuron in a layer is connected to every neuron in the previous layer and every neuron in the subsequent layer. The neurons receive inputs from the previous layer, perform a computation, and pass their output to the next layer.

The computation performed by a neuron in a feed-forward neural network can be described by the following equation:

$$y = h\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

where y is the output of the neuron, x_i are the inputs, w_i are the weights assigned to each input, b is a bias term, and h is an activation function that introduces non-linearity into the computation. Activation functions play a crucial role in neural networks, facilitating fits to highly non-linear data and outputs. Early neural networks often used the sigmoid

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

or hyperbolic tangent

$$\tanh(z)$$

as activation functions (e.g., Hornik, Stinchcombe, and White, 1989). However, these functions suffer from the "vanishing gradient" problem, where gradients become too small for the network to learn effectively, especially in deep networks (Hochreiter, 1998). This problem was addressed by introducing the Rectified Linear Unit (ReLU) function, defined as

$$f(x) = \max(0, x)$$

(Nair and Hinton, 2010). ReLU and its variants, like Leaky ReLU and Parametric ReLU, have since become the de facto standard for many neural network architectures due to their ability to mitigate the vanishing gradient problem and accelerate convergence (Banerjee, Mukherjee, and Pasiliao Jr, 2019).

In neural networks, the importance of an activation function extends beyond the hidden layers and often influences the design of the final layer, especially in classification tasks. In multi-class classification, for example, the *softmax* function is employed in the final layer to convert the raw output scores, known as logits, into a probability distribution over the classes. This transformation ensures that the network's predictions can be interpreted as class probabilities after proper calibration. On the other hand, for regression tasks, the final layer typically remains linear, ensuring that the output is unbounded and can capture the continuous range of target values (though if there is a significant difference in the ranges of the target values in a multi-target regression task, it can be useful to normalize the values to have zero mean and unit variance so the optimization process weights each target equally).

Once the network produces its predictions, the next step is to evaluate how close these predictions are to the true values, i.e. the error. This is where the previously discussed loss

functions come into play. To optimize the parameters of the network to minimize the loss function, the backpropagation algorithm is first used to compute the gradient of the loss function with respect to each weight in the network. The gradient points in the direction of the steepest *increase* in the loss function, so optimization algorithms, such as gradient descent (Cauchy et al., 1847), adjust the model parameters in the *opposite* direction of this gradient to minimize the loss.

In the context of deep learning, given the often massive size of training datasets, processing the entire dataset simultaneously for optimization, known as *batch gradient descent*, can be computationally prohibitive. An alternative is *stochastic gradient descent* (SGD, Robbins and Monro, 1951), where the model's parameters are updated using only a single data point at each iteration:

$$\mathbf{w}^{(t)} \leftarrow \mathbf{w}^{(t-1)} - \epsilon_t \frac{\partial \mathcal{L}(x_t, \mathbf{w})}{\partial \mathbf{w}}$$

where x_t is an example sampled at iteration t and ϵ_t is a hyper-parameter called the learning rate. While SGD offers the advantage of frequent updates, leading to faster convergence, it can be noisy and may not always provide a stable path towards the minimum of the loss function.

To strike a balance between batch gradient descent and SGD, *mini-batch gradient descent* is often employed (Bengio, 2012). Here, the data is divided into smaller subsets or "mini-batches". During each iteration of training, a mini-batch of size B is fed into the model, and the model's predictions for that mini-batch are used to compute the gradient of the loss with respect to the model parameters, and subsequently update the parameters:

$$\mathbf{w}^{(t)} \leftarrow \mathbf{w}^{(t-1)} - \epsilon_t \frac{1}{B} \sum_{t'=Bt+1}^{B(t+1)} \frac{\partial \mathcal{L}(x_{t'}, \mathbf{w})}{\partial \mathbf{w}}$$

A training "epoch" is completed once the network sees all the mini-batches, and the process is repeated for a specified number of epochs. Mini-batching combines the advantages of both batch gradient descent and SGD: it allows for more efficient use of memory, can lead to faster convergence due to more frequent updates (Khirirat, Feyzmahdavian, and Johansson, 2017), and leverages the parallel processing capabilities of modern GPUs. Additionally, the noise introduced by mini-batching in the optimization process can act as a regularizer, potentially preventing overfitting and leading to better generalization (Jain et al., 2018).

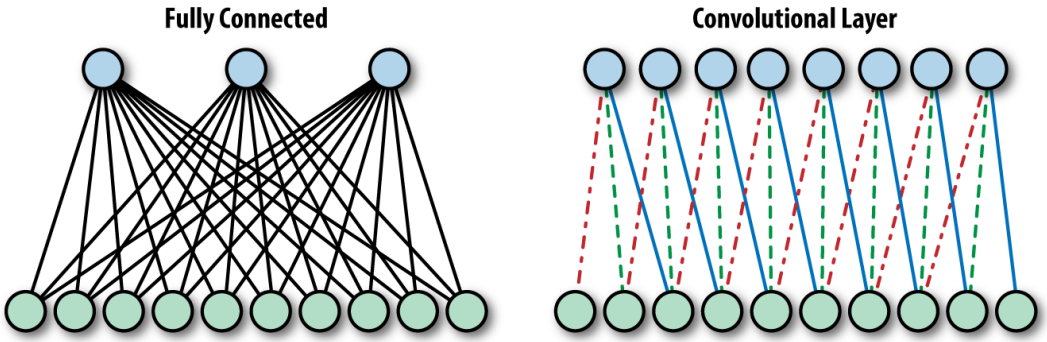

Figure 1.1: A comparison of fully connected and convolutional layers, from Hope, Resheff, and Lieder (2017). In a fully connected layer, each unit is connected to every unit in the previous layer, whereas in a convolutional layer, each unit is connected to only a local region of the previous layer (and each unit actually shares weights, as indicated by the coloured lines).

Modern Architectures

More recently, machine learning and neural network algorithms have undergone significant advancements in both usability and performance. Such improvements include proper initialization (Glorot and Bengio, 2010), advanced activation functions (Nair and Hinton, 2010), and better solvers (Kingma and Ba, 2014). Additionally, high-level user-friendly codes such as Keras (Chollet, 2015) and Pytorch (Paszke et al., 2019b) have been developed to enable easy implementation of these algorithms. Along with the availability of large datasets and the use of GPUs for high performance computing, these developments have allowed for the successful implementation of more complex neural network architectures and the introduction of the *deep learning* era. These architectures have proven to be pivotal in solving difficult tasks such as image recognition and natural language processing.

One such example of a more complex neural network architecture is the *convolutional* neural network (CNN), first effectively implemented by LeCun et al. (1989) for identifying handwritten text. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images. The key idea behind CNNs is the use of local receptive fields and shared weights. Figure 1.1 shows schematically these ideas, and illustrates the difference between "fully connected" and convolutional layers, specifically how each node is connected to the nodes of previous layers.

Mathematically, the convolution operation in a 2-dimensional CNN layer can be represented as:

$$(I * K)(x, y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} I(i, j) \cdot K(x - i, y - j)$$

where I is the input image (or feature map from a previous layer), K is the kernel or filter, and $*$ denotes the convolution operation. The result of this operation is a feature map that highlights specific features in the input, such as edges or textures.

Pooling (subsampling) layers are commonly used in CNNs to reduce the spatial dimensions of the feature maps, thereby reducing the number of parameters and computational cost, and they also introduce a form of translation invariance, ensuring that minor spatial shifts in the input do not drastically alter the network's output (LeCun et al., 1998). This property is particularly beneficial for tasks like image recognition, where the exact position of a feature might be less critical than its presence. The most common pooling operation is max-pooling, where the maximum value is taken from a set of values in a local region:

$$M(x, y) = \max_{i=0}^{a-1} \max_{j=0}^{b-1} I(x + i, y + j)$$

where M is the output of the max-pooling operation, I is the input feature map, and $a \times b$ is the size of the pooling region.

A CNN was used by Krizhevsky, Sutskever, and Hinton (2012) in an annual image classification competition, ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Their particular architecture, now referred to as *AlexNet*, outperformed the competitors of ILSVRC 2012 by a margin of more than 10% in accuracy. This significant achievement led to the widespread use and further development of CNNs in research.

While CNNs have revolutionized image processing, other neural network architectures have been pivotal for different types of data and tasks. Recurrent Neural Networks (RNNs), for instance, were designed for sequential data (like language modeling and speech recognition), continuously maintaining a "memory" of past inputs due to their internal loops, allowing them to make use of historical context (Giles, Kuhn, and Williams, 1994; Schuster and Paliwal, 1997). Traditional RNNs suffer from long-term dependency issues, which were addressed by the introduction of Long Short-Term Memory networks (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Units (Cho et al., 2014), both of which employ gating mechanisms to control the flow of information and mitigate the vanishing gradient problem.

Autoencoders (Hinton and Salakhutdinov, 2006), another type of neural network, were designed for *unsupervised* learning, in which the goal is to learn efficient codings of input data that doesn't have associated labels. By encoding inputs into a compressed representation and then reconstructing the original input as closely as possible, they are useful for dimensionality reduction, feature learning, and even generative tasks (Bengio et al., 2013). Generative models, such as Generative Adversarial Networks (GANs) and diffusion models, take a different approach. GANs, introduced by Goodfellow et al. (2014), consist of two networks trained simultaneously: a generator that creates data, and a discriminator that evaluates its fidelity. This setup allows for the generation of highly realistic data, with applications including image generation, style transfer, and data augmentation (Zhu et al., 2017). Diffusion models, on the other hand, generate data by reversing a diffusion process, typically leading to high-quality samples (Ho et al., 2020).

More recently, the Transformer architecture, introduced by Vaswani et al. (2017), has become prominent, especially in natural language processing. Unlike sequence-based RNNs, Transformers use self-attention mechanisms to weigh the significance of different parts of

the input data, effectively handling long-range dependencies. This architecture forms the backbone of generative models like GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers), which have set new performance benchmarks across a variety of natural language processing tasks (Devlin et al., 2018; Radford et al., 2018); at the time of writing this dissertation, the latest version of GPT (GPT-4) even has some researchers speculating on whether it has achieved aspects of *artificial general intelligence* (Bubeck et al., 2023). Both BERT and GPT employ *self-supervised learning* in which the data itself is used as targets (e.g. BERT learns to predict masked words in a sentence, a task known as the "masked language model" objective), a topic I will return to in Chapter 4.

These diverse architectures underscore the adaptability of neural networks to different data types and tasks, each introducing unique mechanisms and structures to handle specific challenges.

1.1.2 Training a Good, Generalizable Model

To ensure that a model generalizes well to new, unseen data, it's crucial to split the available dataset into three distinct sets: training, validation, and test sets. The **training set** is used to train the model and adjust its parameters. The **validation set** is used to tune hyperparameters, such as learning rate or regularization strength, and to prevent overfitting by providing an unbiased evaluation of the model during training. Finally, the **test set** is used to evaluate the model's performance on entirely new data, giving an indication of how the model will perform in real-world scenarios.

Overfitting can occur when a model becomes too complex and starts to memorize the training data rather than generalizing from it. By evaluating the model on a validation set during training (usually after every epoch), overfitting can be detected early and corrective measures can be taken, such as stopping training early, increasing regularization, or by only saving the model parameters if the validation loss decreases in the current epoch.

When approaching a new modeling problem, it's essential to adopt a systematic strategy to determine the best hyperparameters and model configurations. Here are some general strategies I have found to be useful:

1. **Start Small**:
    - Begin with a simpler model to understand the baseline performance.
    - Train initially on a small subset of the data to ensure quicker iterations and to ascertain if the model can fit this subset well.

2. **Overfitting as a Diagnostic Tool**:
    - Ensure the model can overfit a small subset of the data, indicating that the model and training process are working correctly.
    - Gradually increase the dataset size. If the model stops overfitting, consider the next point:
3. **Increase Model Capacity**:
    - If the model isn't capturing the data patterns, increase its complexity or capacity.
4. **Hyperparameter Search**:
    - Perform a systematic search over hyperparameters using techniques like grid search or random search.
    - Prioritize tuning crucial hyperparameters that have a significant impact on model performance.
5. **Transfer and Multi-task Learning**:
    - Consider leveraging knowledge from related tasks or pre-trained models, especially when training data is limited.
6. **Iterative Refinement**:
    - Model training is often iterative. Based on performance on a validation set, revisit and adjust hyperparameters, training set size, or model configurations.
7. **Monitoring and Evaluation**:
    - Continuously monitor training metrics and performance on validation sets. It helps to use a tool like *Tensorboard* (Abadi et al., 2015) or Weights & Biases (Biewald, 2020).
    - Evaluate the model's generalization capability on a separate test set periodically.

In summary, effective model training requires a systematic and iterative approach, starting with simpler models and gradually refining based on performance metrics. Balancing model complexity with the risk of overfitting is crucial to ensure the model is well-suited to the problem at hand.

1.1.3 The Unreasonable Effectiveness of (High Quality) Data

Even if you have the world's best model, it will be entirely unhelpful if your training set is of poor quality. Indeed, understanding and curating your training set is one of the most important aspects of machine learning. A trained model will be of poor quality if certain aspects of your data are not taken into consideration:

- **Missing Values:** Data can often have missing values, which can lead to inaccurate predictions if not handled properly. Imputation, removal of these values, or proper handling of them in a custom loss function, is crucial.
- **Noisy Data:** Random errors or variances in any dataset can distort the model's view of the underlying pattern, leading to overfitting or poor generalization.
- **Bias and Imbalance:** If certain classes or types of data are overrepresented or underrepresented, the model can become biased. This can lead to poor performance on underrepresented classes.
- **Irrelevant Features:** Including features that don't contribute to the predictive power can introduce noise and reduce the model's accuracy.
- **Outliers:** Extreme values can skew the model's understanding and lead to poor predictions for typical values.
- **Domain mismatch:** The training data might not fully represent the data that you intend to test on. Understanding the distributions of all the data you intend to train and test on is critical.

Sometimes it's infeasible to collect an appropriate data set using real data, in which case it might be better to create your own simulated data. While simulated data can be a valuable resource, especially when real-world data is scarce, there's often a gap between the two; simulated data might not capture all the intricacies and complexities of real-world data. For instance:

- **Model Assumptions:** Simulated data is often generated based on certain assumptions or models, which might not hold true in real-world scenarios.
- **Lack of Noise:** Simulated data can sometimes be too clean, lacking the noise and variability present in real-world data.

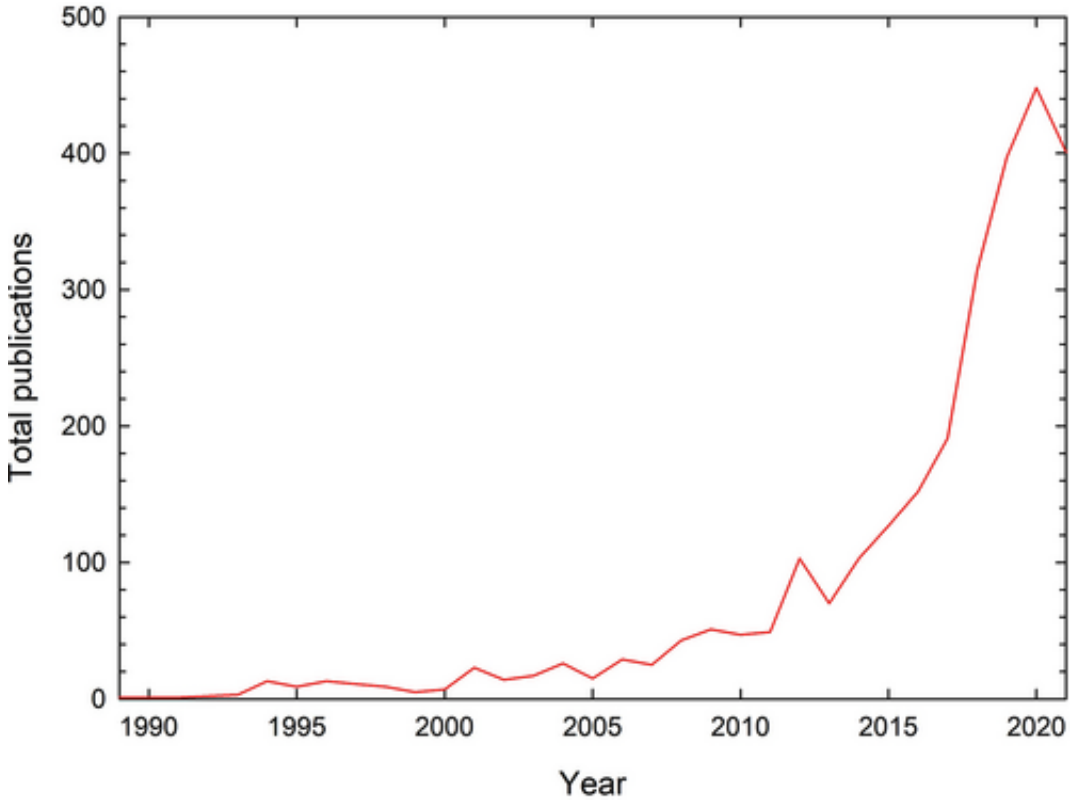

Figure 1.2: A timeline of the number of publications in astronomy using artificial intelligence, machine learning, and related terms, from Rodriguez, Rodríguez-Rodríguez, and Woo (2022).

- **Feature Differences:** There might be features present in real-world data that were not considered or included when generating simulated data.

This subject is known as the *sim2real* gap, an area in machine learning with many proposed solutions (e.g., Doersch and Zisserman, 2019; Weibel, Patten, and Vincze, 2019; Kadian et al., 2020; Rao et al., 2020; O'Briain et al., 2021; Hietala et al., n.d.; Duplevska et al., 2022). Chapters 2 and 3 discuss the gap between simulated and real data in greater detail.

1.2 A Brief History of Machine Learning in Astronomy

There's a kind of symbiosis between machine learning and astronomy. The overwhelming vastness and depth of the universe, combined with our ever-evolving technological capabilities, facilitates a seemingly endless and exponentially growing corpus of rich astronomical datasets to use in machine learning algorithms. From deep photometric sky surveys that collect images of *millions* of galaxies, to large-scale astrometric surveys intended to survey

and collect information from over a *billion* stars in our Galaxy, to next-generation radio observatories that will be collecting *exabytes* of data every day (Scaife, 2020), this exponential rise in dataset size has seen in lockstep a similarly exponential rise in attempts to use machine learning to exploit as much of the data as possible (see Figure 1.2).

In many cases, manual analysis alone and more traditional methods simply cannot handle the enormous volume of data being produced. Consider the case of Galaxy Zoo (Lintott et al., 2008a): it was a novel attempt at harnessing human intelligence to tackle the immense volumes of data generated from galaxy surveys. It hosted an online platform that invited individuals worldwide to assist in classifying millions of galaxies captured in survey images. It was an enormously successful initiative that brought together a community of citizen scientists, fostering a collaborative endeavor that not only accelerated the pace of data analysis but also democratized the process of scientific discovery in astronomy. However, it was designed for the scale of surveys like the Sloan Digital Sky Survey. With the advent of larger-scale surveys like DES (Flaugher, 2005), PanSTARRS (Kaiser et al., 2002), DECaLS (Dey et al., 2019), and especially the upcoming generation of surveys from Euclid (Laureijs et al., 2011) and WFIRST (Spergel et al., 2015), the data generation far outpaces what a limited number of volunteers can classify. Advanced machine learning algorithms have already been designed and implemented to assist in classification of these galaxies in a way no traditional method can compete with (Walmsley et al., 2020; Walmsley et al., 2021).

Machine learning has numerous applications in astronomy, and indeed has been used extensively for decades. Interestingly, one of the first applications was injecting neural networks into numerical simulations of adaptive optics control systems (Angel et al., 1990); the corrections applied to mitigate atmospheric turbulence are highly non-linear, so neural networks were a fitting choice. The first successful test of the system happened only a year later (Sandler et al., 1991), and a prototype system showed reasonable performance in increasing the quality of real images (Lloyd-Hart et al., 1992). I expand on this topic at length in Chapter 3.

The most common early examples of machine learning in astronomy were in the realm of classification. Wide-field CCD images in the early 1990s already contained hundreds of millions of objects so automated methods were developed to differentiate stars from galaxies using decision trees (Weir et al., 1995) and neural networks (Bertin, 1994). Indeed, in the paper which introduced `SExtractor` (Bertin and Arnouts, 1996) – the most common tool used today for extraction of sources in large surveys – a neural network was trained on simulated images to identify stars and galaxies and applied to real images with remarkable success. Once identified, the morphologies of galaxies could be classified by decision trees and

neural networks (Owens, Griffiths, and Ratnatunga, 1996; Kriessler et al., 1998), and stars could be classified by type using principal component analysis, hierarchical clustering, self organising maps, neural networks, or genetic algorithms (Hernandez-Pajares et al., 1992). Stars were then further classified by their spectra using simple neural networks with only one hidden layer (Von Hippel et al., 1994; Bailer-Jones et al., 1997; Bailer-Jones, 2000).

Over time, neural networks have only proven to be increasingly useful in astronomy; discoveries propelled by deep networks have been monumental. For example, 301 previously unidentifiable exoplanets were discovered using a deep learning framework called ExoMiner (Valizadegan et al., 2022), and a further 69 new exoplanets were validated by enhancing ExoMiner with a multiplicity boost framework (Valizadegan et al., 2023). The Nyx stellar stream, discovered in the vicinity of the Sun with the aid of a deep neural network trained on simulations and real data, provided critical insights into the early merger history of the Galaxy (Necib et al., 2020). Kumaran et al. (2023) showed that deep learning could be used to facilitate the reliable identification and discovery of thousands of cosmic objects, including black holes, neutron stars, white dwarfs, and stars using data from the Chandra X-ray Observatory. Large samples of rare types of extragalactic phenomena are now routinely discovered and characterized with neural networks, including galaxy mergers (Bottrell et al., 2019; Pearson et al., 2019; Ferreira et al., 2020; Bickley et al., 2021; Wilkinson et al., 2022; Chang et al., 2022) and gravitational lenses (Davies, Serjeant, and Bromley, 2019; Teimoorinia et al., 2020; Stein et al., 2022; Wilde et al., 2022; Keerthi Vasan et al., 2023; Etsebeth et al., 2023). Excitingly, a recent discovery of a new supernova[1] was announced by a team who developed BTSbot (Rehemtulla et al., 2023), a CNN system designed to automatically request follow-up observations of identified bright transients; BTSbot found a source, automatically alerted a robotic telescope which, on its own, collected a spectrum of the source, reduced it, and used a deep learning system called SNIascore (Fremling et al., 2021) to classify it. It's worth repeating the entire process, from detection to classification, was entirely automated – with no humans in the loop – largely thanks to deep learning.

Beyond discoveries, neural networks have been used for understanding the nature of dark matter (Alexander et al., 2020), dark energy (Escamilla-Rivera, Carvajal Quintero, and Capozziello, 2020), and the cosmic microwave background (Mishra, Reddy, and Nigam, 2019), deblending and deconvolving galaxy images with GANs and diffusion models (Reiman and Göhre, 2019; Xue et al., 2023), forecasting solar flares and solar cycles (Guastavino et al., 2021; Li et al., 2021b), making accurate stellar parameter predictions with data from multiple surveys and instruments (Guiglion et al., 2023), reconstructing images from radio

[1] see the announcement here: https://www.wis-tns.org/astronotes/astronote/2023-265

interferometry (Schmidt et al., 2022), and improving the accuracy of cosmological simulations (Villaescusa-Navarro et al., 2021) with domain adaptation, to name only a few examples. There is no realm in astronomy that machine learning has not enhanced with its ability to properly utilize the large amounts of either simulated or real data.

In one of my own areas of research – studies of stellar spectra – deep neural networks have been used extensively; in Fabbro et al. (2018), I helped pioneer *StarNet*, a CNN that was trained on either real or *synthetic* stellar spectra and made accurate predictions on spectra from the SDSS-III APOGEE Survey in a fraction of the time it took previous methods. I expanded on this work in Bialek et al. (2020) to test several libraries of synthetic spectra within the StarNet framework to understand better the systematic uncertainties in stellar spectral modeling and verify the benefit of using non-local thermodynamic equilibrium in the modeling of stellar atmospheres when making predictions on spectra from the Gaia-ESO Survey. In O'Briain et al. (2021) I additionally helped in the design of a GAN architecture, called *Cycle-StarNet*, that learned how to bridge the gap between theory and data to improve stellar spectra modeling in an automated way (see Wang et al., 2023, for a further application). Leung and Bovy (2018) additionally expanded on our initial work to use a CNN for making estimates of 18 chemical abundances from APOGEE stellar spectra more precisely than the commonly used APOGEE data reduction pipeline, whereas Ting et al. (2019) used a neural network to train on stellar parameters and infer the synthetic spectrum – the generative counterpart to our discriminative network – to fit to an observed spectrum and acquire chemical abundances.

The most recent developments have followed the excitement of Transformer-based large language models (LLMs); there appears to be a growing trend of so-called *foundation models* now being constructed with Transformer architectures. Różański, Ting, and Jabłońska (2023) introduced a stellar spectral foundation model that can operate with spectra of any wavelength sampling and resolution with unprecedentedly small errors in stellar parameter estimates. Leung and Bovy (2023) take a different approach to a "foundation model for stars", training a single model to perform both discriminative (e.g. stellar parameter estimates) and generative (e.g. producing a stellar spectrum from stellar parameters) tasks by being trained in a self-supervised manner on data from several different photometric and spectral surveys, resulting in a model with *general knowledge* about the properties of stars. More related to LLMs, Nguyen et al. (2023) and Ciucă et al. (2023) have created conversational agents for summarizing astronomy papers and generating novel experiments to advance the field. To highlight the astounding ability of LLMs to assist astronomers in obtaining new knowledge, I copy here an LLM-generated research proposal from Ciucă et al.

(2023) which received "high approval from [two domain experts in the field of Galactic Astronomy] due to its broad contextual understanding of Galactic Astronomy, a concrete thesis proposal, appropriate references for idea origins, and building upon existing methodologies and studies" (though I believe at least one of the references provided are in fact indicative of a lack of true understanding):

> *I propose an investigation into the role of dark matter mini-halos in the formation and evolution of globular clusters (GCs) in the Milky Way. This study would combine the high-precision astrometric data from Gaia EDR3 and DR3 (Prusti et al., 2016; Collaboration, 2020) with spectroscopic data from large ground-based surveys like APOGEE, GALAH, SDSS SEGUE, and LAMOST to characterize the dynamics of stars in the peripheral regions of GCs. The aim would be to determine whether these GCs are embedded in dark matter mini-halos, which could provide critical insights into their origins (Peebles, 1984; Peñarrubia et al., 2017). The proposed research would build upon the probabilistic approach developed by (Kuzma, Ferguson, and Peñarrubia, 2021) for studying the peripheral regions of GCs, which utilizes a mixture model in spatial and proper motion space to model cluster, extra-tidal, and contaminant stellar populations. By extending this approach to include the effects of dark matter mini-halos on the kinematics of stars in GC outskirts, we can test the hypothesis that dark matter plays a significant role in the formation and evolution of GCs. Furthermore, this study would provide a better understanding of the distribution and properties of dark matter in the Milky Way, contributing to the broader field of near-field cosmology. Integrating this information with the existing knowledge of the hierarchical assembly of the Milky Way (Viswanathan et al., 2023) and the role of rapid gas accretion in the inner Galactic disc (Snaith et al., 2022) would help paint a more comprehensive picture of our Galaxy's formation history and its underlying dark matter distribution.*

Recognizing the growing utility of machine learning in astronomy, the *Canadian Astronomy Long Range Plan* (Gaensler and Barmby, 2020), and specifically a white paper submitted to it by Venn et al. (2019), identified machine learning as one of the primary reasons for substantially increasing the computing infrastructure available to Canadian astronomers. As the field continues to grow, there will be a need for more astronomers with expertise in machine learning methods to help drive new discoveries and insights in astronomy. I hope

this dissertation serves to further strengthen the synergy between machine learning and astronomy, emphasizing the role of machine learning in the future of astronomical research.

1.3 Dissertation Outline

This dissertation begins in Chapter 2 with a project that was started in response to the growing problem of governments and private companies sending tens of thousands of communications satellites into orbit with very few regulations in place (the only *international* documents which establish the principles of space law include the Outer Space Treaty of 1967 and the International Telecommunication Union's Radio Regulations, with additional *national* regulations coming from, e.g., the Federal Communications Commission in the US which makes laws about allowable radio frequencies for communication and how to properly de-orbit and dispose of satellites). With the assumption that most future regulations will not fully solve the problem, a method was devised to detect and remove the satellite contamination from observations of stellar spectra. Chapter 3 discusses a project which deals with a different kind of contamination, namely the effect of atmospheric turbulence in astronomical images, and how machine learning can be used to mitigate it. Chapter 4 introduces a project that attempts to maximally utilize the vast amounts of photometric data from the UNIONS Survey. Finally, Chapter 5 ends with some conclusions on my work, as well as some musings on the current and future state of the combination of machine learning and astronomy.

Chapter 2

StarUnlink: identifying and mitigating signals from communications satellites in stellar spectral surveys

This chapter contains an edited version – with an expanded introduction – of the paper originally published in Monthly Notices of the Royal Astronomical Society, Volume 524, Issue 1, Pages 529–541 (2023). Authors: Spencer Bialek, Sara Lucatello, Sébastien Fabbro, Kwang Moo Yi, Kim A. Venn.

Personal Contributions

I was the main contributor on this project. I developed the methodologies and wrote the entire Python codebase, and I wrote most of the paper.

Abstract

A relatively new concern for the forthcoming massive spectroscopic sky surveys is the impact of contamination from low earth orbit satellites. Several hundred thousand of these satellites are licensed for launch in the next few years and it has been estimated that, in some cases, up to a few percent of spectra could be contaminated when using wide field, multi-fiber spectrographs. In this paper, a multi-staged approach is used to assess the practicality and limitations of identifying and minimizing the impact of satellite contamination in a WEAVE-like stellar spectral survey. We develop a series of convolutional-network based architectures to attempt identification, stellar parameter and chemical abundances recovery, and source separation of stellar spectra that we artificially contaminate with satellite (i.e. solar-like) spectra. Our results show that we are able to flag 67% of all contaminated sources at a precision level of 80% for low-resolution spectra and 96% for high-resolution spectra. Additionally, we are able to remove the contamination from the spectra and recover the clean

spectra with a <1% reconstruction error. The errors in stellar parameter predictions reduce by up to a factor of 2-3 when either including contamination as an augmentation to a training set or by removing the contamination from the spectra, with overall better performance in the former case. The presented methods illustrate several machine learning mitigation strategies that can be implemented to improve stellar parameters for contaminated spectra in the WEAVE stellar spectroscopic survey and others like it.

2.1 Introduction

The 21st century has ushered in a new era of satellite constellations, marking a significant milestone in space-based technologies and commercial exploitation of the skies above us. The drive towards establishing a robust infrastructure for global low-latency broadband internet and other telecommunications needs has catalyzed the dramatic expansion of satellites, particularly in Low Earth Orbit (LEO). Commercial enterprises have emerged as major players in this realm, leading to an increase in satellite numbers by at least an order of magnitude over the last decade (Lawrence et al., 2022).

Several tens of thousands of LEO satellites from various companies have been licensed for launch by the end of the decade, with announcements of several hundred thousands more on the horizon (see Table 2.1). Projects such as Starlink, OneWeb, Kuiper, and StarNet/GW are at the forefront of this satellite constellation era, with ambitions of providing ubiquitous internet coverage across the globe. It is imperative that astronomers act now to understand and mitigate the full extent of the brewing problem of satellite contamination.

Table 2.1: Satellite constellations currently in orbit or planned for launch in the future. Table re-created from Di Vruno and Tornatore (2023).

Constellation	**Number of Satellites**	**Altitude [km]**
Starlink Phase 1	4,400	550
OneWeb Phase 1	648	1200
Amazon Phase 1	3,200	$\sim$ 600
Guowang (GW)	13,000	590 to 1145
Starlink VLEO	7,600	340
Telesat	298	1000
Starlink Phase 2	30,000	328 to 614
OneWeb Phase 2	6,372	1200
Cinnamon–937	327,320	550 to 643

2.1.1 Impact of Satellites on Astronomical Observations

Besides the potential implications for space debris (Vidaurri and Gilbert, 2023) and polluting our atmosphere with aluminum when de-orbited (Boley and Byers, 2021; Plane, 2023), a growing concern is the contamination of the night sky impacting our observations of astronomical sources. Since this is a novel and evolving problem, most published works have simply attempted to measure and/or forecast the extent of the problem, without a clear direction for mitigation strategies. The magnitude of these projects is well illustrated in Lawler, Boley, and Rein (2022), in which they modeled the orbits of 65,000 artificial satellites soon to inhabit our night sky. The study revealed that at certain points throughout the night, there could be several hundred satellites bright enough to interfere with astronomical observations, accounting for about 7% of the total visible point sources in the sky at $g < 7$ within a couple of hours of sunrise or sunset.

Additionally, reports such as the *SATCON1* and *SATCON2* (Walker and Hall, 2020; Walker and Hall, 2021), the *Report on Mega-Constellations to the Government of Canada and the Canadian Space Agency* (Boley et al., 2021), and the *Dark and Quiet Skies for Science and Society* reports[1] (Walker et al., 2020; Walker et al., 2021), have highlighted the detrimental impacts of satellite constellations on astronomical observations. These comprehensive studies, supported by the National Science Foundation, American Astronomical Society, Canadian Astronomical Society, and the European Astronomical Society, delve into the intricacies of the problem, providing a sobering look at the challenges posed by the onset of the satellite constellation era. In the *SATCON1* report, the myriad ways satellite constellations could impede critical observations of the cosmos were discussed, which I expand on here.

Among the identified vulnerable areas are the observations of rare transient events such as gravitational waves, gamma-ray bursts, and fast radio bursts. These fleeting occurrences provide a window into the extreme conditions and events transpiring across the universe (Zhang, 2016; Wang and Nitz, 2022), and due to their temporary nature, follow-up observations would not be possible if the data were found to be contaminated.

Extensive extragalactic imaging, vital for understanding cosmic phenomena like dark matter and dark energy, is also at risk; this type of imaging relies on capturing weak lensing effects (Heavens, 2009), studying galaxy morphology (Croft et al., 2009), and measuring ultra-faint low surface brightness (Brandt, 2016), all of which demand an unobstructed view of the sky else satellite streaks (see Figure 2.1) – and algorithms that attempt to mask

[1]The reports can be found at this website

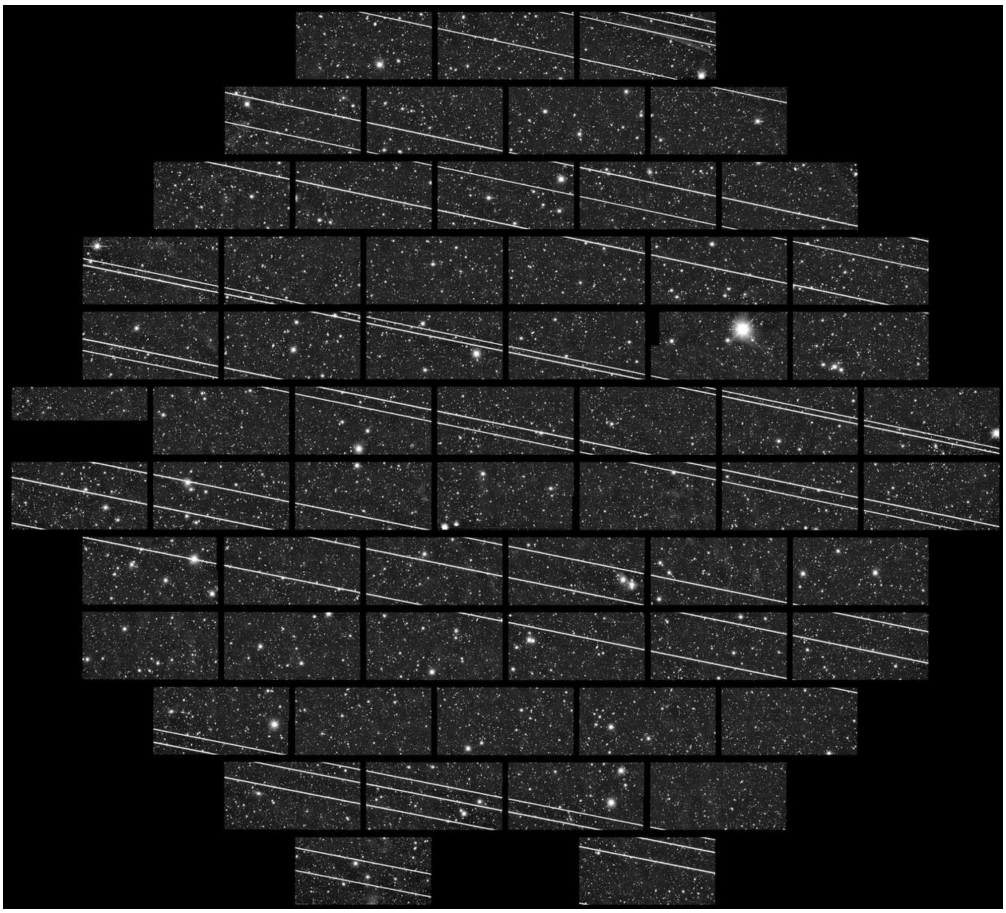

Figure 2.1: Example of an astronomical image affected by satellite streaks. Credit: Cerro Tololo Inter-American Observatory (CTIO).

them – could cause potentially unaccounted systematic errors (Tyson et al., 2020). These satellite streaks can additionally affect the monitoring of Near-Earth Objects (NEOs) – for scientific and planetary defense purposes – since NEOs too leave streaks on images (Ye et al., 2019). Satellite streaks have already been observed to be increasing in frequency across many different imaging facilities (Lang, Spencer, and Mitchell, 2023; Spencer, Lang, and Mitchell, 2023; Kruk et al., 2023).

Aside from the issue of satellite streaks, radio frequency observations face additional challenges. While satellite manufacturers agreed not to transmit within sensitive narrow radio bands, there are occasions where radio observatories may need to observe in broader bands (either for higher sensitivity or to capture specific emission phenomena) which can contain radio frequency interference (RFI) from satellite downlink signals (Massey, Lucatello, and Benvenuti, 2020). Moreover, radio observatories are strategically situated in locations where RFI from terrestrial sources such as phones, microwaves, and computers is minimized. Even if satellite downlink signals did not transmit within bands important for astronomy, the electronic components on board – navigation computers, high speed network switches, Hall effect thrusters, and power supplies – have been found to emit *unintended* electromagnetic radiation strong enough to obscure the faint signatures from astronomical sources (Vruno et al., 2023; Grigg et al., 2023); there is nowhere to hide from the RFI of satellites.

These highlighted domains underline a broad spectrum of scientific endeavors that may encounter hurdles in the face of commercial expansion of satellite constellations. There is an urgent need for coordinated efforts between astronomers and satellite operators to devise strategies that would mitigate these challenges, ensuring that the march of astronomical discovery continues unhindered.

One domain not mentioned above is stellar spectroscopy. Indeed, the issue of satellite contamination in stellar spectral surveys is the main topic of this chapter, and one in which I will devote more attention to in the following sections.

Stellar Spectral Surveys

When a satellite passes through the field-of-view of a telescope collecting a spectrum of a star, it will reflect sunlight and superimpose a solar-like spectrum on the science target. This is a type of contamination that is far more difficult to detect than a streak in an image since it changes the shape of a spectrum in a subtle way; an analogously simple “line finder" algorithm, used for images, does not exist for a contaminated spectrum, and even a trained professional would have difficulties with detection.

Not every spectroscopic facility is necessarily at risk, however. When a high-resolution spectrograph on a small or medium-sized telescope is used, typically the limiting magnitude of the selected targets is bright enough that the faint signal from a satellite will be negligible and undetectable. Conversely, with the same spectrograph on a larger telescope, e.g. the planned 11.25m Maunakea Spectroscopic Explorer (MSE, McConnachie et al., 2016), the fainter limiting magnitude of the targets would introduce two compounding effects: (1) more (faint) targets can be included in a single multi-object exposure, increasing the chances of a single target being contaminated, and (2) the limiting magnitude could be comparable to the effective magnitude of satellites, leading to a contaminated signal that can affect measurements of stellar properties yet be difficult to detect (Bassa, Hainaut, and Galadí-Enríquez, 2022). Indeed, a particular problem is that the contamination from a satellite would likely go unnoticed until after the spectra are recorded, thus can only be detected in *post-processing*. This could be a significant problem for the forthcoming sky surveys.

The expected impact on spectroscopic observations has not been studied extensively; existing work has only attempted to quantify the approximate number of spectra that will be affected at different times throughout the evening. Hainaut and Williams (2020) estimate that, for wide field, multi-fiber spectrographs with a faint limiting magnitude, a few percent of spectra collected during the first and last hours of the night will be impacted by LEO satellite contamination, while Bassa, Hainaut, and Galadí-Enríquez (2022) estimate that up to 0.8% of 4MOST spectra will be affected – though we note these estimates were based on several tens of thousands of LEO satellites which, as previously mentioned, could be an underestimate for when 4MOST and WEAVE are collecting most of their data, and thus are likely lower limits. These surveys are designed to observe thousands of spectra per setting over a large field of view, with typically 10+ settings per night. Over the lifetime of these surveys (∼10 years), then 0.8% per setting x 1000 stars per setting x 10 settings per night x 300 nights per year x 10 years = 240,000 contaminated spectra (out of 30 million spectra in the total survey).

No one has yet quantified the impact the contamination would have on derived stellar properties nor proposed any feasible mitigation strategies (especially for data known to be contaminated, in which case the current plan is to simply discard the data). Because satellites will be abundant, physical mitigation strategies such as avoiding observations cannot be an option – they will sacrifice too much in terms of observing time (Hu et al., 2022). Developing alternative mitigation strategies – along with quantifying the impact on stellar parameter estimation – is therefore the goal of this project. As the light from a bright satellite would primarily be reflected sunlight, then it is easy to simulate a contaminated observation of a

stellar target by modelling it as an additive mix of stellar and solar-like spectra. If there are solar spectra of varying S/N observed with the same instrument collecting stellar spectra, it is possible to create a large data set of artificially contaminated spectra. One can therefore attempt post-processing algorithms to identify which features of the spectra correspond to either solar or stellar light, and thereafter used for contamination mitigation strategies.

In this chapter, I study the practicality of detecting and mitigating satellite contamination using a convolutional neural network and U-Net on *post-processed* spectra. I will focus on spectroscopic surveys that will make use of upcoming and planned wide-field, high multiplexing spectrographs, because of their requirements for extreme precision and accuracy in understanding the dynamics and formation history of the Milky Way (more on this in the next section). In particular I will assess the impacts of LEO satellites on the European WEAVE Survey (Dalton et al., 2018), with implications that carry over to 4MOST (de Jong et al., 2019) (both are on 4m telescopes and are similarly impacted due to having similar characteristics such as diameter, average fiber density on the field-of-view, and spectral range/resolution), and larger telescopes like MSE, which will have both high and low/moderate resolution modes, and the 8.2m Subaru Telescope's Prime Focus Spectrograph (Tamura et al., 2018), which will have only a low-resolution mode and thus a very faint limiting magnitude.

Before I discuss the methods I developed for this purpose, I will first justify why we think stellar spectra are so important, and why we should worry about their contamination, in a large spectroscopic survey like WEAVE.

Why WEAVE?

One of the overarching goals of astronomy is to better understand our place in the universe, propelling us towards fundamental questions: How did the universe originate? How did it evolve to its current state? What factors contributed to our existence here? These questions fuel our motivation for astronomical research. One method to approach these questions is by meticulously studying our own Galaxy. After all, we are offered a unique vantage point – being situated within the Milky Way galaxy itself – to record light from a vast number of its stars. A deep and intricate history of the Galaxy is revealed in its starlight; its formation and evolution imprint specific signatures in the stars' location, kinematics, and chemistry (Freeman and Bland-Hawthorn, 2002).

