Supervised and Unsupervised Deep Learning Methods for Underwater Image Enhancement

by

Alejandro Rico Espinosa

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF APPLIED SCIENCE

in the Department of Electrical and Computer Engineering

University of Victoria

Supervised and Unsupervised Deep Learning Methods for Underwater Image Enhancement

by

Alejandro Rico Espinosa

B.Sc., University of Los Andes (Bogota, Colombia), 2020

Supervisory Committee

Dr. Alexandra Branzan Albu, Supervisor

(Department of Electrical and Computer Engineering)

Dr. George Tzanetakis , Departmental Member

(Department of Electrical and Computer Engineering)

ABSTRACT

Large amounts of underwater imagery are constantly collected for environmental monitoring studies, as they are essential for estimating marine biodiversity and abundance. However, this collected data has variable quality due to uncontrolled environmental factors that cause blur and color casting. We attempt to address this issue by proposing two novel methods for underwater image enhancement.

The first part of the thesis presents a deep learning architecture that integrates elements from classical methods to simultaneously address blurriness and color casting on underwater imagery in real time. We use two parallel architectures trained in a generative adversarial network scheme (GAN) with channel and spatial attention blocks to retrieve color, and discrete wavelength transform to preserve high-frequency components. Our experiments show that our method outperforms the state-of-the-art related works with respect to the structured similarity index metric (SSIM). Qualitative comparisons with color-checkers also show notable improvements over related works.

The second part of the thesis proposes an unsupervised deep-learning approach for underwater image enhancement, which eliminates the need for reference images for training. This is an important step forward as for real (not synthetic) underwater images there is no high-quality reference available. Our method is based on a mathematical model for image dehazing. We use three networks to estimate the transmission map, the atmospheric light, and the enhanced image and propose a new compound loss function. We achieve results comparable to state-of-the-art supervised methods with respect to the SSIM while performing optimally at near real-time inference speeds.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

Throughout this exciting journey, I had the opportunity to closely work with wonderful people. I would like to express my sincere gratitude to my supervisor, Dr. Alexandra Branzan Albu, whose unwavering support in both academic and personal aspects has been invaluable to me throughout this degree. Her guidance, patience, and advice have been instrumental in shaping my research and personal growth.

I am enormously thankful to my family for making it possible for me to be here today. Specially to my parents, sister, grandparents, and uncles. Their unconditional support and love have been a constant source of motivation and strength.

Finally, I extend thanks to all the members of the computer vision lab at the University of Victoria. Declan Mcintosh, Amanda Dash, Rhythm Vohra, Femina Senjaliya, Ali Soltani, and Dr. Mélissa Côté have not only been colleagues but also friends who were always open to listen and help whenever needed.

Chapter 1

Introduction

1.1 Background

Large amounts of underwater imaging data have been collected worldwide from multiple locations for ecological and biological research and monitoring [6, 7, 8]. Underwater data contributes to the analysis of oceans, rivers, and freshwater ecosystems [6, 7]. This facilitates tasks such as species counting, tracking of organism behaviors, and estimation of population stress levels [6, 7, 8]. Hence, fisheries, private companies, and university researchers have collectively acquired growing data, comprising large volumes of underwater imaging. This data continues to increase rapidly due to the ease of camera deployment and the decreasing data storage costs [6, 7, 9].

Despite this abundance of data, underwater images are limited in their utility due to their susceptibility to waterborne effects on images, especially color accuracy, and blurriness [3, 10, 11, 12]. Images affected by these distortions are less likely to fulfill the requirements for optical target recognition tasks, such as underwater object detection [13]. Moreover, the accuracy of object classification in autonomous underwater vehicles (AUV) explorations, is limited since it is not possible to rely on features related to the color of objects [14].

Environmental conditions that affect underwater image quality cannot be controlled during the data acquisition process. For example, objects can be obscured by water turbidity and algae bloom [11, 12], since these phenomena generate suspended particles that introduce background noise to images [15].

Additionally, due to light attenuation and absorption in water, images suffer from low contrast and haze [16, 13]. The propagation of light within the underwater environment is wavelength-dependent, and the red channel is highly attenuated by the absorption effect, resulting in distorted colors compared to ideal lighting conditions [17]. These distortions can often make it challenging to accurately identify species based on their appearance and study marine life behavior, limiting the utility of the imaging data [6, 7, 8].

Therefore, there is a need for image enhancement methods that address light scattering, color distortion, and blurriness effects. These methods also need to operate efficiently at real-time speeds to handle the enormous amount of data and live streams like those used for remotely operated underwater vehicle (ROV) exploration.

Some state-of-the-art methods for image enhancement will be discussed in detail in Chapter 2. These methods could be separated into traditional computer vision [18, 19] and deep learning [2, 11, 20]. Each of these approaches has benefits and limitations in terms of speed, quantitative improvements, and generalization [2, 3, 11, 20]. Traditional computer vision approaches can be highly efficient, but they are typically not easily generalizable [4, 19]. Deep learning methods are generally slow but have an excellent performance in improving image quality [21]. However, the performance of deep-learning methods is limited by the availability of training data.

This is also a problem considered in our research since the lack of real-world underwater data has constrained the training and study of many deep-learning models. It is practically impossible to obtain a real-underwater scene and the corresponding high-quality ground truth [2]. This is without even adding up the complexity of the diverse types of

underwater environments that we could get images of.