In pursuit of these answers, the European Space Agency (ESA) launched the Gaia spacecraft in 2013. This advanced space observatory is tasked with creating a comprehensive

three-dimensional map of our Milky Way galaxy, with a primary mission which involves surveying over a billion stars and carefully recording their positions, distances, motions, and brightness (Prusti et al., 2016). Additionally, Gaia's Radial Velocity Spectrometer (RVS) captures medium-resolution (R $\sim$ 11,000) near-infrared spectra of stars, allowing for the measurement of radial velocities – crucial for understanding their 3D motion – and chemical composition – providing clues about the environment they were born in, especially useful when examining low-metallicity, low-mass stars that were likely formed when the universe was much younger (Frebel and Norris, 2015).

Since its inception, Gaia has profoundly transformed our comprehension of the Milky Way. For example, the Gaia-Enceladus-Sausage kinematic structure in the Galactic halo was identified, shedding light on the Milky Way's last significant merger event which contributed an abundance of stars in the inner halo and thick disk (Helmi et al., 2018). It is important to note that the chemical signatures of the stars, derived from their spectra, provided a significant line of evidence that these stars had a unique origin outside of the Milky Way, and indeed it is this chemical information which helps lift degeneracies when relying on kinematic information alone (Helmi, 2020).

The third release of Gaia data (DR3) provided an all-sky, homogeneous collection of spectra from over 5 million stars in the Milky Way. When combined with earlier releases of astrometric data, the large-scale chemo-dynamic structure and history of the Galaxy could be investigated, revealing symmetries, substructures, and merger debris (Recio-Blanco et al., 2023b). While the spectroscopic capabilities of Gaia are impressive, they are not without their limitations. The observational sensitivity of Gaia's RVS is tailored primarily to stars brighter than a magnitude of G $\sim$ 14, with DR3 only containing 61,000 spectra for magnitudes fainter than 13.6, meaning the vast majority of stars with Gaia kinematic information do not have RVS radial velocity or chemical abundance data. Moreover, while the RVS provides valuable spectral data, its resolution is limited, which can sometimes constrain detailed spectral analysis – though some methods taking full advantage of full spectra analysis may help overcome traditional analysis limitations (Ting et al., 2017). Other instruments are required to fill in this information gap and probe a much larger area of the Milky Way.

Recently installed on the 4.2m William Herschel Telescope (WHT), the WHT Enhanced Area Velocity Explorer (WEAVE) multi-object spectrograph offers a unique capability of capturing up to 1000 individual spectra in a single exposure. Leveraging this capability, the WEAVE Galactic Archaeology survey intends to scan the northern hemisphere sky, collecting both low-resolution (R $\sim$ 5000) and high-resolution (R $\sim$ 20,000) spectra of millions of stars in our Galaxy. The gathered low-resolution spectra will serve to deduce radial

velocities, effective temperatures, surface gravities, and metallicities of $\sim$2 million stars in the fainter Gaia magnitude range ($16 < G < 20.7$), whereas the high-resolution spectra will be used for precise radial velocities and elemental abundances for $\sim$1.5 million stars with $12 < G < 16$ (Jin et al., 2022). These data will augment the Gaia satellite's analysis by providing information on stars too faint for examination by its RVS, probing out to vast distances and acquiring large samples of stars in the halo, thin disk, and thick disk.

It is now clear why spectra from the WEAVE Survey will be so important to a deeper understanding of the formation and history of the Milky Way. If a significant percentage of the spectra are contaminated by satellite reflections, this could cause spurious conclusions by biasing the chemical abundance and stellar parameter estimates. Again, it is imperative that we are able to detect when contamination has occurred, and/or find a way to still use the contaminated data so we do not throw away useful spectra. The rest of the chapter details the method I derived for doing this. In Section 2.2, I clarify how I use the Gaia-ESO FLAMES-UVES spectra as an appropriate *post-processed* test dataset. In Section 2.3, I describe the machine learning methods used. In Section 2.4, I apply the machine learning methods to contaminated Gaia-ESO FLAMES-UVES spectra to detect and remove the contamination and mitigate the errors in stellar parameter estimation. I discuss the results in Section 2.5, including the limitations and best strategies going forward, and further applications of the method for bright sky subtraction and detecting spectroscopic binary stars more generally.

2.2 Spectral Data

A set of 3100 R$\sim$47,000 FLAMES-UVES spectra (Pasquini et al., 2002) from the Gaia-ESO Survey (GES) DR4 (Sacco et al., 2014), in the wavelength range $483 - 530$ nm and with no continuum normalization applied, is used in this study. The GES targeted dwarfs and giants with metallicities between -2.5 and +0.5 dex and spectral types from O to M, so our sample contains a large variety of stars. These spectra were chosen as they cover a broad spectral region at optical-IR wavelengths, with high SNR, and high resolution. To match the resolution and wavelength sampling of the WEAVE stellar survey, they were degraded and re-sampled to R=20,000 for the high-resolution (HR) mode, and R=5000 for the low-resolution (LR) mode, which we respectively refer to as the `WEAVE-GES-HR` sample and `WEAVE-GES-LR` sample hereafter. Note that this spectral range corresponds to the 'green' arm of the WEAVE HR mode and a portion of the 'blue' arm of the WEAVE LR mode; the red ($595 - 685$ nm) and blue ($404 - 465$ nm) arms of the HR mode and the red ($579 - 959$ nm) and entire blue ($366 - 606$ nm) arms of the LR mode will be used in future studies.

Additionally, the spectra are supplemented with a data set of solar/sky spectra[2]. A communications satellite is visible in the sky only because it is reflecting sunlight, thus a collection of solar spectra is crucial for simulating what a contaminated observation would look like. The details of the simulated contamination will be explained in Section 2.3.

2.3 Machine Learning Methods

The application of ML methods for analyzing astrophysical datasets has increased in popularity, particularly as computing power and complex ML algorithms become more powerful and accessible (Fluke and Jacobs, 2020). Large observational surveys, as well as simulations, have provided massive datasets for developing ML tools with astrophysical applications. Large datasets are useful because they help to identify patterns and relationships in the data that might not be apparent with smaller datasets. With more data of high quality, models can learn to generalize better due to the diversity of examples, meaning they can make accurate and robust predictions about new, unseen data.

While other tools could be used to identify solar contamination, e.g. a traditional matched filter with solar spectrum template, we decided to use ML because of its computational efficiency (a dataset of millions of spectra could be analyzed in a couple of minutes) and history of successful implementations in astronomy (see Sen et al. 2022 for a recent review).

A neural network (NN) is a function that transforms an input into a desired output. The function is composed of many parameters, arranged in layers with non-linear activation functions, which form a highly non-linear combination of the input features, and allows for complex mappings to be represented with high precision. A convolutional neural network (CNN), in which a sequence of learned filters followed by a series of learned inter-connected nodes, can transform a stellar spectrum into a prediction of associated stellar parameters. NNs and CNNs have proven to be effective at deconstructing and modelling the components of stellar spectra (e.g., Fabbro et al., 2018; Leung and Bovy, 2019; Ting et al., 2019; Guiglion et al., 2020; Bialek et al., 2020; Sharma et al., 2020; O'Briain et al., 2021; Zhao, Wei, and Jiang, 2022).

In this section, we describe the ML methods we have built to identify and demix contaminated spectra in the *post-processed* spectra in the Gaia-ESO FLAMES-UVES survey.

[2]A set of 232 FLAMES-UVES solar spectra were collected from this website

2.3.1 Creating the Training and Test Sets

The sample of 3100 FLAMES-UVES spectra was split 80%/20% into reference/test sets, with a further 85%/15% split of the reference set to acquire training/validation sets. Within each set, a random stellar spectrum was chosen, contaminated to a random fraction (between 0-50% of the median intensity value) with a randomly chosen solar spectrum from a set of 232 solar spectra, degraded from a resolution of 47,000 to 20,000 for `WEAVE-GES-HR` and to 5000 for `WEAVE-GES-LR`, and finally re-binned to the respective HR or LR WEAVE wavelength grid. Spectra in the reference set were additionally uniformly radial velocity shifted ($|\, v_{rad}| <$ 200km/s) before the contamination was applied. The boundaries on the contamination were chosen based on Figure 6 from Bassa, Hainaut, and Galadí-Enríquez (2022) which shows the effective magnitude of LEO satellites to be around 21 – given the normal exposure times expected for WEAVE (∼600-1200s) – and on the expected limiting magnitude of WEAVE to be G=20.7 in the low-resolution mode (Jin et al., 2023). Note the expected limiting magnitude for the high-resolution spectra is G=16 so the contamination problem is far more problematic in the low-resolution spectra. The contaminated spectrum was obtained with the formula

$$\mathbf{x}_{\text{contaminated}} = \mathbf{x}_{\text{stellar}} + c \times \frac{\tilde{\mathbf{x}}_{\text{stellar}} \mathbf{x}_{\text{solar}}}{\tilde{\mathbf{x}}_{\text{solar}}}$$

where $\mathbf{x}_{\text{stellar}}$ is a radial velocity shifted stellar spectrum, $\mathbf{x}_{\text{solar}}$ is a solar spectrum, c is the contamination fraction, and $\tilde{\mathbf{x}}$ refers to the median. The process was repeated until the final training/validation/test sets contained 35000/5000/10000 spectra (the performance did not increase when training on more samples). All data sets retained the original, clean spectra, such that e.g. the training set included 35000 contaminated and their corresponding 35000 clean spectra, important for corrections when imbalanced training was taking place.

2.3.2 Neural Networks

Classification Network for Detection

A standard CNN was implemented for the classification task (i.e., our StarNet architecture, Fabbro et al. 2018; Bialek et al. 2020). The input was an individual stellar spectrum and the output prediction was a score between 0 (clean) and 1 (contaminated). A binary cross-entropy (BCE) loss function was used to compare the CNN's output prediction with the target label of the input spectrum, such that the output of the trained CNN can be thought of as a prediction of the probability that the spectrum is contaminated (though it is not guaranteed to be a probability - i.e. calibrated). The Adam algorithm (Kingma and Ba,

2014) was used for the gradient-based optimization of the BCE, with a learning rate that was decreased when the performance on the validation set plateaued. The model that scored the lowest BCE on the validation set was saved, to ensure the final model did not overfit to the training data, and used for all tests.

Regression Network for Stellar Parameters

The same CNN and training scheme was used as in the classification task; however, the last layer was changed to linear outputs, and the loss function was changed to the mean-squared error (MSE). These were then used to regress on the stellar parameters of the individual stellar spectra. The mean and standard deviation of the distributions of stellar parameters were used to normalize the labels (for zero mean and unit variance), and then to de-normalize the inferred stellar parameters.

U-Net for Removing Contamination

U-Nets are CNNs that were originally created for biomedical imaging segmentation tasks (Ronneberger, Fischer, and Brox, 2015). These include downsampling and upsampling layers that encode then decode data, and they combine high-resolution information from downsampled layers with subsequent upsampling layers via skip connections which helps to prevent information loss and vanishing gradients.

After a few trials on U-Net architectures, a particular adaptation of the U-Net to one-dimensional domains, *Wave U-Net* (Stoller, Ewert, and Dixon, 2018), was adopted as a base architecture. Wave U-Net was created for audio source separation and introduced architectural improvements such as an upsampling technique and context-aware prediction framework to reduce artifacts in its output. Empirical results of Wave U-Net showed its capacity in isolating raw wave forms from various instruments (e.g. bass, guitar, drums, voice). In a similar manner, Wave U-Net can be given a raw spectrum with LEO satellite contamination and tasked with inferring – and therefore isolating – the pure stellar spectrum. Figure 2.2 shows schematically how Wave U-Net works on a stellar spectrum. We note an improved version of Wave U-Net was used in this study[3].

A similar training scheme as the previous methods was used but a mean absolute error (MAE) regression loss was used instead of an MSE loss because the MSE led to an undesired amount of smoothing in the reconstructed spectra.

[3]details of the architecture improvements can be found at https://github.com/f90/Wave-U-Net-Pytorch

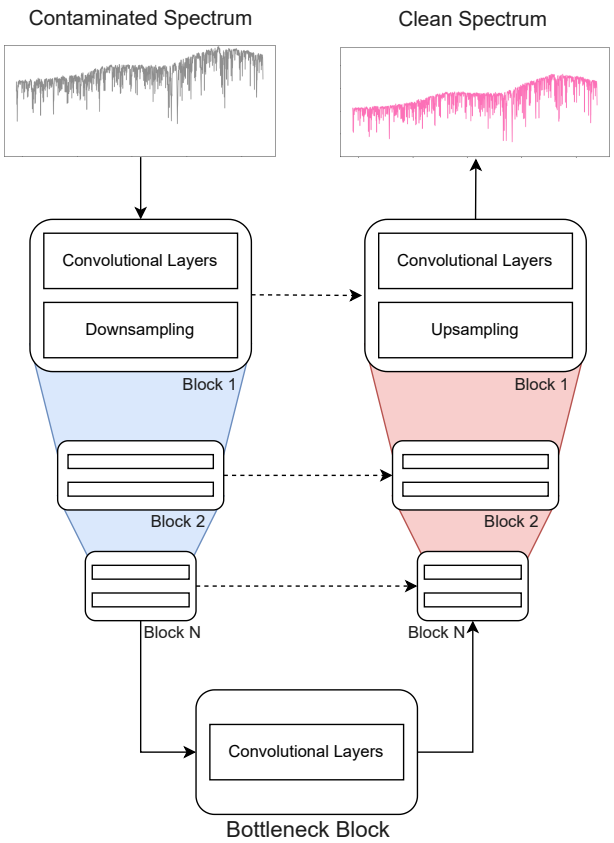

Figure 2.2: A schematic illustration of the Wave U-Net removing satellite contamination from a stellar spectrum. The details of each block can be found in the project's Github repository.

An Ensemble of Models

If an ensemble of NNs with random initializations are independently trained on the same data and given the same task, then every NN will settle to a different minimum of the loss function owing to the stochastic nature of the training procedure. As such, each NN will have learned something unique about the training data, and no single NN in the ensemble can necessarily be considered the best. It is therefore an advantage to utilize each NN in the ensemble to make the most robust predictions. Following our similar procedure as in Bialek et al. (2020), an ensemble of CNNs was trained for the detection and probability of an individual spectrum being contaminated.

2.4 Results

The `WEAVE-GES-HR` and `WEAVE-GES-LR` spectra were used to assess the performance of the CNN and Wave U-Net and showed the following results.

2.4.1 Detection

For the detection task, a spectrum was considered to be contaminated if its contamination level was >1%. In the test set of 10,000 spectra, 1,000 were contaminated and 9,000 were clean, uncontaminated spectra. Figure 2.3 shows a confusion matrix for the contamination detection on the `WEAVE-GES-LR` test set with the decision threshold set to 0.7, i.e. spectra were considered contaminated if the output prediction of the CNN was >0.7. The true negative detection rate was an impressive 97.67%, meaning that, in this particular case, only 2.33% of clean spectra (210 total) were misclassified. Meanwhile, the true positive detection rate was 85.42% such that 14.58% of contaminated spectra (143 total) were misclassified as clean spectra. The results on the `WEAVE-GES-HR` sample were similar. Depending on the decision threshold chosen, these variables can change based on the desired optimization; the next section explores this.

Fine-tuning the Decision Threshold

Since the CNN outputs a prediction between 0 and 1, there must be a choice made about when to consider a prediction to be positive or negative. A natural choice for this *decision threshold* might be 0.5 if calibrated, i.e. predictions above and below this value are the contaminated and clean spectra, respectively. If the goal is to detect contaminated spectra,

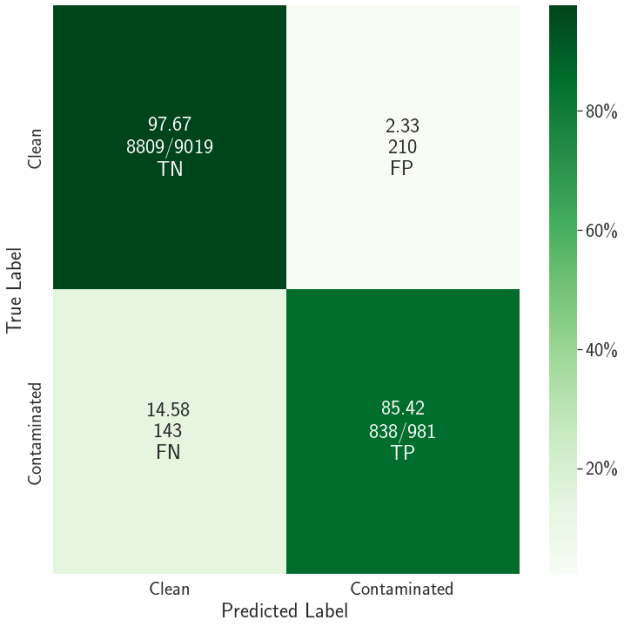

Figure 2.3: A confusion matrix showing information about the rates of true positives, true negatives, false positives, and false negatives of the `WEAVE-GES-LR` sample. The test set of 10,000 spectra was divided such that 10% were contaminated spectra and 90% were uncontaminated (clean) spectra. The model had a true positive detection rate of 85.42% and false positive detection rate of 2.33%. A decision boundary of 0.7 was used for this plot; the false positive rate can be decreased by raising the decision boundary (i.e. accepting positive predictions only when the model is very confident) at the cost of decreasing the true positive rate. Note: similar results were found with the `WEAVE-GES-HR` sample.

the threshold can be lowered to capture fainter contaminations, or raised for only the most confident predictions for contaminated spectra.

To illustrate the prediction dynamics as the decision boundary is modified, a Precision-Recall analysis was conducted. Mathematically, precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

where TP is the true positives, FP is the false positives, and FN is the false negatives. Precision is a measure of the *quality* of the CNN's predictions, whereas recall is a measure of the fraction of contaminated sources that were identified. Figure 2.4 shows how the precision and recall change as a function of decision threshold for the `WEAVE-GES-HR` and `WEAVE-GES-LR` samples: generally as the decision boundary increases, the precision increases and the recall decreases. The performance on the `WEAVE-GES-HR` is better overall, likely because the increased resolution provides more information to distinguish stellar features from solar features.

In the case of identifying contaminated spectra, it is probably desirable to optimize the decision threshold such that there is a high precision while minimally sacrificing the recall, e.g. the values can be optimized such that the model achieves 96% precision and 67% recall if a decision threshold of 0.96 is chosen: while the model misses 33% of the total contaminated sources with this threshold, the trade-off is that there is trust when it does identify contamination. The precision drops to 80% at the same recall level for the `WEAVE-GES-LR` sample.

Undetected Contamination

To understand how the model was failing at properly identifying all of the contaminated sources, we examined how well the model recovered them depending on their contamination level and stellar parameters.

The recall as a function of contamination gives a more detailed understanding of the completeness of the model and is shown in Figure 2.5 for two CNNs trained separately on the `WEAVE-GES-HR` and `WEAVE-GES-LR` samples. Using a decision threshold of 0.95, the `WEAVE-GES-HR` CNN recovered approximately 20% of contaminated sources at a level of 2-3% contamination, but the recovery fraction increases rapidly to 60-80% above 10% contamination. When lowering the decision threshold to 0.80, the recovery rate approximately doubles

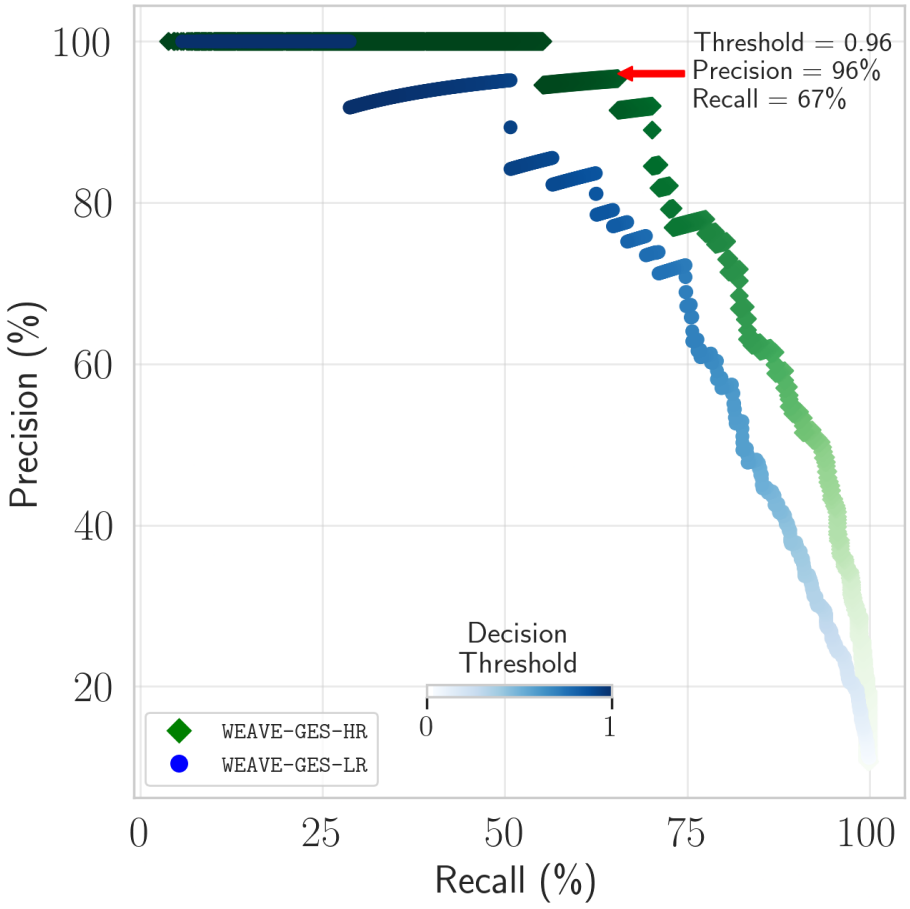

Figure 2.4: A precision-recall plot showing how effectively two CNNs separately trained to identify contaminated `WEAVE-GES-HR` and `WEAVE-GES-LR` spectra were able to identify contaminated spectra depending on the decision threshold chosen. As an example, for the `WEAVE-GES-HR` spectra, the threshold could be chosen such that the precision and recall of the identification procedure are 96% and 67% respectively, i.e. extremely precise yet missing 33% of the true contaminants.

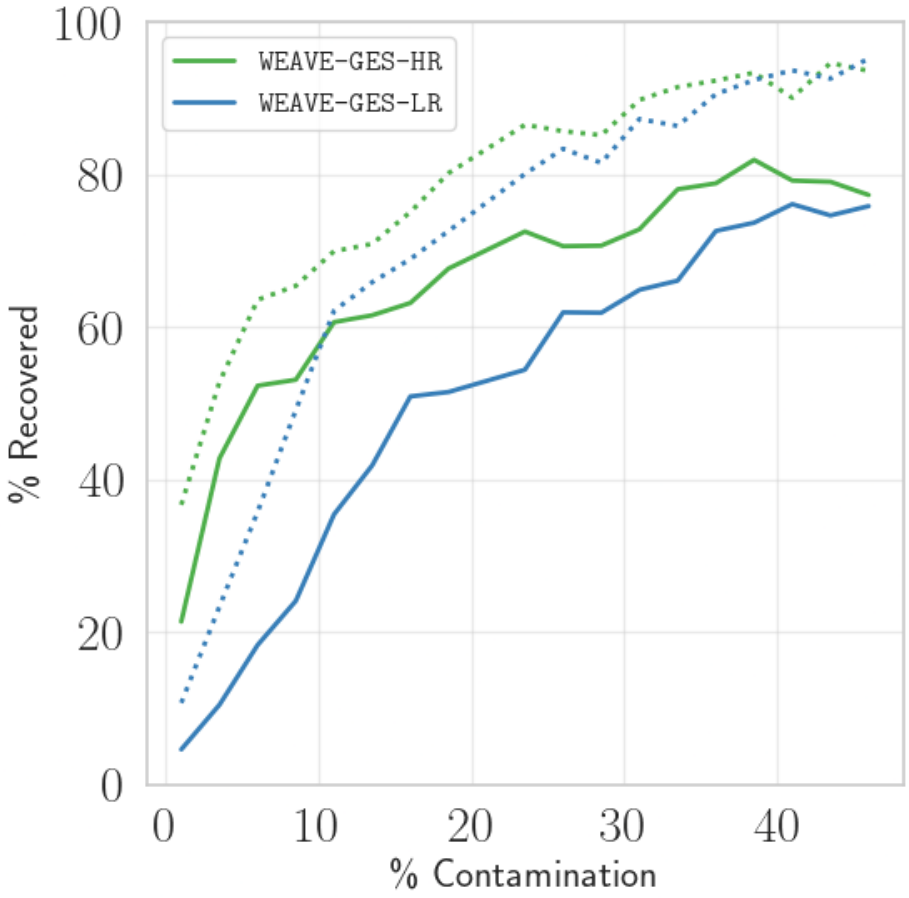

Figure 2.5: Two CNNs were separately trained to identify contaminated `WEAVE-GES-HR` and `WEAVE-GES-LR` spectra and the fraction of sources properly identified as being contaminated (i.e. the recall) was computed in bins of contamination level for decision thresholds of 0.95 (solid line) and 0.80 (dotted line). The majority of false negatives reside in the low (<5%) contamination regime.

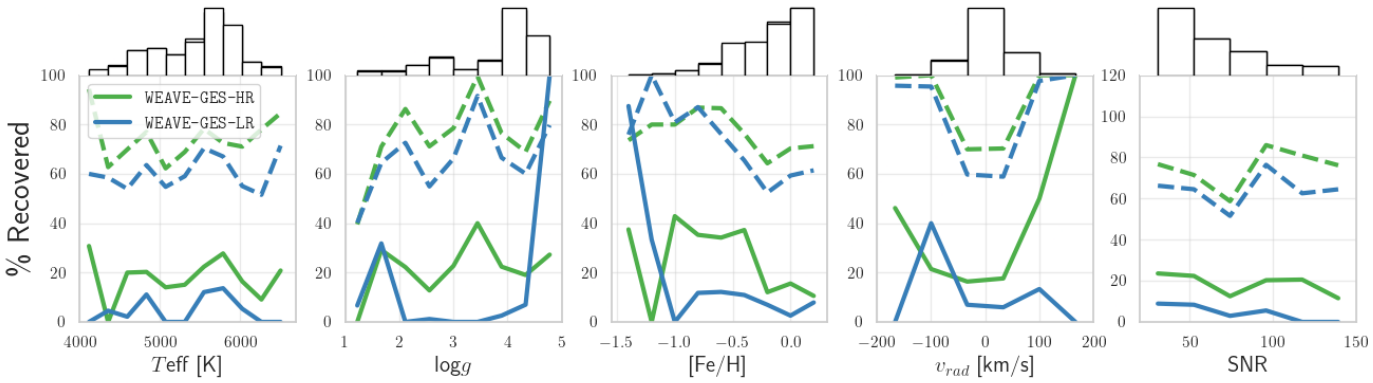

Figure 2.6: The recovery fraction (i.e. recall) of identified contaminated sources as a function of stellar parameter for the `WEAVE-GES-HR` and `WEAVE-GES-LR` samples, split into sources with high contamination (between 5% and 50%; dashed lines) and low contamination ($< 5\%$; solid lines), using a decision threshold of 0.95. Also shown are the histograms of stellar parameters; low number statistics in certain parameter ranges cause spurious features.

in the low contamination regime and rises to 80-90% for higher levels of contamination. The `WEAVE-GES-LR` CNN showed similar trends albeit with a decreased performance overall. The false negatives are concentrated in the low (<5%) contamination regime.

To uncover any potential biases in detection with respect to the stellar parameters, the recall for the `WEAVE-GES-HR` and `WEAVE-GES-LR` CNNs was computed as a function of the parameters $T_{\rm eff}$, log$g$, [Fe/H], $v_{\rm rad}$, and SNR, using a decision threshold of 0.95, as seen in Figure 2.6. One would expect the recovery fraction to decrease around solar values ($T_{\rm eff}$ = 5777 K, log$g$ = 4.44, and [Fe/H] = 0, Mello et al. 2014) since it would be difficult to distinguish a solar spectrum mixed with a solar-like spectrum. There indeed appears to be a local dip at solar log$g$ and a decreasing trend towards solar metallicity. The most obvious trend occurs with radial velocities: the recovery rate increases substantially at higher absolute velocities. Interestingly no clear trends are seen with SNR, perhaps because the SNR of each spectrum is sufficiently high. Lastly, there are several anomalous or sporadic points where e.g. the recovery fraction is zero or nearly zero, but this is likely due to low numbers of samples in those regions (creating problems for both training the model and testing the model with small number statistics).

Taken together, Figures 2.5 and 2.6 imply that (1) a spectrum with solar-like surface gravity and metallicity presents an increased difficulty for the CNN classification scheme, (2) there is no obvious correlation between the recovery rate and effective temperature or SNR, and (3) the most significant factors in detectability are contamination levels and the relative difference between the radial velocity of the stellar and contaminating source.

2.4.2 Stellar Parameters Prediction

The stellar parameters $T_{\rm eff}$, log$g$, and [Fe/H], along with the absolute abundances A(Ca), A(Mg), A(O), A(S), and A(Ti), were predicted with a CNN. For the `WEAVE-GES-LR` sample, two models were trained: one on 35,000 spectra *with* contamination between 0% and 50% (`CNN+`) and one on 35,000 spectra *without* contamination (`CNN-`), and were used in the following tests. Note: the results on `WEAVE-GES-HR` can be seen in the Appendix (Section 2.7).

Removing Contamination First

One method for dealing with the satellite contamination explored in this study was to attempt removing its signature from the observed spectra completely – a process analogous to subtracting off a sky spectrum to isolate the pure stellar spectrum. Once the spectra are

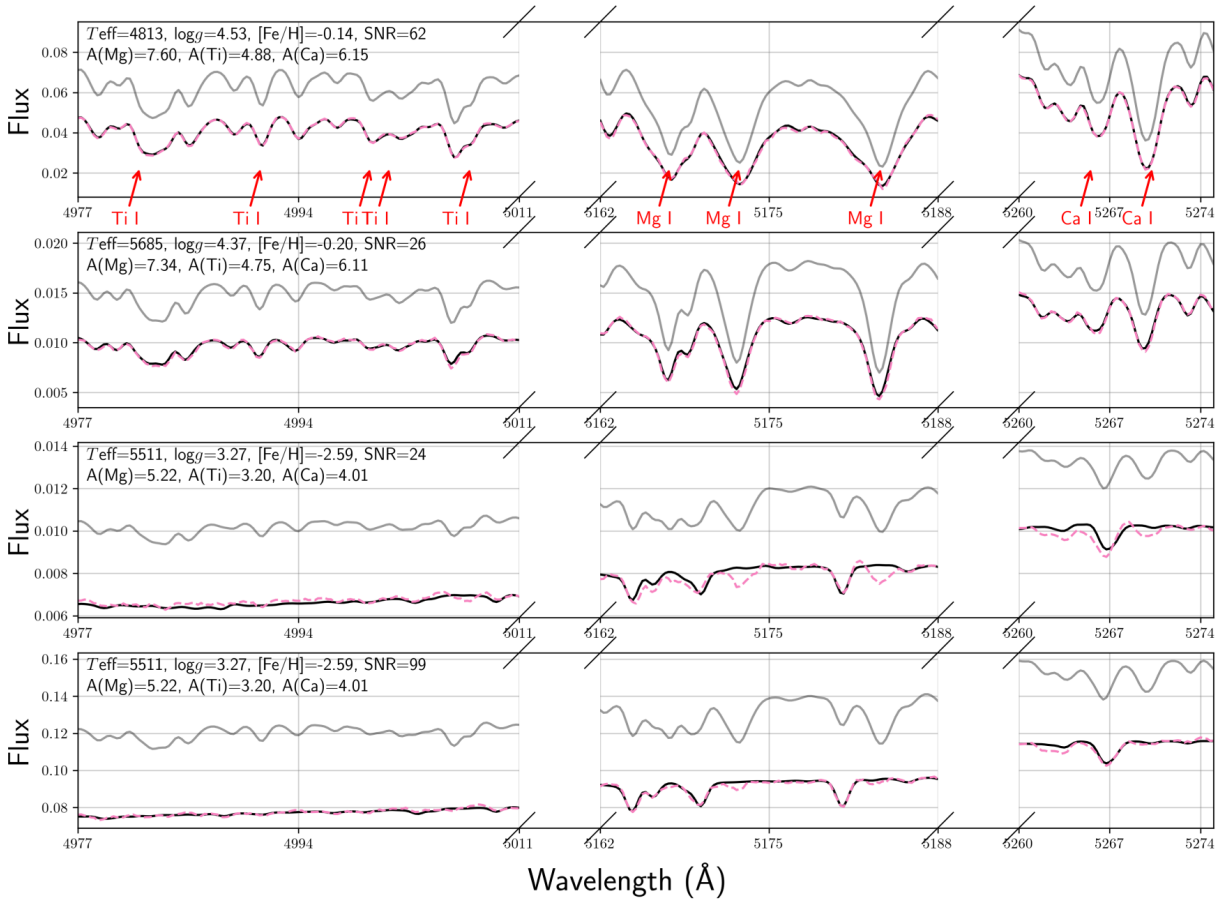

Figure 2.7: A Wave U-Net architecture was tasked with inferring (pink dashed line) the pure stellar spectrum (black solid line) from spectra contaminated by ∼40% with a solar spectrum (grey solid line). Shown here are several examples, in the Mg I *b* triplet region and regions with strong Ti I and Ca I absorption, of the inferred `WEAVE-GES-LR` stellar spectra with their contamination removed. The last two examples are of the same low-metallicity star at different SNR, which highlights the difficulty Wave U-Net has with low-metallicity and low SNR spectra. Aside from that, the predictions closely match the ground truth, though some residuals can be seen in the deeper absorption features.

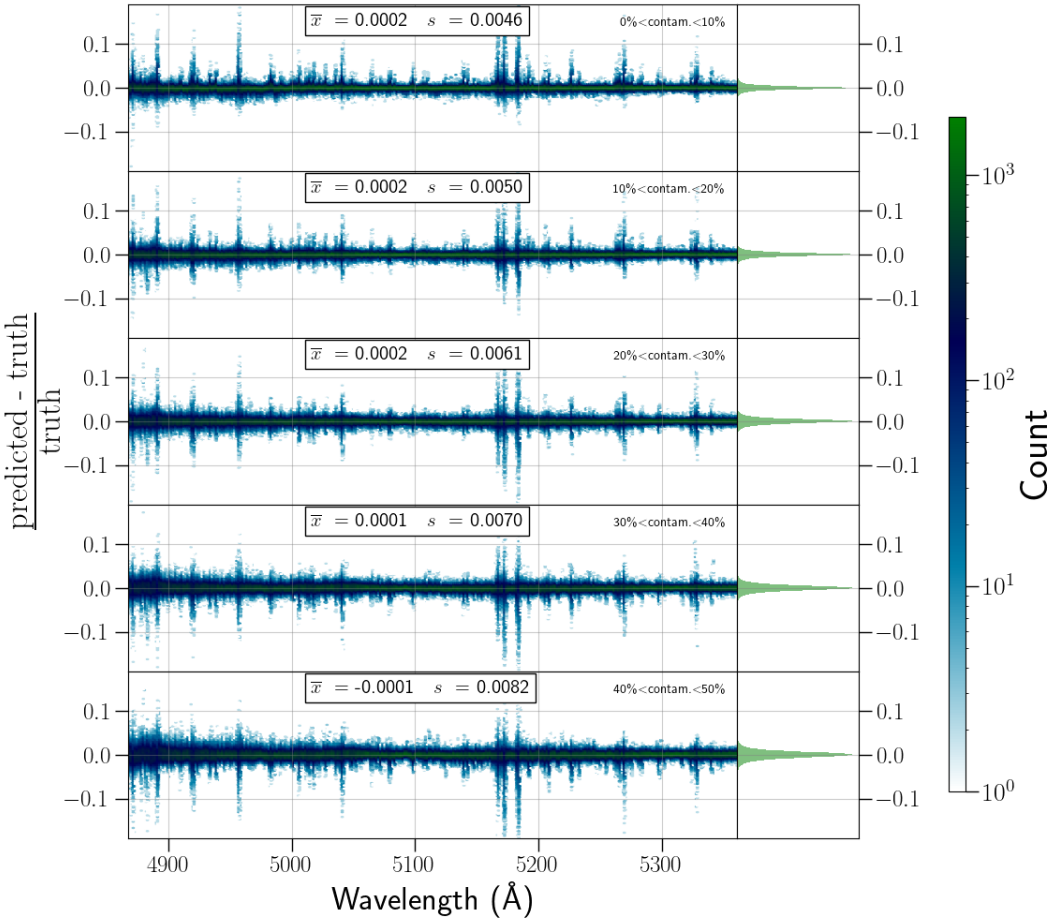

Figure 2.8: Wave U-Net was used to remove solar contamination from the test set of `WEAVE-GES-LR` stellar spectra, and shown here are the distributions of residuals between the predicted spectra and true spectra split into five groups of increasing levels of contamination. The bias, $\overline{x}$, and 1-sigma error, s, were calculated for each group of spectra; the error in all cases is <1% and gradually increases with the level of contamination. The residuals around strong absorption features (e.g. the Mg I b triplet region around 5175 Å) are more severe.

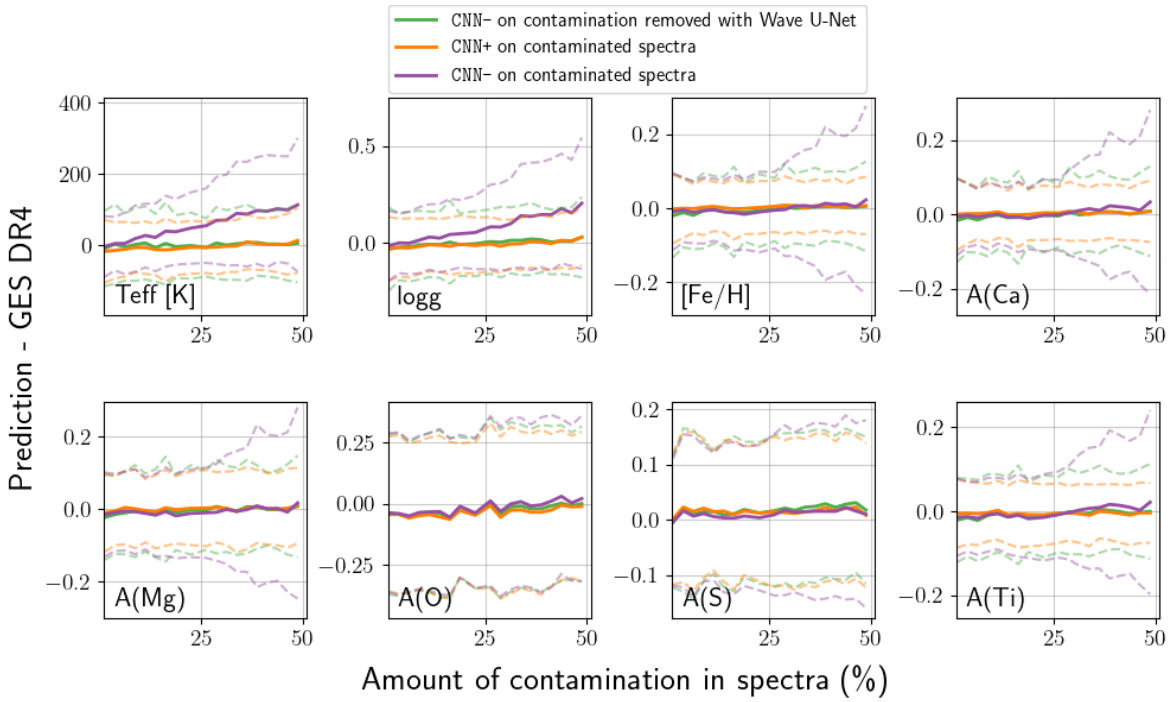

Figure 2.9: The residuals between stellar parameter and abundance predictions and GES catalog values, with the mean (solid lines) and standard deviation (dashed lines) of the residuals binned according to the amount of contamination in the original spectra. Two CNN models were separately trained on `WEAVE-GES-LR` spectra: one with a training set that included contamination (`CNN+`), and one that did not (`CNN-`). Each model was used to predict the stellar parameters *T*eff, logg, [Fe/H], and elemental abundances A(Ca), A(Mg), A(O), A(S), A(Ti) on a test set of `WEAVE-GES-LR` contaminated spectra. Additionally, `CNN-` was used for predictions on a test set of `WEAVE-GES-LR` spectra with their contamination removed by Wave U-Net. Overall better predictions were achieved with the `CNN+`. Predictions on the abundances O and S had large errors in all cases because of the lack of strong absorption features in the limited wavelength range studied.

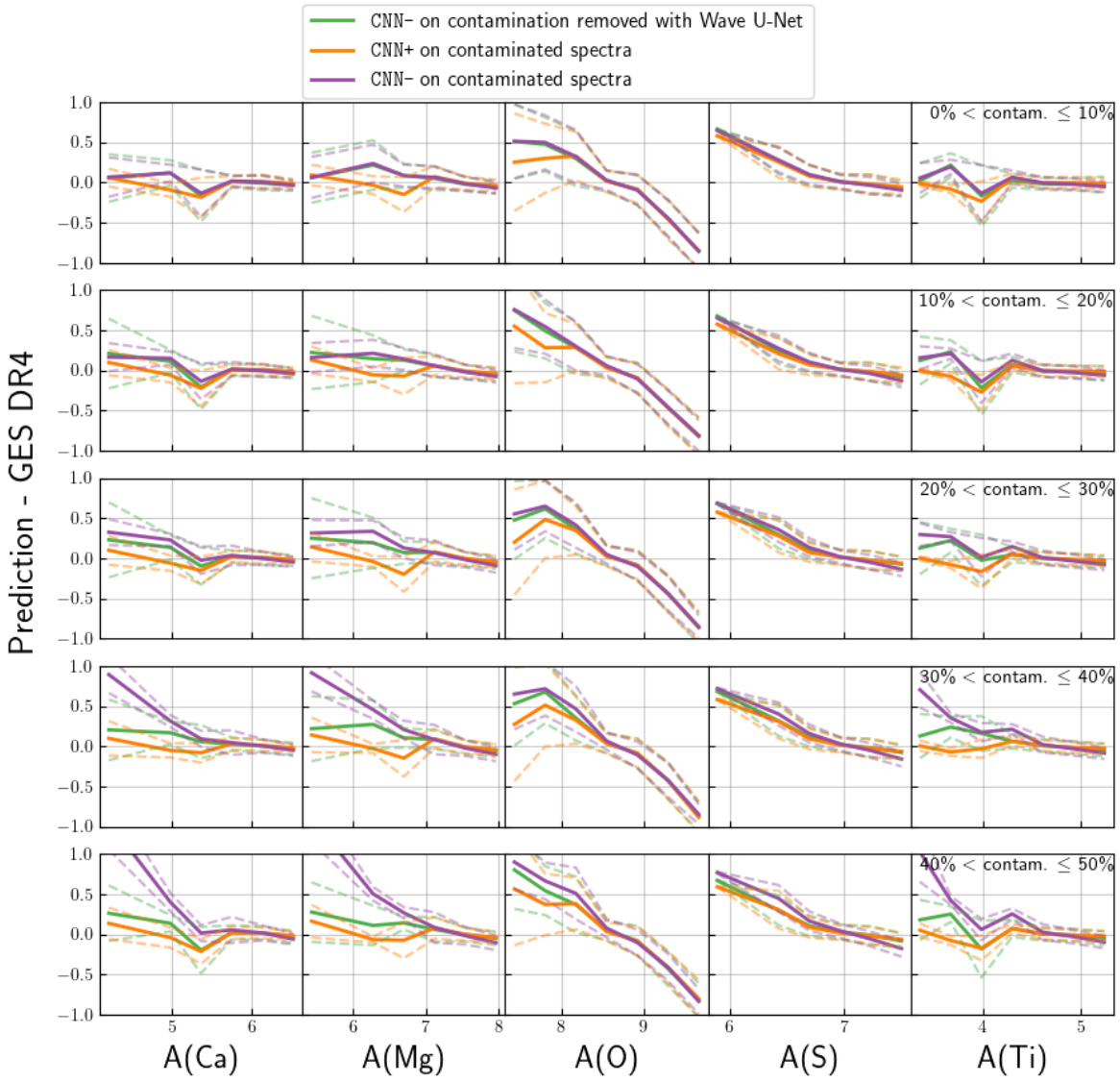

Figure 2.10: The residuals between stellar abundance predictions and GES catalog values. The mean (solid lines) and standard deviation (dashed lines) of the residuals were binned according to the abundance value, and each row corresponds to increasing levels of contamination in steps of 10%. Two CNN models were separately trained on `WEAVE-GES-LR` spectra: one with a training set that included contamination (`CNN+`), and one that did not (`CNN-`). Each model was used to predict the stellar elemental abundances A(Ca), A(Mg), A(O), A(S), A(Ti) on a test set of `WEAVE-GES-LR` contaminated spectra. Additionally, `CNN-` was used for predictions on a test set of `WEAVE-GES-LR` spectra with their contamination removed by Wave U-Net. At higher abundance values and lower contamination levels, there is almost no difference between the three models. The differences become pronounced at lower abundance values and higher contamination levels, where `CNN+` can be seen to perform best.

cleaned of their contamination, they can be sent to any downstream pipeline to estimate stellar parameters. As described in Section 2.3.2, the Wave U-Net architecture was used for this task.