Plenty of datasets do not contain ground truth since there is no way to obtain real underwater images free of disturbances. Therefore, deep-learning approaches have been trained using synthetic datasets [3, 2, 22, 23, 21]. These are normally created by purely intuitive human perspectives of how objects should look [2]. Sometimes their contrast is exaggerated, they contain plenty of artifacts and display the wrong colors, as shown in the examples in Figure 1.1. For example, the coral images exhibit an over-contrast of blues and blacks, accompanied by high sharpness. Conversely, the turtle's shell displays an incorrect red tone. Alternatively, training data can be generated by emulating water disturbances on normal images [24]. However, these are not the real conditions of the environment, which can reduce model performance and bias their results, potentially limiting their use on real applications [25].

Figure 1.1: Examples of underwater images and corresponding groundtruth synthetically generated retrieved from the UIEB (Underwater Image Enhancement Benchmark) dataset [2] for training deep learning models.

Hence, there is a need to develop unsupervised deep-learning methods that allow models to be trained entirely independently of annotated data and reference information to be more robust.

1.2 Thesis Objectives and Contributions

As mentioned before, underwater image enhancement would be highly beneficial for multiple applications such as ROV explorations and marine conservation tasks. The main objective of this thesis is to propose novel methods that improve the overall quality and clarity of underwater images. Specifically, we address the blurriness, contrast issues due to scattering, and color cast issues due to the absorption effect. Additionally, the proposed solutions aim to demonstrate fast performance for real-time systems.

This thesis addresses the limitations of underwater images through two approaches: supervised and unsupervised learning. The supervised approach proposes a model that effectively enhances underwater images by using different key architectural components. Figure 1.2 show some examples of the results for this approach on different types of water and different types of distortions such as haze, contrast, and color casting. On the other side, unsupervised learning will go beyond the first approach, and demonstrate that deep-learning models could successfully be trained without the need of reference images. Figure 1.3 presents some results of the unsupervised approach.

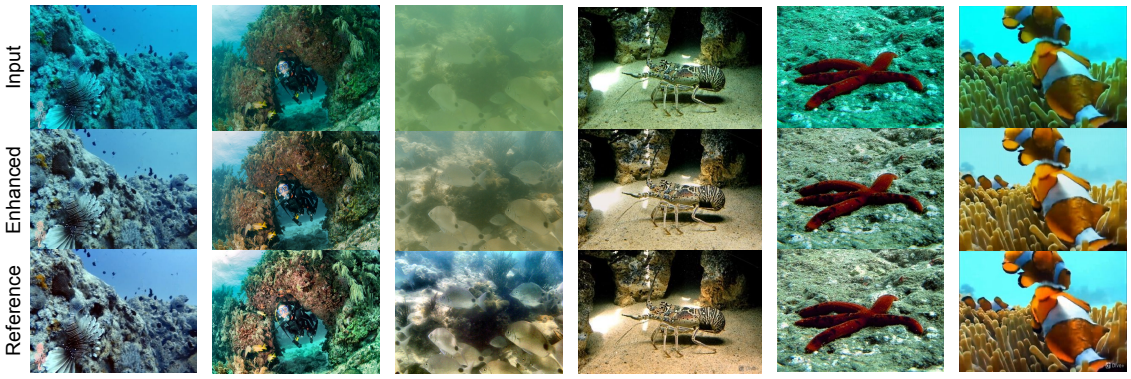

Figure 1.2: Qualitative examples of the results obtained through the supervised approach in the UIEB dataset [2].

The main contributions of this thesis are as follows:

- We propose a supervised deep learning-based method that uses several key archi-

tectural components to encourage good performance in addressing blur and color corruptions simultaneously. Moreover, our method is extremely efficient from a computational viewpoint.

- We present a novel approach based on the mathematical model for image dehazing to train deep-learning architectures for underwater image enhancement in an unsupervised way.

Future discussion and elaboration of these ideas will be found in Chapters. 2, 3, and 4.