Figure 2.7 shows several examples of Wave U-Net recovering `WEAVE-GES-LR` stellar spectra in regions with strong Mg, Ti, and Ca features with ∼40% solar contamination: in most cases, the contamination was almost completely removed and the stellar spectra were recovered but with some more obvious errors around the deeper absorption features, especially for stars with low metallicity and low SNR. To see the overall performance of Wave U-Net the relative residuals of its predictions on the entire test set of `WEAVE-GES-LR` spectra were calculated, and the distributions are shown in Figure 2.8. The 1-sigma errors were under 1% across the entire test set, with a minimum of 0.46% for the spectra with 0-10% contamination, though it was confirmed that the model has larger residuals in the deeper absorption features.

With Wave U-Net operational, we obtained a test set of `WEAVE-GES-LR` spectra with their contamination removed to test the performance of `CNN-`. Figure 2.9 shows the mean and standard deviation of residuals of `CNN-`'s stellar parameter predictions on the test set of `WEAVE-GES-LR` contaminated spectra before and after being processed by Wave U-Net. The residuals on the contaminated spectra are nearly identical to those on the Wave U-Net spectra in the low contamination regime, ∼2-3x higher in the high contamination regime, and significantly increase for the abundances around the 25% contamination level. The residuals on the Wave U-Net spectra are relatively constant across most contamination levels, though a slight increase can be seen at the highest levels.

Training a CNN on Contaminated Spectra

Instead of identifying and throwing away a contaminated spectrum, or attempting to remove its contamination, one might be interested in extracting accurate information *despite* the contamination. One way of accomplishing this task is again to use ML.

A major benefit of the feed-forward CNN architecture used in this study is that the spectra can be augmented with features (e.g. solar contamination) to be identified or, in this case, *ignored*. By feeding artificially contaminated spectra through the NN and giving it the task of predicting stellar parameters of *un*contaminated spectra, the NN learns to be robust to contamination in its predictions.

To demonstrate the ability of a CNN to predict through contamination, the `CNN+` model was given a test set of contaminated spectra, and Figure 2.9 shows the residuals of its predictions on the `WEAVE-GES-LR` spectra. The `CNN+` residuals are constant and overall lower

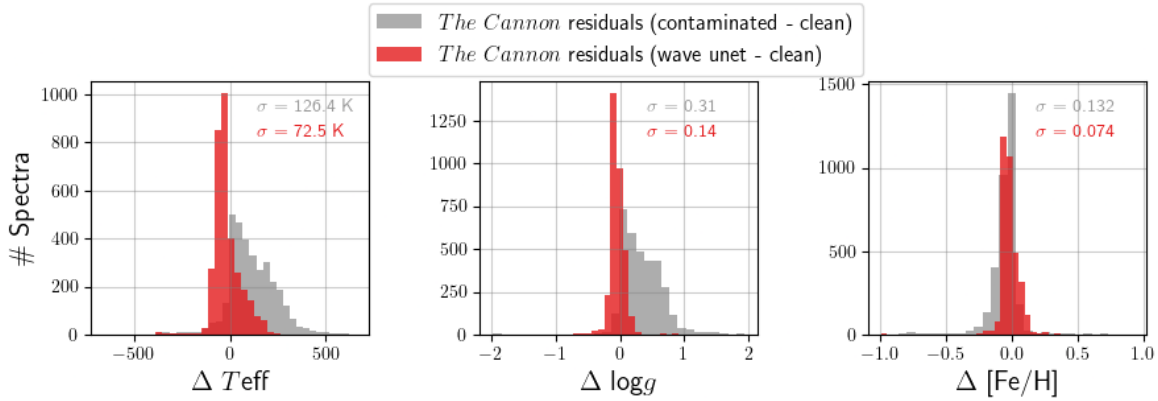

Figure 2.11: *The Cannon* was used to predict the stellar parameters Teff, log$g$, and [Fe/H] on three test sets of `WEAVE-GES-LR` spectra: contaminated (by <50%) spectra, clean spectra, and spectra with contamination removed by Wave U-Net. The residuals between the predictions on the contaminated spectra and clean spectra, as well as the residuals between the predictions on Wave U-Net spectra and clean spectra, were computed. The histograms show the improvements in stellar parameter estimates when Wave U-Net is used to remove the contamination, confirming that Wave U-Net can be used in conjunction with other stellar parameter estimation pipelines.

than `CNN-`'s residuals on both the Wave U-Net spectra and contaminated spectra over the full range of contamination and across all parameters. Even with 0-2% contamination levels, the residuals for $T_{\rm eff}$ and log$g$ are slightly larger in `CNN-` than for `CNN+`. Note that in the wavelength range studied, there are limited oxygen and sulfur features and thus the residuals remain large and somewhat constant for both `CNN+` and `CNN-`.

Figure 2.10 shows a similar comparison but the residuals were instead binned according to the stellar abundances, and split into 5 groups of increasing contamination. At low (0-10%) contamination levels, the results are nearly identical. At lower abundance levels, it can be seen that Wave U-Net has difficulty with properly recovering the `WEAVE-GES-LR` spectra even at lower (10-20%) contamination levels, leading to worse precision than even the `CNN-` on contaminated spectra. The performance of the `CNN+` becomes especially apparent at higher contamination levels and low abundance levels, whereas all three models perform equally well at higher abundance levels.

Removing Contamination for Downstream Tasks

To show the utility of removing contamination outside the scope of neural networks and our pipeline, *The Cannon* (Ness et al., 2015) was used to predict the stellar parameters $T_{\rm eff}$, log$g$, and [Fe/H] on contaminated, clean, and Wave U-Net `WEAVE-GES-LR` spectra. Figure

2.11 shows the substantial reduction in prediction residuals – the standard deviation reduces by a factor of ∼2 – when the contaminated spectra are cleaned by Wave U-Net first.

2.5 Discussion

2.5.1 Limitations

Low Levels of Contamination

The results from our study show that the majority of low-contamination (<5% contamination) spectra are not detected with the presented framework. Since low contamination will likely be the most prevalent level of contamination (especially in the `WEAVE-HR` spectra), this could mean that many contaminated spectra would pass through our current pipeline undetected. This creates a problem if the goal is simply to flag the contaminated spectra for removal or further processing, but there isn't necessarily a problem for stellar parameter and abundance predictions. Indeed, Figures 2.9 and 2.14 show that the errors on abundance predictions due to satellite contamination are insignificant compared to typical spectroscopic uncertainties (e.g., Recio-Blanco et al., 2023a; Accetta et al., 2022; Steinmetz et al., 2020; Aguado et al., 2019; Smiljanic et al., 2016) when contamination is $\lesssim$20% for `WEAVE-GES-LR` and $\lesssim$10% for `WEAVE-GES-HR`.

For temperature and surface gravity predictions, however, Figures 2.9 and 2.14 show that even low levels of contamination can introduce noticeable errors. When Wave U-Net is used to remove the low-level contamination, the resulting predictions are at best marginally better and sometimes marginally worse than those on contaminated spectra; removing very low levels of contamination is difficult for a Wave U-Net. By including contamination as an augmentation to a training set instead of attempting to remove it, such that a CNN learns to bypass (or ignore) the contamination when making stellar parameter predictions, the errors on temperature and surface gravity estimates are reduced. Contamination as an augmentation therefore appears to be the best strategy for dealing with spectra containing low (and high) levels of contamination, though Wave U-Net is only marginally worse.

Realistic Training Set

In this study it was assumed the spectrum reflected by the satellites would be an unmodified solar spectrum, but in reality the satellites from each company will have different albedos and thus reflect a slightly modified solar spectrum. The results in this paper should therefore be interpreted as a best-case scenario of knowing the characteristics of the contamination

completely. There are two options to include more realistic contamination information in the pipeline: (1) use the published materials, if available, from each company regarding the reflection properties of their satellites and include these as a wavelength-dependent data augmentation in the training set, and/or (2) the telescope could intentionally observe stars that either have previously had confirmed satellite contamination in their spectra or are predicted to have a satellite crossing their path, to collect real examples for the training set. Ideally both options would be pursued to curate a realistic training set, however it is not possible – with current data that each company supplies – to predict satellite positions with enough accuracy to know whether they have contaminated or will contaminate a star spectrum at a given time in a given place. We do expect that there will be some studies (both from industry and from astronomers) aimed at characterizing the reflectivity of various classes of satellites (e.g., Žilková et al., 2023).

Solar Twins

A star with solar-like stellar parameters presents an added difficulty for a NN since the contaminating and true spectra are similar and thus harder to disentangle. This was indeed the case for the identification task, where the CNN had a lower recall around solar-like surface gravity and metallicity. For the task of predicting stellar parameters and abundances, however, `CNN+` can help overcome this limitation. A set of five solar twins – defined as having $T_{\rm eff}$ within 100 K, log$g$ within 0.1 dex, and [Fe/H] within 0.04 dex of the solar values (a slightly narrower definition than used by Ramírez, Melendez, and Asplund 2009) – was used to check if their predicted parameters were significantly affected by contamination. Figure 2.12 shows that the distributions of the stellar parameters and abundances for the solar twins are compatible with the overall distributions from the test set with two exceptions: (1) a consistent and significant over-prediction of 0.1 dex for the magnesium abundances, and (2) an insignificant under-prediction of $<$100 K for temperature. We suggest that even low levels of satellite contamination can impact magnesium abundance predictions in solar twins. With a sample size of five, however, the biases we see could also come from an under-representation of solar twins in the test set; a larger test set is required to make statistically significant conclusions.

2.5.2 Further Applications

Another way to view the results in this paper is the application of a CNN to bright sky subtraction. Typically high resolution spectroscopy is carried out at observatories during

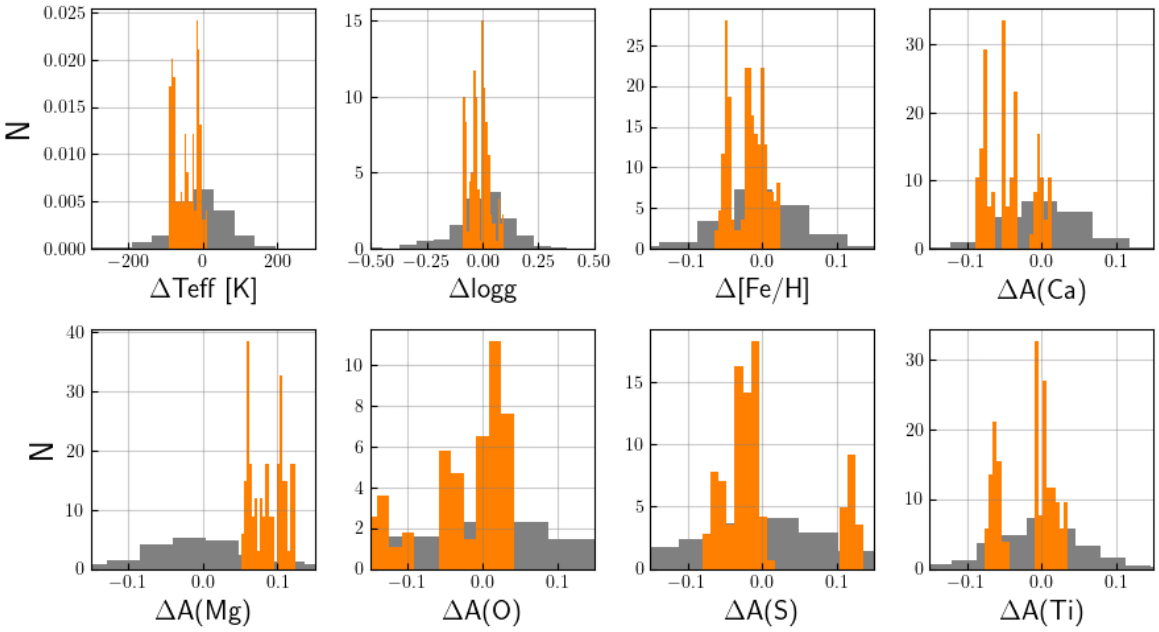

Figure 2.12: A set of `WEAVE-GES-LR` solar twins (defined as having $T_{\rm eff}$ within 100 K, log$g$ within 0.1 dex, and [Fe/H] within 0.04 dex of the solar values) was collected and `CNN+` was used to predict their stellar parameters and abundances and compare to GES catalog values. Shown here are the normalized distributions of residuals (predicted minus catalog values) on the solar twins (orange) and the rest of the test set (grey). The residuals are mostly compatible with the overall distribution, but there is a consistent over-prediction of magnesium abundance and a small under-prediction of temperature. Note: only five solar twins were found in the test set, each with multiple instances of contamination and all with $|v_{\rm rad}| < 15$km/s.

“bright time", when the face of the Moon is $> 50\%$ illuminated, leading to stellar spectra with higher levels of solar contamination. There is nothing fundamentally different about a satellite reflecting solar light and the moon reflecting solar light (aside from wavelength-dependent reflectance properties) so the methods outlined in this paper could be directly applied to bright time stellar spectra.

A similar method could also be used to detect whether a measured spectrum contains light from a binary stellar system. A training data set could be created in which stellar spectra are "contaminated" with a library of artificial binary counterparts; a CNN could be trained to detect the binary systems while a U-Net could be trained to separate the two spectra. One would need to carefully consider realistic values for the relative differences in radial velocity, temperature, surface gravity, abundances, and brightness between the artificially created binary stellar spectra.

2.6 Conclusion

The detection and mitigation of contaminated spectra are important tasks in astronomical research, as contamination can lead to inaccurate conclusions and bias in data analysis. In this study, we applied a convolutional neural network (CNN) for both detection and stellar parameter prediction (specifically stellar parameters $T_{\rm eff}$, $\log g$, and [Fe/H], and the abundances A(Ca), A(Mg), A(O), A(S), and A(Ti)), and a Wave U-Net for removal of satellite contamination in a WEAVE-like spectral survey. We determined the limitations and feasibility of a machine learning-based pipeline for these tasks.

In a test set of 10,000 artificially contaminated `WEAVE-GES-HR` spectra, the decision threshold of a CNN was adjusted such that a detection precision of 96% and a detection recall of 67% were achieved, which means that while the model misses 33% of the total contaminated sources (mostly those with $<10\%$ contamination), it can confidently identify the contamination it does detect. The results on `WEAVE-GES-LR` spectra were slightly poorer, with a precision of 80% at the same recall level. It was also found that a contaminated spectrum is harder to identify if it has solar-like surface gravity, solar-like metallicity, or low radial velocity.

A Wave U-Net model was trained to recover clean spectra from contaminated spectra and was able to recover them with 1-σ errors of 0.46% in the low ($< 10\%$) contamination regime and $<1\%$ in the high (40-50%) contamination regime. Two CNNs were separately trained on contaminated (`CNN+`) and clean (`CNN-`) spectra, and it was shown that the accuracy in predictions, when compared to `CNN-` predictions on contaminated spectra, was improved by

up to a factor of 2-3 when using `CNN-` on the Wave U-Net spectra and when using `CNN+` on contaminated spectra, with overall better results from the `CNN+`. Including an expected contamination as an artificially augmented training set therefore appears to be the best strategy for mitigating the effects of contamination in our pipeline.

Acknowledgements

We thank Federico Sestito, John Pazder, and Ted Grosson for the helpful comments and discussions about bright time sky subtraction, satellite coatings, and LSST satellite contamination. We also thank the anonymous referee for comments which helped strengthen and clarify our results. SB and KAV thank the Natural Sciences and Engineering Research Council for funding through the Discovery Grants program and the CREATE program in New Technologies for Canadian Observatories. SL acknowledges the support by PRIN INAF 2019 grant ObFu 1.05.01.85.14 ("Building up the halo: chemo-dynamical tagging in the age of large surveys", PI. S. Lucatello).

Software: `astropy` (Robitaille et al., 2013), `matplotlib` (Hunter, 2007), `numpy` (Harris et al., 2020), `pytorch` (Paszke et al., 2019c), `scipy` (Virtanen et al., 2020), `seaborn` (Waskom, 2021).

Code Availability

All code for this project can be found at the project's Github repository: https://github.com/astroai/StarUnLink.

Data Availability

All data underlying this article will be shared on reasonable request to the corresponding author.

2.7 Appendix

2.7.1 Results on high-resolution spectra

The stellar parameters $T_{\rm eff}$, log$g$, and [Fe/H], along with the absolute abundances A(Ca), A(Mg), A(O), A(S), and A(Ti), were predicted with a CNN. For the `WEAVE-GES-HR` sample, two models were trained: one on 35,000 spectra *with* contamination between 0% and 50% (`CNN+`) and one on 35,000 spectra *without* contamination (`CNN-`). Additionally, a Wave U-Net was used to remove the contamination from the `WEAVE-GES-HR` sample. Figure 2.13 shows the residuals in the reconstructed Wave U-Net spectra, and Figure 2.14 shows the residuals in the stellar parameter and abundance predictions of `CNN+` and `CNN-`.

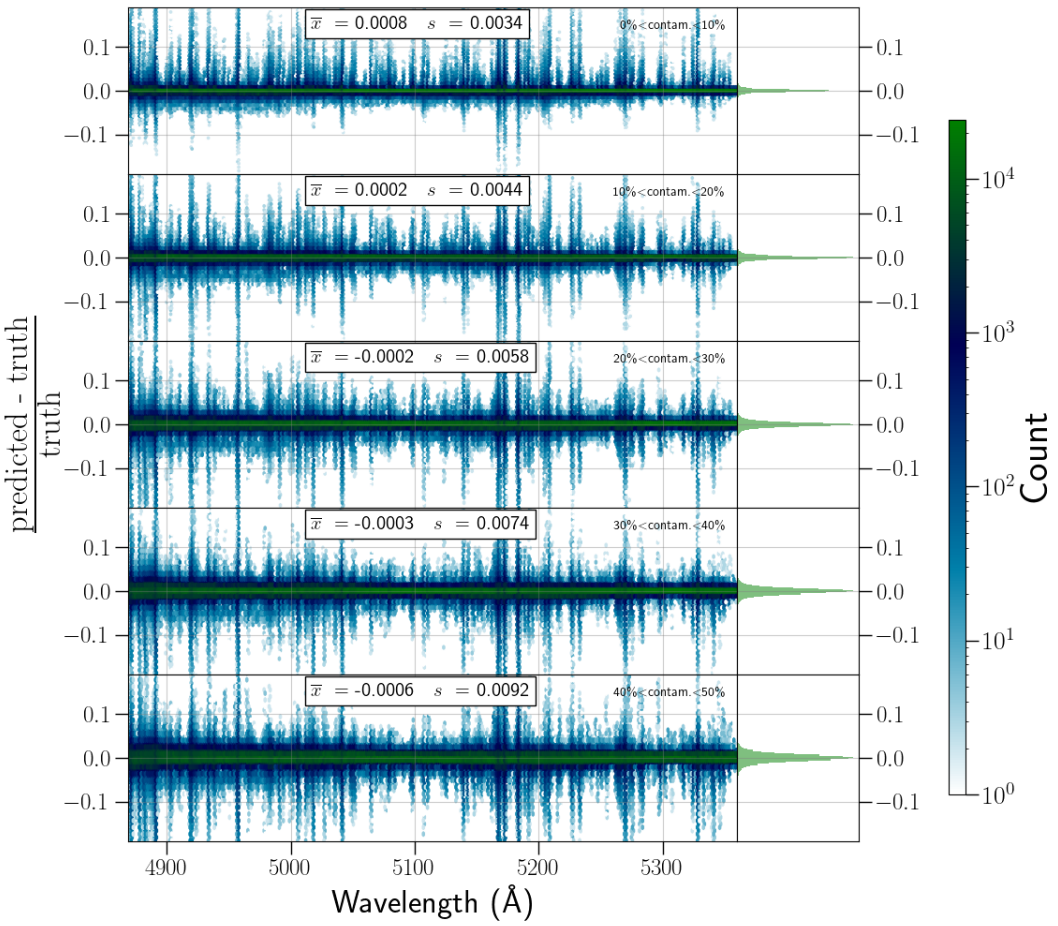

Figure 2.13: Wave U-Net was used to remove solar contamination from the test set of `WEAVE-GES-HR` stellar spectra, and shown here are the distributions of residuals between the predicted spectra and true spectra split into five groups of increasing levels of contamination. The bias, $\overline{x}$, and 1-sigma error, s, were calculated for each group of spectra; the error in all cases is <1% and gradually increases with the level of contamination. The residuals around strong absorption features (e.g. the Mg I b triplet region around 5175 Å) are more severe.

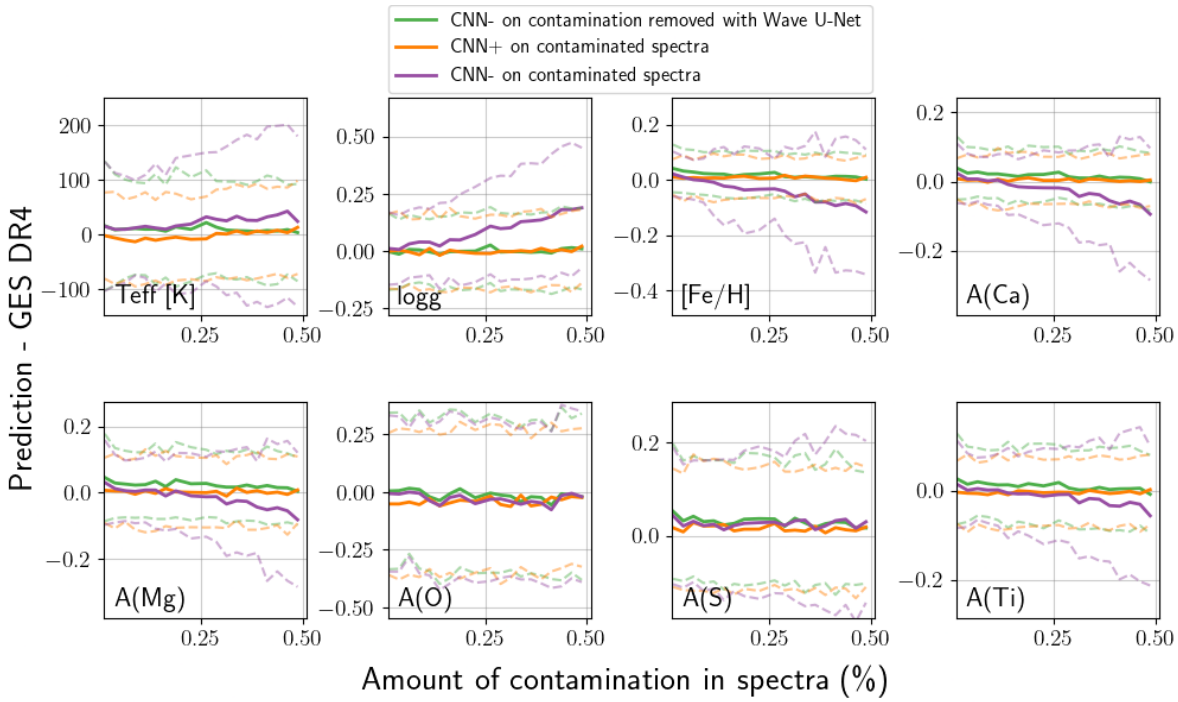

Figure 2.14: The residuals between stellar parameter and abundance predictions and GES catalog values, with the mean (solid lines) and standard deviation (dashed lines) of the residuals binned according to the amount of contamination in the original spectra. Two CNN models were separately trained on `WEAVE-GES-HR` spectra: one with a training set that included contamination (`CNN+`), and one that did not (`CNN-`). Each model was used to predict the stellar parameters *T*eff, logg, [Fe/H], and elemental abundances A(Ca), A(Mg), A(O), A(S), A(Ti) on a test set of `WEAVE-GES-HR` contaminated spectra. Additionally, `CNN-` was used for predictions on a test set of `WEAVE-GES-HR` spectra with their contamination removed by Wave U-Net. Overall better predictions were achieved with the `CNN+`. Predictions on the abundances O and S had large errors in all cases because of the lack of strong absorption features in the limited wavelength range studied.

Chapter 3

DanceCam: atmospheric turbulence mitigation in wide-field astronomical images with short-exposure video streams

This chapter contains the nearly completed draft of a paper – with an expanded introduction for this dissertation – that will be submitted to the Monthly Notices of the Royal Astronomical Society by January, 2024. Authors: Spencer Bialek, Emmanuel Bertin, Sébastien Fabbro, Hervé Bouy, Jean-Pierre Rivet, Olivier Lai.

Personal Contributions

I was one of the main contributors on this project, helping to strengthen the simulation pipeline and develop most of the training and testing framework. I wrote the first draft of this paper.

Abstract

We introduce a novel technique to mitigate the adverse effects of atmospheric turbulence on astronomical imaging. Utilizing a video-to-image neural network trained on simulated data, our method processes a sliding sequence of short-exposure ($\sim$0.2s) stellar field images to reconstruct an image devoid of both turbulence and noise. We demonstrate the method with simulated and observed stellar fields, and show that the brief exposure sequence allows the network to accurately associate speckles to their originating stars and effectively disentangle light from adjacent sources across a range of seeing conditions, all while preserving flux to a lower signal-to-noise than an average stack. This approach results in a marked improvement in angular resolution without compromising the astrometric stability of the final image.

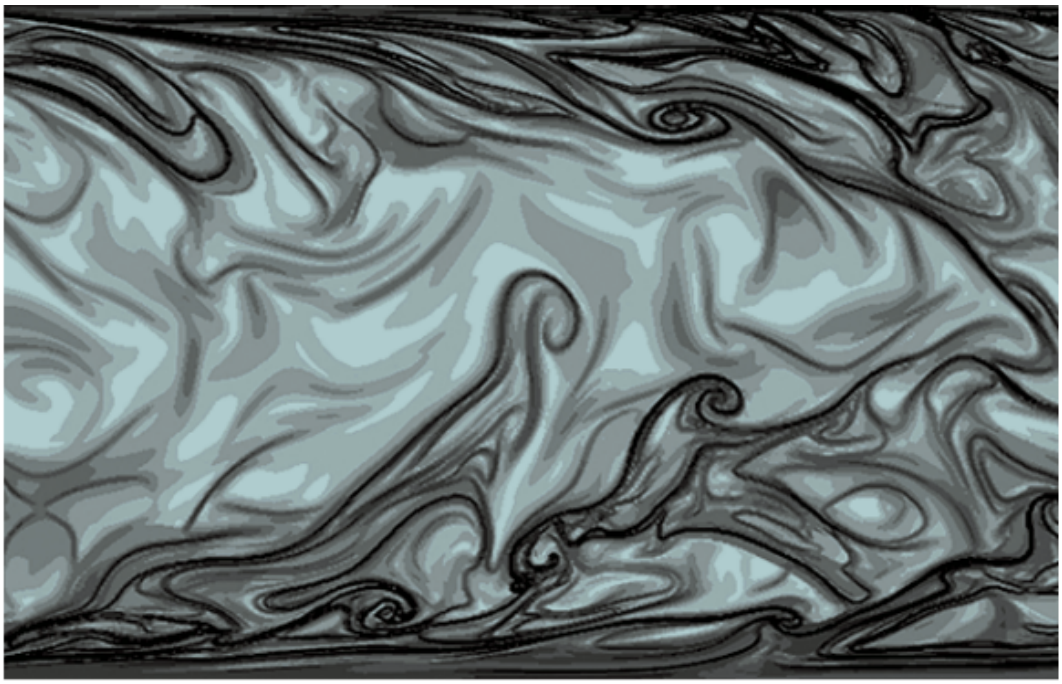

Figure 3.1: Example of turbulence occurring between a layer of air and its boundaries (Smits and Marusic, 2013)

3.1 Introduction

The Earth's atmosphere, a dynamic and complex system, poses significant challenges for astronomical observations. Influenced by factors such as the Earth's rotation and the presence of geographical features like mountains and oceans, the atmosphere is in perpetual motion. This motion varies across different atmospheric layers, leading to intricate interactions at their boundaries. Atmospheric turbulence arises from a combination of these dynamics and other specific causes:

1. *Temperature gradients*: Temperature gradients in the atmosphere causes the air to expand and contract, leading to fluctuations in the density and refractive properties of the air.

2. *Wind shear*: Differences in wind speed and direction at different altitudes can cause rapidly fluctuating turbulence.

3. *Water vapour*: Light can be absorbed or scattered by water vapor in the atmosphere, leading to additional distortion of the light.

4. *Solar heating*: Heating of the atmosphere by the sun can cause thermal currents which

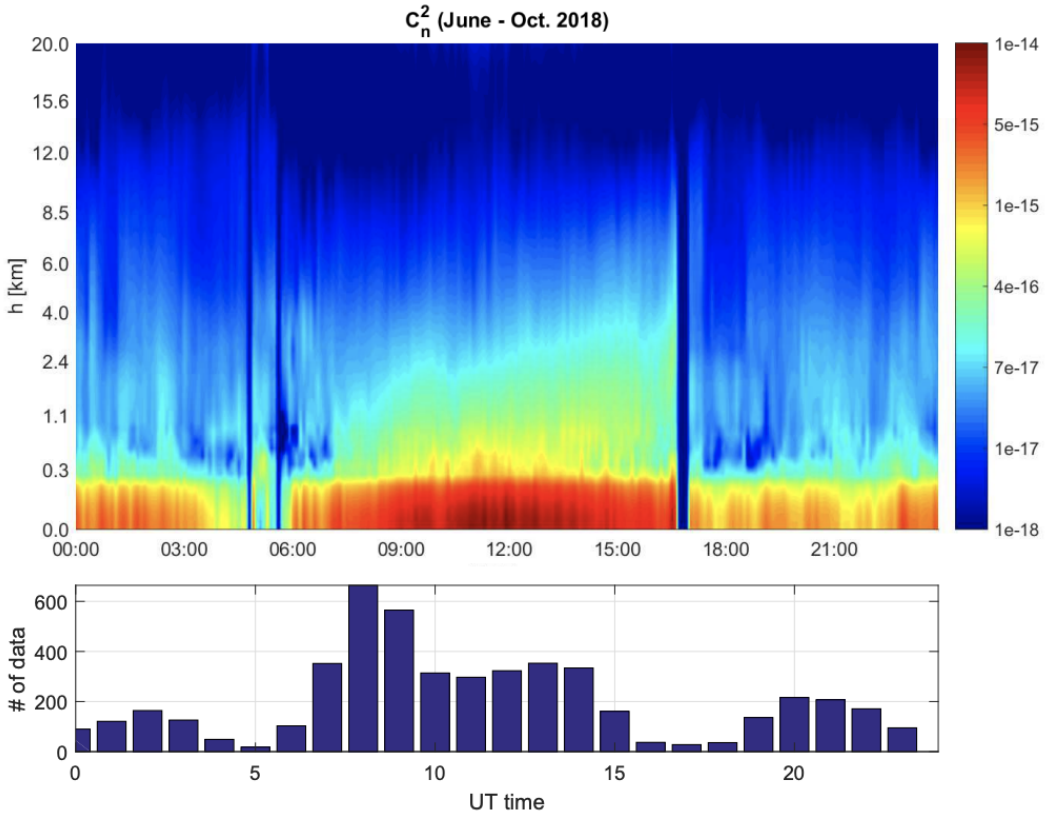

Figure 3.2: The turbulence profile measured at the Plateau de Calern in France in bins of altitude and time, where C_n^2 is a measure of the strength of turbulence, from Aristidi et al. (2020).

leads to laminar flows breaking into large cells that roll over into themselves to produce eddies (see Figure 3.1).

These factors, when combined, result in a highly non-linear and rapidly changing refractive index distribution in the atmosphere. Indeed, the severity of turbulence and its causes can fluctuate over both short and long timescales; Figure 3.2 illustrates a time-series of turbulence strength as a function of altitude at the Plateau de Calern in France. Notably, turbulence is primarily concentrated in lower altitude regions, where solar heating plays a significant role (especially during daylight hours), though even at altitudes of 20 km there can be a non-negligible amount of turbulence.

As starlight traverses this turbulent medium, its wavefront becomes dynamically distorted, causing the familiar twinkle of stars. Short exposure images reveal the starlight to be composed of “speckles" with randomly varying shapes and positions around a central point, ultimately degrading the quality of long-exposure images by modifying the point source of a

star to be spread out into a broader Gaussian shape described by the point spread function (PSF). The result is a significant degradation in the spatial resolution of astronomical images. Fine details, such as the intricate structures of distant galaxies, become harder to discern. Additionally, atmospheric turbulence can introduce errors in measurements of a star's position, brightness, and colour, posing a major challenge to obtaining precise and high-quality astronomical data (Roddier, 1981). Indeed, turbulence has profound implications for various branches of astrophysics and observational astronomy. The challenges introduced by turbulence are not merely technical but have direct consequences on our ability to decipher the mysteries of the universe. I highlight a few impacted areas here, along with some of the research that has employed turbulence-mitigating technology to advance our understanding (note: adaptive optics is one such technology I mention, which I go into much greater detail in the next section):

Galaxy morphologies and their role in cosmic evolution: Galaxy morphologies, or the shapes and structures of galaxies, offer a wealth of information about the universe's past, present, and future. These morphologies provide crucial insights into the processes of galaxy formation, their interactions with other galaxies, and their evolutionary histories, and often need high-resolution imaging to understand in detail. For example, Rouan et al. (2007) used deep K-band adaptive optics high-resolution ($\sim$0.1 arcsec) imaging to explore the morphologies of galaxies at redshift $z \sim 1$, while Laag et al. (2006) observed previously unidentified far-infrared galaxies at $z < 0.3$ with unprecedented clarity, revealing that a significant fraction of these galaxies are involved in tidal interactions or are clear mergers. Kubo et al. (2017) studied the diverse morphologies – such as compact ellipticals and massive star-forming galaxies – observed in very distant ($z = 3.09$) extremely dense clusters of galaxies, which have implications for our understanding of early-universe dynamics, including the beginning of morphology-density relationships and massive galaxy formation. High-resolution imaging was critical for the accurate classification and study of these galaxy morphologies and phenomena.

Furthermore, Tacchella et al. (2015) utilized Hubble Space Telescope imaging, in addition to integral field spectroscopy from the Very Large Telescope using adaptive optics, to emphasize the importance of high spatial resolution for both imaging and spectroscopy in observing galaxies on the star-forming main sequence at $z \sim 2$. Their study highlighted the significance of resolving ongoing star formation and ionized gas kinematics on scales of 1–2 kpc – with a pixel scale of 0.05 arcsec and a full width at half maximum (FWHM) of 0.10-0.17 arcsec – which is essential for understanding the internal dynamics and star formation

processes within these galaxies.

As we strive to understand the early and large-scale universe, the need for high-resolution imaging in studying galaxy morphologies becomes increasingly evident. Such imaging not only allows for a detailed examination of individual galaxies but also provides a broader perspective on galaxy interactions and evolutions.

Exoplanetary studies and the quest for alien worlds: Directly imaging exoplanets presents significant challenges due to their faintness and close proximity to their luminous host stars. Atmospheric turbulence further complicates these observations, adding noise and reducing the clarity of potential planetary signals. High-resolution imaging, which mitigates the effects of the atmosphere, is crucial for studying these worlds. For example, results from Dekany et al. (2013) demonstrated the necessity of high-contrast, high-resolution imaging to resolve the circumstellar disks involved in planet formation. Jovanovic et al. (2015) designed the Subaru Coronagraphic Extreme Adaptive Optics instrument for the discovery and detailed characterization of exoplanetary systems, making use of light in the bandpass 600-2500 nm to allow for direct exoplanet imaging and spectroscopy close to their host stars by achieving incredible angular resolutions of $\sim$0.009 arcsec. The Robo-AO Kepler Planetary Candidate Survey (Baranec et al., 2016; Ziegler et al., 2017) utilized high-resolution ground-based imaging to observe every Kepler planet candidate host star, aiming to identify blended nearby stars that might be responsible for false-positive exoplanet detections.

High-resolution imaging is paramount not only for confirming the presence of exoplanets but also for delving deeper into their atmospheres, potentially even detecting signs of life or habitability. Yes, we must mitigate the influence of *our* atmosphere to study *their* atmospheres.

Microlensing and dark matter: Microlensing events of distant stars can be helpful for understanding the nature of dark matter and the mass distribution in the Galaxy and universe (Paczynski, 1986; Griest et al., 1991). These phenomena arise when a massive object, such as a star or planet (or dark matter object), magnifies the light of a further background star through gravitational lensing. It therefore helps to observe dense clusters of stars, as your chances of detecting microlensing events increases. Two common directions to look for microlensing events, owing to their density of stars, are towards the Galactic bulge and towards the Magellanic clouds. Indeed, Pietrukowicz et al. (2005) observed a *probable* microlens event in the globular cluster M22 (with the Galactic bulge as a back-drop), but it wasn't until higher-resolution imaging, which mitigated the atmospheric turbulence, became

available that the lens and source stars were characterized and it was finally confirmed as the first detection of a microlensing event in a globular cluster (Pietrukowicz et al., 2011). Similarly, Lu et al. (2016) used follow-up high-resolution imaging to constrain the masses of lenses detected in the Optical Gravitational Lensing Experiment (OGLE, Udalski et al., 1992). Surveys like OGLE depend on wide fields and medium- to high-cadence imaging to monitor a large number of stars and gather numerous lens candidates, so it would be beneficial to equip a monitoring telescope with the capacity to correct for atmospheric turbulence over wide fields (the topic of this chapter) to increase the image quality, find fainter sources, and better discern the properties of sources. The potential of turbulence mitigation in resolving the lens and source of microlensing events prompted Henderson et al. (2014) to curate a list of microlensing events toward the Galactic bulge that were suitable for constraining the lens mass with future high-resolution imaging.

Faint objects and faint transients in the solar system: High spatial resolution imaging has been instrumental in the detection and understanding of faint objects within our solar system, such as asteroids, comets, and moons. These celestial bodies, often elusive due to their faintness or small size, hold valuable clues about the early solar system, its formation, and its dynamical evolution (DeMeo and Carry, 2014). For example, Hartung et al. (2004) successfully mapped the surface of Saturn's moon Titan, which has an angular diameter of 0.8 arcsec, using adaptive optics, providing unprecedented insights into the moon's surface features and seasonal atmospheric conditions. Bolin et al. (2020) utilized observations of comet 2I/Borisov to acquire an estimate of the diameter of its nucleus, shedding light on the physical properties of this interstellar visitor. Close et al. (2000) and Marchis et al. (2006) employed turbulence-mitigating high-resolution imaging to detect the fist binary asteroid system and reveal that some asteroids are multiple systems, with moons. Finally, Descamps and Marchis (2008) conducted a comprehensive study of the physical and orbital characteristics of extensively observed main-belt and trojan binaries, finding that the total angular momentum serves as a valuable indicator to assess the internal structure of such asteroidal systems, providing insights into their potential origins.

It is clear from this brief summary of a few important research topics in astronomy that atmospheric turbulence negatively and significantly impacts the kinds of research that can be conducted due to the loss of angular/spatial resolution and the inability to detect faint objects/features lost to background noise. I will now introduce how we have come to understand the impacts of turbulence, and the technologies we have created to mitigate it

as much as possible. The limitations of the current methods will lead into the motivations for the novel method I helped create for atmospheric turbulence mitigation.

3.1.1 Understanding and Mitigating Atmospheric Turbulence

In the realm of astronomical observations, the terms "diffraction-limited" and "seeing-limited" play crucial roles in defining the quality and clarity of images obtained from telescopes. A *diffraction-limited* telescope is one where the primary limitation to the resolution is the diffraction of light at the telescope's aperture, estimated by the Rayleigh criterion (Rayleigh, 1879):

$$\theta \approx 1.22\lambda/D \tag{3.1}$$

where θ is the angular resolution, λ is the wavelength of light, and D is the diameter of the primary lens or mirror. This is the theoretical maximum resolution a telescope can achieve, given its aperture size, without considering external factors like atmospheric turbulence. On the other hand, a *seeing-limited* telescope is one where the resolution is primarily limited by atmospheric effects, rather than the diffraction of light. Figure 3.3 shows the seeing (i.e. a measure of how wide the PSF of a star would be due to the atmosphere when taking long-exposure images) at various altitudes at Plateau de Calern, indicating an average seeing level of about 0.8 arcsec. Using Equation 3.1 with a 0.8 arcsec resolution, we find that any telescope with a mirror larger than 0.2m in diameter observing at optical wavelengths would be seeing-limited. This is the primary reason astronomers want to build telescopes at a site like Mauna Kea where the average seeing is $\sim$0.4 arcsec (Bely, 1987), but even then you run into seeing limitations above a 0.4m diameter telescope.

Central to understanding these concepts is the Fried parameter, r_0 (Fried, 1965; Fried, 1967), which characterizes the coherence length of the atmospheric turbulence and effectively describes the size of the telescope aperture for which atmospheric turbulence and diffraction effects are equally significant. When the diameter of a telescope is smaller than the Fried parameter, the telescope is essentially diffraction-limited. Conversely, when the diameter is much larger than r_0, the telescope is seeing-limited.

Astronomers have known for a long time that the primary issue with seeing limitations is long exposures. Part of (and often a significant portion of) the distortion of the wavefront is simply a random tip or tilt which displaces the light but does not decrease its sharpness (Fried, 1965). Only when you take a long exposure does this cumulative tip-tilting of the wavefront, which happens on tens of millisecond time-scales, result in a spreading of the PSF

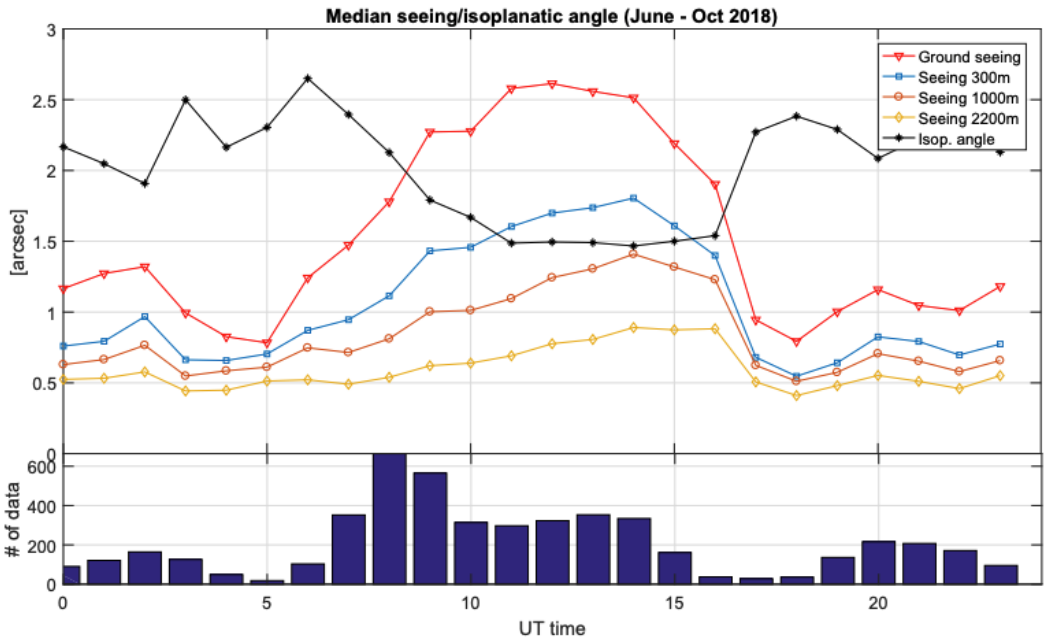

Figure 3.3: The median seeing of different layers of the atmosphere measured at the Plateau de Calern in France in bins of time, from Aristidi et al. (2020).

and loss of sharpness. If one were able to capture short-exposures, effectively causing the image to be insensitive to tilt, the resolution of the imaging system could be increased.