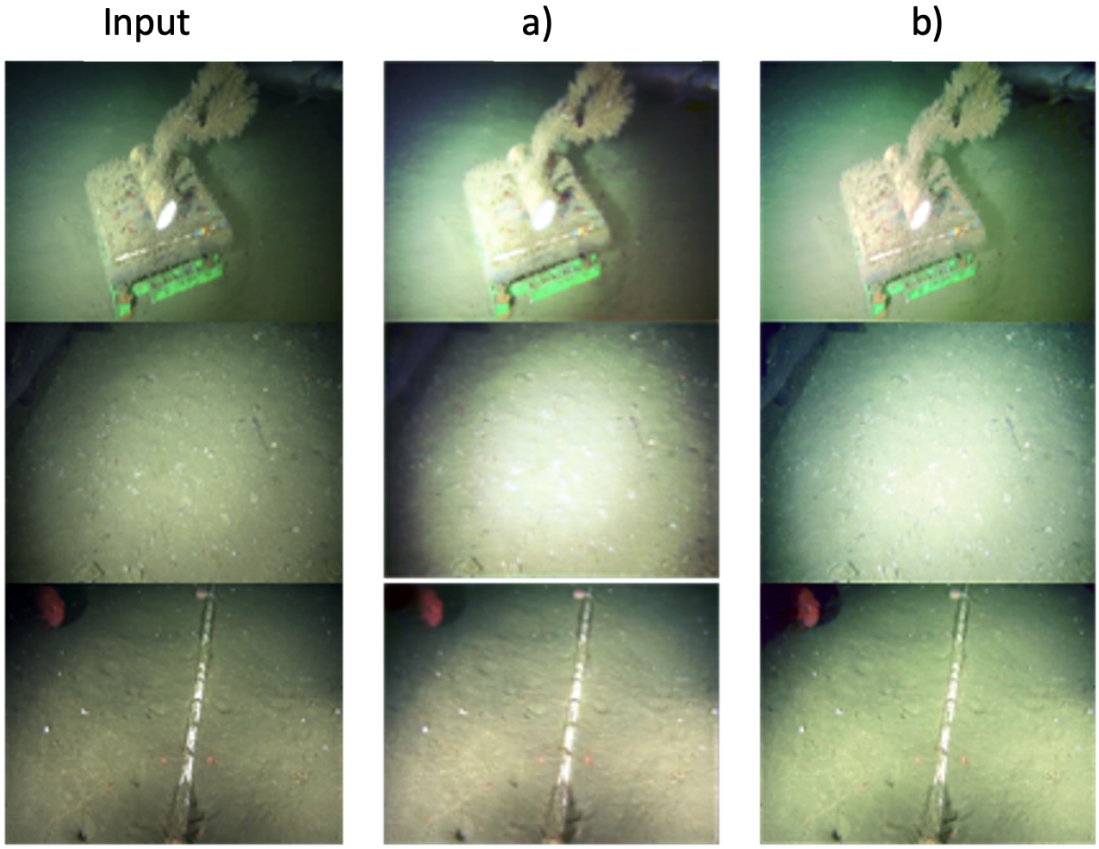

Figure 1.3: Qualitative examples of the results obtained through the unsupervised approach in the OceanDark dataset [3] where a) and b) are different configurations of the atmospheric network.

1.3 Publications

The contributions of this study led to one paper published in the Workshop on Maritime Computer Vision (MaCVi) 2023 at the Winter Conference on Applications of Computer Vision (WACV). Additionally, another paper was recently submitted to The International Symposium on Visual Computing (ISVC) conference.

1. **An Efficient Approach for Underwater Image Improvement: Deblurring, Dehazing, and Color Correction** (published in [26]),
2. **Unsupervised Deep-Learning Approach for Underwater Image Enhancement** (Submitted)

I was the first author of these works. The co-authors helped me write and review the papers during the submission process. Also, they guided me through the research process. The first publication [26] will be discussed in Chapter 3 and the details of the unsupervised study will be detailed in Chapter 4.

Chapter 2

Literature Review

In this chapter, we review the state of the art and the current progress in underwater image enhancement. This literature review was extracted and slightly modified from the publications that are further discussed in chapters 3 and 4. Section 2.1 presents the Dark Channel Prior, which is needed to explain since our unsupervised approach in Chapter 4 is based on this model. Further, Sections 2.2 and 2.3 review some related work based on traditional computer vision and deep learning approaches respectively for underwater and terrestrial image enhancement. Finally, the last section of this chapter slightly discusses some motivation behind our research.

There is a growing range of underwater image applications in increasingly challenging visual environments, particularly in shallow waters. These applications include environmental monitoring, research underwater explorations, fisheries, underwater infrastructure inspection, among others. As a result, there has been an increase in underwater image enhancement research [6, 7]. Attention has been directed to two important and challenging components of enhancement: dehazing, a particular case of non-homogeneous deblurring, and color correction. These image corruptions exist in all underwater imaging but are especially prevalent in naturally lit shallow water applications. As deblurring and dehaz-

ing corruptions have terrestrial analogs, we included terrestrial and underwater deblurring methods in this section [27, 28].

2.1 Dark Channel Prior

A widely used model to describe a hazy image in computer vision [29, 30, 31] is shown in equation 2.1 where J is the scene radiance, A is the global atmospheric light, t is the medium of transmission and I is the observed intensity [32]. The goal is to extract the individual components J, A, and T from the observed image I. The first multiplicative term describes how the radiance decays in the medium (direct attenuation), while the additive one describes the shift of the scene color due to the scattered light (airlight).

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{2.1}$$

A classical single image dehazing removal was described in [32], where they use a "dark channel" observation. Authors argue that in a patch of a haze-free image at least one channel in the RGB space will have pixels who's intensity is very low (close to zero). This means, the dark channel is the outcome of two minimum operators, as shown in equation 2.2 where x and y are the coordinates in a patch $\mu(x)$ of a channel c of the image.

$$J_{dark}(x) = \min_{y \epsilon \mu(x)}(\min_{c \epsilon (R,G,B)} I_c(y)) \tag{2.2}$$

Based on their observation, authors describe that the intensity of the Dark Channel is rough approximation of thickness of the haze, since a hazy image is usually brighter. However, for very bright pixels in haze-free images (i.e. sky), the dark channel is not a good prior. Fortunately, authors highlight that these pixels are very similar to the global atmospheric light descriptor A. Additionally, the presence of haze is fundamental for human to perceive depth [33]. Therefore, the estimated transmission map suggested in [32] includes

a constant parameter w (between 0 and 1) to keep more haze in for the distant objects. Based on the initial observation of the dark channel prior that the dark channel of J will be close to 0, equation 2.2 can be simplified and reorganize to define final estimation of the transmission map as shown in 2.3, where A is a normalized atmospheric light in the range [0,255].

$$t(x) = 1 - w \min_{y \epsilon \mu(x)} (\min_{c \epsilon (R,G,B)} \frac{I_c(y)}{A_c}) \tag{2.3}$$

As previously mentioned, the pixels with highest intensity in the observed image I are selected as the atmospheric light. Then, a haze free version of the original image can be obtained with equation 2.4

$$J_c(x) = \frac{I_c(x) - A_c}{max(t(x), t_0)} + A_c \tag{2.4}$$

Note that in equation 2.4 includes a term t_0. Since the transmission map could be close to zero, this additional term allows to preserve small amounts of haze in these regions acting as a lower bound.

2.2 Classical computer vision approaches

Multiple approaches to underwater image enhancement diverge from different prior-based classical computer vision terrestrial models, some showing better performances than other techniques [34]. Recently, [35] presented an efficient underwater image restoration using the dark channel prior with some variations. They proposed to use arithmetic Mode operation as a new global atmospheric light estimator. They show better results than other methods with evaluation on non-reference metrics such as underwater color image qualitative evaluation (UCIQE) and underwater image quality metric (UIQM). UCIQE is a linear combination of saturation and contrast, proposed to quantify image quality primarily based

on its color [36]. On the other side, the UIQM combines image colorfulness, sharpness, and contrast [37]. Recall that these metrics are build based on subjective evaluation data [38].

Additionally, [11] proposed a fusion-based method only relying on the information gained from the degraded image. They separately enhance the color and contrast of the image and then incorporate several other weight maps to account for the non-linear image corruption from long-distance objects [11]. These weight maps and improved images are then fused to generate an enhanced image. The method is a classical computer vision method and it is agnostic to image scene structure or the specific underwater conditions [11].

2.3 Deep learning methods

Also, multiple deep-learning approaches have been suggested for the underwater image quality improvement. For instance, [23] suggest the use of a dual network to handle the color cast and blurry details in underwater images. The dual network architecture consist of two individual networks which separately estimates the structure of the image and the wavelet sub-images. They use the discrete wavelet transform (DWT) to extract the different frequency sub bands and used them as inputs for their networks. One of the models target the color recovery by employing feature representations from multiple color spaces. The other architecture is designed to enhance the image details by using the high frequency sub-bands. Their model was trained under synthetic dataset due the lack of ground truths for underwater images. Furthermore, with the proliferation of deep learning-based statistical methods, the classical approaches based on visual priors and physical models have become less prevalent while still being used in applications lacking ground truth data.

This was not the first time the DWT was used with deep learning models, previous

studies also proposed the use of DWT feature spaces for better performance. For example, with the use of a wavelet residual network [28] discovered the benefit of learning on subbands. Wavelet residual networks are able to learn on additional frequency subbands, increasing model representational power. Also, [39] propose a discrete wavelet transform (DWT) based GAN for dehazing: DW-GAN . DW-GAN consists of two parallel networks seeking to improve details in small datasets, more specifically, using non-homogeneous hazed images. The first network method uses DWTs to downsample the feature maps generated by the CNN while maintaining more of the high-frequency information in additional skip connections [39]. This provides the decoder with extra high-frequency feature map information. These additional features allow the network to better reconstruct the deblurring image, specifically by creating a sharper reconstruction. Then, the predictions of the two parallel CNNs are averaged [39]. The second shallower knowledge adaptation branch uses pre-trained weights from a classification problem to increase model performance on smaller datasets [39]. This is motivated by the relative rarity of ground-truth information for dehazing problems.

Moreover, a method with significant success in deblurring using a terrestrial dataset was DeblurGAN from [27]. They present a conditional GAN loss in conjunction with a reconstruction loss [27]. This loss improves perceptual losses by penalizing the model for generating reconstructions that can score well on perceptual metrics but do not accurately sharpen the image. This improves the overall sharpness of the final images [27]. They also propose using a secondary detector's performance on the enhanced images as a metric or real-world applicability over previously used metrics like PSNR or SSIM [27]. To our knowledge, this method was not used in underwater image enhancement previous to [4].

Another deep learning approach was introduced by [40], by presenting a U-shaped transformer for the underwater image enhancement task. Their proposed method uses a channel-wise multi-scale feature fusion transformer and a novel spatial-wise global feature

modeling transformer [40]. The latter is designed to increase the model's attention to regions of the image with significant attenuation and the relationship between color channels in the overall image [40]. The authors also propose a multi-space loss function that considers the reconstructed image in multiple color spaces to improve the qualitative results [40].

Finally, another interesting approach was described by [5], who proposed the use of three networks and the mathematical model of a terrestrial haze image as an enhancement unsupervised method. Two of the models avoid downsampling to retain image details, while the third one is based on the latent Gaussian distribution that global atmospheric light has on terrestrial images. It is possible to assume this distribution for A since the atmospheric is global and independent from the image content. They achieve state-of-the-art results and beat other unsupervised deep-learning approaches in the SSIM and PSNR.

The SSIM aims to compare structures of the reference and distorted signals to determine how similar two images are based on its luminance, contrast and structure with reference of the human visual system. [41, 42]. SSIM perceptual metric evaluates the visible structures of the image and separates the influence of luminance of the surface of an object that is being observed [41]. On the other side PSNR is directly related to the mean square error while comparing pixel wise two images. This metric is not upper bounded and implies numerical comparison, therefore a highly difference between the images would result in a score near zero [42]. PSNR is more sensitive to Gaussian noise, while SSIM is more sensitive to compression [42].