It was shown by Fried (1966) that, in the case of removing tip-tilt distortions alone, the optimal image quality happens when $D \approx 4r_0$; for telescopes smaller than this the resolution is diffraction-limited and for larger telescopes the higher-order wavefront distortions are more significant than the tip-tilt distortions (Kaiser, Tonry, and Luppino, 2000). For many ground-based observatories, the Fried parameter typically lies in the range of 10 to 20 centimetres under average seeing conditions, so a $\sim$1 m telescope is about the maximum size of telescope you could use with tip-tilt correction to get optimal imaging resolution, a size that by today's standards is relatively small. You could of course use a larger telescope if you're more concerned about capturing the light of faint objects and not so much the resolution, but what if you want both? Because of the demands of modern astronomy requiring data of unparalleled clarity, resolution, and accuracy to further its insights and discoveries, it was crucial for astronomers and physicists to understand turbulence and mitigate it as much as possible. Indeed, many solutions have been proposed and implemented. They broadly fall into two categories: those focusing primarily on deformable mirror solutions and those focusing primarily on post-processing solutions.

Deformable Mirror Solutions to Atmospheric Turbulence

As previously mentioned, the only way for a ground-based telescope to combat turbulence is by recording its effects in real-time and using the information from these snapshots to move the light back to a central point. Adaptive optics (AO), first envisioned by Babcock (1953) and subsequently built in the 1970s (e.g. Hardy, Lefebvre, and Koliopoulos, 1977), is one such method that has emerged as a revolutionary tool to counteract these disturbances, enabling telescopes to achieve near diffraction-limited observations (Beckers, 1993; Hardy, 1998). AO systems combine the following pieces to mitigate atmospheric turbulence:

1. **Wavefront sensor (WFS):** This is the "eye" of the AO system, designed to measure the wavefront phase aberrations introduced by atmospheric turbulence in real-time. The most common type of WFS is the Shack-Hartmann sensor (SH-WFS, Shack, 1971), which divides the incoming wavefront into a grid of smaller sections using an array of lenslets. Each lenslet focuses the light onto a detector, producing an array of focal spots. By comparing the positions of these spots with those from a reference wavefront, the local tilt (gradient) of the wavefront over each lenslet can be determined. The SH-WFS can accurately measure a wide variety of wavefront distortions, but its ability to detect details is limited by the fixed number of small lenses it has (Brunner et al., 2021). Additionally, under certain conditions, such as the low wind effect at the aperture of large telescopes, the SH-WFS can struggle to provide accurate wavefront corrections (Pourré et al., 2022).

   A more recent development, and a growing alternative to the SH-WFS, is the Pyramid WFS (P-WFS, Ragazzoni, 1996). It operates by splitting the incoming wavefront into four separate pupil images using a pyramidal prism. The resulting images are either processed as such, with the difference between them providing information about the wavefront phase errors, or are merged into a single format that gives information similar to the slope or gradient of the wavefront errors (Vérinaud, 2004). On-sky tests reveal the P-WFS to have better sensitivity than the SH-WFS (Ghedina et al., 2003; Peter et al., 2009), though the design of it can often be more complex and thus more expensive than the well-established SH-WFS. Regardless, it is often the WFS of choice for instruments on the next generation of extremely large telescopes (Shatokhina, Hutterer, and Ramlau, 2020).

2. **Deformable mirror (DM):** These specialized mirrors are designed to adjust their shape in real-time, compensating for wavefront distortions introduced by atmospheric

turbulence or even internal heating of optical components. In the paper introducing the idea of AO, Babcock (1953) presented his initial idea of an “Eidophor" DM device: a mirror covered in oil that changed shape when electric charges were applied to it. Although the design was impractical, being slow and sensitive to its environment, it established the basic idea of a flexible membrane being locally “actuated" to achieve a desired shape.

Due to the demands of DMs requiring both high actuator count and high frame rates (to achieve high-order deformations in real-time), today’s DMs predominantly utilize microelectromechanical systems (MEMS) technology, employing hundreds to thousands of tiny actuators behind the mirror surface to precisely adjust and achieve intricate shapes. All MEMS DMs have features like sub-millimeter spacing, low energy use, the ability to support a high number of actuators, and lightweight design (Madec, 2012), and have been found to work reliably for years despite the physical demands of, e.g., high frame rates ($\sim$10 kHz, Hartzell et al., 2010). Although more exotic technologies are being developed to possibly replace DMs – e.g. liquid crystal AO (Gourlay et al., 1997), which has seen impressive on-sky results (Zhang et al., 2020a) – the more established MEMS DMs will be used for the next generation of telescopes.

3. **Control system:** This is the “brain" of the AO system, responsible for processing measurements from the wavefront sensor and determining the necessary corrections. Once these corrections are calculated, the control system sends commands to the DM, instructing it to adjust its shape to counteract the atmospheric distortions. This entire process unfolds at often hundreds or even thousands of times per second, to keep pace with the ever-changing atmospheric conditions.

   A significant aspect of the control system’s calculations involves the use of Zernike polynomials (Noll, 1976). These polynomials, first developed by Zernike (1934), are orthogonal over a unit circle and have found extensive applications in optics, vision sciences, and image processing (Niu and Tian, 2022). In the context of AO, Zernike polynomials are used to represent wavefront aberrations. By fitting these polynomials to the wavefront errors measured by the sensor, the control system can determine the necessary corrections. For a tip-tilt AO system, only the second and third polynomials are required since they represent the vertical and horizontal tilts. Higher-order corrections often need to be made on faster timescales because they tend to represent finer-scale turbulence structures that can evolve more rapidly than the larger-scale, lower-order aberrations, but unfortunately their computation is inefficient, thus requir-

ing specialized methods for estimating them (e.g., Hwang and Kim, 2006). Modern AO control systems also rely heavily on GPUs for low-latency calculations (Dipper et al., 2013; Venugopalan, 2014; Sevin et al., 2014).

4. **Reference source:** For the AO system to work, it needs a reference source of light to measure the wavefront distortions. This can be a bright star located near the astronomical object of interest (natural guide star, or NGS). However, in many cases, there might not be a sufficiently bright star nearby. Foy and Labeyrie (1985) envisioned that, in such situations, a laser guide star (LGS) could be used used; a laser beam is projected into the atmosphere, creating an artificial "star" by causing a layer of sodium atoms in the mesosphere to glow. The AO system then uses this artificial star as a reference to measure and correct for wavefront distortions.

   The Keck Observatory was among the first to adopt LGS technology (Max, Gavel, and Olivier, 1995), and it is now common on modern observatories (e.g., Bonaccini et al., 2003; Fusco et al., 2010; Martinez et al., 2022). Indeed, a lot of work has gone into fixing some of the problems of LGS systems for the next generation of extremely large telescopes, such as LGS spot elongation (Fusco et al., 2022) and the persistent challenge of achieving sufficient return flux from a LGS (Hickson, Hellemeier, and Yang, 2021).

The performance of AO has seen remarkable improvements over time, with modern systems capable of delivering images with resolutions that rival those from space telescopes (Davies and Kasper, 2012). However, as briefly mentioned above, AO is not without its limitations. Traditional Single Conjugate AO (SCAO) and Single Laser AO (SLAO) systems often have a limited field of view, correcting only a small region (a few tens of arcseconds) around the guide star. Their effectiveness is also contingent on the presence of a *bright* NGS or the use of a LGS (which comes with it the "cone effect"), limiting where and when they can be used (NGS's and LGS's have a sky coverage of $\sim$1% and $\sim$30%, respectively (Rigaut and Neichel, 2020)). To address these limitations, several "flavours" of AO have been developed to incorporate advanced techniques, a few of which are discussed here.

Ground layer AO (GLAO) is a technique for correcting atmospheric turbulence that addresses some of the limitations of traditional AO. Rather than correcting only for the turbulence in a small volume of the atmosphere, GLAO aims to provide a uniform correction over a wide field – up to 15 arcminutes – by primarily correcting the turbulence close to the ground (Rigaut, Ellerbroek, and Flicker, 2000; Tokovinin, 2004). In many observing locations, the majority of atmospheric turbulence is concentrated near the ground, so GLAO

effectively enhances the resolution of imaging systems by addressing this ground-layer turbulence. Its performance is modest, however, when compared to other systems because of its ineffectiveness at mitigating turbulence in the mid- to upper atmosphere; on-sky experiments have demonstrated that GLAO typically offers a two to three-fold improvement in the average full width at half maximum (e.g., Minowa et al., 2017; Abdurrahman et al., 2018). Certain instruments and telescopes with expansive fields-of-view remain beyond the corrective reach of GLAO. For example, the forthcoming Vera Rubin Observatory has an impressive 9.62 square degree field. Regardless, many next generation instruments and telescopes will have a GLAO option (e.g., MOSAIC on the E-ELT, ULTIMATE on Subaru, WFOS on TMT, and the AO system on the GMT, see Hammer et al., 2016; Bouchez et al., 2018; Kupke et al., 2018; Minowa et al., 2020).

Multi-Conjugate Adaptive Optics (MCAO) uses multiple DMs conjugated to specific layers of the atmosphere and multiple guide stars to correct over a broader field of view than SCAO (Beckers, 1988; Johnston and Welsh, 1994; Rigaut, Ellerbroek, and Flicker, 2000). The use of multiple guide stars in MCAO allows for a tomographic reconstruction of the turbulence volume above the telescope, meaning that instead of just correcting for the turbulence along a single line of sight (as in SCAO), MCAO can correct for turbulence across a volume of the atmosphere (Tokovinin et al., 2001). This comprehensive correction approach allows MCAO to enhance image quality across a wider field, while also counteracting the "cone effect" inherent to using SLAO (Meyer et al., 2006). While offering a wider correction, MCAO is inherently more complex than SCAO due to its need for sophisticated control algorithms, though successful implementations of it exist (Marchetti et al., 2007; Rigaut et al., 2014). MCAO will be used extensively in future instrumentation (e.g., the AO system on the GMT, NFIRAOS on TMT, and MORFEO (formerly MAORY) on E-ELT, see Johns et al., 2004; Herriot et al., 2006; Diolaiti et al., 2016).

One more variant on traditional AO systems is Multi-Object AO (MOAO, Hammer et al., 2004), an advanced technique designed to provide high-resolution corrections over fields of view even larger than MCAO. MOAO integrates the tomographic capabilities of MCAO, using multiple DMs with "pick-off arms" centred on guide stars, to reconstruct the turbulence volume and apply corrections to multiple objects within that volume simultaneously. This method allows for targeted corrections on specific objects of interest within a large field, making it especially advantageous for observations where multiple targets are spread out across the sky, e.g. clusters of galaxies. Despite its technical complexity and the costs associated with multiplexing, it has been successfully demonstrated on Subaru with the Raven instrument (Lardière et al., 2014), and there are plans to implement it on other

facilities, e.g. GIRMOS on Gemini (Chapman et al., 2018).

While all these upgrades to SCAO and SLAO have achieved impressive corrections over wider fields, there remain many challenges related to the cost and complexity of implementing and maintaining advanced AO systems. Additionally, as mentioned, there are some instruments with wide enough fields that not even GLAO can correct. This is why other, less expensive and less complex, methods have been developed, such as those relying on the post-processing of images.

Post-processing Solutions to Atmospheric Turbulence

An alternative method that has gained traction in the astronomical community is high cadence imaging, in particular popularised with the “lucky imaging" technique. Lucky imaging capitalizes on the intermittency of turbulence and the brief moments of atmospheric stability. By taking a rapid series of short-exposure images, only the sharpest frames – those taken during moments of optimal seeing conditions – are selected. These “lucky" frames are then aligned and combined to produce a single high-resolution image. While lucky imaging can achieve impressive resolutions over wide fields (Mackay et al., 2018), especially for brighter targets, its effectiveness is inherently tied to the whims of the atmosphere. In poor seeing conditions, the probability of capturing lucky frames decreases, making the technique less effective (Faedi et al., 2013).

As the field of astronomy continues to push the boundaries of observational capabilities, there is a pressing need for more versatile and cost-effective solutions that can operate under a broader range of conditions. With the rapidly evolving landscape of digital technology, there is a growing opportunity for machine learning (ML) methods to meet these demands and help fill in some of the gaps of AO and lucky imaging systems.

ML has already been successfully employed to mitigate atmospheric turbulence effects in long-range imaging applications (Nieuwenhuizen and Schutte, 2019; Vint et al., 2020; Hoffmire et al., 2021; Zhang et al., 2022). These methodologies primarily rely on training ML algorithms on a collection of artificially distorted images, enabling them to predict the known ground truth and rectify turbulence-induced aberrations, though there are other methods which attempt to accomplish this in an unsupervised way (e.g., Li et al., 2021a). In essence, ML can be used as a digital counterpart to lucky imaging, where instead of relying on elusive high quality exposures, algorithms *infer* the turbulence-free image from a sequence of short-exposure images. This approach offers the potential to consistently produce sharp, high-resolution images, even in less-than-ideal seeing conditions and with a more efficient

use of telescope time.

Despite the success in long-range imaging, the adaptation and application of ML techniques to turbulence in astronomical images remains largely uncharted, presenting a promising frontier for future research. The uniqueness of astronomical imaging poses specific challenges and requirements that differentiate it from long-range imaging. Characteristics like the extreme range of object brightness, varied object sizes and scales, and the requirement for ultra-high resolution and precision make astronomical imaging a domain where a distinct approach is required (Tyson and Frazier, 2022).

In this chapter, I present a novel method to combat the deleterious effects of atmospheric turbulence in astronomical images using ML trained on simulations of turbulent and noisy video-streams of stellar fields. Section 3.3 outlines the method used to simulate atmospheric turbulence and the ML methods used in this study. Section 3.2 describes both how the simulated datasets were created and how real testing data was collected from the C2PU telescope. Section 3.4 is an overview of the main results from evaluating the proposed method on simulated and real data. Section 3.5 summarizes the strengths and limitations of the proposed method, and concluding remarks are in Section 3.6.

3.2 Data

3.2.1 Synthetic Data Generation

Each video sequence in a training dataset was created with one of two methods:

1. *Random field*: The number of stars to be simulated was uniformly sampled from [3, 2000] and given random x and y coordinates. An exponential law was used to sample magnitudes for each star: $m \sim m_{max} - e^{1/(\alpha \ln(10))}$ where m_{max} was 21 in this study and α (the slope of the differential source counts, i.e. dlogN/dmag) was 0.4.

2. *Realistic field*: To include realistic priors in our simulation pipeline, we incorporated the data from real star clusters in the Milky Way. The catalog from Kharchenko et al. (2013) includes the sky coordinates of thousands of clusters, which we used to query the Gaia DR3 database (Prusti et al., 2016; Vallenari, Brown, and Prusti, 2022) to obtain the RA, Dec, and `phot_g_mean_mag` values of each star. A random offset in both RA and Dec was uniformly sampled from $[-FOV/2, FOV/2]$, where FOV is the field-of-view of the telescope, and applied to get a shifted image, i.e. a more diverse training set.

The dataset therefore contained both randomly generated and realistic stellar fields.

We generated a training dataset containing 40,000 6-second video sequences, noting that 6 seconds was chosen as a compromise between GPU memory constraints and collecting enough information about the turbulence and faint stars. Each frame was 256x256 pixels and, to match the properties of the C2PU Telescope (see Section 3.2.2), we used a 1-metre diameter telescope aperture, a central wavelength of 650 nm, a pixel scale of 0.235"/pixel, and a sampling rate of 5.25s^{-1} (i.e. sampled every $\sim$200 ms) – noting that while this sampling rate is generally insufficient at capturing the quickly evolving turbulence, it does offer an advantage of having higher signal-to-noise in each frame, and in any case we plan on increasing the sampling rate in future iterations of the method. In each video sequence, the Fried parameter and windspeeds for each layer of the atmosphere were sampled from a normal distribution. Along with each video sequence, we generated the corresponding ground truth frame in which we disabled contributions from the atmosphere and any sources of noise in our simulation pipeline. The datasets were split 90%/10% into training/validation sets.

3.2.2 Data from the C2PU Telescope

The Centre Pédagogique Planète et Univers (C2PU) facility (Bendjoya et al., 2012) is located on the Plateau de Caussols at an elevation of 1260 metres, approximately 50 km from Nice in southern France. The site benefits from good seeing in summer time, with a median of 1.06", down to 0.8" at the end of the nights (Aristidi et al., 2020).

On the night of May 27, 2022, we collected around 60 seconds of short-exposure (200 ms) images of the globular cluster M92 in the SDSS r bandpass from the wide-field camera installed at the prime focus of the C2PU Omicron 1.04m telescope. No guiding was used as the tracking accuracy of the telescope allows for unguided exposures up to a few minutes without image degradation. The wide field camera provides a $37.6' \times 25.2'$ field of view with excellent image quality, through a three-lens Wynne coma corrector, at a resulting F/3.17 focal ratio. One important feature of this optical setup for our project is its relatively low obstruction (30% linear) for a wide-field instrument, which preserves more than 80% of the central peak intensity of the telescopic point spread function, compared to an unobstructed aperture. The camera at the C2PU Omicron prime focus is a QHY600Pro equipped with a Sony IMX 455M sCMOS sensor. Despite its modest quantum efficiency at redder wavelengths, this generation of sensors has proven to be competitive for quantitative astronomy (Betoule et al., 2023; Alarcon et al., 2023). The sequence images have a size of 1024x1024 pixels and a pixel scale of 0.235".

3.3 Methods

3.3.1 Simulating Atmospheric Turbulence

We implemented a variation of the "split-step" method (e.g. Chatterjee and Mohamed, 2014) for simulating the propagation of a wavefront through the atmosphere, in which the atmosphere is decomposed into several distinct layers which perturb the wavefront as it passes through. This section describes the theoretical framework used for the split-step simulations.

The Kolmogorov model is a theoretical model that describes the statistics of atmospheric turbulence. It assumes that the turbulence is isotropic, homogeneous, and fully developed, which means that the turbulence has reached a statistically steady state. The Kolmogorov model is characterized by a power-law scaling of the spatial frequency spectrum of the turbulence, and the refractive-index power spectral density is given by:

$$\Phi_n = 0.033\ C_n^2\ k^{-11/3} \quad \text{for } \frac{1}{L_0} \ll k \ll \frac{1}{l_0} \tag{3.2}$$

where L_0 is the outer scale, i.e. the average size of the largest eddies, l_0 is the inner scale, i.e. the average size of the smallest eddies, C_n^2 is the refractive index structure constant, which is a measure of the strength of the turbulence, and k is the angular spatial frequency. The spatial frequency spectrum in the Kolmogorov model has a power-law behavior with an exponent of $-11/3$, which means that high spatial frequencies are strongly attenuated by atmospheric turbulence.

There exist more sophisticated models which include inner-scale and outer-scale factors to improve fits between theory and experiment. In the modified von Karman model, for example, the refractive-index power spectral density is given by:

$$\Phi_n = 0.033\ C_n^2\ \frac{\exp\left(-k^2/k_m^2\right)}{(k^2 + k_0^2)^{11/6}} \tag{3.3}$$

where $k_m = 5.92/l_0$ and $k_0 = 2\pi/L_0$. The modified von Karman model is used in the atmospheric turbulence simulations in this study because it provides a more realistic description of the statistical properties of the turbulence than the simple Kolmogorov model.

Atmospheric turbulence is usually modeled using phase screens. A phase screen is a two-dimensional map of the phase distortion introduced by the turbulence at a particular altitude, and is typically represented as a matrix of random phase values. In the modified von Karman model, the statistical properties of the phase power spectral density are given

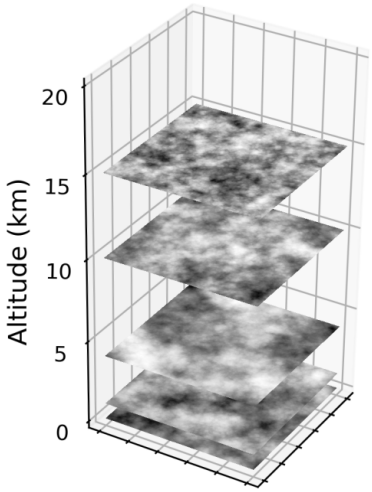

Figure 3.4: An example of the phase screens used in the simulation pipeline. For every simulated video sequence, each layer is initialized with a different r_0 and wind speed to account for varying seeing conditions.

(a) (b)

Figure 3.5: The PSFs of the same star imaged with a 1m telescope in (a) mild turbulence ($D/r_0 \approx 2$) and (b) strong turbulence ($D/r_0 \approx 5$). In the case of mild turbulence, the PSF is mostly concentrated in the centre with a clear Airy pattern around it, whereas with the strong turbulence, several Fried parameter length-scales can fit within the area of the telescope's aperture and so the light is distributed away from the centre in multiple "speckles".

by[1]:

$$\Phi_\phi = 0.023\ r_0^{-5/3}\ (f + 1/L_0^2)^{-11/6} \exp\left(-1.126\ l_0^2\ f\right) \tag{3.4}$$

where f is the frequency in cycles/m, and r_0 is the Fried parameter, which represents the size of the region over which the wavefront distortion due to atmospheric turbulence is roughly constant. Because r_0, L_0, and l_0 change as a function of altitude, the phase screens change with altitude as well. Figure 3.4 shows an example of the phase screens computed in our simulation software.

To calculate the impact of the phase screens on the wavefront of the light, the mathematical framework of Fourier optics, which describes the propagation of light through optical systems, was used. This calculation involves taking the Fourier transform of the phase screen and multiplying it by the Fourier transform of the incident wavefront. The resulting wavefront is then inverse Fourier transformed to obtain the distorted view.

In the case of atmospheric turbulence, the phase screen is represented as a two-dimensional function $h(x, y)$, where x and y are the spatial coordinates. The Fourier transform of the phase screen is given by:

$$\tilde{h}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) e^{-2\pi i(k_x x + k_y y)} dx dy \tag{3.5}$$

where k_x and k_y are the spatial frequencies in the x and y directions, respectively.

A monochromatic wavefront with complex amplitude, $A(x, y)$, where x and y are the spatial coordinates, was used in this study. The Fourier transform of the incident wavefront is given by:

$$\tilde{A}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) e^{-2\pi i(k_x x + k_y y)} dx dy \tag{3.6}$$

The wavefront distortion due to the phase screen can then be calculated by multiplying the Fourier transform of the phase screen by the Fourier transform of the incident wavefront:

$$\tilde{W}(k_x, k_y) = \tilde{h}(k_x, k_y) \cdot \tilde{A}(k_x, k_y) \tag{3.7}$$

The distorted wavefront can be obtained by taking the inverse Fourier transform of $\tilde{W}(k_x, k_y)$:

[1] See Schmidt (2010) for the full derivation.

$$W(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{W}(k_x,k_y)e^{2\pi i(k_x x + k_y y)}dk_x dk_y \tag{3.8}$$

The distorted wavefront can then be propagated through the rest of the optical system using Fresnel propagation, a mathematical method used to calculate the propagation of a wavefront through a distance in free space. The wavefront is propagated a distance z using the phase shift given by:

$$\phi(k_x,k_y) = e^{-i\pi\lambda z(k_x^2+k_y^2)}, \tag{3.9}$$

where λ is the wavelength of the wave. The spatial frequency spectrum of the wavefront after propagation is then given by:

$$\tilde{A}'(k_x,k_y) = \tilde{A}(k_x,k_y)\phi(k_x,k_y). \tag{3.10}$$

Finally, the wavefront in the output plane can be obtained by taking the inverse Fourier transform:

$$A'(x',y') = \frac{1}{2\pi}\iint_{-\infty}^{\infty} \tilde{A}'(k_x,k_y)e^{i(k_x x' + k_y y')}dk_x dk_y, \tag{3.11}$$

where x' and y' are the spatial coordinates in the output plane.

With the Fourier optics and Fresnel propagation implemented, we have a full description of how to propagate a wavefront from one atmospheric layer to the next, as well as how to distort the wavefront when it interacts with an atmospheric layer. Starting with a cosine-model wavefront that encodes the location of a star – represented by a point source – in the field of view, we can calculate the distorted wavefront from the star before it enters the telescope optics.

To get the final wavefront after it passes through the telescope with aperture D, we compute the Fourier transform of the wavefront as follows:

$$\tilde{A}(k_x,k_y) = \frac{1}{D^2}\int_{-\frac{D}{2}}^{\frac{D}{2}}\int_{-\frac{D}{2}}^{\frac{D}{2}} A(x,y)e^{-2\pi i\left(\frac{k_x x}{D}+\frac{k_y y}{D}\right)}dxdy$$

We then compute the pupil function $P(x,y)$, which acts as a circular window at the

entrance aperture of the telescope. The pupil function is defined as:

$$P(x,y) = \begin{cases} 1, & \text{if } x^2 + y^2 \leq \left(\frac{D}{2}\right)^2 \\ 0, & \text{otherwise} \end{cases}$$

Note: the same technique can be applied to an obstruction (such as that holding a secondary mirror) by defining a circular mask instead. We multiply the Fourier transforms of the wavefront amplitude and pupil function to obtain the Fourier transform of the wavefront that has passed through the telescope aperture:

$$\tilde{W}(u,v) = \tilde{A}(u,v)\tilde{P}(u,v)$$

And finally we compute the inverse Fourier transform of $\tilde{W}(u,v)$, obtaining the image of the stars in the focal plane of the telescope, which is proportional to the Point Spread Function (PSF):

$$\text{PSF}(x,y) = \frac{1}{D^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{W}(k_x,k_y)e^{2\pi i\left(\frac{k_x x}{D}+\frac{k_y y}{D}\right)}dk_x dk_y$$

The PSF describes the response of the telescope and atmosphere to a point source object, and can be used to estimate the resolution of the imaging system. Figure 3.5 shows what the PSF of a star looks like when imaged with a 1m aperture telescope through an atmosphere with different strengths of turbulence (characterized by r_0 in Equation 3.4).

To include realistic observational effects, the simulation pipeline includes the ability to set a varying r_0, wind speed, and wind direction for very phase screen so that they change at every time step. Additionally, the images can include a background level, Poisson noise, and Gaussian electron readout noise.

The entire simulation pipeline was written with Pytorch so that GPUs could be maximally utilized with Fast Fourier Transforms (Brigham and Morrow, 1967). This resulted in the capability to render ∼150,000 PSFs per second, which is a couple orders of magnitude faster than other similar implementations (e.g. Hardie et al., 2017). Maximizing efficiency in the simulation pipeline was required because the datasets we generated contained hundreds of thousands of frames.

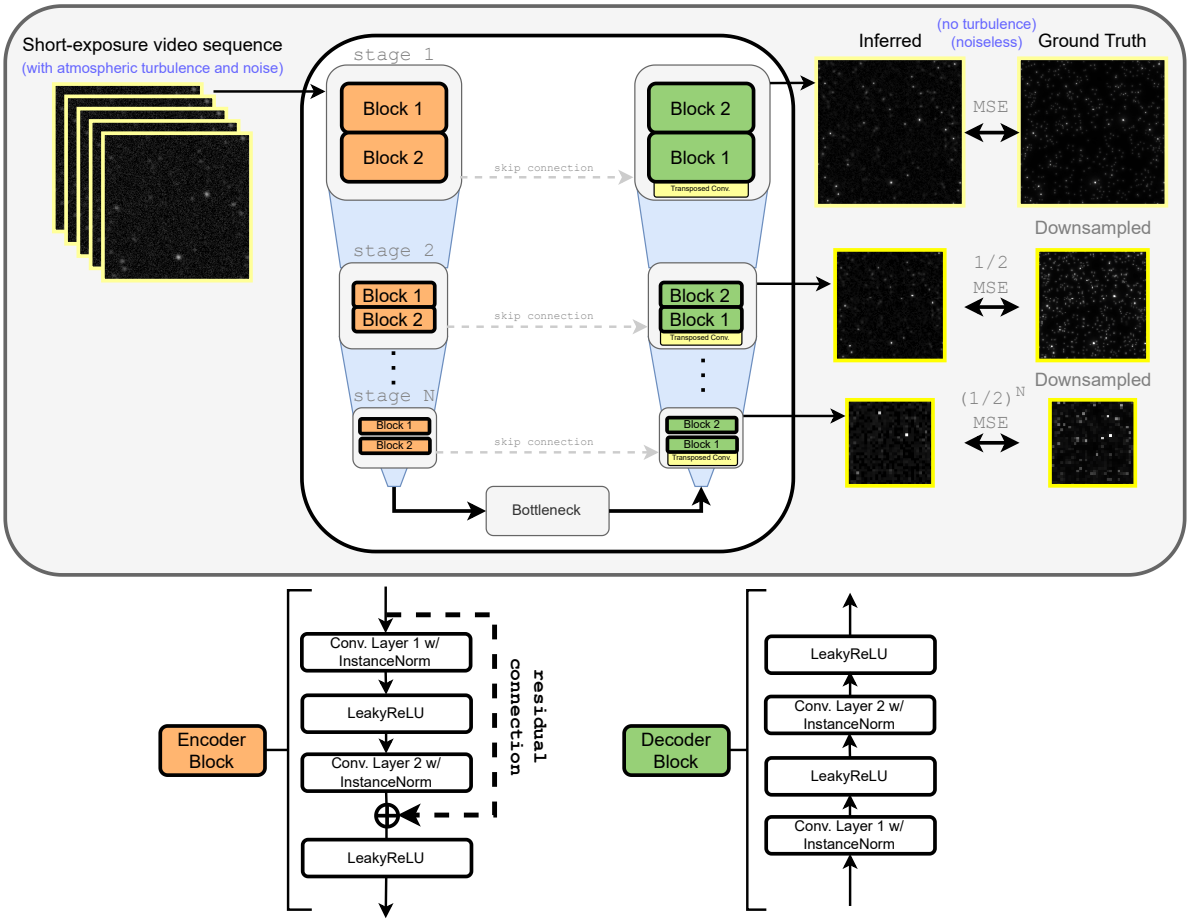

Figure 3.6: The DanceCam Residual U-Net architecture. A set of simulated short-exposure video streams of stellar fields – with turbulence and noise – along with their corresponding ground truth frames – with no turbulence or noise – is used to train the model. Instead of a single output, the model additionally has outputs from each stage in the decoder which are compared to downsampled versions of the ground truth using a weighted mean-squared error (MSE) loss function. Once trained, either a simulated or real video stream can be used as input and only a single (not downsampled) inferred image is retrieved.

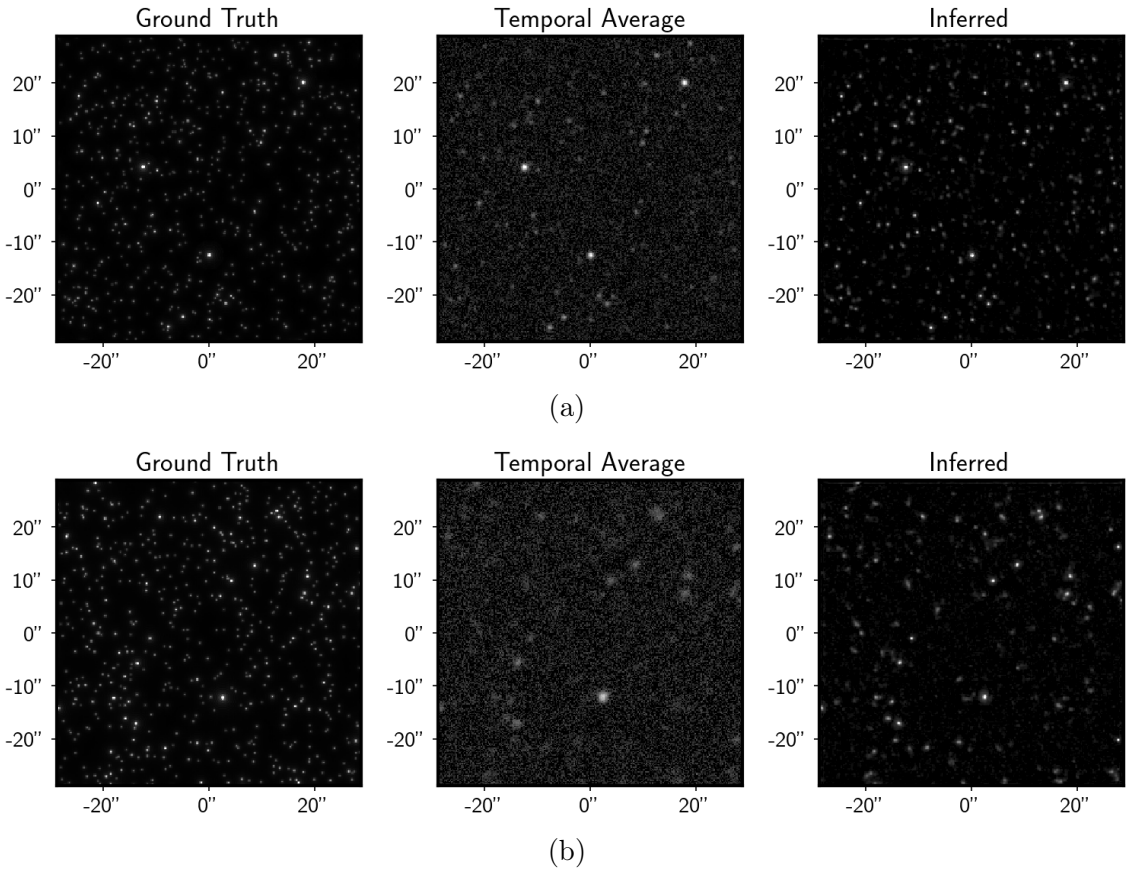

Figure 3.7: Two examples highlighting the ability of the proposed method to remove the effects of atmospheric turbulence and produce a sharp, clear image. 6-second sequences of random stellar fields were simulated with (a) 0.7" seeing and (b) 1.4" seeing, and the ground truth, temporally averaged sequence, and inferred frames are shown here.

3.3.2 Deep Learning Inference of Turbulence-free Images

Model Architecture

The cornerstone of our proposed method is the application of the Residual U-Net, a variant of the traditional U-Net architecture known for its proficiency in semantic segmentation and image reconstruction tasks (Ronneberger, Fischer, and Brox, 2015; Çiçek et al., 2016; Yao et al., 2018; Zhang, Liu, and Wang, 2018; Mizusawa et al., 2021). The model used in this study, developed by the Medical Imaging Computing Group at the German Cancer Research Center (MIC-DKFZ)[2], is part of a widely-used Python package used for deep learning-based biomedical image segmentation (Isensee et al., 2021). It was selected based on its unique architecture and inherent properties – detailed below – that make it suitable for our image reconstruction task.

A U-Net can be defined as a combination of a contracting path (encoder) and an expansive path (decoder), bridged by a bottleneck which helps to reduce the computational complexity of the model. The encoder performs consecutive convolutions and downsampling to output *feature maps*, learning the contextual information while decreasing the spatial dimension of the input. The expansive path, on the other hand, upsamples the feature maps and then concatenates them with the corresponding feature maps from the encoder (skip connections), allowing the decoder to merge high-level features with preserved local information.

Each layer of the U-Net consists of *blocks*. Each block includes convolutional layers with activation and normalization functions – in our case, `LeakyReLU` activation and `InstanceNorm` normalization were used, noting that `InstanceNorm` was chosen for its ability to perform well on smaller batch sizes (Kolarik, Burget, and Riha, 2020), which we were limited to due to GPU memory constraints. The encoder block reduces the spatial dimension while increasing the feature channels progressively, and the decoder block does the inverse operation.

The Residual U-Net uses *residual* blocks to help alleviate the vanishing gradient problem (He et al., 2015; Borawar and Kaur, 2023) and allow for deeper networks. For an input feature map X, each residual block in the architecture performs two convolutional operations F_1 and F_2, separated by a `LeakyReLU` activation function. The output of these operations is added to the input feature map, and another `LeakyReLU` function is applied. This can be mathematically represented as:

$$Y = g_2(F_2(g_1(F_1(X))) + W(X)) \tag{3.12}$$

[2] https://github.com/MIC-DKFZ/dynamic-network-architectures

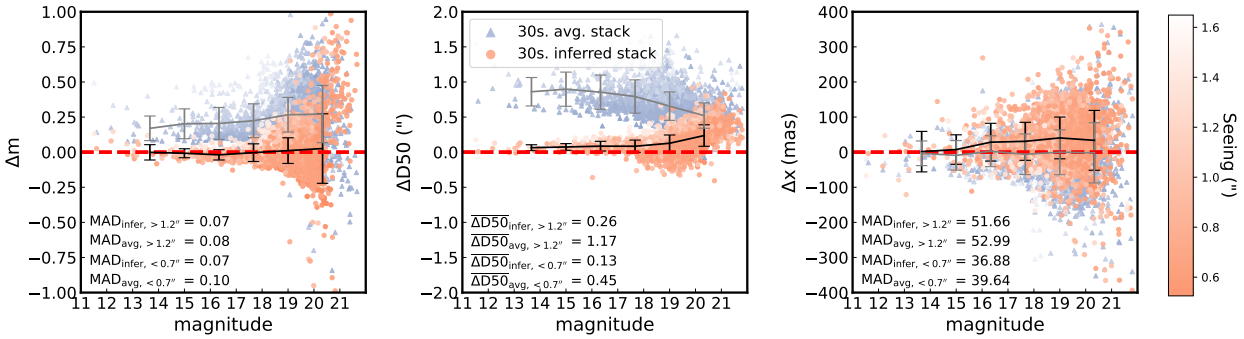

Figure 3.8: A series of quality assurance tests were made to validate the image reconstructions made by the U-Net. Hundreds of 30-second simulated observations of random stellar fields, with varying seeing conditions, were created and two images were made for each example: a stack made from the U-Net inferred images and a simple averaged stack of the raw frames. `SExtractor` was run on each frame, along with their corresponding ground truth frame, to collect information about each detected star's (*left panel*) magnitude – using a 10 pixel diameter aperture – to test for flux conservation, (*middle panel*) flux radius – defined as the radius in which 50% of the light from a star is contained within (D50) – to test for improvements in image quality, and (*right panel*) centroid coordinates, to test for astrometric stability. Shown here are the residuals of those metrics for the inferred stack (red circles) and simple averaged stack (blue triangles) when compared to the matching stars in the ground truth frames as a function of magnitude, along with their binned means and standard deviations (shown as error bars) – where the black and grey lines correspond to the inferred and averaged stack values, respectively. Also shown are the computed means for "bad seeing" and "good seeing" subsets of the data (> 1.2" and < 0.7", respectively). The "fainter fatter" effect can be seen for the inferred stacks in the D50 figure, wherein the U-Net tends to smooth the fainter stars (see Section 3.5.1 for a further discussion).

where F_1 and F_2 denote 2D convolutional operations, W is a transformation operation that adjusts the input feature map's dimensions and/or number of channels to match $F_2(g_1(F_1(X)))$, and g_1 and g_2 represent the `LeakyReLU` functions. Each layer consists of two such residual blocks along with down- or upsampling operations.

Following the principle of *deep supervision* (Wang et al., 2015; Li et al., 2022b), which enables the backpropagation of gradients from the deeper layers of the network to the earlier layers more effectively, the output feature maps of all the layers in the decoder part are connected to a final convolution layer to produce the final image. The entire network is shown schematically in Figure 3.6.

Training Configuration

During training, as a batch of data is loaded, the background is first subtracted by iteratively determining the median and excluding points that deviate more than three times the Mean Absolute Deviation (MAD) from the median. Such an approach effectively minimizes the impact of outliers, such as stars, in the estimation of a statistically robust background level, and ensures the varied background levels of any test data will not negatively influence predictions.

When working with image reconstruction or denoising, it's often beneficial to transform the input data such that the noise level is approximately constant across the image. This can help algorithms, like those used in U-Net architectures, to perform more consistently. We performed experiments with cross-validation on various scaling transformations. After evaluating the results, in particular the flux conservation, we converged on the Anscombe transformation for the input images. Not only does it stabilize the variance in Poisson-distributed data, making the noise homoscedastic, but it also scales the data. This scaling is particularly advantageous as it amplifies signals in dim regions, enhancing the visibility of faint features. The transformation is given by:

$$f(x) = 2\sqrt{x + \frac{3}{8}} \tag{3.13}$$

For its inversion, we adopt the closed-form approximation of the exact unbiased inverse as defined in Makitalo and Foi (2011):

$$f^{-1}(x) = \frac{1}{4}x^2 + \frac{1}{4}\sqrt{\frac{3}{2}}x^{-1} - \frac{11}{8}x^{-2} + \frac{5}{8}\sqrt{\frac{3}{2}}x^{-3} - \frac{1}{8} \tag{3.14}$$

For the optimization process, we employed the Adaptive Moment Estimation (Adam)

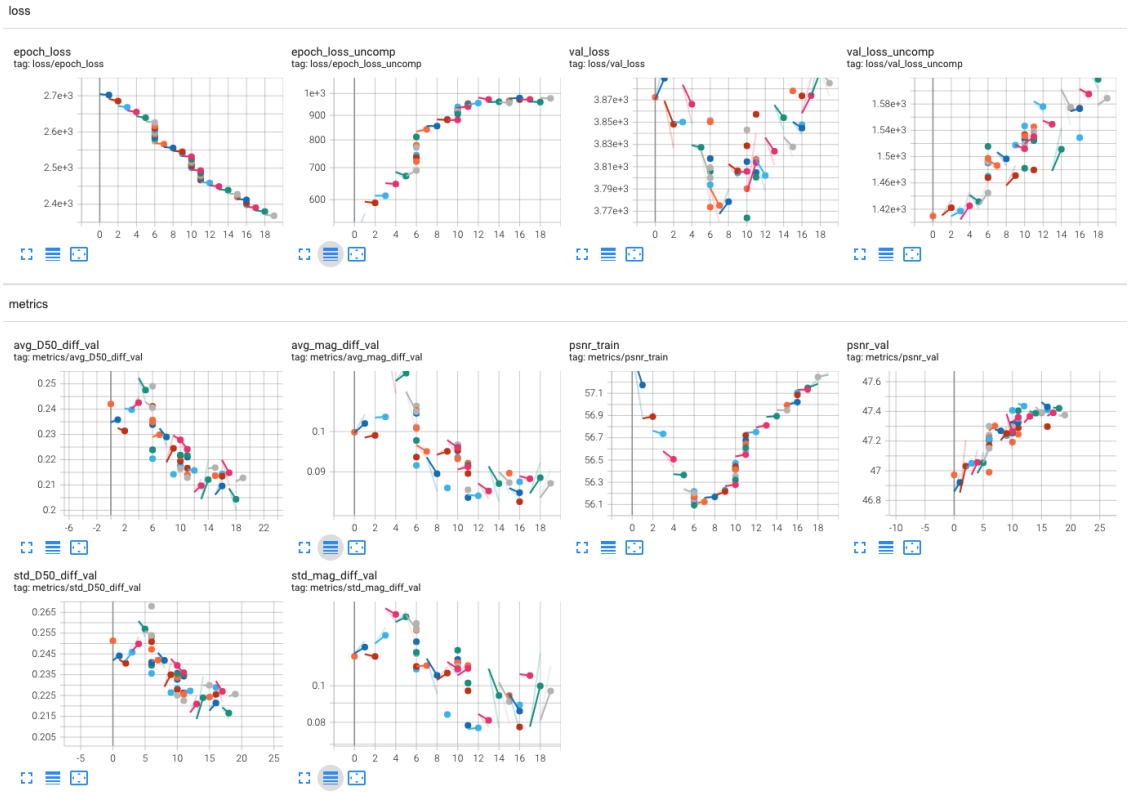

Figure 3.9: *Tensorboard* was used to track several losses and metrics throughout the training process, and shown here is the dashboard for part of a training run for a single model.

optimizer, a popular first-order stochastic gradient descent algorithm using estimates of the first and second moments of the gradients (Kingma and Ba, 2014). In our implementation, the default values were used such that the initial learning rate was set to 0.001, and beta values of 0.9 and 0.999 were used for the exponential decay rates of the gradient and squared gradient, respectively.