2.4 Motivation

State-of-the-art approaches have acquired good-quality results. Nevertheless, we consider that underwater image enhancement requires further research. For example, some of the

successful studies to enhance images such as [39] and [27] are designed specifically for terrestrial purposes, targeting individual disturbances one at a time, and do not consider underwater specific issues. On the other side, studies specific to these underwater images could have disadvantages. For instance, the transformer proposed by [40] for color recovery in underwater atmospheres requires a large amount of training data and other methods such as [20] and [11] can be relatively slow. Also, the approach proposed on [5] only trains in a single image. This means it is designed to enhance in a non-general way, which is not efficient for real-time applications or large amounts of video data.

Chapter 3

An Efficient Approach for Underwater Image Improvement: Deblurring, Dehazing, and Color Correction

This chapter presents a supervised deep-learning approach to correct blurriness and color casting in underwater images. Our method addresses these two conditions simultaneously while having fast performance. Our focus was not on building a whole new architecture from scratch, instead, we selected and integrated key components to construct an architecture capable of improving image quality in real-time speed performances. This chapter illustrates design choices and how each component improves the overall performance of the task while ensuring a highly efficient response. The content of this chapter was originally submitted, peer-reviewed, and published to the Workshop on Maritime Computer Vision (MaCVi) 2023, held during the Winter Conference on Applications of Computer Vision (WACV). Minimal modifications were made such as updating references and rephrasing some statements for clarity.

3.1 Introduction

Due to light attenuation and absorption in water, images suffer from low contrast and haze [16, 13]. For this, we introduce a deep learning-based method that uses several key architectural features to encourage good performance in blur and color corruptions while remaining very efficient. Our proposed method uses a U-Net style encoder-decoder architecture which is trained in a GAN structure with a second adversarial network [43, 44].

GANs (generative adversarial networks) were proposed by [45]. The idea of using GAN is to define a game between two competitive networks, the discriminator and the generator [27]. In the context of image enhancement, the generator receives a noise image as input and attempts to generate a clean output. Meanwhile, the discriminator receives generator output (improved image) as input along with a real clean image and tries to distinguish between them. The reward for the generator is associated with fooling the discriminator and the reward for the discriminator is associated with correctly distinguishing the improved image against the target goal image.

Our initial network is designed to be efficient and small [44]. We then present specific modifications of this base training regimen and architecture, which are well motivated by the image corruptions we consider more damaging to underwater images: atmospheric scattering, blur, and color absorption. The modifications minimally degrade the model's efficiency as a core goal of our work is to present greater than real-time performance, basing its effectiveness primarily on the structural similarity index (SSIM) score. First, we include Discrete Wavelet Transform (DWT) skip connections [39]. We apply the DWT to all channels of our feature maps before down-sampling and skip connections to increase the accessible information for the decoder. The DWT increases the number of feature channels, separating them into multiple representations at different frequency bands [39]. We also propose to use the channel and spatial attention blocks (CBAM) as an effective way to increase the color recovery of our model. This is because CBAM helps to represent

multiple feature channels into relative color channels [46]. Finally, we add a gradient penalty to our GAN discriminator to enforce a Lipschitz constraint; this makes the training of our reconstruction network more stable [43].

Our proposed method shows state-of-the-art results in SSIM and inference speed on the UIEB test set [2]. Also, our model is competitive with PSNR with the current state-of-the-art methods. In this chapter, we compare our method with existing classical computer vision, statistical deep learning methods, and hybrid methods. Section 3.2 detail our proposed approach and each of its principal components, Section 3.3 shows the quantitative and qualitative results compared with other previous approaches, and Section 3.4 finally state our conclusions. The ideas that will be found in this chapter can be summarized as:

- Propose a novel architecture that achieves state-of-the-art performance while being extremely fast.
- Propose a modified version of the initial architecture that may involve increased computational costs but demonstrates the best performance among current state-of-the-art approaches
- Demonstrate the benefits of each implemented component in enhancing underwater images, highlighting their relevance to the task.

3.2 Proposed Approach

3.2.1 Initial Approach

Our proposed approach consists of an end-to-end CNN-based image enhancement method. We modified the classical U-Net encoder-decoder structure incorporating several key components. U-Net is composed with repetitive blocks of convolutions, ReLu and max pooling

during the downsampling and upsampling. The bottleneck is also particular of this structure, forcing the network to learn the compression for the input data. Each of these target specific corruptions of underwater images. This ground-up design from a lightweight architecture allows us to have greater than real-time efficiency beyond comparable previous works. First, we incorporate Discrete Wavelet Transform (DWT) skip connections while down-sampling in our encoder [39]. This preserves texture details for sharper output images during the dehazing task by maintaining high-frequency feature map components. Next, we adjust the design of our architecture using spatial and channel-wise attention blocks between convolutions. Specifically, we incorporate Channel Based Attention Module (CBAM) blocks which better exploit relationships between feature maps. This lead to the effect of creating significantly improved color accuracy. Hence, we use them in a novel way to address the absorption effect of the red channel found in the underwater images. The entire model is presented in Figure 3.1.

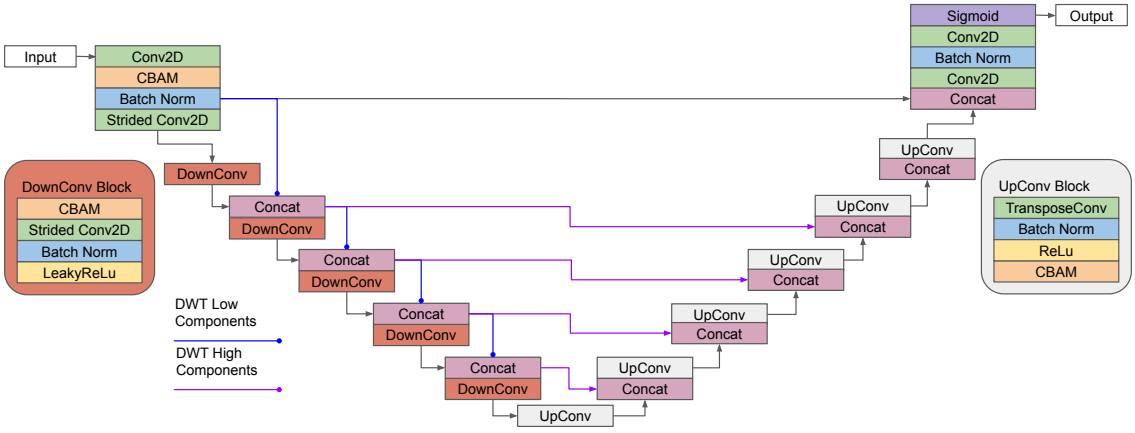

Figure 3.1: Proposed architecture based on the U-Net with multiple skip connections, DWT algorithm and CBAM blocks. The low frequency DWT information paths are highlighted in blue and the corresponding high frequency components are highlighted in purple.

Additionally, we also present a modification of our model that incorporates a second parallel deblurring Res2Net architecture with an additional global skip connection resulting in better generalization and faster training than the conventional Res2Net [27]. This

modification of our proposed approach does come at the cost of model efficiency. Figure 3.2 shows this variation.

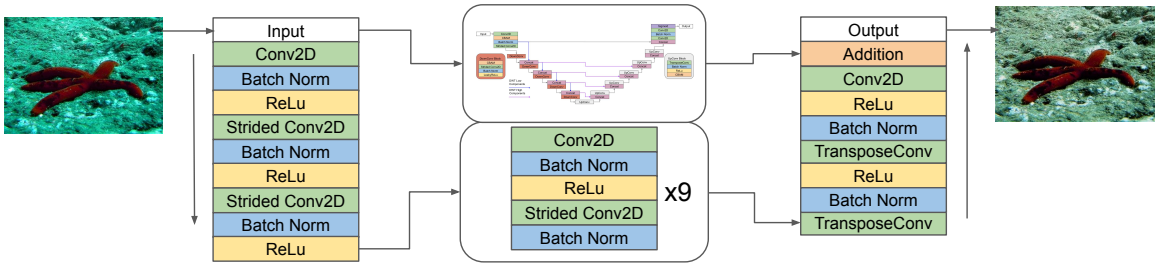

Figure 3.2: Proposed architecture variant with Res2Net second branch in parallel. The parallel Branch is simply added with the final activation function or Sigmoid for final image reconstruction. Notably the processing only down-samples the input image by a factor of 4, then symmetrically up-samples to the original resolution.

Finally, we formulate the training of our network against a second adversarial network in a GAN based structure. The specific training structure we select is the Wasserstein variation because it helps mitigate issues with Jenssen-Shannon divergence. More details of these issues in the context of image reconstruction can be seen in Section 3.2.3. We modify this variation further by including a gradient penalty in the discriminator rather than clipping the values to enforce a Lipschitz constraint [43]. This modification increases the sharpness and color restoration in the final images while making the model more robust when the second deblurring branch is added. The following sections will discuss each of these contributions to our final novel architecture in detail.

Discrete Wavelet Transform Skip Connections

It is very likely that the filters learned by convolutional neural networks for denoising and smoothing tasks act as low-pass filters (gaussian), completely removing high-frequency information. Hence, the formulation of the DWT as a method for preserving high-frequency features was used in skip connections [39]. Retaining more information from the feature maps in the high-frequency ranges is of particular concern for the reconstruction of hazy

images.

The DWT decomposes the image into high and low frequency sub-images which as a collection retain more of the original information than standard down-sampling. Specifically, the 2D DWT in the first level is implemented by performing the convolution operation with four different filters f_{LL} low-low-pass, f_{LH} low-high-pass, f_{HL} high-low-pass, and f_{HH} high-high-pass. The low-frequency components are comprised of the downsampling feature outputs obtained through the convolutional layers. The high-frequency components are derived using the 1D Haar filter bank. The 2D problem is then formulated as a sequential repetition of the 1D procedure, first applying the filter in rows and then in columns [47]. You can continue applying the same procedure in the resulting LL sub-band to increase in the DWT levels. Equations for the first level can be formulated as:

$$f_{LL}(x, y) = f_L(x) f_L(y) \tag{3.1}$$

$$f_{LH}(x, y) = f_L(x) f_H(y) \tag{3.2}$$

$$f_{HL}(x, y) = f_H(x) f_L(y) \tag{3.3}$$

$$f_{HH}(x, y) = f_H(x) f_H(y) \tag{3.4}$$

Where the second filter is applied after downsampling the signal by 2. Figure 3.3 illustrates this process. Despite testing our architecture with another filter such as biorthogonal (LeGall 5-3), we find that conventional Haar filters performed better and faster in our enhancement task. Therefore, the filters we use in our approach are defined by:

$$f_L = \frac{1}{\sqrt{2}}(1 + z^{-1}) \tag{3.5}$$

$$f_H = \frac{1}{\sqrt{2}}(1 - z^{-1}) \tag{3.6}$$

Where z represents the z transform.