To expedite the training process, we utilized multiple GPUs – specifically 32GB NVIDIA Tesla V100s – made available by the Digital Research Alliance of Canada on their distributed network. On average, each epoch took approximately 14 minutes to complete. The mean-squared error (MSE) loss was used, and a learning rate scheduler was incorporated which adjusted the learning rate when the training MSE loss plateaued.

Since we were particularly concerned with the image reconstructions conserving the flux of stars, `SExtractor` (Bertin and Arnouts, 1996) was run on the inferred and ground truth images in the validation set after every epoch to track the magnitude estimates of detected stars. Several other losses and metrics were tracked with *Tensorboard* (Abadi et al., 2015) during the training process to better understand what the model was learning and how it

was performing at every epoch. Figure 3.9 shows an example of the *Tensorboard* terminal for part of the training process of a single model, highlighting the complexity of the problem: not all of the metrics are optimized in the same way or at the same time. For example, while the validation loss (`val_loss`) decreases for 8 epochs and then reverses direction to increase for the following 10 epochs, the MAE of the magnitudes on the validation set (`avg_mag_diff_val`) generally decreases for all of the epochs. Meanwhile, the MSE loss on the unscaled validation data (`val_loss_uncomp`) continually increases. The final model was chosen such that the mean absolute error (MAE) between the magnitudes of the inferred and ground truth stars in the validation set was minimized. I note here that I tried many different loss functions, including the MAE, Huber loss, and SSIM, but it was only the MSE that provided flux conservation. Further testing of different loss functions will be done.

3.4 Experimental Evaluation

Visually, the proposed method does an excellent job at taking in a short sequence of turbulent images and producing a clear, sharp, noise- and turbulence-free image. For example, Figure 3.7 shows the massive improvement – compared to temporally averaging the sequence – in image quality of frames inferred from sequences with good (0.7") and bad (1.4") seeing conditions. Some stars which are barely visible (or not visible at all) in the averaged frame appear quite clearly in the inferred frame.

Validating the method *quantitatively* in addition to qualitatively, however, required analyzing the performance metrics of a large number of simulated observations, as detailed in the following section.

3.4.1 Quality Assurance

To quantitatively assess the performance of the Residual U-Net, a series of quality assurance tests were implemented. These tests were designed to evaluate the improvements in image quality, flux conservation, and astrometric stability in comparison to simply stacking the images.

A 30-second video sequence of a stellar field was simulated following the *random field* process described in Section 3.2.1, including the random sampling of Fried parameters and wind speeds for each atmospheric layer. We iteratively step along the sequence one frame at a time and partition a 6-second subset to obtain an inferred U-Net frame until a full sequence of inferred frames was collected. The inferred sequence was stacked by taking the mean along the temporal dimension, resulting in a single frame representing a 30-second inferred

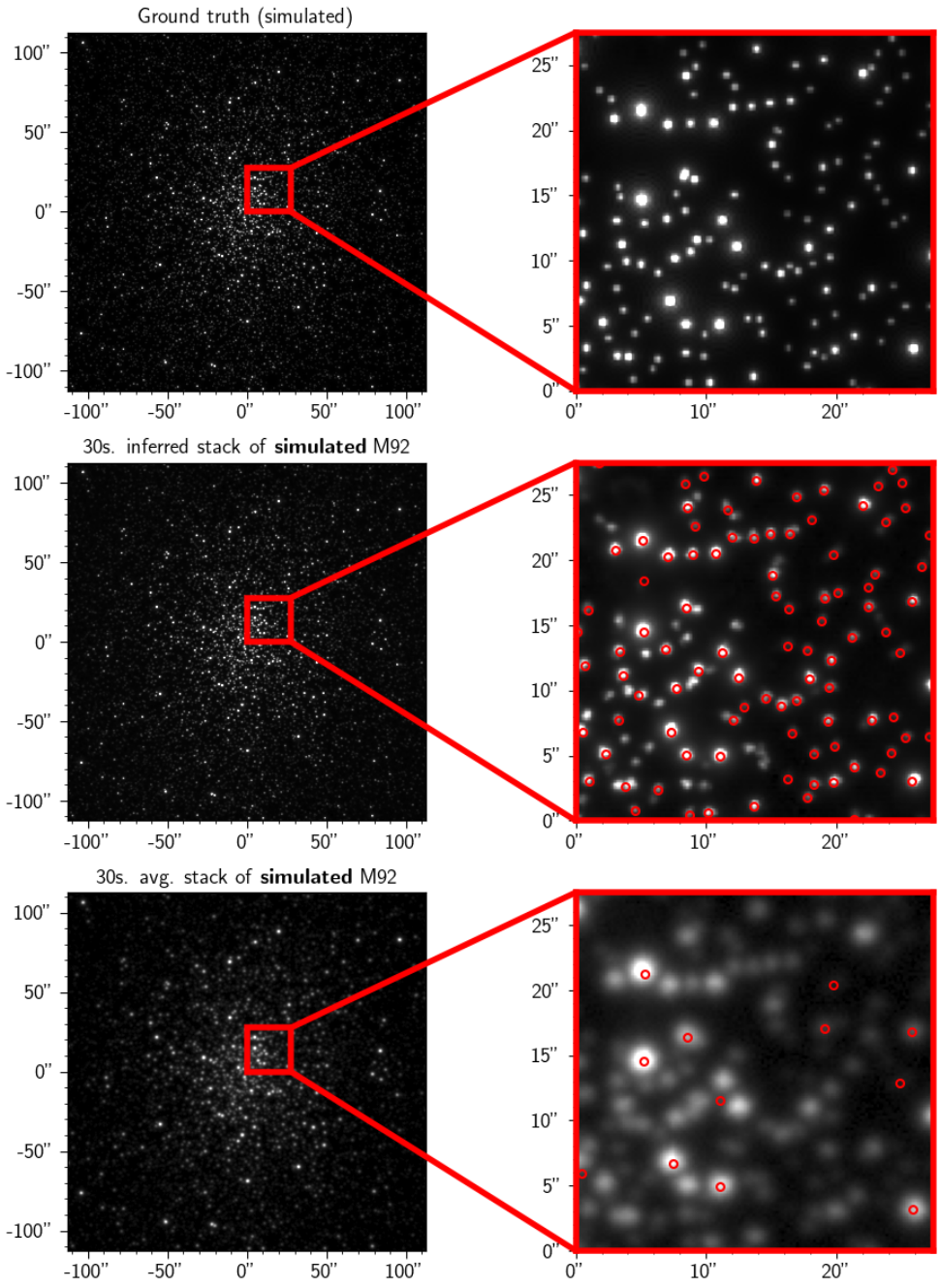

Figure 3.10: Our simulation pipeline was used to create a 30 second video stream – 160 frames in total @ 5.25 frames/sec – of the globular cluster M92 using Gaia positions and G magnitudes and a total seeing of 1.36". Shown here is a comparison of the simulated ground truth, a stack made from inferred images, and a simple averaged stack of the frames. The red circles indicate stars that were detected by `SExtractor`.

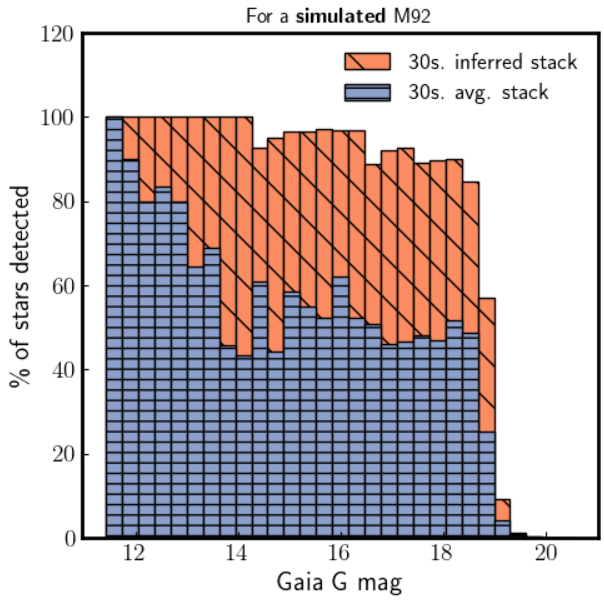

Figure 3.11: `SExtractor` was run on the ground truth frame, inferred stack, and simple averaged stack of the simulated M92 shown in Figure 3.10. The stars that were identified in the inferred stack and averaged stack were matched to the stars in the ground truth frame (by ensuring their measured positions differed by less than 2 pixels), and shown here is the completeness of the detected stars as a function of Gaia G magnitude. Note that the false positive rate for both frames was $\sim$2%, i.e. of all the stars `SExtractor` identified in the inferred and averaged frames, 2% were not matched to those in the ground truth frame.

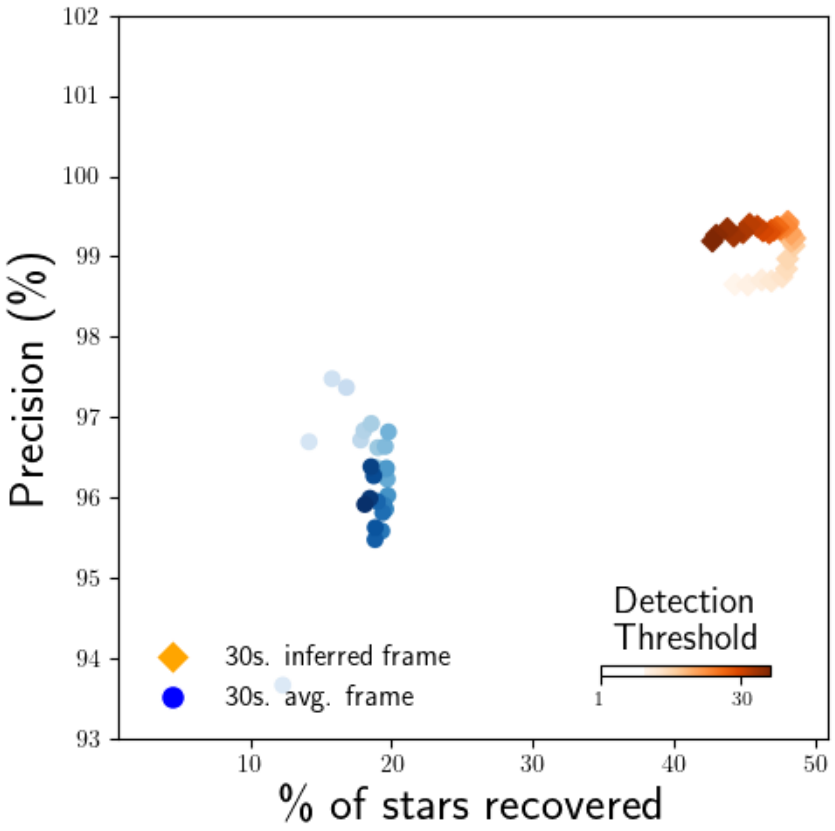

Figure 3.12: `SExtractor` was run on the inferred stack and simple averaged stack of the simulated M92 shown in Figure 3.10. Shown here are the precision and recovery rate of the identified stars compared to the stars in the ground truth frame as a function of the detection threshold used in `SExtractor`.

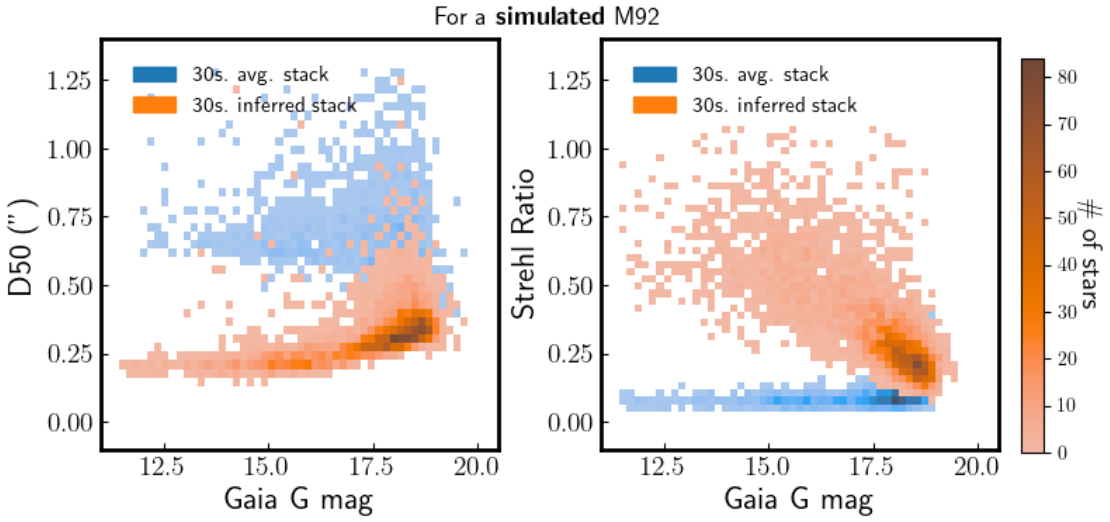

Figure 3.13: `SExtractor` was run on the inferred stack and simple averaged stack of the simulated M92 shown in Figure 3.10. Shown here are the distributions of D50 (left plot) and Strehl ratio (right plot) estimates of the detected stars as a function of Gaia G magnitude.

observation. The same 30-second sequence was used to obtain a single averaged frame. We additionally save the ground truth frame (simulated with no turbulence or noise) for each sequence. The entire procedure was repeated 500 times, resulting in a "quality assurance" dataset of 500 inferred frames, 500 averaged frames, and 500 ground truth frames.

`SExtractor` was then used on all of the frames in the dataset to identify the stars and calculate for each one the magnitude (by using an aperture diameter of 10 pixels to measure the total flux), the angular diameter (in terms of the D50, which measures the diameter at which 50% of the light is contained within an aperture), and the x-position, for the purpose of showing flux conservation, improved image quality, and astrometric stability, respectively. Figure 3.8 shows the residuals of these estimates, as a function of magnitude, on the inferred and averaged frames compared to the ground truth frames. The magnitude estimates of stars from the inferred frames can be seen to have overall a smaller dispersion throughout the magnitude range than stars in the averaged stack (which is evidenced by the median absolute deviation (MAD) being 0.01" - 0.03" lower across the seeing range). It is also the case that the magnitude residuals are closer to ground truth for the inferred frames, but in practice this could easily be resolved by using an adaptive aperture size on the averaged frames; we emphasize the importance of the magnitude residual plot is showing that flux is conserved in the inferred frames with lower dispersion than for the averaged frames. The lowered dispersion is also shown in the astrometric stability plot, where the MAD is 1 - 3 mas less for the inferred frames than the averaged frames, a very modest reduction but importantly there is no apparent degradation in astrometric stability. The strength of our proposed method is showcased in the angular diameter plot, where the average D50 residual is a factor of 3 - 4x lower in the inferred frames than the averaged frames, and for bright stars can even be a factor of about 8x less.

In all cases, the performance on the inferred frames gets worse towards fainter magnitudes: at a magnitude of 16, the standard deviations for magnitude residuals and D50 residuals are 0.01 and 0.07", respectively, whereas at a magnitude of 19, the residuals increase to 0.08 and 0.1". It appears the U-Net tends to "smooth" fainter stars, leading to a "fainter fatter" effect, contrasting with the "brighter fatter" effect in CCDs where brighter stars appear larger due to charge spillover in the sensors, further discussed in Section 3.5.1.

3.4.2 Test Case: M92

After ensuring that the trained model performed well in conserving flux, increasing image quality, and stabilizing the astrometry, it was time to task it with a more realistic scenario:

inferring the turbulence-free image of the globular cluster M92. With the high stellar densities in the core of the cluster, this would be a true test of our method's ability to disentangle the light from nearby stars and improve the spatial resolution. Testing was performed on both simulated data and real data obtained with the C2PU telescope.

Simulated Data

We began the test case of M92 with a simulated observation of it (30 seconds total @ 5.25 frames/second for a total of 160 frames) to obtain a baseline to compare to. A wide field of view (∼4 arcmin, 1024x1024 pixels) was used to showcase the ability of our method to perform on larger images than it was trained on. As described in Section 3.2.1, Gaia coordinates and magnitudes were used for each star. We set a maximum magnitude of 21, consistent with the model's training parameters, leading to a total of ∼12,000 stars being simulated. Relatively poor seeing of 1.36" was chosen to test the method's proficiency under sub-optimal conditions.

Since the model was trained on frames of size 256×256 pixels, the full 1024x1024 frames of M92 could not be used as input. Instead, each frame was split into tiles of size 256×256 with 50% overlap, resulting in the full 1024×1024×160 video sequence being split into 49 overlapping sequences of size 256×256×160. As described in Section 3.4.1, the U-Net was used on each of these sequences to produce a series of inferred images along the temporal axis which were then averaged together to produce a single 256×256 inferred image, thereby reducing the 49 sequences of size 256×256×160 into overlapping inferred tiles of size 256×256. In the final reconstruction, the overlapping regions of these tiles were blended smoothly: masks were generated for each tile, consisting of a central region with full contribution (value 1), and edge regions with a linear gradient from 0 to 1, reflecting the degree of overlap. The final image B at each pixel was reconstructed by calculating the normalized weighted sum of the n overlapping tiles, expressed as $B(x,y) = \frac{\sum_{i=1}^{n} T_i(x,y) \cdot M_i(x,y)}{\sum_{i=1}^{n} M_i(x,y)}$, where $T_i(x,y)$ and $M_i(x,y)$ are the value and mask at position (x,y) in the i^{th} tile, respectively. This ensured a seamless integration of the tiles, with the centre of each tile retaining its original value and the overlapping edges merging smoothly, resulting in the final 1024×1024 inferred image.

This tiling strategy introduces potential limitations, particularly for stars located on the edges of the tiles. Stars at these edges may suffer from partial coverage, potentially impacting the accuracy of their reconstruction in the inferred images. Additionally, the challenge of ensuring accurate flux conservation for these edge-located stars is of concern, as inference could be split across multiple overlapping tiles. If this poses a substantial issue, repositioning

a tile so that a star of particular interest moves away from the edge and into the centre of the tile could be considered as a mitigative measure.

Figure 3.10 shows the ground truth frame, the inferred frame, and the temporally averaged frame, along with zoomed in images of the central region of M92 which show, with red circles, stars identified with `SExtractor`. It is visually clear that the inferred frame is (∼2x) sharper and had significantly more stars (∼3x as many in the central regions) identified than the averaged frame, so a more quantitative analysis was conducted to confirm this.

Figure 3.11 shows the relative number of stars identified by `SExtractor` in both the inferred and averaged frames in bins of magnitude, where true positives were defined as being no more than 2 pixels away from the corresponding star in the ground truth frame. We note that the detection threshold for `SExtractor` was selected to optimize both the precision and recovery rate of its star identification. Figure 3.12 shows how the precision and recovery rate changed as the detection threshold was varied; the final detection threshold was chosen such that the point in precision-recall space was furthest to the top-right in the plot. For the averaged frame, the chosen detection threshold resulted in a precision of 97% (i.e. 3% false positive rate), and 19% of stars were successfully recovered. In contrast, for the inferred frame, the detection precision was 99.5% (i.e. false positive rate of 0.5%), and the recovery rate was 49%, meaning ∼2.5x more stars were identified in the inferred frame.

To test for the improvements in image quality, the D50 measurements from `SExtractor` were used and, because it is a more common diagnostic metric in AO analyses, the Strehl ratio, S, was calculated for each star according to Roberts Jr et al. (2004):

$$S = \frac{I(x=0)}{\sum I} \frac{\sum P}{P(x=0)}, \tag{3.15}$$

where x is the position vector, $I(x=0)$ is the maximum of the measured PSF, $P(x=0)$ is the maximum of the diffraction limited PSF, and $\sum I$ and $\sum P$ are computed over squares of size 10x10 pixels (centred on the coordinates extracted by `SExtractor`) and used to normalize the ground truth PSF to have the same total intensity as the observed PSF. This method likely contributes uncertainty in the measurements due to photon noise, background levels, and possible contamination from nearby stars, but this is okay because we are more interested in the *relative* improvements. Figure 3.13 shows how these values change over the magnitude range. Below a Gaia G magnitude of 17.5, the inferred stars have a 3x average reduction in D50 and an average of ∼5x improvement in Strehl ratio. Beyond a magnitude of 17.5, the performance on inferred stars drops rapidly yet still maintains a 1.5-3x improvement on D50 measurements, and 2-4x improvement on Strehl ratios; again we find that the model

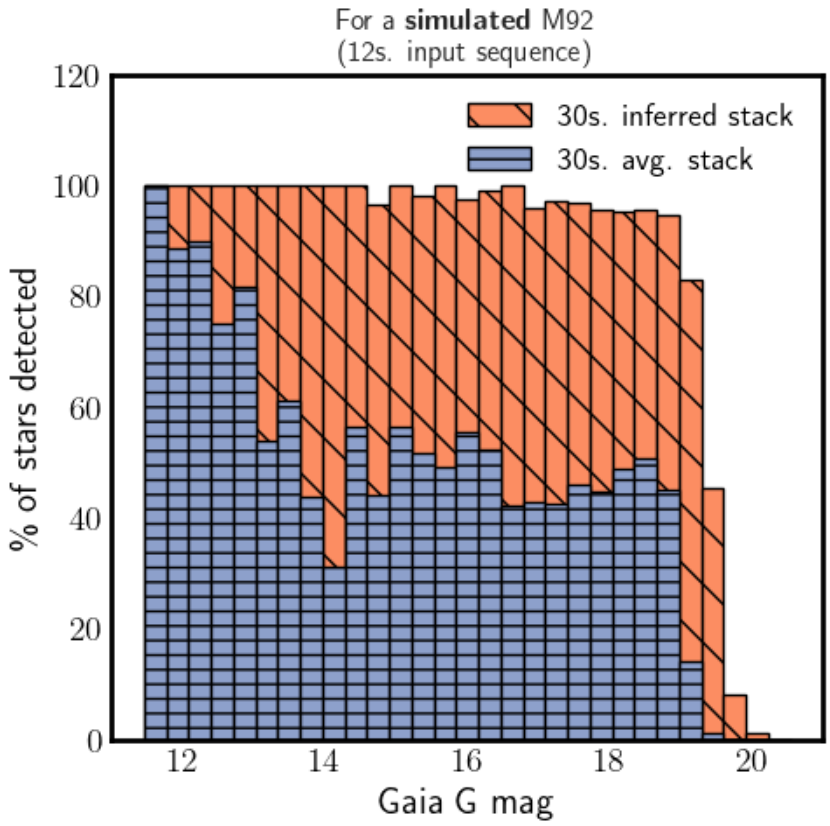

Figure 3.14: Same as Figure 3.11 except using a 12 second input context to the U-Net instead of a 6 second input.

struggles with fainter stars, tending to smooth their PSFs.

Finally, I tested how the input temporal context affected the U-Net inferences: instead of a 6 second input, I doubled it to 12 seconds. Figures 3.14 and 3.15 show the drastic improvements in the results. Many more stars were recovered (62% in total) – with several stars a half magnitude fainter compared to both the averaged frame and the 6s. temporal context inferred frame – and the average width of the stars remains lower, especially out to fainter magnitudes. Unfortunately, when using this same model on real data, the results were worse than for the model using a 6 second context, likely because of the limitations discussed in Section 3.5. Therefore, the model trained with a 6 second context is used for all tests on real data.

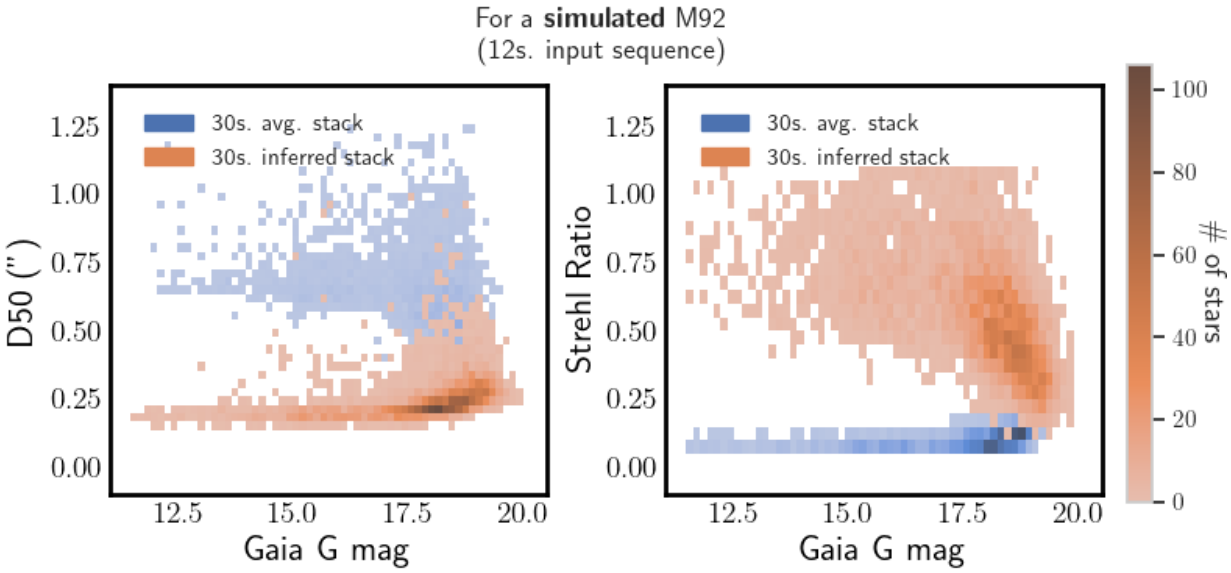

Figure 3.15: Same as Figure 3.13 except using a 12 second input context to the U-Net instead of a 6 second input.

C2PU Telescope Data

All of the steps in 3.4.2 were repeated for the real M92 data collected by the C2PU telescope. Figure 3.16 shows a comparison between the ground truth (simulated with Gaia coordinates and magnitudes), inferred, and temporally averaged images of a 30 second observation of M92 containing 160 total frames of size 1024x1024. The zoomed in areas of the central regions of M92 show that the brighter stars in the inferred image appear substantially sharper, though there are clearly some spurious effects in the fainter parts; in the video stream of M92, there were indications that an intermittently very turbulent ground layer was causing all the stars in the field to move in lockstep by up to a couple arcseconds, an effect which is not currently accounted for in the simulations. This and other limitations, as discussed in Section 3.5.1, contributed to a decreased visual performance on real data.

As for the number of stars correctly identified, Figure 3.17 shows that slightly more stars were found in the inferred frame than the averaged frame ($\sim$12% vs. $\sim$10% of the total stars in the ground truth image, both with a $\sim$92% precision). More bright stars were found in the inferred frame, whereas a few more faint stars were found in the averaged frame, which prompted us to investigate where the stars were being more readily identified in both images. Figure 3.18 shows that the inferred image works particularly well at de-blending the crowded central regions of M92, recovering about 2x more stars within 50 arcsec from the centre than the averaged image.

Finally, Figure 3.19 shows quantitatively how the D50 and Strehl ratio measurements on the inferred and averaged images change across the magnitude range. Again we find that, for the inferred image, the performance on brighter stars ($G < 17$) is enhanced, with an average of 2.5x reduction in D50 measurements and 5x increase in Strehl ratio measurements. For fainter stars, the performance drops in a similar manner as in the case of a simulated M92, leading to only slightly better average PSF width than for the averaged frame.

3.5 Discussion

The proposed method introduces a novel approach to handling the challenges posed by atmospheric turbulence in astronomical imaging, especially when dealing with wide fields. By segmenting the observed field into smaller overlapping tiles, producing an inferred frame for each tile, and merging the inferred tiles, the method allows for an effective mitigation of atmospheric turbulence effects in arbitrarily large images without compromising the quality of turbulence correction.

An important feature of the approach is its independence from the traditional reliance

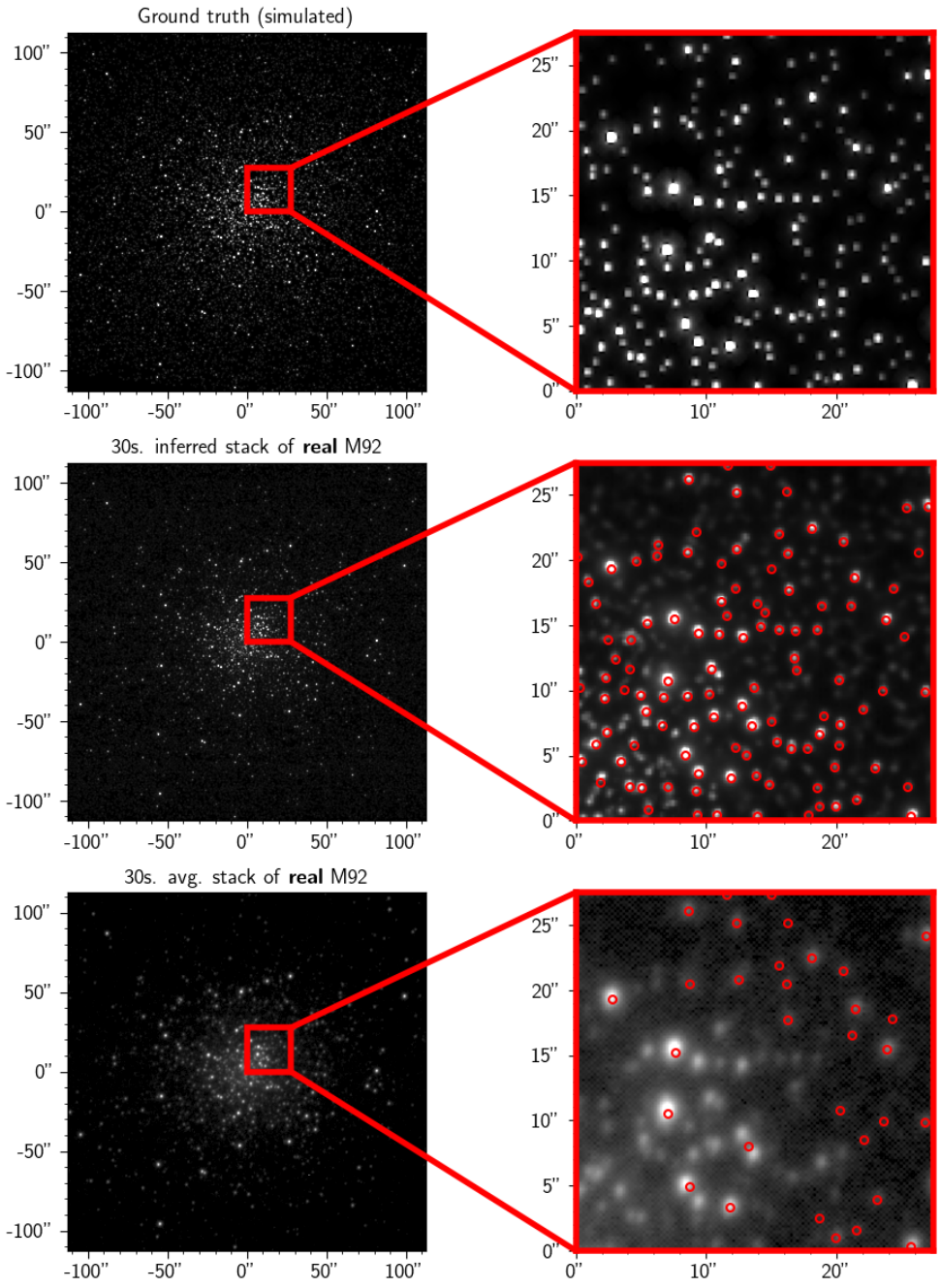

Figure 3.16: The C2PU telescope was used to obtain a 30 second video stream – 160 frames in total @ 5.25 frames/sec – of the globular cluster M92. Shown here is a comparison of the simulated ground truth using Gaia positions and magnitudes, a stack made from inferred images, and a simple averaged stack of the frames. The red circles indicate stars that were detected by `SExtractor`.

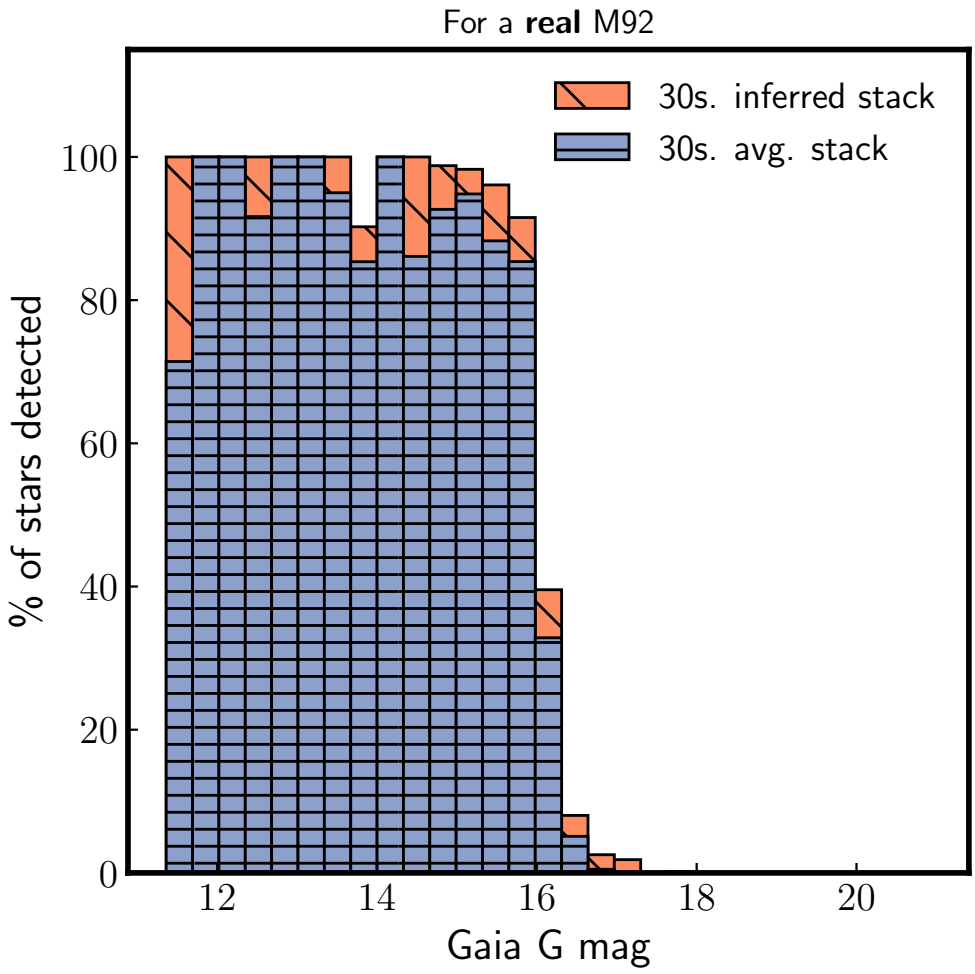

Figure 3.17: `SExtractor` was run on the ground truth frame, inferred stack, and simple averaged stack of the real M92 shown in Figure 3.16. The stars that were identified in the inferred stack and averaged stack were matched to the stars in the ground truth frame, and shown here is the completeness of the detected stars as a function of Gaia G magnitude. Note that most of the stars in M92 are fainter than $G \sim 18$, so the majority of stars (>90%) not detected had $G > 18$.

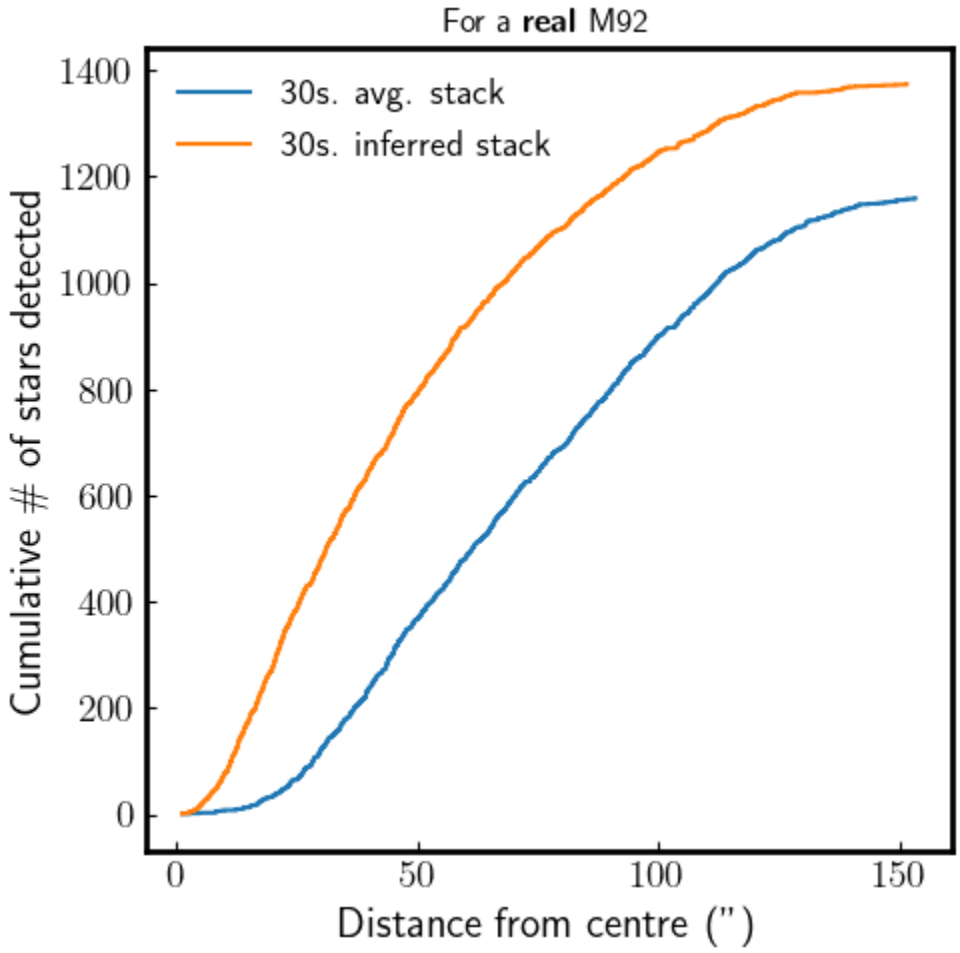

Figure 3.18: `SExtractor` was run on the inferred and simple averaged stack of the real M92 shown in Figure 3.16. Shown here is the cumulative number of stars found as a function of distance from the centre of the cluster for each stack, highlighting the ability of the proposed method to work in the crowded central regions of a stellar cluster.

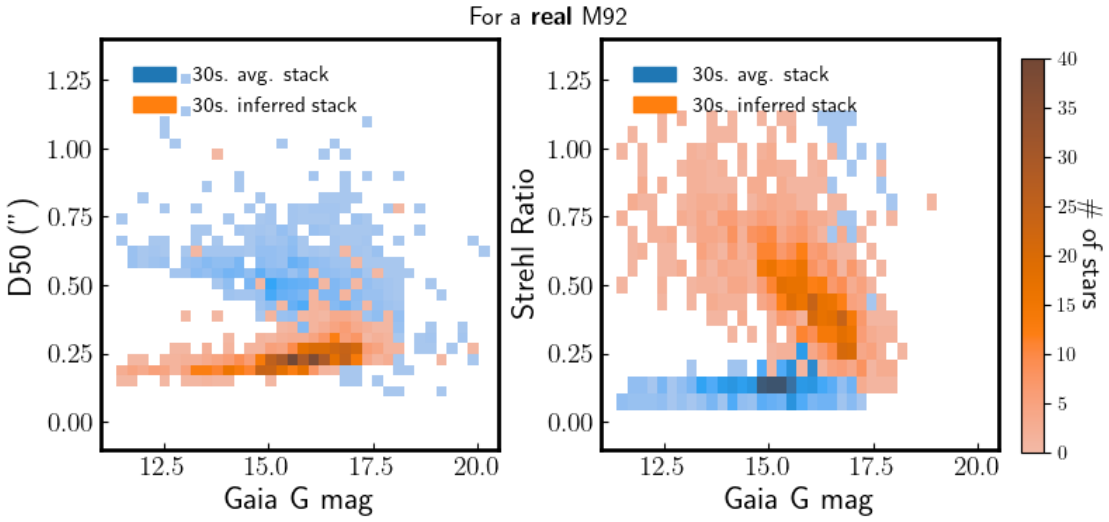

Figure 3.19: `SExtractor` was run on the inferred stack and simple averaged stack of M92 shown in Figure 3.10. Shown here are the distributions of D50 (left plot) and Strehl ratio (right plot) estimates of the detected stars as a function of Gaia G magnitude.

on guide stars, instead using the information from the entire field to understand the turbulence properties. While AO systems are tethered to either natural or artificial guide stars for measuring and correcting atmospheric distortions, our method's flexibility broadens its applicability, offering the potential to correct turbulence even in regions where suitable guide stars might be elusive. Moreover, the ability to analyze distortions in captured images paves the way for reverse-engineering the atmospheric layers and their respective turbulence properties. Such insights can be beneficial for our understanding of atmospheric conditions and refining observation strategies.

Modern technological advancements further strengthen the efficacy of the method. With the requirement of short exposures to capture the near real-time turbulence, most CCDs would produce images with too much noise, effectively drowning out the faint targets. The integration of CMOS cameras, on the other hand – renowned for their rapid readout speeds, reduced noise, and enhanced quantum efficiency (Guidash et al., 2016; Zhang et al., 2020b; Zhu et al., 2022) – complements the turbulence mitigation technique. Modern GPUs, additionally, allow for an impressive processing speed; the time required to produce a single inferred frame from a fully trained model is mere fractions of a second, allowing for the correction of a video stream comprising hundreds of short-exposure wide-field images in near real-time.

The proposed method is also appealing due to its minimalistic hardware requirements. Eschewing the intricate setups of traditional AO systems, which involve expensive deformable mirrors and wavefront sensors, our approach is predominantly software-centric. This orientation not only simplifies its implementation but also offers significant cost savings, making it an attractive solution for budget-conscious observatories and researchers, and even amateur astronomers.

3.5.1 Limitations

The proposed method offers a novel approach to mitigating the effects of atmospheric turbulence in astronomical imaging, yet it is essential to acknowledge its current limitations and explore potential avenues for improvement:

- **Intermittent Turbulence:** One of the primary constraints is the absence of intermittent turbulence in the current implementation. Intermittent turbulence can introduce sporadic and unpredictable distortions, which the method might not handle effectively in its present form. Interestingly, it could even enhance the method's effectiveness by occasionally providing "lucky frames" in a sequence. A straightforward solution

could involve incorporating dynamic changes to r_0 for each atmospheric layer in the simulation pipeline.

- **Motion Blurring:** Another limitation is the lack of motion blurring implementation. In real-world scenarios, the shutter collects light over a duration (e.g., 0.2 seconds), whereas the simulation pipeline currently takes instantaneous snapshots at regular intervals. Addressing this could involve splitting each frame into N sub-frames, simulating turbulence at every sub-timestep, and averaging the sub-frames together. This approach would effectively simulate an open shutter, offering a more realistic representation. Alternatively, using shorter exposures could also address this limitation.

- **Monochromatic Light:** The method's current reliance on simulating monochromatic light simplifies the computational process but overlooks chromatic aberrations introduced by atmospheric turbulence. Enhancing the simulation pipeline to cover a broader band-pass of light could improve the method's accuracy.

- **Rolling shutter:** Today's astronomical sCMOS sensors are usually operating in rolling shutter mode: lines of pixels are read and reset in sequence. The efficiency of photon collection is thereby optimized, but this has the inconvenience that two successive sensor lines are recorded at slightly different times. The maximum time difference amounts to only a few milliseconds inside the current 256×256 analysis window. However this delay would reach $\sim$100ms if distant parts of the full sensor were to be analyzed jointly, in which case it could not be neglected anymore.

- **GPU Constraints:** Temporal context is currently limited due to GPU memory constraints. Optimizing GPU usage, such as parallelizing the pipeline to distribute computational load across multiple GPUs, could address this limitation. Additionally, as GPUs continue to technologically advance, the method's capabilities will naturally expand.

- **Only simulating stars:** A major limitation of the method is that it is currently tailored to simulate stars only, excluding other celestial objects like galaxies or planets. Integrating diverse light profiles and spatial structures associated with a variety of astronomical objects would enhance its scope and versatility.

- **Transient Objects:** The current system does not handle transient objects that appear temporarily and unpredictably, e.g. satellites or other moving objects, or objects

that change in brightness. These types of objects could be included in the simulation pipeline. Alternatively, or additionally, owing to the high cadence video streams, frames containing transients of interest could be detected and saved for later processing using automated methods (e.g., Cabrera-Vives et al., 2017; Gieseke et al., 2017). Deleterious transients (specifically telecommunications satellites) could be removed before processing. This is indeed a major benefit of the video stream approach to observations, since with long exposures the best you can do is mask satellite streaks (e.g., Paillassa, Bertin, and Bouy, 2019). See Beskin et al. (2023) for a more in-depth look at the benefit of wide-field, high cadence imaging for transient detection and analysis.

- **High-Volume Data Streams:** Given our focus on wide fields and short exposures, the method generates a significant volume of data, which can be challenging to process and store efficiently. To manage these massive data streams, it will likely be crucial to process and analyze the collected data in quasi-real-time using one or several GPUs, helping to filter and prioritize data for storage and detailed examination.

- **Circular Stationarity of Phase Screens:** In the current scheme, we initialize a static phase screen and "roll" the values along for each time step, where the values at the edge roll over to the opposite edge. This approach, while efficient, isn't a problem when using short temporal contexts for training. However, it could become increasingly problematic when using longer video streams, as it introduces artificial periodicity and may not accurately represent the evolving nature of atmospheric turbulence. To address this, a more dynamic model of phase screens could be developed, where new turbulence patterns are continuously generated rather than recycled. This would better mimic the natural, non-repetitive behavior of atmospheric turbulence over extended periods. Implementing algorithms that can generate realistic, time-evolving turbulence patterns without significantly increasing computational load would be key.

In addition to these limitations, there are other effects not accounted for in the simulations, including 16-bit quantization, dark currents, quantum efficiency, malfunctioning pixels, and filter variations. These effects can all be added in future iterations of the method, either in simulations or by using a semi-supervised approached to jointly train on real observations.

Two notable issues that have emerged are the "fainter fatter" effect, where fainter stars are excessively smoothed, and the "hallucination" effect, where the U-Net erroneously interprets noise as non-existent stars (as evidenced by imperfect precision of detected sources). These phenomena are partially attributable to the U-Net's limited temporal context. Atmospheric

turbulence disperses light from stars, reducing the signal-to-noise ratio per pixel, and in some cases, burying the signal entirely within noise. Given such limited information, the U-Net's tendency to misinterpret noise either as part of a faint star's light profile or as an entirely new star is enhanced. This effect is particularly pronounced in fainter stars, where the signal is inherently weaker.

This observed limitation aligns with the inherent characteristics of the U-Net architecture. While effective at enhancing brighter point sources, the U-Net struggles with fainter stars due to its design, which essentially provides a maximum a posteriori (MAP) estimate. The model's performance is robust above a certain flux threshold but degrades below it, as it cannot effectively disentangle weak stellar signals from the surrounding noise.

Addressing these limitations could involve expanding the temporal context of the observations, where an increase in temporal context correlates with improvements in precision, Strehl ratio, and D50 measurements (see Figures 3.14 and 3.15). This suggests that a broader temporal scope allows for a more accurate signal reconstruction, particularly for fainter stars.

Future enhancements could also include integrating uncertainty predictions into the U-Net's inferences (e.g., Angelopoulos and Bates, 2021), providing astronomers with guidance on the reliability of the model's output. Indeed, to extend the capabilities of the model to even fainter stars, a shift towards a more generative approach may be beneficial. Such an approach would not just attempt to provide a single, 'best guess' reconstruction of the image (as in MAP estimation) but would rather aim to generate realistic samples drawn from the posterior distribution of reconstructed images. This would involve a probabilistic framework that accounts for the uncertainties and variabilities inherent in observations of fainter stars, potentially leveraging techniques like Bayesian neural networks, generative adversarial networks (GANs), or diffusion-based models.

Another crucial limitation of our current method is its inability to improve astrometric precision despite sharper images. This shortcoming likely stems from the residual jitter caused by atmospheric turbulence, as the U-Net, though capable of tracking the turbulent PSF across exposures, cannot determine the true position of stars. Averaging the centroids of these wandering PSFs over the 32 or 64 frames is not significantly more beneficial than averaging the light distribution, leaving a gap in astrometric accuracy. Addressing this issue will be crucial for showing the scientific validity of the method.

While the method has its limitations, many of these challenges offer exciting avenues for future research and development. Addressing these constraints can evolve the method into a more robust and comprehensive solution for astronomical imaging in the presence of atmospheric turbulence. Indeed, we would like to emphasize that the work presented in

this paper represents only the first iteration of the proposed method. There are numerous optimizations which can be implemented to enhance the performance by, e.g., incorporating more realism into the simulations, broadening the temporal and spatial contexts over which the U-Net makes its inferences, using shorter duration images to capture more information about the turbulence, and implementing more advanced machine learning methodologies and architectures.

3.6 Conclusion

In this study, we introduced a novel machine learning-based approach to counteract the challenges posed by atmospheric turbulence in astronomical imaging. By utilizing a U-Net architecture, we have demonstrated the potential to significantly enhance the sharpness of astronomical images. Our method, trained on simulated observations, is adept at inferring a turbulence- and noise-free image from a sequence of short-exposure observations of a stellar field, effectively associating speckles with their source star and disentangling light from proximate sources, while conserving flux.

Visually, the method showcased an enhancement in image clarity, especially under sub-optimal seeing conditions. Quantitatively, our results have been compelling: when tested on the simulated M92 dataset, the inferred frames exhibited an average reduction in D50 measurements by a factor of 3 for stars brighter than a Gaia G magnitude of 17.5, and an average 5x improvement in the Strehl ratio. This performance, however, tapered for fainter stars, indicating areas for further refinement. Furthermore, our quantitative analysis using `SExtractor` revealed that, when using the inferred frame, up to 2.5 times more stars were identified compared to the averaged frame, with a precision rate of $\sim$99%. This is a testament to the model's ability to enhance image quality and resolution, even in densely populated stellar fields.

In real-world tests, using a 30 sec. video sequence of the globular cluster M92 as a case study, our method demonstrated its ability to de-blend crowded regions. Specifically, the inferred image recovered about 50% more stars within 70 arcsec from the center of M92 compared to the temporally averaged image. However, the performance metrics, particularly for fainter stars, indicated a more rapid decline in the real data scenario than in the simulated one. This highlights the challenges posed by real-world conditions, such as the intermittent turbulent ground layer observed in the M92 video stream. The model's tendency to "smooth" and/or possibly "hallucinate" fainter stars is an expected limitation that warrants further investigation. The performance on real data, while promising, highlights areas of improve-

ment, particularly in handling effects not currently accounted for in simulations. Indeed, given the suite of current limitations in the simulations mentioned in Section 3.5.1, it was not granted that meaningful inferences could be made on real data.

In conclusion, this study presents a machine learning-based approach to addressing the challenge of atmospheric turbulence in astronomical imaging. The results obtained from both simulated and real data demonstrate the capabilities and potential of this method. However, it is important to acknowledge that further development and refinement are necessary, particularly in enhancing the reconstruction of fainter stars and improving astrometric precision. The method, in its current form, shows promise for contributing to the field of astronomical research. Approaches like ours, while still in their nascent stages, could play a role in improving the clarity and accuracy of wide-field ground-based observations. Nevertheless, we recognize the need for cautious optimism regarding the method's scientific potential, and we encourage continued exploration and testing to fully ascertain its efficacy in diverse observational scenarios.

Acknowledgements

We acknowledge and respect the lək̓ʷəŋən peoples on whose traditional territory the University of Victoria stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are very fortunate to have had the opportunity to conduct observations from this mountain.

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 682903, P.I. H. Bouy), and from the French State in the framework of the "Investments for the future" Program, IdEx Bordeaux, reference ANR-10-IDEX-03-02 .

We thank the National Sciences and Engineering Research Council of Canada for funding through the Discovery Grants and CREATE programs. We also thank Prof. Kim A. Venn for helpful comments and support on this project. This research was enabled in part by support provided by Calcul Québec (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

Data Availability

All data underlying this article will be shared on reasonable request to the corresponding author.

Chapter 4

Self-supervised contrastive learning of galaxies

This chapter contains the beginning of a project focused on using the representations obtained from self-supervised learning to automate the learning of galaxy morphologies and improve the precision of photometric redshift estimates.

Personal Contributions

I was the main contributor on this project, writing all of the code for retrieving, processing, and curating the data, as well as the entire training and testing framework. Everything here was written by myself.

4.1 Introduction

4.1.1 Classifying Galaxy Morphology

Observations of galactic phenomena and morphologies have been a dominant source of scientific discovery and understanding of the cosmos and galactic evolution for the past century (Conselice, 2014). Hubble's classification system (Hubble, 1926) made it clear that galaxies could be separated and understood in terms of their morphology, and that they broadly fell into three separate populations: those with spheroidal components, those with disk-like components, and those with so-called irregular components. Astronomers suggested this classification scheme reflected an evolutionary sequence, with elliptical galaxies representing "early-type" galaxies that evolved into spiral galaxies which represented "late-type" galaxies, though Hubble himself was adamant that his sequence represented no such temporal interpretation (Hubble, 1927). Astronomers today agree that it is likely not an evolutionary sequence.

In the 1960s and 1970s, the idea that galaxy morphology was linked to galaxy evolution gained further support through studies of galaxy clustering and the discovery of peculiar galaxies. Toomre and Toomre (1972) provided observational evidence for galaxy mergers, and their importance in shaping morphology and dynamics, with the discovery of peculiar galaxies with features such as “bridges" and “tails" which formed due to interactions with other galaxies.

The connection between galaxy morphology and evolution was further strengthened through the discovery of the colour-magnitude relation (Faber and Jackson, 1976) and the morphological-density relation (Dressler, 1980) which suggested that the morphology of galaxies was linked to their environment, with early-type galaxies (ellipticals) being more common in dense environments, and late-type galaxies (spirals and irregulars) being more common in low-density environments. Several studies used galaxy morphology to investigate the assembly history and growth of galaxies, including the discovery of the so-called “green valley" of galaxies, which may represent a transitional stage in the evolution of galaxies from blue, star-forming spirals to red, passive ellipticals (Balogh et al., 2004).

While the early progress in galactic morphology was made possible by experts examining individual examples on photographic plates and CCDs, this strategy became untenable in the era of large photometric surveys. An automated way of identifying galaxies and quantifying some aspect of their morphology was needed. Early software, such as FOCAS (Jarvis and Tyson, 1981), was able to automatically produce galaxy catalogs from astronomical images, but there were concerns about the robustness of the extraction process, e.g. how reliably the software could identify a galaxy from a star. Since most studies were conducted on the catalogs themselves and not on the images, this was a problem. Source Extractor (SExtractor, Bertin and Arnouts 1996) was a milestone in extraction software because the developers prioritized robustness – using a neural network trained on simulated images for reliable separation of galaxies and stars – and efficiency – capable of quickly operating on very large survey images.

As for quantifying morphological properties, there had been important work in parameterizing the radial distribution of light in galaxies by e.g. de Vaucouleurs, who formulated the relationship for elliptical galaxies (Vaucouleurs, 1948; De Vaucouleurs, 1959), and later by Sérsic, who generalized de Vaucouleurs’ law (Sérsic, 1963). Sérsic profiles are widely used for analyzing the surface brightness profiles of galaxies, accomplished by fitting the observed profile with a model that describes the distribution of light as a function of radius from the center of the galaxy. The Sérsic model has been found to be a good fit to the surface brightness profiles of many galaxies, particularly elliptical galaxies and bulges of spiral galaxies,

and indeed usually involves fitting two (or more) components that can characterize e.g. the bulge and disk of a disk galaxy (Simard et al., 2011).

There are however several limitations to Sérsic fitting, particularly when it comes to characterizing the structural properties of galaxies. For example, it is a parametric model that assumes a specific functional form for the surface brightness profile of galaxies, which may not accurately capture the true distribution of light for all types of galaxies, particularly irregular and asymmetric galaxies, and can indeed bias size measurements of galaxies (Davari et al., 2014). Additionally, Sérsic fitting only provides information about the light distribution of a galaxy, and does not capture other important properties, such as the presence of clumps or asymmetries which might be helpful for identifying mergers.

The CAS system (which includes information about the Concentration, Asymmetry, and Smoothness of light distributions in galaxies) introduced a non-parametric way of quantifying galaxy morphology (Conselice, 2003). The CAS system proved especially useful for identifying galaxy mergers, though there are non-negligible dependencies on the quality of data, such as the spatial resolution and signal-to-noise (Thorp et al., 2021).

To summarize, organizing and classifying large amounts of galaxies in photometric surveys is a challenging task due to several reasons, including:

1. *Scale*: Modern astronomical surveys can produce datasets with millions or even billions of galaxies, which can be overwhelming for manual classification by astronomers.

2. *Complexity*: Galaxies exhibit a wide range of morphologies, including elliptical, spiral, and irregular, and can also exhibit different stages of evolution, such as mergers and interactions. Additionally, there are instances of distorted galaxies in lensed systems.

3. *Variability*: The properties of galaxies can vary with respect to several factors, including redshift, stellar mass, and environment, making it difficult to build a classification scheme that works well across different datasets.

4. *Incomplete knowledge*: Despite decades of study, our knowledge of the physical processes that govern the formation and evolution of galaxies is still incomplete. This makes it difficult to identify all the relevant features that are important for classification.

These problems necessitate more advanced solutions, such as those involving machine learning algorithms which can learn patterns in a large dataset of galaxies without prior assumptions of which features are important. While the vast amounts of data in sky surveys

can be cumbersome to the traditional methods of galaxy analysis, this in particular is what allows machine learning algorithms to excel. The data can be leveraged to learn patterns and relationships between the observable properties of galaxies and their physical properties, such as masses, merger history, age, and star formation rates.

Supervised learning has been the main focus of astronomers looking to automate galaxy identification and morphological classification (e.g. Khalifa et al., 2018; Dai and Tong, 2018; Zhu et al., 2019; Ghosh et al., 2020; Cheng et al., 2021b; Li et al., 2022a). These methods all rely on there being a large set of labeled examples used to train a machine learning model to recognize patterns and classify new examples. The labeled examples typically consist of galaxy images along with corresponding labels that indicate the objects or features present in the images, though the methods differ in the way they train their neural networks and collect examples with labels. It is often the case that labels for real images are scarce, leading to some astronomers opting to use simulated observations instead. For example, Bickley et al. (2021) used IllustrisTNG (Marinacci et al., 2018) to identify galaxy post-mergers in simulations and trained a CNN on the simulated dataset to then find post-mergers in real data, a method that required substantial work in adding realism to the simulations (Bottrell et al., 2019). Walmsley et al. (2020) adopted an alternative approach to acquiring labeled data: the responses from people participating in Galaxy Zoo (Lintott et al., 2008b) – a widely acclaimed citizen science project in which millions of classifications were contributed by more than 150,000 participants tasked with identifying galactic features – were used to train Bayesian CNNs and predict the probabilities of unlabeled galaxies containing features such as bars and bulges. While supervised learning can achieve high accuracy, its requirement of large amounts of labeled data, which can be expensive and time-consuming to obtain, imposes a significant limitation. It also depends on the labels in the dataset: it generally cannot be used to find extremely rare galaxies or galaxies with features that are difficult to identify.

The goal of *unsupervised learning* is to identify and learn the underlying structure or distribution in a dataset without the use of labels. In other words, unsupervised learning algorithms try to learn the patterns and relationships in the input data without any explicit instruction on what to look for or how to categorize the data. Unsupervised learning methods include clustering, dimensionality reduction, and generative models. Clustering methods group similar examples together based on similarity metrics, such as Euclidean distance, while dimensionality reduction techniques aim to compress the data into a lower-dimensional space while preserving enough information to accurately reconstruct it. Generative models, such as GANs and variational autoencoders, learn to generate new images that are similar

to the input images. The representations generated by some unsupervised learning methods can be used in downstream tasks such as classification, e.g. by *pretraining* a model before fine-tuning it using labeled data in a supervised learning framework. In this case, the major limitation of supervised learning – requiring a large corpus of labeled data – is reduced; the model is able to first acquire a meaningful representation of the data during pretraining, drastically reducing the number of labeled samples required (Calhoun et al., 2022). The amount of useful information contained in the representations is therefore a critical component for the quality of the overall results.

These unsupervised approaches have been widely used in astronomical image analysis, where large amounts of unlabeled data are often available. For example, unsupervised methods have recently been used for automated galaxy morphology classification (e.g. Hocking et al., 2018; Cheng et al., 2021a; Fielding, Nyirenda, and Vaccari, 2022), although in most computer vision tasks, including these, it has been the case that unsupervised methods lag behind supervised in terms of their performance in most downstream tasks (excluding out-of-distribution detection (Hendrycks et al., 2019)), so indeed new methods that can better handle large amounts of unlabeled data are needed.

At the moment, some of the largest improvements in machine learning methods are in the area of self-supervised learning, where a model learns to extract meaningful features and relationships from the data by *defining its own training objectives* based on the inherent structure of the data itself (e.g., Paszke et al., 2019a; Goyal et al., 2019). In self-supervised learning, the data is usually *transformed* in some way, such as by rotating, flipping, or masking certain parts of the image, and often the model is trained to predict the original data from these transformed versions. For example, in the case of image data, a model could be trained to predict the orientation of an image after it has been rotated by a certain angle. By training on these self-supervised tasks, the model learns to extract low-level features and relationships from the data that can be used for downstream tasks such as classification, segmentation, or object detection. Self-supervised learning has become increasingly popular in recent years, especially in the fields of natural language processing and computer vision, and has been used to achieve state-of-the-art results on a variety of tasks (see Ericsson et al. 2022 for a review). Indeed, Chen et al. (2020) demonstrated that self-supervised learning could be used to outperform supervised methods on image classification metrics, such as accuracy, when trained on the same number of samples. Their method, called SimCLR, was recently used by Hayat et al. (2021) on 1.2 million SDSS *ugriz* galaxy images (Gunn et al., 1998) for automated morphology classification and downstream tasks like photometric redshift prediction (which achieved a higher precision than by training on the

images themselves). Indeed, acquiring precise redshifts of galaxies is another motivation of this project.

4.1.2 Galactic redshifts

As the universe expands, the space between galaxies stretches, causing the light emitted by those galaxies to stretch as well. When we observe the light from these distant galaxies, we therefore find that their spectral lines are shifted towards the red end of the spectrum. This phenomenon, aptly known as redshift – often denoted as z – is a direct consequence of the expansion of the universe, and is a cornerstone in the study of galaxies and the broader universe (Giovanelli and Haynes, 1991). It can be computed as:

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}} \tag{4.1}$$

where λ_{obs} is the observed wavelength and λ_{em} is the emitted wavelength. The magnitude of this effect can tell us how fast an object is moving away from us and, importantly, give us information about its age and distance (Hubble, 1929), assuming a cosmological model. The relationship between redshift and distance, known as Hubble's Law, laid the foundation for the Big Bang theory of cosmology, suggesting that the universe is expanding from an initial point of singularity.

The study of redshifts added another dimension to cosmological analyses, allowing for the study of galactic properties as a function of both distance and time, and the identification and characterization of entire clusters of galaxies with similar redshifts (Hubble and Humason, 1931; Humason, 1936). More recent studies have been able to study galaxies across a wide span of the age of the universe. For example, Sonnenfeld et al. (2013) examined the relationship between redshift and the internal structure of massive early-type galaxies over half the age of the universe and found that individual galaxies largely maintain their total mass density profiles over time, suggesting growth through minor mergers without significantly altering the mass density. Shen et al. (2023) used data from the Hubble Space Telescope (HST) and James Webb Space Telescope to collect mid-infrared and near-UV galaxy morphologies from $0.2 < z < 2.5$ (spanning nearly 9 billion years) to reveal that more massive galaxies increasingly have a higher obscured fraction of star formation in their inner regions.

Beyond the study of galaxy morphologies over time, redshift measurements are also crucial for understanding the largest scale properties of the universe. Lewis and Bridle (2002) used cosmic microwave background and supernovae type Ia data, along with data from the 2dF Survey (Colless et al., 2001) – which included redshift values for ∼250,000 galaxies – and

the HST, to conduct a Monte Carlo exploration of cosmological parameter space, helping to constrain several cosmological parameters including the equation of state of dark energy. The Baryon Oscillation Spectroscopic Survey (BOSS), part of the Sloan Digital Sky Survey (SDSS) III, and extended BOSS program (eBOSS), part of SDSS IV, collected spectra, and hence redshifts (distances), of *millions* of galaxies to map out a massive volume of space (out to $z \sim 3$, i.e. $\sim$80% of cosmic history) to further constrain cosmological parameters to a precision of one percent (Alam et al., 2017; Alam et al., 2021).

The upcoming surveys from the Vera C. Rubin Observatory (Ivezić et al., 2019) and Euclid (Racca et al., 2016) will collectively observe billions of galaxies, however it is infeasible to collect spectroscopic measurements for such a large sample. There will instead be photometric measurements – photons collected in certain band-pass filters which can be converted to magnitudes and colours – so the success of these surveys depends critically on the precision of the *photometric redshift* estimates, which are typically a factor 10-100 less precise than spectroscopic redshifts. Precise photometric redshifts are crucial for several reasons, especially when considering the study of millions or billions of galaxies:

1. Cosmological Surveys: In large-scale cosmological surveys, it's often impractical or too expensive in terms of time and resources to obtain spectroscopic redshifts for every galaxy. Photometric redshifts provide a way to estimate the redshifts for a vast number of galaxies quickly, enabling the study of the large-scale structure of the universe, the nature of dark energy, and the evolution of galaxies over cosmic time (Di Dio et al., 2014; Alonso et al., 2015; Balaguera-Antolínez et al., 2018).

2. Depth and Volume: Photometric surveys can more readily reach deeper into space (i.e., they can detect fainter galaxies) than spectroscopic surveys. This means that with photometric redshifts, we can study galaxies that are further away, and thus, look further back in time, providing insights into the early universe.

3. Gravitational Lensing Studies: Photometric redshifts are essential for weak gravitational lensing studies, which probe the distribution of dark matter in the universe. Knowing the redshifts of background galaxies accurately helps in understanding the lensing signal and the mass distribution of foreground structures (Köhlinger, Hoekstra, and Eriksen, 2015), and can even help to constrain dark energy properties (Zhan and Knox, 2006).

4. Enhancing Spectroscopic Surveys: While photometric redshifts are typically less accurate than spectroscopic ones, they can be used to prioritize targets for follow-up

spectroscopic observations, ensuring that telescope time is used efficiently.

Photometric redshifts can be estimated by comparing the observed magnitudes and colours of galaxies, which are affected by the redshift of their spectral features, with the colours predicted by models or templates of galaxy spectra at different redshifts. Since these magnitudes are effectively coarse sampling of the underlying spectral energy distribution (SED), this method, known as *template fitting*, begins with the creation of SED templates which encapsulate the intrinsic luminosity of galaxies at zero redshift. They are either derived from theoretical models that simulate galaxy spectra based on our understanding of stellar populations and galaxy evolution (e.g., Bruzual and Charlot, 1993) or are constructed from empirical data of well-studied galaxies (e.g., Coleman, Wu, and Weedman, 1980).

The template fitting method has a long history of implementation (e.g., Puschell, Owen, and Laing, 1982; Arnouts et al., 1999; Bolzonella, Miralles, and Pelló, 2000; Feldmann et al., 2006; Mobasher et al., 2007; Dobos et al., 2012) and has achieved impressive precision in some cases (e.g., De Vicente, Sánchez, and Sevilla-Noarbe, 2016; Eriksen et al., 2019), but there are some notable limitations to the method. One significant assumption is that the templates used are representative of the full diversity of galaxy types and luminosities. If the templates are incomplete or biased, the resulting photometric redshifts can be skewed. Moreover, the effects of reddening and extinction due to dust within galaxies can alter their observed colours, potentially mimicking the effects of redshift. This requires careful modeling to disentangle the effects of dust from those of the cosmological redshift. Additionally, there is the issue of degeneracy, where different combinations of galaxy types, redshifts, and dust content can result in similar observed colours, making it challenging to determine the exact redshift (Cunha et al., 2010). Finally, comparing the templates to observed magnitudes requires considerations of effects from the Earth's atmosphere, the telescope optics, the shape of the filter transmission curves, and the efficiency and transmission of the camera itself (Salvato, Ilbert, and Hoyle, 2019). Limitations like these are why machine learning methods have been explored for photometric redshift estimates.

Similar to the template fitting method, supervised machine learning methods require a representative sample of galaxies to train on. The machine learning model uses the photometric observations – which have been paired with spectroscopic observations – to infer the spectroscopic redshifts. Once trained, the model can then infer redshifts for galaxies which have no spectroscopic measurements. In this data-driven context, the model learns all of the complexities of the modeling process of the template fitting method from the data itself, allowing for a mapping from photometry directly to redshifts. While earlier works

typically used the measured fluxes/magnitudes of galaxies as input data (e.g., Tagliaferri et al., 2003; Collister and Lahav, 2004; Bonnett, 2015; Sadeh, Abdalla, and Lahav, 2016), the success of image processing in machine learning has led to modern methods using the multi-band imaging data itself as input to neural networks (Hoyle, 2016; D'Isanto and Polsterer, 2018; Pasquet et al., 2019; Schuldt et al., 2021) and have been shown to outperform non-image-based methods.

One of the biggest limitations of supervised approaches is that they typically require a large number of spectroscopic redshifts, which might not always be available, to train on. They are also typically ineffective at extrapolating beyond the domain of redshift values trained on. In contrast, and as described in the previous section, self-supervised approaches can utilize the entire data set of unlabeled galaxies to learn physically meaningful representations of galaxies in a pretraining phase which also helps to make the model more robust to out-of-distribution uncertainty quantification (Hendrycks et al., 2019). Ideally, the trained model will contain an internal representation of redshift without ever explicitly being told about it, as shown in Hayat et al. (2021). It is then possible to fine-tune the trained model on a redshift estimation task using a smaller dataset than is required for a purely supervised method, and indeed can achieve higher precision and accuracy when using the same size dataset (Hayat et al., 2021).

I have now provided a summary of the importance of both morphologies and redshift values for a large sample of galaxies, and how self-supervised methods applied to multi-band imaging data can extract physically meaningful information about galaxies that is helpful for both tasks. The aim of this project was to verify the utility of self-supervised methods (specifically SimCLR – described in Section 4.3.1) for the UNIONS Survey (Chambers, Team, and Team, 2020), assess the feasibility of using the learned representations for automatically clustering and classifying galaxy types, improve photometric redshift estimations, and use these methods to help in finding more rare objects such as ultra-faint dwarf galaxies, gravitational lenses, and merging galaxies.

4.2 Data

The Ultraviolet Near-Infrared Optical Northern Survey (UNIONS) project is a large-scale multi-band astronomical survey aimed at securing optical imaging to complement the Euclid space mission, aimed at maximizing the science return of these large and deep surveys of the northern skies for both near- and far-field cosmology. One of its stated goals is studying the properties and evolution of galaxies over cosmic time by providing measurements of

weak lensing and photometric redshifts for millions of galaxies, ultimately helping to answer fundamental questions about dark matter and dark energy. It is a collaboration between several telescopes in Hawaii:

1. *Subaru Telescope*: a large optical and infrared telescope with an 8.2-metre primary mirror, it is equipped with several instruments, including the Hyper Suprime-Cam (HSC) – a wide-field 870-megapixel imaging camera with a field-of-view of 1.77 square degrees and which can observe in five different bands from the blue to the near-infrared. The g-band (referred to as `HSC-g` hereafter) was used in this study.

2. *Canada-France-Hawaii Telescope (CFHT)*: a 3.6-metre optical and near-infrared telescope jointly operated by the National Research Council of Canada, the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. One of the key instruments on the CFHT is MegaCam, a wide-field optical imaging camera with 40 CCD detectors giving it a field of view of 1 square degree with 380 megapixels. MegaCam is mapping the sky in the u- and r-bands as part of the Canada France Imaging Survey (CFIS), and both bands were used in this study (referred to as `CFIS-u` and `CFIS-LSBr` hereafter, noting that LSB refers to the low-surface-brightness version of the data reduction pipeline)

3. *Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)*: a wide-field imaging facility with a 1.8-metre primary mirror and 1.4-gigapixel camera that surveys in five wavelength bands, from visible to near-infrared, with a primary goal of conducting a comprehensive survey of the northern sky to search for near-Earth objects that threaten impacts. Pan-STARRS consists of two separate but identical telescopes, PS1 and PS2, and the i- and z-bands of PS1 (hereafter referred to as `PS1-i` and `PS1-z`) were used in this study.

UNIONS combines the strengths of each telescope to collect multi-band photometric imaging data covering 4800 square degrees in the northern hemisphere, organized in (mostly) overlapping tiles of the same size which can be processed independently – though not every tile has observations from every telescope. All of the images from the three telescopes were resampled and stacked onto the same sky projection, giving a consistent plate scale of 0.2"/pixel. For the images to be used in a self-supervised framework, it was necessary to extract cutouts of each galaxy in each band to be used as input to the model. `SExtractor` was utilized for this task, and it was used to extract the coordinates of 600 million identified sources and generate a catalogue of right ascension, declination, and multi-band magnitude

data for each one. A couple cuts were made to this catalogue to curate two smaller catalogues of data:

1. Only sources with simultaneous observations in `CFIS-LSBr`, `CFIS-u`, `HSC-g`, `PS1-i`, and `PS1-z` bands were kept, resulting in a *test catalogue* of just over 7 million sources.

2. The $m_{\texttt{CFIS-LSBr}}$ magnitudes were split into 8 bins in the range $14 \geq m_{\texttt{CFIS-LSBr}} \geq 23$, and a similar number of sources ($\sim$125,000) were collected from each bin to ensure a more uniform distribution of galaxy magnitudes were selected (otherwise most of the galaxies would be faint and small, only occupying a few pixels). This resulted in a *training catalogue* of 900,000 sources, itself a subset of the test catalogue.

From these curated catalogues, the `SExtractor` coordinates were used to extract 200pix^2 cutouts – corresponding to a 40" field-of-view – of each source, in each band, to be used as training and test sets in the machine learning pipeline. Because there were millions of cutouts to extract, the operation was parallelized using the Slurm scheduler on Compute Canada servers, which temporarily downloaded the relevant UNIONS tiles from VOSpace, made cutouts from each band of each source, and saved them to hdf5 files.

4.3 Methods

4.3.1 Contrastive Learning

Contrastive learning is a type of self-supervised learning that is trained to be invariant to different views (i.e. transformations) of the input data, without using any labels (see Huertas-Company, Sarmiento, and Knapen, 2023, for a review of contrastive learning in astrophysics). This approach involves learning a function, $f : \mathcal{X} \to \mathbb{R}^d$ that maps an input x from the space $\mathcal{X}$ to a lower-dimensional embedding vector $z = f(x)$. To do this, the model is trained on *positive* pairs of examples (x_i, x_i'), where x_i and x_i' are transformed versions of the same input, and *negative* pairs of examples (x_i, x_k), where x_i and x_k are dissimilar examples.

The training process for contrastive self-supervised learning typically involves minimizing a *contrastive* loss function (Hadsell, Chopra, and LeCun, 2006), encouraging the model to learn representations of positive pairs which are projected into the same region in $\mathbb{R}^d$, and representations of negative pairs which are separated in $\mathbb{R}^d$. This is what allows the model to be invariant to different augmentations/transformations of the same image, while also being discriminative between different images.

The SimCLR model

The Simple Framework for Contrastive Learning of Representations (SimCLR; Chen et al. 2020) framework is a prominent example of contrastive self-supervised learning, and was adopted for this project. It simplifies the process of learning representations by eliminating the need for specialized architectures or a memory bank of representations, a major limitation of earlier methods (Wu et al., 2018; Bachman, Hjelm, and Buchwalter, 2019). Instead, SimCLR uses a standard neural network base encoder to acquire embedded representations and a small neural network projection head to map the representations to the space where contrastive loss is applied.

The loss function used by SimCLR is a specific form of the noise contrastive estimation (NCE) loss, namely InfoNCE. The InfoNCE loss aims to learn representations such that positive pairs are closer together in the embedding space, while negative pairs are farther apart, and is defined as follows:

$$\mathcal{L}_{InfoNCE} = -\log \frac{\exp\left(\text{sim}(z_i, z_i')/\tau\right)}{\sum_{j=1}^{2N} \mathbb{1}_{[j\neq i]} \exp\left(\text{sim}(z_i, z_j)/\tau\right)}$$

where z_i and z_i' are two embedding vectors generated from augmented views of the same image, $\text{sim}(z_i, z_j) = \frac{z_i^T z_j}{|z_i||z_j|}$ is the cosine similarity between the two vectors, τ is a temperature parameter that controls the softness of the comparison, and $\mathbb{1}_{[i\neq j]}$ is an indicator function that ensures that the similarity between z_i and itself is not included in the denominator. It encourages the model to increase the similarity between the positive pairs since higher similarity leads to a higher numerator and thus a lower loss. It also encourages the model to decrease the similarity between the negative pairs since lower similarity decreases the denominator and thus decreases the loss.

The loss is computed again with the positions of the images interchanged, and then computed over all pairs in the batch of size N and the average is taken, resulting in the batch loss:

$$\mathcal{L}_{InfoNCE,batch} = \frac{1}{2N} \sum_{k=1}^{N} [\mathcal{L}_{InfoNCE}(2k-1, 2k) + \mathcal{L}_{InfoNCE}(2k, 2k-1)]$$

The transformations used in this study were chosen so that the images changed in ways that were physically plausible, while still preserving the identity and overall morphology of the objects being studied. By using a diverse set of data augmentation techniques, the contrastive self-supervised learning model can learn to identify the underlying physical properties

of the objects, rather than simply memorizing non-physical features of the images, though of course residual non-informative features may still be captured if the transformations are not exhaustive. Without these transformations the model might, for example, treat galaxies with a similar signal-to-noise level as similar to each other, which would be undesirable. The transformations are as follows:

1. *Galactic extinction*: the E(B-V) values of each source were obtained via their sky coordinates and the 2013 dust map of the Planck Collaboration (Abergel et al., 2014), then fit with a log-normal distribution which was uniformly sampled from. The reddening factors in each band are calculated using the conversion factors from Schlafly and Finkbeiner (2011) and the sampled E(B-V) via $10^{-(\text{conversion factor})\times(\text{sampled E(B-V)})/2.5}$. The images are multiplied by this factor and used to scale the brightness of images in each band to simulate the effect of dust.

2. *Translation*: The image cutout was shifted by up to 6.5% its width/height to account for any errors in centering.

3. *Rotation*: An angle was uniformly sampled from 0 to 180 degrees and applied to the image to ensure the orientation of the galaxy did not matter.

4. *Centre Crop*: The previous transformations caused some pixels to be blank, so the image was centre cropped from 200x200 pixels to 140x140 pixels.

5. *Horizontal Flip*: The image was flipped horizontally with a 50% probability.

6. *Gaussian blur*: A blurring kernel with a sigma uniformly sampled between 0.01 and 0.7 pixels was applied to account for differences in image quality.

7. *Gaussian noise*: The median absolute deviation of pixel values in each band across all image cutouts was calculated and used to create a normal distribution that was sampled from and added to the image.

An example of the transformations applied is shown in Figure 4.1.

The base encoder/model used for mapping the input to z in this study was the `ResNet-50` architecture (He et al., 2016), and the VISSL library (Goyal et al., 2021) was used with default settings for the training process (e.g., temperature of 0.1, an output size of 2048 for the base encoder, an embedding dimension of 128 for the multi-layer perceptron (MLP) projection head, stochastic gradient descent with a cosine learning rate scheduler and initial learning rate of 0.6). Training the model for a total of 100 epochs took approximately 96 hours using two NVIDIA V100 Volta GPUs.

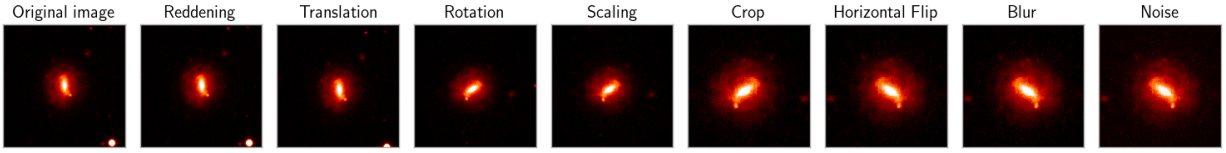

Figure 4.1: The transformations applied to each galaxy cutout for the contrastive self-supervised model

4.4 Results

Once the model was trained on the 900,000 curated multi-band UNIONS cutouts, the embedded representations were computed and saved for all cutouts in the dataset of 7 million sources. The representations were then compared via the cosine similarity metric in Pytorch:

$$\text{similarity} = \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\max(\|\mathbf{x}_1\|_2 \cdot \|\mathbf{x}_2\|_2, \epsilon)}$$

where $\|\mathbf{x}_1\|_2$ and $\|\mathbf{x}_2\|_2$ are the L2 norms of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, and ϵ is a small positive number to prevent division by zero in case the norms of the vectors are zero. A value of 1 means the representations are identical, whereas a value of -1 means they are diametrically opposed, so it can be used as a way of finding, for a given galaxy cutout, the most similar galaxies in the dataset.

Figures 4.2 and 4.3 show examples of the top 9 most similar galaxies to two different reference galaxies. In Figure 4.2 the model correctly identified that the loose spiral arms of the galaxy were important, and found several other spiral galaxies with loosely bound arms. In Figure 4.3 the compact core of the galaxy was a strong identifier, and indeed many examples of spiral galaxies with compact cores were found.

I believe these limited results are encouraging, showing that the method is working to identify galaxies with similar morphologies independent of their orientation – at least for brighter galaxies over a modest range of angular sizes (2" - 6") – however some of the transformations can be modified to better represent the data. For example, for the sake of brevity and ease of implementation, the same kernel was used on all 5 bands during the Gaussian blur transformation. This is obviously not physically justified since the images were taken with different telescopes at different times such that the point-spread function (PSF) is uncorrelated between bands, and there is additionally known to be wavelength-dependent PSF effects (Xin et al., 2018). The range in sigmas can likely be better constrained to fit the measured PSF distribution of the dataset.

4.5 Future Work

Once the model has been trained optimally with more realistic and physically motivated augmentations, the embedded representations can be used to locate specific objects like merged galaxies, gravitational lenses, and dwarf galaxies in a self-similarity search. This approach can be further enhanced by fine-tuning the model in a classification scheme by e.g., using known mergers from datasets like UNIONS. Both the self-similarity search and the fine-tuning method offer promising avenues for discovering new galaxies, and their effectiveness could be benchmarked against existing identification techniques. For example, Wilkinson et al. (2022) find that their CNN classification method misses about 70% of known post-mergers in a simulated dataset, so there is a lot of room for improvement.

The quest for identifying ultra-faint dwarf (UFD) galaxies around the Milky Way provides an interesting context for applying these methods. Traditionally, the "matched filter" technique (MFT) has been highly successful in this domain (e.g., Martin, Jong, and Rix, 2008; Bechtol et al., 2015; Drlica-Wagner et al., 2015; Smith et al., 2023), in which an isochrone iteratively stepped to larger distances is used to select stars from pre-detected and classified catalogues based on both their proximity to it in a colour-magnitude diagram (CMD) and their statistical spatial overdensity with respect to the local stellar density. In the current iteration of the SimCLR method, the angular extent of the cutouts (28 arcsec) is smaller than the size of some of the UFD galaxies found (e.g., ∼180 arcsec; Smith et al. 2023), so at best the cutouts would only capture the inner regions of these objects and potentially result in difficulties in proper identification. Nevertheless, the SimCLR method has an advantage of using imaging data in multiple bands (and is not dependent on the output of a source finding algorithm), so it utilizes more information than the MFT and indeed may be able to identify faint or diffuse features not accounted for in the MFT. UFDs, likely occupying a unique region in the "embedded representation space" of the SimCLR model, could be identified through self-similarity searches. It should be interesting to compare the MFT and SimCLR methods, and indeed larger cutouts could be used if the performance is unsatisfactory.

The trained SimCLR model will also be fine-tuned on spectroscopic redshifts that have been matched to objects in the UNIONS Survey, creating a label transfer pipeline that can predict accurate photometric redshifts directly from multi-band images. This task is a priority since acquiring accurate photometric redshifts is one of the goals of the UNIONS survey.

Finally, a hierarchical clustering algorithm (e.g. Chami et al., 2020) will be applied to the

embedded representations so that similar galaxies will be clustered together in branches of increasing specificity, resulting in an automated hierarchical classification scheme. A similar method has been used successfully for Hubble Space Telescope data (e.g., Peth et al., 2016; Hocking et al., 2018) in which the automated classifications closely resembled those used in Galaxy Zoo, and was used as a way to identify the nodes in the hierarchy that contained rare types of galaxies and/or anomalies.

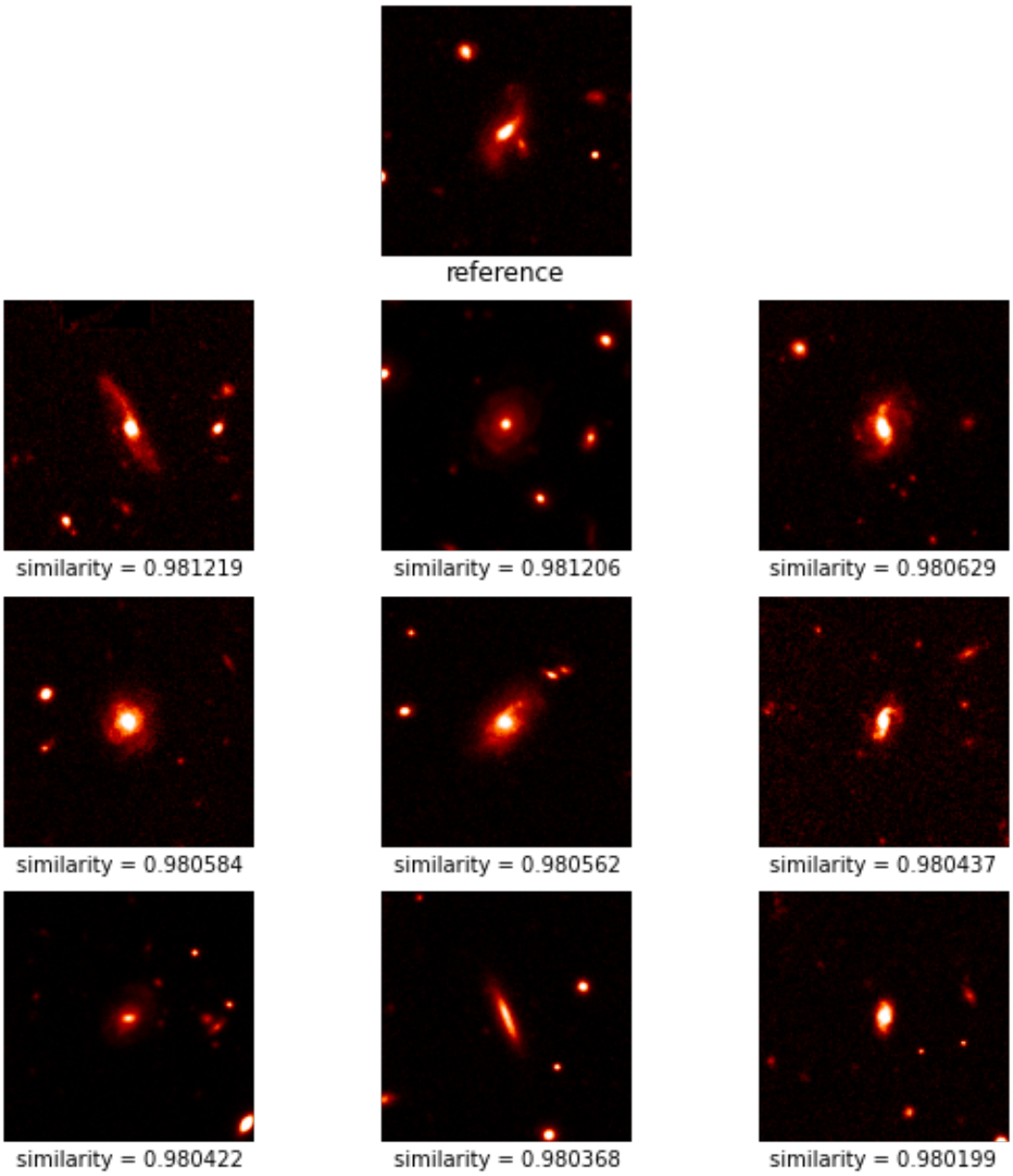

Figure 4.2: The top 9 similar galaxies to a chosen reference, selected through comparing cosine similarity of the embedded vectors generated by a trained SimCLR model.