As a whole understanding of the process, the low-frequency components generated from the strided convolutions are passed down in the encoder and the results of upsampling from our decoder transpose convolutions are concatenated with the high frequency components retrieved from the skip connections. In this way, the network is forced to learn from both the spatial and the frequency domain, retaining abundant high frequency feature content which showed to improve the sharpness and contrast in the reconstructed images while learning the color mapping from hazy to haze-free images [39].

CBAM

The convolution block attention module was initially proposed by [46] for the domains of image classification and object detection. Their main purpose is to infer attention maps along the channel and spatial dimensions without modifying the input dimension. This helps the network to improve the representation of interest by focusing on the important features and suppressing the unnecessary ones. The process is described in Equations 3.7 and 3.8, where F is the input feature map, M_c is a 1D channel attention map, M_s is a 2D spatial attention map, F' is the feature map defined after the channel attention mapping, F''

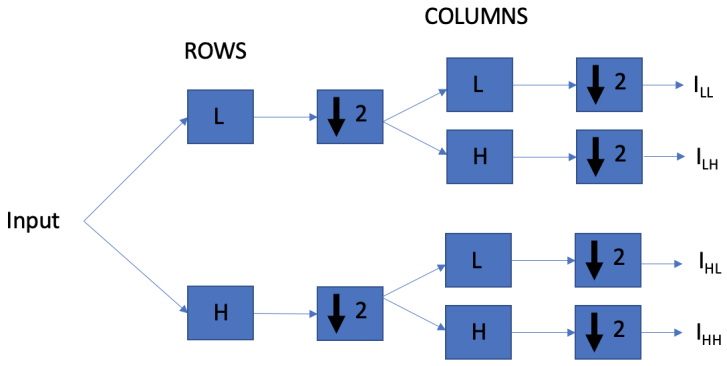

Figure 3.3: Two dimension DWT algorithm illustrated as a repetition of one dimension DWT first applied into rows and then into columns. L makes reference to the low-pass filter, H to the high-pass filter and 2 means down-sampling by two

is the final refined output, and $*$ denotes the convolution operation.

$$F' = M_c(F) * F \tag{3.7}$$

$$F'' = M_s(F') * F' \tag{3.8}$$

Differing a bit from the original purpose of the module, we propose the use of CBAM (especially channel attention) to improve color correction. Strong color correction in the CNNs requires good integration of multiple feature channels together. Nevertheless, the integration of channels is highly dependent on the weighting through the network. Hence, using the CBAM helps to learn the most effective relative weighting of these channels for reconstruction. CBAM modules are better suited for strongly channel-dependent applications. Since we are trying to address different problems simultaneously, underwater image improvement was a clear candidate for the use of this attention module.

Instead of using the CBAM in the U-Net throughout the skip connections like suggested by [48], we decide to use a similar approach to the one suggested by [49]. Specifically, we place these blocks at the encoder while down-sampling before each convolution layer, to ensure a more optimal feature extraction and make the network focus its attention on the main characteristics before losing information during down-sampling. Additionally, we use the CBAM in the decoder as well. We apply the attention module after we concatenate all input skip connections to ensure we have a good synthesis of these channels from multiple sources (DTW skip connections with high frequency components, and upsampled feature maps) for the transpose convolution.

3.2.2 Deblurring branch (Model variant)

Inspired by [50] and [39] we propose an augmentation to our initial structure to improve performance at the detriment of some efficiency. We use a second parallel CNN in conjunction with our proposed architecture and then synthesize our final reconstruction from both of these models. The idea for this is to predict a residual image, similarly to methods previously used in physical models, for the dehazing task and input this knowledge to the DWT branch with the simple summation, see Figure 3.2. Due to the deblurring nature, the output image quality might be reduced. Then, we found that including the gradient penalty will compensate for this effect making the model more robust and getting sharper images. This approach increases network complexity and processing time per image but overall achieves better performance in the underwater image enhancement metrics, as shown in Table 3.2.

For this branch, we use the lightweight CNN proposed by [27] which is similar to [51] initially presented for the style transfer task. With this architecture, we aim to integrate the residual image knowledge and shallower encoder features to enhance the underwater images, specifically their sharpness by only processing the image at a higher resolution. It contains nine residual blocks (convolutional layer, instance normalization layer, ReLu activation function, and Dropout with probability of 0.5), two stride convolution blocks with $\frac{1}{2}$ stride, and two transposed convolution blocks. This architecture was found to train faster and generalize better due to a unique global skip connection [27]. Furthermore, since back scattering is usually homogeneous in our domain, we avoid unnecessary pixel shuffling layers previously suggested by [39] for terrestrial dehazing.

3.2.3 Gradient penalty (Wasserstein GAN variation)

Motivated by issues with GAN convergence in this formulation we utilize the contributions of [52] who described GANs training difficulties caused by the Jenssen-Shannon divergence. To solve this divergence [43] proposed the addition of a gradient penalty into the

discriminator to enforce the Lipschitz constraint. This constraints limits the predictions of the discriminator by controlling the gradient of its output to prevent training instability on the generator. Compared to other methods to solve this training instability (like gradient clipping), this variation makes the training more robust and requires almost no hyperparameter tuning [27].

3.2.4 Loss Function

We denote the recovered enhanced image as I^e, the underwater original input image as I^u, the groundtruth image as I^g, and the GAN generator and discriminator as G and D respectively.