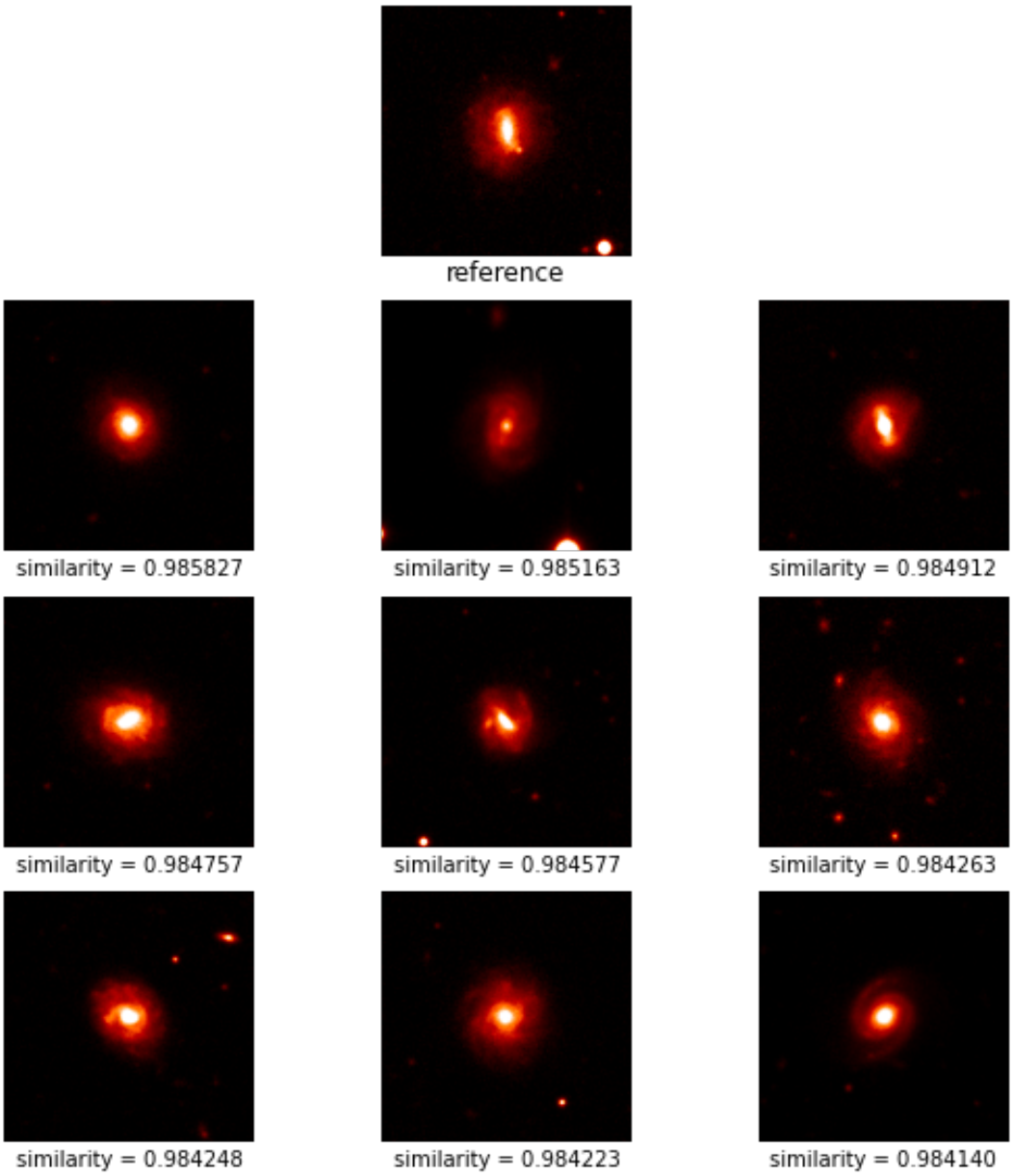

Figure 4.3: The top 9 similar galaxies to a chosen reference, selected through comparing cosine similarity of the embedded vectors generated by a trained SimCLR model.

Chapter 5

Conclusion

In this dissertation, I presented three projects that harness the power of machine learning to address pressing challenges in the realm of astronomy. Each endeavor, while distinct in its challenges and methodologies, is bound by a common thread: the pursuit of harnessing computational intelligence to help decipher the mysteries of the cosmos. In my first project, I confront the imminent challenge of contamination from low earth orbit satellites in upcoming massive spectroscopic sky surveys. By employing convolutional neural networks and a Wave U-Net, I not only detect and predict stellar parameters but also effectively remove satellite contamination. Thanks to my research on this topic, I was invited to join the International Astronomical Union's Centre for the Protection of the Dark and Quiet Sky from Satellite Constellation Interference (IAU-CPS), and I presented my research at their first symposium on the topic[1]. It seems machine learning will play an important role in helping to solve this issue, and my methods can be directly implemented in the WEAVE Survey and others like it.

In the same vein, my second project addresses "contamination" of astronomical images by the atmosphere. Here, a U-Net architecture helps demonstrate that with the right training and approach, machine learning can significantly enhance image quality, even in the most challenging conditions. We currently have a prototype system set up at the C2PU telescope on the Plateau de Calern in Frace, and more planned observations at the end of November will allow testing the model on a larger variety of real data. Our code will be open-sourced on Github, enabling these methods to be adopted by others in the astronomical community and even further developed for a variety of sites.

My final project aims to extract meaningful representations of galaxies in the UNIONS Survey with self-supervised learning to automate the morphological classification of galaxies

[1] see symposium website: https://research.iac.es/congreso/iaus385/

and predict precise photometric redshifts. The results, while promising, also highlight the iterative nature of machine learning – a continuous cycle of learning, refining, and reapplying. I foresee this method being a very useful tool for the UNIONS Survey, as well as the upcoming Euclid and LSST Surveys, helping researchers find more galaxies and galactic phenomena of interest, as well as fully exploiting the spectroscopic and photometric data together to make precise redshift predictions for billions of galaxies.

5.0.1 Future Applications and Reflections

As I reflect on the trajectory of machine intelligence in astronomy, I am both awed and contemplative. The advent of Large Language Models (LLMs) and agents with multimodal capacities (e.g., Gato and PaLM-E, Reed et al., 2022; Driess et al., 2023), along with the proven ability of deep learning to both significantly advance long-standing scientific problems (e.g., protein folding and fusion reactor design, Jumper et al., 2021; Degrave et al., 2022) and out-perform humans at tasks previously thought to forever be solely human-centric (e.g., the game of Go, and analysis of medical data, Silver et al., 2017; Morrow and Sormani, 2020), it is hard to imagine a future in which astronomy and machine learning do not become further intricately entwined, reshaping the landscape of astronomical research. LLMs in particular, with their vast knowledge bases and intricate pattern recognition capabilities, are showing promise in assisting astronomers in formulating research questions and knowledge acquisition. Their ability to rapidly assimilate, process, and generate insights from vast amounts of data is unparalleled.

While the capabilities of LLMs and other machine learning tools are potentially profound, I believe they will not replace astronomers but rather augment their capabilities, perhaps helping to define new ways of conducting research. The human element – an innate curiosity, the ability to ask impactful questions, and the passion for discovery – currently remains irreplaceable. What LLMs offer is a partnership; they serve as invaluable assistants, helping to refine hypotheses, expedite data analysis, and even suggest novel avenues of exploration. In essence, they will amplify the capabilities of astronomers, allowing us to delve deeper and reach further.

This process will not happen without a lot of hesitation, hurtles, and doubt. Having made my foray into machine learning in 2016, at a time when not every astronomer had heard of neural networks, I witnessed firsthand the strong skepticism and push-back of the integration of these novel methodologies in existing problems in astronomy. The skepticism was of course warranted; the excitement of new technologies sometimes results in people over-

selling their significance and blindly applying the methods without understanding them. Indeed, a workshop[2] hosted last year focused on addressing the "reproducibility crisis" in science that has resulted from improper utilization of machine learning techniques (e.g. no train-test splits, duplicate samples in the train-test splits, and the model using illegitimate features). Unfortunately, machine learning has often been used "just because" and the results accepted based on excitement alone. There needs to be a clear reason – beyond following the trends of the day – one would adopt machine learning for their problem, and standards of implementation need to be adhered to[3]. It is especially problematic when the "black-box" nature of machine learning models is ignored in any implementation. Aside from rigorous analysis of results, skepticism will only be reduced if a robust estimation of uncertainty and/or explainability are incorporated. Luckily these are active areas of development in machine learning (e.g., Došilović, Brčić, and Hlupić, 2018; Xu et al., 2019; Abdar et al., 2021), and no doubt the research will be incorporated into the physical sciences.

I am thrilled to continue witnessing the transformative power of machine learning in astronomy and society at large, and I feel extremely grateful that I can contribute to that journey, but it is imperative to approach this union with both enthusiasm and caution. The lessons learned, both from successes and missteps, have shaped my perspective and reinforced the importance of responsible and informed application of these tools. The future beckons with endless possibilities, and as researchers, we must tread this path with a commitment to integrity, collaboration, and continuous learning. As I close this chapter of my academic journey, I am filled with optimism for the road ahead, eager to further bridge the worlds of astronomy and machine learning, and hopeful that our collective efforts will illuminate the mysteries of the cosmos in ways we have yet to imagine.

[2] see workshop website: https://sites.google.com/princeton.edu/rep-workshop

[3] see the "rules of machine learning" as laid out by Google for an example: https://developers.google.com/machine-learning/guides/rules-of-ml

Bibliography

Abadi, Martín et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: http://tensorflow.org/.

Abdar, Moloud et al. (2021). "A review of uncertainty quantification in deep learning: Techniques, applications and challenges". In: *Information fusion* 76, pp. 243–297.

Abdurrahman, FN et al. (2018). "Improved Image Quality over 10 Fields with the Imaka Ground-layer Adaptive Optics Experiment". In: *The Astronomical Journal* 156.3, p. 100.

Abergel, Ade et al. (2014). "Planck 2013 results. XI. All-sky model of thermal dust emission". In: *Astronomy & Astrophysics* 571, A11.

Accetta, Katherine et al. (2022). "The seventeenth data release of the Sloan Digital Sky Surveys: Complete release of MaNGA, MaStar, and APOGEE-2 data". In: *The Astrophysical Journal Supplement Series* 259.2, p. 35.

Aguado, David S et al. (2019). "The Pristine survey–VI. The first three years of medium-resolution follow-up spectroscopy of Pristine EMP star candidates". In: *Monthly Notices of the Royal Astronomical Society* 490.2, pp. 2241–2253.

Alam, Shadab et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". In: *Monthly Notices of the Royal Astronomical Society* 470.3, pp. 2617–2652.

Alam, Shadab et al. (2021). "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory". In: *Physical Review D* 103.8, p. 083533.

Alarcon, Miguel R. et al. (May 2023). "Scientific CMOS Sensors in Astronomy: IMX455 and IMX411". In: 135.1047, 055001, p. 055001. DOI: 10.1088/1538-3873/acd04a. arXiv: 2302.03700 [astro-ph.IM].

Alexander, Stephon et al. (Apr. 2020). "Deep Learning the Morphology of Dark Matter Substructure". In: 893.1, 15, p. 15. DOI: 10.3847/1538-4357/ab7925. arXiv: 1909.07346 [astro-ph.CO].

Alonso, David et al. (2015). "Ultra-large-scale cosmology in next-generation experiments with single tracers". In: *The Astrophysical Journal* 814.2, p. 145.

Angel, J. R. P. et al. (Nov. 1990). "Adaptive optics for array telescopes using neural-network techniques". In: 348.6298, pp. 221–224. DOI: 10.1038/348221a0.

Angelopoulos, Anastasios N and Stephen Bates (2021). "A gentle introduction to conformal prediction and distribution-free uncertainty quantification". In: *arXiv preprint arXiv:2107.07511*.

Aristidi, Eric et al. (2020). "Monitoring daytime and nighttime optical turbulence profiles with the PML instrument". In: *arXiv preprint arXiv:2002.04947*.

Aristidi, Eric et al. (Dec. 2020). "Turbulence monitoring at Calern observatory with the Generalised Differential Image Motion Monitor". In: *arXiv e-prints*, arXiv:2012.08818, arXiv:2012.08818. DOI: 10.48550/arXiv.2012.08818. arXiv: 2012.08818 [astro-ph.IM].

Arnouts, Stephane et al. (1999). "Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North". In: *Monthly Notices of the Royal Astronomical Society* 310.2, pp. 540–556.

Austin, Melissa A et al. (1988). "Low-density lipoprotein subclass patterns and risk of myocardial infarction". In: *Jama* 260.13, pp. 1917–1921.

Babcock, Horace W (1953). "The possibility of compensating astronomical seeing". In: *Publications of the Astronomical Society of the Pacific* 65.386, pp. 229–236.

Bachman, Philip, R Devon Hjelm, and William Buchwalter (2019). "Learning representations by maximizing mutual information across views". In: *Advances in neural information processing systems* 32.

Bailer-Jones, Coryn AL (2000). "Stellar parameters from very low resolution spectra and medium band filters: teff, logg and [m/h] using neural networks". In: *arXiv preprint astro-ph/0003071*.

Bailer-Jones, Coryn AL et al. (1997). "Physical parametrization of stellar spectra: the neural network approach". In: *Monthly Notices of the Royal Astronomical Society* 292.1, pp. 157–166.

Balaguera-Antolínez, A et al. (2018). "Extracting cosmological information from the angular power spectrum of the 2MASS Photometric Redshift catalogue". In: *Monthly Notices of the Royal Astronomical Society* 476.1, pp. 1050–1070.

Balogh, Michael L et al. (2004). "The bimodal galaxy color distribution: dependence on luminosity and environment". In: *The Astrophysical Journal* 615.2, p. L101.

Banerjee, Chaity, Tathagata Mukherjee, and Eduardo Pasiliao Jr (2019). "An empirical study on generalizations of the ReLU activation function". In: *Proceedings of the 2019 ACM Southeast Conference*, pp. 164–167.

Baranec, Christoph et al. (2016). "Robo-AO Kepler Planetary Candidate Survey. II. Adaptive Optics Imaging of 969 Kepler Exoplanet Candidate Host Stars". In: *The Astronomical Journal* 152.1, p. 18.

Bassa, C. G., O. R. Hainaut, and D. Galadí-Enríquez (Jan. 2022). "Analytical simulations of the effect of satellite constellations on optical and near-infrared observations". In: 657, A75, A75. DOI: 10.1051/0004-6361/202142101. arXiv: 2108.12335 [astro-ph.IM].

Bechtol, Keith et al. (2015). "Eight new Milky Way companions discovered in first-year Dark Energy Survey data". In: *The Astrophysical Journal* 807.1, p. 50.

Beckers, Jacques M (1988). "Increasing the size of the isoplanatic patch with multiconjugate adaptive optics". In: *Very Large Telescopes and their Instrumentation, Vol. 2*. Vol. 30, p. 693.

— (1993). "Adaptive optics for astronomy: principles, performance, and applications". In: *Annual review of astronomy and astrophysics* 31.1, pp. 13–62.

Bely, Pierre-Yves (1987). "Weather and seeing on Mauna Kea". In: *Publications of the Astronomical Society of the Pacific* 99.616, p. 560.

Bendjoya, Ph et al. (2012). "C2PU: An original mix of research and pedagogy at Observatoire de la Côte d'Azur". In: *proc. of the SF2A-2012*, pp. 643–648.

Bengio, Yoshua (2012). "Practical recommendations for gradient-based training of deep architectures". In: *Neural networks: Tricks of the trade* 7700, pp. 437–478. URL: https://arxiv.org/abs/1206.5533.

Bengio, Yoshua et al. (2013). "Generalized Denoising Auto-Encoders as Generative Models". In: *arXiv preprint arXiv:1305.6663*. URL: https://arxiv.org/abs/1305.6663.

Bertin, E. (July 1994). "Classification of Astronomical Images with a Neural Network". In: 217.1-2, pp. 49–51. DOI: 10.1007/BF00990023.

Bertin, Emmanuel and Stephane Arnouts (1996). "SExtractor: Software for source extraction". In: *Astronomy and astrophysics supplement series* 117.2, pp. 393–404.

Beskin, Grigory et al. (Dec. 2023). "SAINT (Small Aperture Imaging Network Telescope) – a wide-field telescope complex for detecting and studying optical transients at times from milliseconds to years". In: arXiv:2312.01369. arXiv:2312.01369 [astro-ph]. URL: http://arxiv.org/abs/2312.01369.

Betoule, Marc et al. (Feb. 2023). "StarDICE. I. Sensor calibration bench and absolute photometric calibration of a Sony IMX411 sensor". In: 670, A119, A119. DOI: 10.1051/0004-6361/202244973. arXiv: 2211.04913 [astro-ph.IM].

Bialek, Spencer et al. (Nov. 2020). "Assessing the performance of LTE and NLTE synthetic stellar spectra in a machine learning framework". In: 498.3, pp. 3817–3834. DOI: 10.1093/mnras/staa2582. arXiv: 1911.02602 [astro-ph.IM].

Bickley, Robert W. et al. (June 2021). "Convolutional neural network identification of galaxy post-mergers in UNIONS using IllustrisTNG". In: 504.1, pp. 372–392. DOI: 10.1093/mnras/stab806. arXiv: 2103.09367 [astro-ph.GA].

Bickley, Robert W et al. (2021). "Convolutional neural network identification of galaxy post-mergers in UNIONS using IllustrisTNG". In: *Monthly Notices of the Royal Astronomical Society* 504.1, pp. 372–392.

Biewald, Lukas (2020). *Experiment Tracking with Weights and Biases*. Software available from wandb.com. URL: https://www.wandb.com/.

Boley, A. et al. (2021). *Report on Mega-Constellations to the Government of Canada and the Canadian Space Agency*. Tech. rep. https://arxiv.org/abs/2104.05733.

Boley, Aaron C and Michael Byers (2021). "Satellite mega-constellations create risks in Low Earth Orbit, the atmosphere and on Earth". In: *Scientific Reports* 11.1, p. 10642.

Bolin, Bryce T et al. (2020). "Characterization of the nucleus, morphology, and activity of interstellar comet 2I/Borisov by optical and near-infrared GROWTH, Apache Point, IRTF, ZTF, and Keck observations". In: *The Astronomical Journal* 160.1, p. 26.

Bolzonella, Micol, Joan-Marc Miralles, and Roser Pelló (2000). "Photometric redshifts based on standard SED fitting procedures". In: *Arxiv preprint astro-ph/0003380*.

Bonaccini, Domenico et al. (2003). "The VLT laser guide star facility". In: *Adaptive Optical System Technologies II*. Vol. 4839. SPIE, pp. 381–392.

Bonnett, Christopher (2015). "Using neural networks to estimate redshift distributions. An application to CFHTLenS". In: *Monthly Notices of the Royal Astronomical Society* 449.1, pp. 1043–1056.

Borawar, Lokesh and Ravinder Kaur (2023). "ResNet: Solving Vanishing Gradient in Deep Networks". In: *Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022*. Springer, pp. 235–247.

Bottrell, Connor et al. (Dec. 2019). "Deep learning predictions of galaxy merger stage and the importance of observational realism". In: 490.4, pp. 5390–5413. DOI: 10.1093/mnras/stz2934. arXiv: 1910.07031 [astro-ph.GA].

Bottrell, Connor et al. (2019). "Deep learning predictions of galaxy merger stage and the importance of observational realism". In: *Monthly Notices of the Royal Astronomical Society* 490.4, pp. 5390–5413.

Bouchez, Antonin H et al. (2018). "An overview and status of GMT active and adaptive optics". In: *Adaptive Optics Systems VI* 10703, pp. 284–299.

Brandt, Timothy D (2016). "Constraints on MACHO dark matter from compact stellar systems in ultra-faint dwarf galaxies". In: *The Astrophysical Journal Letters* 824.2, p. L31.

Breiman, Leo (2001). "Random forests". In: *Machine learning* 45, pp. 5–32.

Brigham, E Oran and RE Morrow (1967). "The fast Fourier transform". In: *IEEE spectrum* 4.12, pp. 63–70.

Brunner, Elisabeth et al. (2021). "Retinal adaptive optics imaging with a pyramid wavefront sensor". In: *Biomedical Optics Express* 12.10, pp. 5969–5990.

Bruzual, Gustavo and Stephane Charlot (1993). "Spectral evolution of stellar populations using isochrone synthesis". In: *The Astrophysical Journal* 405, p. 538.

Bubeck, Sébastien et al. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. arXiv: 2303.12712 [cs.CL].

Burges, Christopher J and Bernhard Schölkopf (1996). "Improving the accuracy and speed of support vector machines". In: *Advances in neural information processing systems* 9.

Cabrera-Vives, Guillermo et al. (2017). "Deep-HiTS: Rotation Invariant Convolutional Neural Network for Transient Detection". In: *The Astrophysical Journal* 836.1, p. 97.

Calhoun, Zachary D et al. (2022). "Self-Supervised Encoders Are Better Transfer Learners in Remote Sensing Applications". In: *Remote Sensing* 14.21, p. 5500.

Cauchy, Augustin et al. (1847). "Méthode générale pour la résolution des systemes d'équations simultanées". In: *Comp. Rend. Sci. Paris* 25.1847, pp. 536–538.

Chambers, K, Unions Team Including Pan-Starrs Team, and Cfis Team (2020). "UNIONS—the Ultraviolet Near Infrared Optical Northern Survey". In: *American Astronomical Society meeting Abstracts# 235*. Vol. 235, pp. 154–04.

Chami, Ines et al. (2020). "From trees to continuous embeddings and back: Hyperbolic hierarchical clustering". In: *Advances in Neural Information Processing Systems* 33, pp. 15065–15076.

Chang, Yu-Yen et al. (Oct. 2022). "SDSS-IV MaNGA: Unveiling Galaxy Interaction by Merger Stages with Machine Learning". In: 937.2, 97, p. 97. DOI: 10.3847/1538-4357/ac8c27. arXiv: 2208.11132 [astro-ph.GA].

Chapman, Scott C et al. (2018). "The multi-object adaptive optics system for the GIRMOS spectrograph on Gemini-South". In: *Adaptive Optics Systems VI*. Vol. 10703. SPIE, pp. 527–535.

Chatterjee, Monish R and Fathi HA Mohamed (2014). "Split-step approach to electromagnetic propagation through atmospheric turbulence using the modified von Karman spectrum and planar apertures". In: *Optical Engineering* 53.12, pp. 126107–126107.

Chen, Ting et al. (2020). "A simple framework for contrastive learning of visual representations". In: *International conference on machine learning*. PMLR, pp. 1597–1607.

Cheng, Ting-Yun et al. (2021a). "Beyond the hubble sequence–exploring galaxy morphology with unsupervised machine learning". In: *Monthly Notices of the Royal Astronomical Society* 503.3, pp. 4446–4465.

Cheng, Ting-Yun et al. (2021b). "Galaxy morphological classification catalogue of the Dark Energy Survey Year 3 data with convolutional neural networks". In: *Monthly Notices of the Royal Astronomical Society* 507.3, pp. 4425–4444.

Cho, Kyunghyun et al. (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation". In: *arXiv preprint arXiv:1406.1078*.

Chollet, François (2015). *keras*. https://github.com/fchollet/keras.

Çiçek, Özgün et al. (2016). "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer, pp. 424–432.

Ciucă, Ioana et al. (June 2023). "Harnessing the Power of Adversarial Prompting and Large Language Models for Robust Hypothesis Generation in Astronomy". In: arXiv:2306.11648. URL: http://arxiv.org/abs/2306.11648.

Close, Laird Miller et al. (2000). "Search for asteroidal satellites using adaptive optics". In: *Adaptive Optical Systems Technology*. Vol. 4007. SPIE, pp. 796–802.

Coleman, GD, C-C Wu, and DW Weedman (1980). "Colors and magnitudes predicted for high redshift galaxies". In: *Astrophysical Journal Supplement Series, vol. 43, July 1980, p. 393-416. Research sponsored by the Koninklijke Nederlandse Akademie van Wetenschappen;* 43, pp. 393–416.

Collaboration, Gaia (2020). "VizieR Online Data Catalog: Gaia EDR3". In: *I/350yCatI/350*.

Colless, Matthew et al. (2001). "The 2df galaxy redshift survey: spectra and redshifts". In: *Monthly Notices of the Royal Astronomical Society* 328.4, pp. 1039–1063.

Collister, Adrian A and Ofer Lahav (2004). "ANNz: estimating photometric redshifts using artificial neural networks". In: *Publications of the Astronomical Society of the Pacific* 116.818, p. 345.

Conselice, Christopher J (2003). "The relationship between stellar light distributions of galaxies and their formation histories". In: *The Astrophysical Journal Supplement Series* 147.1, p. 1.

— (2014). "The evolution of galaxy structure over cosmic time". In: *Annual Review of Astronomy and Astrophysics* 52, pp. 291–337.

Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine learning* 20, pp. 273–297.

Croft, Rupert AC et al. (2009). "Galaxy morphology, kinematics and clustering in a hydrodynamic simulation of a Λ cold dark matter universe". In: *Monthly Notices of the Royal Astronomical Society* 400.1, pp. 43–67.

Cunha, Elisabete da et al. (2010). "New insight into the relation between star formation activity and dust content in galaxies". In: *Monthly Notices of the Royal Astronomical Society* 403.4, pp. 1894–1908.

Dai, JM and J Tong (2018). "Galaxy Morphology Classification with Deep Convolutional Neural Networks National Space Science Center". In: *Chinese Academy of Sciences, China*. University of Chinese Academy of Sciences.

Dalton, Gavin et al. (2018). "Construction progress of WEAVE: the next generation wide-field spectroscopy facility for the William Herschel Telescope". In: *Ground-based and Airborne Instrumentation for Astronomy VII*. Vol. 10702. SPIE, pp. 388–397.

Davari, Roozbeh et al. (2014). "How robust are the size measurements of high-redshift compact galaxies?" In: *The Astrophysical Journal* 787.1, p. 69.

Davies, Andrew, Stephen Serjeant, and Jane M. Bromley (Aug. 2019). "Using convolutional neural networks to identify gravitational lenses in astronomical images". In: 487.4, pp. 5263–5271. DOI: 10.1093/mnras/stz1288. arXiv: 1905.04303 [astro-ph.IM].

Davies, Richard and Markus Kasper (2012). "Adaptive optics for astronomy". In: *Annual Review of Astronomy and Astrophysics* 50, pp. 305–351.

De Jong, Kenneth (1988). “Learning with genetic algorithms: An overview”. In: *Machine learning* 3, pp. 121–138.

de Jong, R. S. et al. (2019). “4MOST: Project overview and information for the First Call for Proposals”. In: *The Messenger* 175, pp. 3–11. DOI: 10.18727/0722-6691/5117. arXiv: 1903.02464 [astro-ph.IM].

De Vaucouleurs, Gérard (1959). “Classification and morphology of external galaxies”. In: *Astrophysik iv: Sternsysteme/astrophysics iv: Stellar systems*, pp. 275–310.

De Vicente, Juan, Eusebio Sánchez, and Ignacio Sevilla-Noarbe (2016). “DNF–Galaxy photometric redshift by Directional Neighbourhood Fitting”. In: *Monthly Notices of the Royal Astronomical Society* 459.3, pp. 3078–3088.

Degrave, Jonas et al. (2022). “Magnetic control of tokamak plasmas through deep reinforcement learning”. In: *Nature* 602.7897, pp. 414–419.

Dekany, Richard et al. (2013). “PALM-3000: exoplanet adaptive optics for the 5 m Hale telescope”. In: *The Astrophysical Journal* 776.2, p. 130.

DeMeo, Francesca E and Benoıt Carry (2014). “Solar System evolution from compositional mapping of the asteroid belt”. In: *Nature* 505.7485, pp. 629–634.

Descamps, Pascal and Franck Marchis (2008). “Angular momentum of binary asteroids: Implications for their possible origin”. In: *Icarus* 193.1, pp. 74–84.

Devlin, Jacob et al. (2018). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *arXiv preprint arXiv:1810.04805*.

Dey, Arjun et al. (May 2019). “Overview of the DESI Legacy Imaging Surveys”. In: 157.5, 168, p. 168. DOI: 10.3847/1538-3881/ab089d. arXiv: 1804.08657 [astro-ph.IM].

Di Dio, Enea et al. (2014). “Cosmological parameter estimation with large scale structure observations”. In: *Journal of Cosmology and Astroparticle Physics* 2014.01, p. 042.

Di Vruno, Federico and Vincenza Tornatore (2023). “Large Satellite Constellations and Their Potential Impact on VGOS Operations”. In: *International VLBI Service for Geodesy and Astrometry 2022 General Meeting Proceedings*, pp. 19–28.

Diolaiti, EMILIANO et al. (2016). “MAORY: adaptive optics module for the E-ELT”. In: *Adaptive Optics Systems V*. Vol. 9909. SPIE, pp. 768–774.

Dipper, Nigel et al. (2013). "Adaptive optics real-time control systems for the E-ELT". In: *Proceedings of the Third AO4ELT Conference*. Vol. 1, p. 41.

Dobos, László et al. (2012). "A high-resolution atlas of composite Sloan Digital Sky Survey galaxy spectra". In: *Monthly Notices of the Royal Astronomical Society* 420.2, pp. 1217–1238.

Doersch, Carl and Andrew Zisserman (2019). *Sim2real transfer learning for 3D human pose estimation: motion to the rescue*. arXiv: 1907.02499 [cs.CV].

Došilović, Filip Karlo, Mario Brčić, and Nikica Hlupić (2018). "Explainable artificial intelligence: A survey". In: *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*. IEEE, pp. 0210–0215.

Dressler, Alan (1980). "Galaxy morphology in rich clusters-Implications for the formation and evolution of galaxies". In: *Astrophysical Journal, Part 1, vol. 236, Mar. 1, 1980, p. 351-365.* 236, pp. 351–365.

Driess, Danny et al. (2023). "Palm-e: An embodied multimodal language model". In: *arXiv preprint arXiv:2303.03378*.

Drlica-Wagner, Alex et al. (2015). "Eight ultra-faint galaxy candidates discovered in year two of the dark energy survey". In: *The Astrophysical Journal* 813.2, p. 109.

Duplevska, Diana et al. (2022). "Sim2Real image translation to improve a synthetic dataset for a bin picking task". In: *IEEE*. DOI: 10.1109/ETFA52439.2022.9921431. URL: https://dx.doi.org/10.1109/ETFA52439.2022.9921431.

D'Isanto, Antonio and Kai Lars Polsterer (2018). "Photometric redshift estimation via deep learning-generalized and pre-classification-less, image based, fully probabilistic redshifts". In: *Astronomy & Astrophysics* 609, A111.

Ericsson, Linus et al. (2022). "Self-supervised representation learning: Introduction, advances, and challenges". In: *IEEE Signal Processing Magazine* 39.3, pp. 42–62.

Eriksen, Martin et al. (2019). "The PAU Survey: early demonstration of photometric redshift performance in the COSMOS field". In: *Monthly Notices of the Royal Astronomical Society* 484.3, pp. 4200–4215.

Escamilla-Rivera, Celia, Maryi A. Carvajal Quintero, and Salvatore Capozziello (Mar. 2020). "A deep learning approach to cosmological dark energy models". In: 2020.3, 008, p. 008. DOI: 10.1088/1475-7516/2020/03/008. arXiv: 1910.02788 [astro-ph.CO].

Etsebeth, Verlon et al. (2023). *Astronomaly at Scale: Searching for Anomalies Amongst 4 Million Galaxies*. arXiv: 2309.08660 [astro-ph.IM].

Fabbro, S. et al. (2018). "An application of deep learning in the analysis of stellar spectra". In: 475.3, pp. 2978–2993. DOI: 10.1093/mnras/stx3298. arXiv: 1709.09182 [astro-ph.IM].

Faber, SM and Robert E Jackson (1976). "Velocity dispersions and mass-to-light ratios for elliptical galaxies". In: *The Astrophysical Journal* 204, pp. 668–683.

Faedi, Francesca et al. (2013). "Lucky imaging of transiting planet host stars with LuckyCam". In: *Monthly Notices of the Royal Astronomical Society* 433.3, pp. 2097–2106.

Feldmann, R et al. (2006). "The zurich extragalactic bayesian redshift analyzer and its first application: COSMOS". In: *Monthly Notices of the Royal Astronomical Society* 372.2, pp. 565–577.

Ferreira, Leonardo et al. (June 2020). "Galaxy Merger Rates up to z $\sim$ 3 Using a Bayesian Deep Learning Model: A Major-merger Classifier Using IllustrisTNG Simulation Data". In: 895.2, 115, p. 115. DOI: 10.3847/1538-4357/ab8f9b. arXiv: 2005.00476 [astro-ph.GA].

Fielding, Ezra, Clement N Nyirenda, and Mattia Vaccari (2022). "The Classification of Optical Galaxy Morphology Using Unsupervised Learning Techniques". In: *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, pp. 1–6.

Flaugher, Brenna (Jan. 2005). "The Dark Energy Survey". In: *International Journal of Modern Physics A* 20.14, pp. 3121–3123. DOI: 10.1142/S0217751X05025917.

Fluke, Christopher J and Colin Jacobs (2020). "Surveying the reach and maturity of machine learning and artificial intelligence in astronomy". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.2, e1349.

Foy, R and Astron Labeyrie (1985). "Feasibility of adaptive telescope with laser probe". In: *Astronomy and Astrophysics (ISSN 0004-6361), vol. 152, no. 2, Nov. 1985, p. L29-L31.* 152, pp. L29–L31.

Frebel, Anna and John E Norris (2015). "Near-field cosmology with extremely metal-poor stars". In: *Annual Review of Astronomy and Astrophysics* 53, pp. 631–688.

Freeman, Ken and Joss Bland-Hawthorn (2002). "The new galaxy: Signatures of its formation". In: *Annual Review of Astronomy and Astrophysics* 40.1, pp. 487–537.

Fremling, Christoffer et al. (Aug. 2021). "SNIascore: Deep-learning Classification of Low-resolution Supernova Spectra". In: 917.1, L2, p. L2. DOI: 10.3847/2041-8213/ac116f. arXiv: 2104.12980 [astro-ph.IM].

Freund, Yoav and Robert E Schapire (1995). "A desicion-theoretic generalization of on-line learning and an application to boosting". In: *European conference on computational learning theory*. Springer, pp. 23–37.

Fried, David L (1965). "Statistics of a geometric representation of wavefront distortion". In: *JoSA* 55.11, pp. 1427–1435.

— (1966). "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures". In: *JOSA* 56.10, pp. 1372–1379.

— (1967). "Optical heterodyne detection of an atmospherically distorted signal wave front". In: *Proceedings of the IEEE* 55.1, pp. 57–77.

Fu, KC (1968). *Sequential methods in pattern recognition and machine learning*. Academic press.

Fusco, Thierry et al. (2010). "ATLAS: the E-ELT laser tomographic adaptive optics system". In: *Adaptive Optics Systems II*. Vol. 7736. SPIE, pp. 151–162.

Fusco, Thierry et al. (2022). "Key wavefront sensors features for laser-assisted tomographic adaptive optics systems on the Extremely Large Telescope". In: *Journal of Astronomical Telescopes, Instruments, and Systems* 8.2, pp. 021514–021514.

Gaensler, Bryan M and Pauline Barmby (2020). "The Canadian Astronomy Long Range Plan 2020-2030". In: *American Astronomical Society Meeting Abstracts# 235*. Vol. 235, pp. 290–01.

Geem, Zong Woo, Joong Hoon Kim, and Gobichettipalayam Vasudevan Loganathan (2001). "A new heuristic optimization algorithm: harmony search". In: *simulation* 76.2, pp. 60–68.

Ghedina, Adriano et al. (2003). "On-sky test of the pyramid wavefront sensor". In: *Adaptive Optical System Technologies II*. Vol. 4839. SPIE, pp. 869–877.

Ghosh, Aritra et al. (2020). "Galaxy Morphology Network: A Convolutional Neural Network Used to Study Morphology and Quenching in 100,000 SDSS and 20,000 CANDELS Galaxies". In: *The Astrophysical Journal* 895.2, p. 112.

Gieseke, Fabian et al. (2017). "Convolutional neural networks for transient candidate vetting in large-scale surveys". In: *Monthly Notices of the Royal Astronomical Society* 472.3, pp. 3101–3114.

Giles, Colin, G. Kuhn, and Ronald J. Williams (1994). "Dynamic recurrent neural networks: Theory and applications". In: *IEEE Transactions on Neural Networks*. DOI: 10.1109/TNN.1994.8753425.

Giovanelli, Riccardo and Martha P Haynes (1991). "Redshift surveys of galaxies". In: *Annual review of astronomy and astrophysics* 29.1, pp. 499–541.

Glorot, Xavier and Yoshua Bengio (2010). "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256.

Goodfellow, Ian et al. (2014). "Generative adversarial nets". In: *Advances in neural information processing systems*. Vol. 27.

Gourlay, James et al. (1997). "A real-time closed-loop liquid crystal adaptive optics system: first results". In: *Optics communications* 137.1-3, pp. 17–21.

Goyal, Priya et al. (2019). "Scaling and benchmarking self-supervised visual representation learning". In: *Proceedings of the ieee/cvf International Conference on computer vision*, pp. 6391–6400.

Goyal, Priya et al. (2021). *VISSL*. https://github.com/facebookresearch/vissl.

Griest, Kim et al. (1991). "Gravitational microlensing as a method of detecting disk dark matter and faint disk stars". In: *The Astrophysical Journal* 372, pp. L79–L82.

Grigg, Dylan et al. (Sept. 2023). "Detection of intended and unintended emissions from Starlink satellites in the SKA-Low frequency range, at the SKA-Low site, with an SKA-

Low station analog". In: arXiv:2309.15672. arXiv:2309.15672 [astro-ph]. DOI: 10.48550/arXiv.2309.15672. URL: http://arxiv.org/abs/2309.15672.

Guastavino, S. et al. (2021). "Implementation paradigm for supervised flare forecasting studies: a deep learning application with video data". In: URL: https://dx.doi.org/10.1051/0004-6361/202243617.

Guidash, Michael et al. (2016). "Reduction of CMOS image sensor read noise to enable photon counting". In: *Sensors* 16.4, p. 517.

Guiglion, G. et al. (2023). *Beyond Gaia DR3: tracing the* $[\alpha/M]-[M/H]$ *bimodality from the inner to the outer Networks*. arXiv: 2306.05086 [astro-ph.GA].

Guiglion, Guillaume et al. (2020). "The RAdial Velocity Experiment (RAVE): Parameterisation of RAVE spectra based on convolutional neural networks". In: *Astronomy & Astrophysics* 644, A168.

Gunn, James E et al. (1998). "The Sloan digital sky survey photometric camera". In: *The Astronomical Journal* 116.6, p. 3040.

Hadsell, Raia, Sumit Chopra, and Yann LeCun (2006). "Dimensionality reduction by learning an invariant mapping". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. IEEE, pp. 1735–1742.

Hainaut, Olivier R. and Andrew P. Williams (Apr. 2020). "Impact of satellite constellations on astronomical observations with ESO telescopes in the visible and infrared domains". In: 636, A121, A121. DOI: 10.1051/0004-6361/202037501. arXiv: 2003.01992 [astro-ph.IM].

Haji, Saad Hikmat and Adnan Mohsin Abdulazeez (2021). "Comparison of optimization techniques based on gradient descent algorithm: A review". In: *PalArch's Journal of Archaeology of Egypt/Egyptology* 18.4, pp. 2715–2743.

Hammer, Francois et al. (2004). "FALCON: a concept to extend adaptive optics corrections to cosmological fields". In: *Second backaskog workshop on extremely large telescopes*. Vol. 5382. SPIE, pp. 727–736.

Hammer, François et al. (2016). "The E-ELT multi-object spectrograph: latest news from MOSAIC". In: *Ground-based and Airborne Instrumentation for Astronomy VI*. Vol. 9908. SPIE, pp. 688–697.

Hardie, Russell C et al. (2017). "Simulation of anisoplanatic imaging through optical turbulence using numerical wave propagation with new validation analysis". In: *Optical Engineering* 56.7, pp. 071502–071502.

Hardy, John W (1998). *Adaptive optics for astronomical telescopes*. Vol. 16. Oxford Optical and Imaging Sci.

Hardy, John W, J E_ Lefebvre, and CL Koliopoulos (1977). "Real-time atmospheric compensation". In: *JOSA* 67.3, pp. 360–369.

Harris, Charles R et al. (2020). "Array programming with NumPy". In: *Nature* 585.7825, pp. 357–362.

Hartung, M et al. (2004). "A new VLT surface map of Titan at 1.575 microns". In: *Astronomy & Astrophysics* 421.1, pp. L17–L20.

Hartzell, Allyson L et al. (2010). "Reliability of MEMS deformable mirror technology used in adaptive optics imaging systems". In: *MEMS Adaptive Optics IV*. Vol. 7595. SPIE, pp. 73–85.

Hayat, Md Abul et al. (2021). "Self-supervised representation learning for astronomical images". In: *The Astrophysical Journal Letters* 911.2, p. L33.

He, Kaiming et al. (2015). *Deep residual learning for image recognition. CoRR abs/1512.03385 (2015)*.

— (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Heavens, Alan (2009). "Weak lensing: Dark matter, dark energy and dark gravity". In: *Nuclear Physics B-Proceedings Supplements* 194, pp. 76–81.

Helmi, Amina (2020). "Streams, substructures, and the early history of the Milky Way". In: *Annual Review of Astronomy and Astrophysics* 58, pp. 205–256.

Helmi, Amina et al. (2018). "The merger that led to the formation of the Milky Way's inner stellar halo and thick disk". In: *Nature* 563.7729, p. 85.

Henderson, CB et al. (2014). "Candidate gravitational microlensing events for future direct lens imaging". In: *The Astrophysical Journal* 794.1, p. 71.

Hendrycks, Dan et al. (2019). "Using self-supervised learning can improve model robustness and uncertainty". In: *Advances in neural information processing systems* 32.

Hernandez-Pajares, M. et al. (Sept. 1992). "Classifying Stars: A Comparison between Classical, Genetic and Neural Network Algorithms". In: *European Southern Observatory Conference and Workshop Proceedings*. Vol. 43. European Southern Observatory Conference and Workshop Proceedings, p. 325.

Herriot, Glen et al. (2006). "NFIRAOS: TMT narrow field near-infrared facility adaptive optics". In: *Advances in Adaptive Optics II*. Vol. 6272. SPIE, pp. 228–239.

Hickson, Paul, Joschua Hellemeier, and Rui Yang (2021). "Can amplified spontaneous emission produce intense laser guide stars for adaptive optics?" In: *Optics Letters* 46.8, pp. 1792–1795.

Hietala, Julius et al. (n.d.). "Closing the Sim2Real Gap in Dynamic Cloth Manipulation". In: *arXiv preprint* (). URL: https://dblp.org/rec/journals/corr/abs-2109-04771.html.

Hinton, Geoffrey E and Ruslan R Salakhutdinov (2006). "Reducing the dimensionality of data with neural networks". In: *Science* 313.5786, pp. 504–507.

Ho, Jonathan et al. (2020). "Denoising diffusion probabilistic models". In: *arXiv preprint arXiv:2006.11239*.

Hochreiter, Sepp (1998). "The vanishing gradient problem during learning recurrent neural nets and problem solutions". In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6.02, pp. 107–116.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Hocking, Alex et al. (2018). "An automatic taxonomy of galaxy morphology using unsupervised machine learning". In: *Monthly Notices of the Royal Astronomical Society* 473.1, pp. 1108–1129.

Hoffmire, Matthew A et al. (2021). "Deep learning for anisoplanatic optical turbulence mitigation in long-range imaging". In: *Optical Engineering* 60.3, pp. 033103–033103.

Hope, Tom, Yehezkel S Resheff, and Itay Lieder (2017). *Learning tensorflow: A guide to building deep learning systems.* " O'Reilly Media, Inc.".

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989). "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5, pp. 359–366.

Hoyle, Ben (2016). "Measuring photometric redshifts using galaxy images and Deep Neural Networks". In: *Astronomy and Computing* 16, pp. 34–40.

Hu, Jinghan Alina et al. (2022). "Satellite Constellation Avoidance with the Rubin Observatory Legacy Survey of Space and Time". In: *The Astrophysical Journal Letters* 941.1, p. L15.

Hubble, Edwin (1929). "A relation between distance and radial velocity among extra-galactic nebulae". In: *Proceedings of the national academy of sciences* 15.3, pp. 168–173.

Hubble, Edwin and Milton L Humason (1931). "The velocity-distance relation among extra-galactic nebulae". In: *Astrophysical Journal, vol. 74, p. 43* 74, p. 43.

Hubble, Edwin P (1926). "Extragalactic nebulae." In: *Astrophysical Journal, 64, 321-369 (1926)* 64.

Hubble, EP (1927). "The classification of spiral nebulae". In: *The Observatory* 50, pp. 276–281.

Huber, Peter J. (1964). "Robust Estimation of a Location Parameter". In: *The Annals of Mathematical Statistics* 35.1, pp. 73–101.

Huertas-Company, Marc, Regina Sarmiento, and Johan H Knapen (2023). "A brief review of contrastive learning applied to astrophysics". In: *RAS Techniques and Instruments* 2.1, pp. 441–452.

Humason, ML (1936). "The apparent radial velocities of 100 extra-galactic nebulae". In: *Astrophysical Journal, vol. 83, p. 10* 83, p. 10.

Hunter, John D (2007). "Matplotlib: A 2D graphics environment". In: *Computing in science & engineering* 9.03, pp. 90–95.

Hwang, Sun-Kyoo and Whoi-Yul Kim (2006). "A novel approach to the fast computation of Zernike moments". In: *Pattern Recognition* 39.11, pp. 2065–2076.

Isensee, Fabian et al. (2021). "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2, pp. 203–211.

Ivezić, Željko et al. (2019). "LSST: from science drivers to reference design and anticipated data products". In: *The Astrophysical Journal* 873.2, p. 111.

Jain, Prateek et al. (2018). "Parallelizing Stochastic Gradient Descent for Least Squares Regression: Mini-batching, Averaging, and Model Misspecification". en. In.

Jarvis, JF and JA Tyson (1981). "FOCAS-Faint object classification and analysis system". In: *Astronomical Journal, vol. 86, Mar. 1981, p. 476-495.* 86, pp. 476–495.

Jin, Shoko et al. (Dec. 2022). "The wide-field, multiplexed, spectroscopic facility WEAVE: Survey design, overview, and simulated implementation". In: arXiv:2212.03981. arXiv:2212.03981 [astro-ph]. URL: http://arxiv.org/abs/2212.03981.

Jin, Shoko et al. (2023). "The wide-field, multiplexed, spectroscopic facility WEAVE: Survey design, overview, and simulated implementation". In: *Monthly Notices of the Royal Astronomical Society*, stad557.

Johns, Matt et al. (2004). "Status of the giant magellan telescope (GMT) project". In: *Ground-based Telescopes*. Vol. 5489. SPIE, pp. 441–453.

Johnston, Dustin C and Byron M Welsh (1994). "Analysis of multiconjugate adaptive optics". In: *JOSA A* 11.1, pp. 394–408.

Jovanovic, N et al. (2015). "The Subaru coronagraphic extreme adaptive optics system: enabling high-contrast imaging on solar-system scales". In: *Publications of the Astronomical Society of the Pacific* 127.955, p. 890.

Jumper, John et al. (2021). "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873, pp. 583–589.

Kadian, Abhishek et al. (2020). "Sim2Real Predictivity: Does Evaluation in Simulation Predict Real-World Performance?" In: *IEEE Robotics and Automation Letters* 5.4, pp. 6670–6677. DOI: 10.1109/lra.2020.3013848. URL: https://doi.org/10.1109%2Flra.2020.3013848.

Kaiser, N, JL Tonry, and GA Luppino (2000). "A New Strategy for Deep Wide-Field High-Resolution Optical Imaging". In: *Publications of the Astronomical Society of the Pacific* 112.772, p. 768.

Kaiser, Nicholas et al. (Dec. 2002). "Pan-STARRS: A Large Synoptic Survey Telescope Array". In: *Survey and Other Telescope Technologies and Discoveries*. Ed. by J. Anthony Tyson and Sidney Wolff. Vol. 4836. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 154–164. DOI: 10.1117/12.457365.

Keerthi Vasan, G. C. et al. (Oct. 2023). "Optimizing machine learning methods to discover strong gravitational lenses in the deep lens survey". In: 524.4, pp. 5368–5390. DOI: 10.1093/mnras/stad1709. arXiv: 2211.00047 [astro-ph.GA].

Khalifa, Nour Eldeen et al. (2018). "Deep galaxy V2: Robust deep convolutional neural networks for galaxy morphology classifications". In: *2018 International Conference on Computing Sciences and Engineering (ICCSE)*. IEEE, pp. 1–6.

Kharchenko, NV et al. (2013). "Global survey of star clusters in the Milky Way-II. The catalogue of basic parameters". In: *Astronomy & Astrophysics* 558, A53.

Khirirat, Sarit, H. Feyzmahdavian, and M. Johansson (2017). "Mini-batch gradient descent: Faster convergence under data sparsity". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE. DOI: 10.1109/CDC.2017.8264077. URL: https://dx.doi.org/10.1109/CDC.2017.8264077.

Kingma, Diederik P and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Köhlinger, F, H Hoekstra, and M Eriksen (2015). "Statistical uncertainties and systematic errors in weak lensing mass estimates of galaxy clusters". In: *Monthly Notices of the Royal Astronomical Society* 453.3, pp. 3107–3119.

Kolarik, Martin, Radim Burget, and Kamil Riha (2020). "Comparing normalization methods for limited batch size segmentation neural networks". In: *2020 43rd international conference on telecommunications and signal processing (TSP)*. IEEE, pp. 677–680.

Kriessler, J. R. et al. (Dec. 1998). "Automated Morphological Classification of Galaxies and the Morphology-Density Relation". In: *American Astronomical Society Meeting Abstracts*. Vol. 193. American Astronomical Society Meeting Abstracts, 38.20, p. 38.20.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105.

Kruk, Sandor et al. (Mar. 2023). "The impact of satellite trails on Hubble Space Telescope observations". en. In: *Nature Astronomy* 7.33, 262–268. ISSN: 2397-3366. DOI: 10.1038/s41550-023-01903-3.

Kubo, Mariko et al. (2017). "Bimodal morphologies of massive galaxies at the core of a protocluster at z= 3.09 and the strong size growth of a brightest cluster galaxy". In: *Monthly Notices of the Royal Astronomical Society* 469.2, pp. 2235–2250.

Kumaran, Shivam et al. (Apr. 2023). "Automated classification of Chandra X-ray point sources using machine learning methods". In: 520.4, pp. 5065–5076. DOI: 10.1093/mnras/stad414. arXiv: 2302.09008 [astro-ph.HE].

Kupke, Renate et al. (2018). "The wide field optical spectrograph (WFOS) for TMT: fiber-WFOS optical design (Conference Presentation)". In: *Ground-based and Airborne Instrumentation for Astronomy VII*. Vol. 10702. SPIE, p. 1070221.

Kuzma, Pete B, AMN Ferguson, and Jorge Peñarrubia (2021). "Detecting globular cluster tidal extensions with Bayesian inference–I. Analysis of ω Centauri with Gaia EDR3". In: *Monthly Notices of the Royal Astronomical Society* 507.1, pp. 1127–1137.

Laag, Edward A et al. (2006). "Adaptive Optics Imaging Survey of Luminous Infrared Galaxies". In: *The Astronomical Journal* 131.6, p. 2877.

Lang, Thomas, Samuel T. Spencer, and Alison M. W. Mitchell (Sept. 2023). "Impact of Satellite Trails on H.E.S.S. Astronomical Observations". In: *Astronomy Astrophysics* 677. arXiv:2307.13293 [astro-ph], A141. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202347200.

Lardière, Olivier et al. (2014). "Multi-object adaptive optics on-sky results with Raven". In: *Adaptive Optics Systems IV*. Vol. 9148. SPIE, pp. 527–540.

Laureijs, R. et al. (2011). "Euclid Definition Study Report". In: *arXiv preprint*. URL: https://arxiv.org/abs/1110.3193.

Lawler, Samantha M., Aaron C. Boley, and Hanno Rein (Jan. 2022). "Visibility Predictions for Near-future Satellite Megaconstellations: Latitudes near 50° Will Experience

the Worst Light Pollution". In: 163.1, 21, p. 21. DOI: 10.3847/1538-3881/ac341b. arXiv: 2109.04328 [astro-ph.EP].

Lawrence, Andy et al. (2022). "The case for space environmentalism". In: *Nature Astronomy* 6.4, pp. 428–435.

LeCun, Yann et al. (1989). "Handwritten digit recognition with a back-propagation network". In: *Advances in neural information processing systems* 2.

LeCun, Yann et al. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

Leung, Henry W and Jo Bovy (2018). "Deep learning of multi-element abundances from high-resolution spectroscopic data". In: *Monthly Notices of the Royal Astronomical Society* 483.3, pp. 3255–3277.

Leung, Henry W. and Jo Bovy (2019). "Deep learning of multi-element abundances from high-resolution spectroscopic data". In: 483.3, pp. 3255–3277. DOI: 10.1093/mnras/sty3217. arXiv: 1808.04428 [astro-ph.GA].

Leung, Henry W. and Jo Bovy (Aug. 2023). "Towards an astronomical foundation model for stars with a Transformer-based model". In: arXiv:2308.10944. arXiv:2308.10944 [astro-ph]. URL: http://arxiv.org/abs/2308.10944.

Lewis, Antony and Sarah Bridle (2002). "Cosmological parameters from CMB and other data: A Monte Carlo approach". In: *Physical Review D* 66.10, p. 103511.

Li, Nianyi et al. (2021a). "Unsupervised non-rigid image distortion removal via grid deformation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2522–2532.

Li, Qiang et al. (2021b). "Predicting the 25th solar cycle using deep learning methods based on sunspot area data". In: URL: https://dx.doi.org/10.1088/1674-4527/21/7/184.

Li, R et al. (2022a). "Galaxy Light Profile Convolutional Neural Networks (GaLNets). I. Fast and Accurate Structural Parameters for Billion-galaxy Samples". In: *The Astrophysical Journal* 929.2, p. 152.

Li, Renjie et al. (2022b). "A comprehensive review on deep supervision: Theories and applications". In: *arXiv preprint arXiv:2207.02376*.

Lintott, C. et al. (2008a). "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey". In: *Monthly Notices of the Royal Astronomical Society* 389.3, pp. 1179–1189. DOI: 10.1111/j.1365-2966.2008.13689.x. URL: https://academic.oup.com/mnras/article/389/3/1179/1017183?login=false.

Lintott, Chris J et al. (2008b). "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey". In: *Monthly Notices of the Royal Astronomical Society* 389.3, pp. 1179–1189.

Lloyd-Hart, M. et al. (Aug. 1992). "High resolution imaging at the Multiple Mirror Telescope using adaptive optics". In: *Atmospheric Propagation and Remote Sensing*. Ed. by Anton Kohnle and Walter B. Miller. Vol. 1688. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 442–452. DOI: 10.1117/12.137911.

Lu, JR et al. (2016). "A search for stellar-mass black holes via astrometric microlensing". In: *The Astrophysical Journal* 830.1, p. 41.

Mackay, C et al. (2018). "GravityCam: Wide-field high-resolution high-cadence imaging surveys in the visible from the ground". In: *Publications of the Astronomical Society of Australia* 35, e047.

Madec, P-Y (2012). "Overview of deformable mirror technologies for adaptive optics and astronomy". In: *Adaptive Optics Systems III*. Vol. 8447. SPIE, pp. 22–39.

Makitalo, Markku and Alessandro Foi (2011). "A closed-form approximation of the exact unbiased inverse of the Anscombe variance-stabilizing transformation". In: *IEEE transactions on image processing* 20.9, pp. 2697–2698.

Marchetti, Enrico et al. (2007). "On-sky testing of the multi-conjugate adaptive optics demonstrator". In: *The Messenger* 129.8.

Marchis, Franck et al. (2006). "Shape, size and multiplicity of main-belt asteroids: I. Keck Adaptive Optics survey". In: *Icarus* 185.1, pp. 39–63.

Marinacci, Federico et al. (2018). "First results from the IllustrisTNG simulations: radio haloes and magnetic fields". In: *Monthly Notices of the Royal Astronomical Society* 480.4, pp. 5113–5139.

Martin, Nicolas F, Jelte TA de Jong, and Hans-Walter Rix (2008). "A comprehensive maximum likelihood analysis of the structural properties of faint Milky Way satellites". In: *The Astrophysical Journal* 684.2, p. 1075.

Martinez, Noelia et al. (2022). "Preliminary design of the Laser Guide Star Facility for the ULTIMATE-Subaru Ground Layer Adaptive Optics system." In: *Adaptive Optics Systems VIII*. Vol. 12185. SPIE, pp. 2355–2363.

Massey, Robert, Sara Lucatello, and Piero Benvenuti (2020). "The challenge of satellite megaconstellations". In: *Nature astronomy* 4.11, pp. 1022–1023.

Max, Claire E, Donald T Gavel, and Scot S Olivier (1995). "Near-infrared astronomy with adaptive optics and laser guide stars at the Keck Observatory". In: *Adaptive Optical Systems and Applications*. Vol. 2534. SPIE, pp. 412–422.

McConnachie, Alan W. et al. (May 2016). "A concise overview of the Maunakea Spectroscopic Explorer". In: *arXiv e-prints*, arXiv:1606.00060, arXiv:1606.00060. arXiv: 1606.00060 [astro-ph.IM].

Mello, GF Porto de et al. (2014). "A photometric and spectroscopic survey of solar twin stars within 50 parsecs of the Sun-I. Atmospheric parameters and color similarity to the Sun". In: *Astronomy & Astrophysics* 563, A52.

Meyer, E et al. (2006). "Multiconjugated adaptive optics for ELTs: an enhancement of the PIGS setup". In: *Advances in Adaptive Optics II*. Vol. 6272. SPIE, pp. 1216–1223.

Minowa, Yosuke et al. (2017). "Ultimate-subaru: Wide-field near-infrared surveyor with glao at subaru telescope". In: *Adaptive Optics for Extremely Large Telescopes V (AO4ELT5)],(Jun 2017)*.

Minowa, Yosuke et al. (2020). "ULTIMATE-Subaru: enhancing the Subaru's wide-field capability with GLAO". In: *Advances in Optical Astronomical Instrumentation 2019*. Vol. 11203. SPIE, pp. 21–22.

Mishra, Amit, Pranath Reddy, and Rahul Nigam (Mar. 2019). "Baryon density extraction and isotropy analysis of Cosmic Microwave Background using Deep Learning". In: *arXiv e-prints*, arXiv:1903.12253, arXiv:1903.12253. DOI: 10.48550/arXiv.1903.12253. arXiv: 1903.12253 [astro-ph.CO].

Mizusawa, Satoru et al. (2021). "Computed tomography image reconstruction using stacked U-Net". In: *Computerized Medical Imaging and Graphics* 90, p. 101920.

Mobasher, B et al. (2007). "Photometric redshifts of galaxies in COSMOS". In: *The Astrophysical Journal Supplement Series* 172.1, p. 117.

Morrow, Jasper M and Maria Pia Sormani (2020). *Machine learning outperforms human experts in MRI pattern analysis of muscular dystrophies*.

Nair, Vinod and Geoffrey E Hinton (2010). "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.

Necib, Lina et al. (July 2020). "Evidence for a vast prograde stellar stream in the solar vicinity". In: *Nature Astronomy* 4, pp. 1078–1083. DOI: 10.1038/s41550-020-1131-2. arXiv: 1907.07190 [astro-ph.GA].

Ness, Melissa et al. (2015). "The cannon: A data-driven approach to stellar label determination". In: *The Astrophysical Journal* 808.1, p. 16.

Nguyen, Tuan Dung et al. (Sept. 2023). "AstroLLaMA: Towards Specialized Foundation Models in Astronomy". In: arXiv:2309.06126. DOI: 10.48550/arXiv.2309.06126. URL: http://arxiv.org/abs/2309.06126.

Nieuwenhuizen, Robert and Klamer Schutte (2019). "Deep learning for software-based turbulence mitigation in long-range imaging". In: *Artificial Intelligence and Machine Learning in Defense Applications*. Vol. 11169. SPIE, pp. 153–162.

Niu, Kuo and Chao Tian (2022). "Zernike polynomials and their applications". In: *Journal of Optics*.

Noll, Robert J (1976). "Zernike polynomials and atmospheric turbulence". In: *JOsA* 66.3, pp. 207–211.

Owens, E. A., R. E. Griffiths, and K. U. Ratnatunga (July 1996). "Using oblique decision trees for the morphological classification of galaxies". In: 281, pp. 153–157. DOI: 10.1093/mnras/281.1.153. arXiv: astro-ph/9603004 [astro-ph].

O'Briain, Teaghan et al. (2021). "Cycle-StarNet: Bridging the Gap between Theory and Data by Leveraging Large Data Sets". In: *The Astrophysical Journal* 906.2, p. 130.

Paczynski, Bohdan (1986). "Gravitational microlensing by the galactic halo". In: *The Astrophysical Journal* 304, pp. 1–5.

Paillassa, Maxime, Emmanuel Bertin, and Herve Bouy (2019). "MAXIMASK and MAXITRACK: Two new tools for identifying contaminants in astronomical images using convolutional neural networks". In: *arXiv preprint arXiv:1907.08298*.

Pasquet, Johanna et al. (2019). "Photometric redshifts from SDSS images using a convolutional neural network". In: *Astronomy & Astrophysics* 621, A26.

Pasquini, L et al. (2002). "Installation and commissioning of FLAMES, the VLT Multifibre Facility". In: *The Messenger* 110, pp. 1–9.

Paszke, Adam et al. (2019a). "Advances in neural information processing systems 32". In: *Curran Associates, Inc*, pp. 8024–8035.

Paszke, Adam et al. (2019b). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Paszke, Adam et al. (2019c). "Pytorch: An imperative style, high-performance deep learning library". In: *Advances in neural information processing systems* 32.

Pearson, W. J. et al. (June 2019). "Identifying galaxy mergers in observations and simulations with deep learning". In: 626, A49, A49. DOI: 10.1051/0004-6361/201935355. arXiv: 1902.10626 [astro-ph.GA].

Peebles, Phillip James Edwin (1984). "Dark matter and the origin of galaxies and globular star clusters". In: *The Astrophysical Journal* 277, pp. 470–477.

Peñarrubia, Jorge et al. (2017). "Stellar envelopes of globular clusters embedded in dark minihaloes". In: *Monthly Notices of the Royal Astronomical Society: Letters* 471.1, pp. L31–L35.

Peter, D et al. (2009). "PYRAMIR: exploring the on-sky performance of the world's first near-infrared pyramid wavefront sensor". In: *Publications of the Astronomical Society of the Pacific* 122.887, p. 63.

Peth, Michael A et al. (2016). "Beyond spheroids and discs: classifications of CANDELS galaxy structure at 1.4< z< 2 via principal component analysis". In: *Monthly Notices of the Royal Astronomical Society* 458.1, pp. 963–987.

Pietrukowicz, P et al. (2005). "Cluster AgeS Experiment (CASE): Dwarf Novae and a Probable Microlensing Event in the Globular Cluster M22". In: *arXiv preprint astro-ph/0509797*.

Pietrukowicz, Pawel et al. (2011). "The first confirmed microlens in a globular cluster". In: *The Astrophysical Journal Letters* 744.2, p. L18.

Plane, J (2023). "Metals from spacecraft reentry in stratospheric aerosol particles". In: *Proceedings of the National Academy of Sciences of USA*.

Pourré, N et al. (2022). "Low-wind-effect impact on Shack-Hartmann-based adaptive optics-Partial control solution in the context of SPHERE and GRAVITY+". In: *Astronomy & Astrophysics* 665, A158.

Prusti, Timo et al. (2016). "The gaia mission". In: *Astronomy & astrophysics* 595, A1.

Puschell, JJ, FN Owen, and RA Laing (1982). "Near-infrared photometry of distant radio galaxies-Spectral flux distributions and redshift estimates". In: *Astrophysical Journal, Part 2-Letters to the Editor, vol. 257, June 15, 1982, p. L57-L61. NSF-NASA-USAF-supported research.* 257, pp. L57–L61.

Quinlan, J. Ross (1986). "Induction of decision trees". In: *Machine learning* 1, pp. 81–106.

Racca, Giuseppe D et al. (2016). "The Euclid mission design". In: *Space telescopes and instrumentation 2016: optical, infrared, and millimeter wave*. Vol. 9904. SPIE, pp. 235–257.

Radford, Alec et al. (2018). "Improving language understanding by generative pre-training". In.

Ragazzoni, Roberto (1996). "Pupil plane wavefront sensing with an oscillating prism". In: *Journal of modern optics* 43.2, pp. 289–293.

Ramírez, I, Jorge Melendez, and Martin Asplund (2009). "Accurate abundance patterns of solar twins and analogs-Does the anomalous solar chemical composition come from planet formation?" In: *Astronomy & Astrophysics* 508.1, pp. L17–L20.

Rao, Kanishka et al. (2020). *RL-CycleGAN: Reinforcement Learning Aware Simulation-To-Real*. arXiv: 2006.09001 [cs.RO].

Rayleigh (1879). "XXXI. Investigations in optics, with special reference to the spectroscope". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 8.49, pp. 261–274.

Recio-Blanco, A et al. (2023a). "Gaia Data Release 3-Analysis of RVS spectra using the General Stellar Parametriser from spectroscopy". In: *Astronomy & Astrophysics* 674, A29.

Recio-Blanco, A et al. (2023b). "Gaia Data Release 3-Chemical cartography of the Milky Way". In: *Astronomy & Astrophysics* 674, A38.

Reed, Scott et al. (2022). "A generalist agent". In: *arXiv preprint arXiv:2205.06175*.

Rehemtulla, Nabeel et al. (July 2023). "`BTSbot`: A Multi-input Convolutional Neural Network to Automate and Expedite Bright Transient Identification for the Zwicky Transient Facility". In: *arXiv e-prints*, arXiv:2307.07618, arXiv:2307.07618. DOI: 10.48550/arXiv.2307.07618. arXiv: 2307.07618 [astro-ph.IM].

Reiman, David M. and Brett E. Göhre (May 2019). "Deblending galaxy superpositions with branched generative adversarial networks". In: 485.2, pp. 2617–2627. DOI: 10.1093/mnras/stz575. arXiv: 1810.10098 [astro-ph.IM].

Rigaut, Francois and Benoit Neichel (Mar. 2020). "Multiconjugate Adaptive Optics for Astronomy". In: arXiv:2003.03097. arXiv:2003.03097 [astro-ph]. URL: http://arxiv.org/abs/2003.03097.

Rigaut, François et al. (2014). "Gemini multiconjugate adaptive optics system review–I. Design, trade-offs and integration". In: *Monthly Notices of the Royal Astronomical Society* 437.3, pp. 2361–2375.

Rigaut, Francois J, Brent L Ellerbroek, and Ralf Flicker (2000). "Principles, limitations, and performance of multiconjugate adaptive optics". In: *Adaptive Optical Systems Technology*. Vol. 4007. SPIE, pp. 1022–1031.

Robbins, H. and S. Monro (1951). "A stochastic approximation method". In: *Annals of Mathematical Statistics* 22, pp. 400–407.

Roberts Jr, Lewis C et al. (2004). "Is that really your Strehl ratio?" In: *Advancements in Adaptive Optics*. Vol. 5490. SPIE, pp. 504–515.

Robitaille, Thomas P et al. (2013). "Astropy: A community Python package for astronomy". In: *Astronomy & Astrophysics* 558, A33.

Roddier, François (1981). "V the effects of atmospheric turbulence in optical astronomy". In: *Progress in optics*. Vol. 19. Elsevier, pp. 281–376.

Rodriguez, José-Víctor, Ignacio Rodríguez-Rodríguez, and Wai Lok Woo (Sept. 2022). "On the application of machine learning in astronomy and astrophysics: A text-mining-based scientometric analysis". In: *WIREs Data Mining and Knowledge Discovery* 12.5, e1476, e1476. DOI: 10.1002/widm.1476.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.

Rosenblatt, Frank (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, p. 386.

Rouan, D et al. (2007). "Morphological evolution of z˜ 1 galaxies from deep K-band AO imaging in the COSMOS deep field". In: *Astronomy & Astrophysics* 468.3, pp. 937–950.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1986). "Learning representations by back-propagating errors". In: *nature* 323.6088, pp. 533–536.

Różański, Tomasz, Yuan-Sen Ting, and Maja Jabłońska (June 2023). "Toward a Spectral Foundation Model: An Attention-Based Approach with Domain-Inspired Fine-Tuning and Wavelength Parameterization". In: arXiv:2306.15703. arXiv:2306.15703 [astro-ph]. DOI: 10.48550/arXiv.2306.15703. URL: http://arxiv.org/abs/2306.15703.

Sacco, GIUSEPPE GERMANO et al. (2014). "The Gaia-ESO Survey: processing FLAMES-UVES spectra". In: *Astronomy & Astrophysics* 565, A113.

Sadeh, Iftach, Filipe B Abdalla, and Ofer Lahav (2016). "ANNz2: photometric redshift and probability distribution function estimation using machine learning". In: *Publications of the Astronomical Society of the Pacific* 128.968, p. 104502.

Salvato, Mara, Olivier Ilbert, and Ben Hoyle (2019). "The many flavours of photometric redshifts". In: *Nature Astronomy* 3.3, pp. 212–222.

Sandler, D. G. et al. (May 1991). "Use of a neural network to control an adaptive optics system for an astronomical telescope". In: 351.6324, pp. 300–302. DOI: 10.1038/351300a0.

Scaife, AMM (2020). "Big telescope, big data: towards exascale with the Square Kilometre Array". In: *Philosophical Transactions of the Royal Society A* 378.2166, p. 20190060.

Schlafly, Edward F and Douglas P Finkbeiner (2011). "Measuring reddening with Sloan Digital Sky Survey stellar spectra and recalibrating SFD". In: *The Astrophysical Journal* 737.2, p. 103.

Schmidt, Jason Daniel (2010). "Numerical simulation of optical wave propagation: With examples in MATLAB". In: SPIE.

Schmidt, K. et al. (2022). "Deep learning-based imaging in radio interferometry". In: URL: https://dx.doi.org/10.1051/0004-6361/202142113.

Schuldt, S et al. (2021). "Photometric redshift estimation with a convolutional neural network: NetZ". In: *Astronomy & Astrophysics* 651, A55.

Schuster, M. and K. Paliwal (1997). "Bidirectional recurrent neural networks". In: *IEEE Transactions on Signal Processing* 45.11, pp. 2673–2681.

Sen, Snigdha et al. (2022). "Astronomical big data processing using machine learning: A comprehensive review". In: *Experimental Astronomy* 53.1, pp. 1–43.

Sérsic, JL (1963). "Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy". In: *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* 6, pp. 41–43.

Sevin, A et al. (2014). "Enabling technologies for GPU driven adaptive optics real-time control". In: *Adaptive Optics Systems IV*. Vol. 9148. SPIE, pp. 853–861.

Shack, Roland V (1971). "Production and use of a lenticular Hartmann screen". In: *Spring Meeting of Optical Society of America, 1971*. Vol. 656.

Sharma, Kaushal et al. (2020). "Application of convolutional neural networks for stellar spectral classification". In: *Monthly Notices of the Royal Astronomical Society* 491.2, pp. 2280–2300.

Shatokhina, Iuliia, Victoria Hutterer, and Ronny Ramlau (2020). "Review on methods for wavefront reconstruction from pyramid wavefront sensor data". In: *Journal of Astronomical Telescopes, Instruments, and Systems* 6.1, pp. 010901–010901.

Shen, Lu et al. (2023). "CEERS: Spatially Resolved UV and Mid-infrared Star Formation in Galaxies at 0.2< z< 2.5: The Picture from the Hubble and James Webb Space Telescopes". In: *The Astrophysical Journal* 950.1, p. 7.

Silver, David et al. (2017). "Mastering the game of go without human knowledge". In: *nature* 550.7676, pp. 354–359.

Simard, Luc et al. (2011). "A catalog of bulge+ disk decompositions and updated photometry for 1.12 million galaxies in the sloan digital sky survey". In: *The Astrophysical Journal Supplement Series* 196.1, p. 11.

Smiljanic, R et al. (2016). "The Gaia-ESO Survey: Sodium and aluminium abundances in giants and dwarfs-Implications for stellar and Galactic chemical evolution". In: *Astronomy & Astrophysics* 589, A115.

Smith, Simon ET et al. (2023). "Discovery of a New Local Group Dwarf Galaxy Candidate in UNIONS: Boötes V". In: *The Astronomical Journal* 166.2, p. 76.

Smits, Alexander J and Ivan Marusic (2013). "Wall-bounded turbulence". In: *Phys. Today* 66.9, pp. 25–30.

Snaith, Owain et al. (2022). "Rapid early gas accretion for the inner Galactic disc-A case for a short accretion timescale". In: *Astronomy & Astrophysics* 659, A64.

Sonnenfeld, Alessandro et al. (2013). "The SL2S Galaxy-scale Lens Sample. IV. The dependence of the total mass density profile of early-type galaxies on redshift, stellar mass, and size". In: *The Astrophysical Journal* 777.2, p. 98.

Soydaner, Derya (2020). "A comparison of optimization algorithms for deep learning". In: *International Journal of Pattern Recognition and Artificial Intelligence* 34.13, p. 2052013.

Spencer, Samuel T., Thomas Lang, and Alison M. W. Mitchell (Aug. 2023). "The Influence of Satellite Trails on H.E.S.S. Gamma-Ray Astronomical Observations". In: arXiv:2308.01509. arXiv:2308.01509 [astro-ph]. DOI: 10.48550/arXiv.2308.01509. URL: http://arxiv.org/abs/2308.01509.

Spergel, D. et al. (2015). "Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report". In: URL: https://arxiv.org/abs/1503.03757.

Stein, George et al. (June 2022). "Mining for Strong Gravitational Lenses with Self-supervised Learning". In: 932.2, 107, p. 107. DOI: 10.3847/1538-4357/ac6d63. arXiv: 2110.00023 [astro-ph.IM].

Steinmetz, Matthias et al. (2020). "The Sixth Data Release of the Radial Velocity Experiment (Rave). II. Stellar Atmospheric Parameters, Chemical Abundances, and Distances". In: *The Astronomical Journal* 160.2, p. 83.

Stoller, Daniel, Sebastian Ewert, and Simon Dixon (2018). "Wave-u-net: A multi-scale neural network for end-to-end audio source separation". In: *arXiv preprint arXiv:1806.03185*.

Tacchella, Sandro et al. (2015). "THE SINS/zC-SINF survey of z 2 galaxy kinematics: Rest-frame morphology, structure, and colors from near-infrared hubble space telescope imaging". In: *The Astrophysical Journal* 802.2, p. 101.

Tagliaferri, Roberto et al. (2003). "Neural networks for photometric redshifts evaluation". In: *Neural Nets: 14th Italian Workshop on Neural Nets, WIRN VIETRI 2003, Vietri sul Mare, Italy, June 4-7, 2003. Revised Papers 14*. Springer, pp. 226–234.

Tamura, Naoyuki et al. (July 2018). "Prime Focus Spectrograph (PFS) for the Subaru telescope: ongoing integration and future plans". In: *Ground-based and Airborne Instrumentation for Astronomy VII*. Ed. by Christopher J. Evans, Luc Simard, and Hideki Takami. Vol. 10702. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 107021C, p. 107021C. DOI: 10.1117/12.2311871.

Teimoorinia, Hossen et al. (Apr. 2020). "Comparison of Multi-class and Binary Classification Machine Learning Models in Identifying Strong Gravitational Lenses". In: 132.1010, 044501, p. 044501. DOI: 10.1088/1538-3873/ab747b. arXiv: 2002.11849 [astro-ph.GA].

Thorp, Mallory D et al. (2021). "Towards robust determination of non-parametric morphologies in marginal astronomical data: resolving uncertainties with cosmological hydrodynamical simulations". In: *Monthly Notices of the Royal Astronomical Society* 507.1, pp. 886–903.

Ting, Yuan-Sen et al. (2017). "Prospects for measuring abundances of> 20 elements with low-resolution stellar spectra". In: *The Astrophysical Journal* 843.1, p. 32.

— (2019). "The Payne: self-consistent ab initio fitting of stellar spectra". In: *The Astrophysical Journal* 879.2, p. 69.

Tokovinin, A (2004). "Seeing improvement with ground-layer adaptive optics". In: *Publications of the Astronomical Society of the Pacific* 116.824, p. 941.

Tokovinin, Andrei et al. (2001). "Optimized modal tomography in adaptive optics". In: *Astronomy & Astrophysics* 378.2, pp. 710–721.

Toomre, Alar and Juri Toomre (1972). "Galactic bridges and tails". In: *Astrophysical Journal, Vol. 178, pp. 623-666 (1972)* 178, pp. 623–666.

Tyson, J Anthony et al. (2020). "Mitigation of LEO satellite brightness and trail effects on the Rubin Observatory LSST". In: *The Astronomical Journal* 160.5, p. 226.

Tyson, Robert K and Benjamin West Frazier (2022). *Principles of adaptive optics*. CRC press.

Udalski, Andrzej et al. (1992). "The optical gravitational lensing experiment". In: *Acta Astronomica (ISSN 0001-5237), vol. 42, no. 4, p. 253-284.* 42, pp. 253–284.

Valizadegan, Hamed et al. (2022). "ExoMiner: A Highly Accurate and Explainable Deep Learning Classifier That Validates 301 New Exoplanets". In: *The Astrophysical Journal* 926.2, p. 120. DOI: 10.3847/1538-4357/ac4399. URL: https://doi.org/10.3847%2F1538-4357%2Fac4399.

Valizadegan, Hamed et al. (2023). "Multiplicity Boost of Transit Signal Classifiers: Validation of 69 New Exoplanets using the Multiplicity Boost of ExoMiner". In: *The Astronomical Journal* 166.1, p. 28. DOI: 10.3847/1538-3881/acd344. URL: https://doi.org/10.3847%2F1538-3881%2Facd344.

Vallenari, A, AGA Brown, and T Prusti (2022). "Gaia Data Release 3. Summary of the content and survey properties". In: *Astronomy & Astrophysics*.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems*. Vol. 30.

Vaucouleurs, Gerard de (1948). “Recherches sur les nebuleuses extragalactiques”. In: *Annales d’Astrophysique*. Vol. 11, p. 247.

Venn, KA et al. (2019). “Lrp2020: Machine learning advantages in canadian astrophysics”. In: *arXiv preprint arXiv:1910.00774*.

Venugopalan, Vivek (2014). “Evaluating latency and throughput bound acceleration of FPGAs and GPUs for adaptive optics algorithms”. In: *2014 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, pp. 1–6.

Vérinaud, Christophe (2004). “On the nature of the measurements provided by a pyramid wave-front sensor”. In: *Optics Communications* 233.1-3, pp. 27–38.

Vidaurri, Monica R. and Alexander Q. Gilbert (June 2023). “Environmental Considerations in the age of Space Exploration: the Conservation and Protection of Non-Earth Environments”. In: arXiv:2306.05594. arXiv:2306.05594 [astro-ph]. DOI: 10.48550/arXiv.2306.05594. URL: http://arxiv.org/abs/2306.05594.

Villaescusa-Navarro, Francisco et al. (2021). “The camels project: Cosmology and astrophysics with machine-learning simulations”. In: *The Astrophysical Journal* 915.1, p. 71.

Vint, David et al. (2020). “Analysis of deep learning architectures for turbulence mitigation in long-range imagery”. In: *Artificial Intelligence and Machine Learning in Defense Applications II*. Vol. 11543. SPIE, p. 1154303.

Virtanen, Pauli et al. (2020). “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature methods* 17.3, pp. 261–272.

Viswanathan, Akshara et al. (2023). “Hidden deep in the halo: selection of a reduced proper motion halo catalogue and mining retrograde streams in the velocity space”. In: *Monthly Notices of the Royal Astronomical Society* 521.2, pp. 2087–2102.

Von Hippel, T et al. (1994). “Automated classification of stellar spectra–I. Initial results with artificial neural networks”. In: *Monthly Notices of the Royal Astronomical Society* 269.1, pp. 97–104.

Vruno, F. Di et al. (Aug. 2023). “Unintended electromagnetic radiation from Starlink satellites detected with LOFAR between 110 and 188 MHz”. en. In: *Astronomy Astrophysics* 676, A75. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202346374.

Walker, C. and J. Hall (2020). *Impact of Satellite Constellations on Optical Astronomy and Recommendations Toward Mitigations*. Tech. rep. NSF's NOIRLab (Communications, Education & Engagement division) Workshop Report.

— (2021). *SATCON2 Executive Summary Report*. Tech. rep. NSF's NOIRLab (Communications, Education & Engagement division) Workshop Report.

Walker, Connie et al. (Oct. 2020). *Dark & Quiet Skies I (2020)*. Dark & Quiet Skies I (2020), Report of the conference held 5-9 October, 2020. DOI: 10.5281/zenodo.5898785.

Walker, Connie et al. (Oct. 2021). *Dark & Quiet Skies II (2021)*. Dark & Quiet Skies II (2021), Report of the conference held 3-7 October, 2021. DOI: 10.5281/zenodo.5874725.

Walmsley, Mike et al. (2020). "Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning". In: *Monthly Notices of the Royal Astronomical Society* 491.2, pp. 1554–1574.

Walmsley, Mike et al. (2021). "Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies". In: *Monthly Notices of the Royal Astronomical Society* 509.3, pp. 3966–3988. DOI: 10.1093/mnras/stab2093. URL: https://doi.org/10.1093%2Fmnras%2Fstab2093.

Wang, Liwei et al. (2015). "Training deeper convolutional networks with deep supervision". In: *arXiv preprint arXiv:1505.02496*.

Wang, Rui et al. (May 2023). "Stellar Parameters and Chemical Abundances Estimated from LAMOST-II DR8 MRS based on Cycle-StarNet". In: arXiv:2305.05854. arXiv:2305.05854 [astro-ph]. URL: http://arxiv.org/abs/2305.05854.

Wang, Yi-Fan and Alexander H Nitz (2022). "Search for Coincident Gravitational-wave and Fast Radio Burst Events from 4-OGC and the First CHIME/FRB Catalog". In: *The Astrophysical Journal* 937.2, p. 89.

Wang, Zhou et al. (2004). "Image quality assessment: from error visibility to structural similarity". In: *IEEE transactions on image processing* 13.4, pp. 600–612.

Waskom, Michael L (2021). "Seaborn: statistical data visualization". In: *Journal of Open Source Software* 6.60, p. 3021.

Weibel, Jean-Baptiste, Timothy Patten, and Markus Vincze (2019). *Addressing the Sim2Real Gap in Robotic 3D Object Classification*. arXiv: 1910.12585 [cs.CV].

Weir, Nicholas et al. (Dec. 1995). "The SKICAT System for Processing and Analyzing Digital Imaging Sky Surveys". In: 107, p. 1243. DOI: 10.1086/133683.

Wilde, Joshua et al. (May 2022). "Detecting gravitational lenses using machine learning: exploring interpretability and sensitivity to rare lensing configurations". In: 512.3, pp. 3464–3479. DOI: 10.1093/mnras/stac562. arXiv: 2202.12776 [astro-ph.GA].

Wilkinson, Scott et al. (Nov. 2022). "The merger fraction of post-starburst galaxies in UNIONS". In: 516.3, pp. 4354–4372. DOI: 10.1093/mnras/stac1962. arXiv: 2207.04152 [astro-ph.GA].

Wilkinson, Scott et al. (2022). "The merger fraction of post-starburst galaxies in UNIONS". In: *Monthly Notices of the Royal Astronomical Society* 516.3, pp. 4354–4372.

Wu, Zhirong et al. (2018). "Unsupervised feature learning via non-parametric instance discrimination". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3733–3742.

Xin, Bo et al. (2018). "A Study of the Point-spread Function in SDSS Images". In: *The Astronomical Journal* 156.5, p. 222.

Xu, Feiyu et al. (2019). "Explainable AI: A brief survey on history, research areas, approaches and challenges". In: *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8*. Springer, pp. 563–574.

Xue, Zhiwei et al. (2023). "Diffusion Models for Probabilistic Deconvolution of Galaxy Images". In: URL: https://dx.doi.org/10.48550/arXiv.2307.11122.

Yao, Wei et al. (2018). "Pixel-wise regression using U-Net and its application on pansharpening". In: *Neurocomputing* 312, pp. 364–371.

Ye, Quanzhi et al. (2019). "Toward efficient detection of small Near-Earth asteroids using the Zwicky Transient Facility (ZTF)". In: *Publications of the Astronomical Society of the Pacific* 131.1001, p. 078002.

Zaheer, Raniah and Humera Shaziya (2019). "A study of the optimization algorithms in deep learning". In: *2019 third international conference on inventive systems and control (ICISC)*. IEEE, pp. 536–539.

Zernike, Frits (1934). "Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode". In: *physica* 1.7-12, pp. 689–704.

Zhan, Hu and Lloyd Knox (2006). "How tomographic cosmic shear maps lead to constraints on dark energy properties". In: *arXiv preprint astro-ph/0611159*.

Zhang, Bing (2016). "Mergers of charged black holes: Gravitational-wave events, short gamma-ray bursts, and fast radio bursts". In: *The Astrophysical Journal Letters* 827.2, p. L31.

Zhang, Lin et al. (2012). "A comprehensive evaluation of full reference image quality assessment algorithms". In: *2012 19th IEEE International Conference on Image Processing*. IEEE, pp. 1477–1480.

Zhang, Xingguang et al. (2022). "Imaging through the atmosphere using turbulence mitigation transformer". In: *arXiv preprint arXiv:2207.06465*.

Zhang, Xingyun et al. (2020a). "Progress of liquid crystal adaptive optics for applications in ground-based telescopes". In: *Monthly Notices of the Royal Astronomical Society* 494.3, pp. 3536–3540.

Zhang, Yi-hao et al. (2020b). "Design of a large-format high-rate scientific CMOS camera". In: *X-Ray, Optical, and Infrared Detectors for Astronomy IX*. Vol. 11454. SPIE, pp. 19–25.

Zhang, Zhengxin, Qingjie Liu, and Yunhong Wang (2018). "Road extraction by deep residual u-net". In: *IEEE Geoscience and Remote Sensing Letters* 15.5, pp. 749–753.

Zhao, Hang et al. (2016). "Loss functions for image restoration with neural networks". In: *IEEE Transactions on computational imaging* 3.1, pp. 47–57.

Zhao, Zhuang, Jiyu Wei, and Bin Jiang (2022). "Automated Stellar Spectra Classification with Ensemble Convolutional Neural Network". In: *Advances in Astronomy* 2022, pp. 1–7.

Zhu, Jie et al. (2022). "Design of a scientific CMOS camera for astronomical observation". In: *X-Ray, Optical, and Infrared Detectors for Astronomy X*. Vol. 12191. SPIE, pp. 743–748.

Zhu, Jun-Yan et al. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232.

Zhu, Xiao-Pan et al. (2019). "Galaxy morphology classification with deep convolutional neural networks". In: *Astrophysics and Space Science* 364, pp. 1–15.

Ziegler, Carl et al. (2017). "Robo-AO Kepler Planetary Candidate Survey. III. Adaptive optics imaging of 1629 Kepler exoplanet candidate host stars". In: *The Astronomical Journal* 153.2, p. 66.

Žilková, Danica et al. (2023). "Space debris spectroscopy: Specular reflections at LEO regime". In: *Advances in Space Research* 71.8, pp. 3249–3261.