Smooth L1 Loss

Let $I_c(i)$ be the intensity in pixel i of the c-th channel and N the total number of pixels.

$$L1_{smooth} = \frac{1}{3N} \sum_{i=1}^{N} \sum_{c=1}^{3} \alpha(e) \tag{3.9}$$

Where e is the error $(I_c^e(i) - I_c^g(i))$, and $\alpha(e)$ is a function of it defined as:

$$\alpha(e) = \begin{cases} 0.5e^2, & \text{if } |e| < 1 \\ |e| - 0.5, & \text{otherwise} \end{cases} \tag{3.10}$$

MS-SSIM Loss

As explained by [39], denotes two windows of common size E and U centered at pixel i in the recovered enhanced image and the underwater input image respectively. Then, a Gaussian filter is applied to each window, and means μ_E,μ_U standard deviation σ_E,σ_U, and covariance σ_{EU} are computed. The SSIM for pixel i is defined as:

$$SSIM(i) = \frac{2\mu_E\mu_U + C_1}{\mu_E^2 + \mu_U^2 + C_1} * \frac{2\sigma_{EU} + C_2}{\sigma_E + \sigma_U + C_2} = l(i) * cs(i) \tag{3.11}$$

where $l(i)$ represents luminance and $cs(i)$ represent contract and structure measures. Also, C_1 and C_2 are constraints to stabilize division with a weak denominator. The MS-SSIM loss uses M levels of SSIM (1-MS-SSIM). This means using multiple scales of the images and averaging the SSIM of each of them, computing global and local similarity and making the metric more robust.

$$MS(i) - SSIM(i) = l_M^{\alpha}(i) * \prod_{m=1}^{M} cs_m^{\beta_m}(i) \tag{3.12}$$

with α and β default parameters suggested in the original study [53].

Adversarial loss

Since we are using WGAN-GP for training, the adversarial loss is computed as:

$$L_{GAN} = \sum_{n=1}^{N} -D(G(I^u)) \tag{3.13}$$

Total Loss

The total loss is the combination of losses defined as:

$$L_{total} = L1_{smooth} + R1 * L_{Ms-SSIM} + R2 * L_{GAN} \tag{3.14}$$

Where R1=0.2 and R2=0.005 are hyperparameters inspired by fast performance showed in [39].

3.3 Results and Discussion

We present state of the art results on the UIEBD dataset [2] while operating at a greater than real time inference speed (30 frames per second on images of size 1600x1200 pixels). Details of the dataset, hardware, and experiments are described in this section.

3.3.1 Dataset

For our experiments, we used a real-world underwater dataset: UIEBD [2]. This dataset contains 890 underwater images and their corresponding groundtruth. These groundtruth were obtained through subjective selection after studying 12 different underwater images enhances. We use standard splits for direct comparison [15]. The first 700 images are used for training or validation and the remaining 190 for testing. Also, for the qualitative color comparison, we used the Color Checker dataset [54].

3.3.2 Training Details

We use Adam optimizer with an initial learning rate 1e-4 for 400 epochs to train the generator (our model) and the discriminator. We decrease the learning rate by half on epochs 250 and 350. Our models were trained on a GPU NVIDIA GeForce RTX 3060 with 12GB of memory and used a batch size of 2. For the dataset, each image was randomly cropped in patches of 256x256. Then, augmentation was performed by randomly rotating it 90, 180, or 270 degrees and horizontal flipping. No other augmentations were performed to ensure the model was specifically tailored to underwater image corruption.

3.3.3 Ablation

We performed different experiments to show that each module of our proposed architecture helps to improve the enhancement of underwater images. The multiple variations of the

model are shown in Table **??**. It is possible to appreciate that using a different DWT filter increases problem complexity and does not improve the scores. Also, note that the deblurring branch introduces noise (PSNR decreases). Then, it is necessary to use the gradient penalty while training to compensate for the sharpness and color quality, leading to higher scores.

CBAM	DWT	Deblurring	Grad Penalty	Average SSIM	Red SSIM	PSNR (dB)	Inference time (s)
✓	Haar	✓	✓	**0.8802**	**0.8465**	**20.9226**	0.3266
✓	Haar	✓		0.8651	0.8264	19.8227	0.3267
✓	Haar			0.8703	0.8434	20.8694	0.0248
✓				0.8683	0.8265	20.3843	0.02324
				0.8637	0.8126	19.4034	**0.01457**
✓	Biorthogonal			0.8606	0.8046	19.4747	0.3094

Table 3.1: Ablation study showing individual improvement of each module of the proposed architecture using UIEBD dataset. The base model (row 5) makes reference to the U-Net architecture.

3.3.4 Quantitative Results

Our proposed method shows state-of-the-art results on this dataset in both SSIM and inference speed. The details of our results are shown in Table 3.2.

We present two configurations of our method: one minimal configuration and a configuration with a second parallel deblurring branch. Our minimal implementation shows state-of-the-art SSIM results on UIEBD task with a score of 0.8703. This is only a slight improvement over other methods such as the methods proposed by DWG but we show a nearly 18x speed-up over their method while improving the SSIM scores. Further, when we introduce the second deblurring branch we show a larger gap to other methods achieving an SSIM score of 0.8802, almost 2 percentage points higher than the next comparable method. This moderate increase in scores over our minimal method does come with significant costs with regards to speed.

Notably, our method does not gain state of the art results in PSNR in either configu-

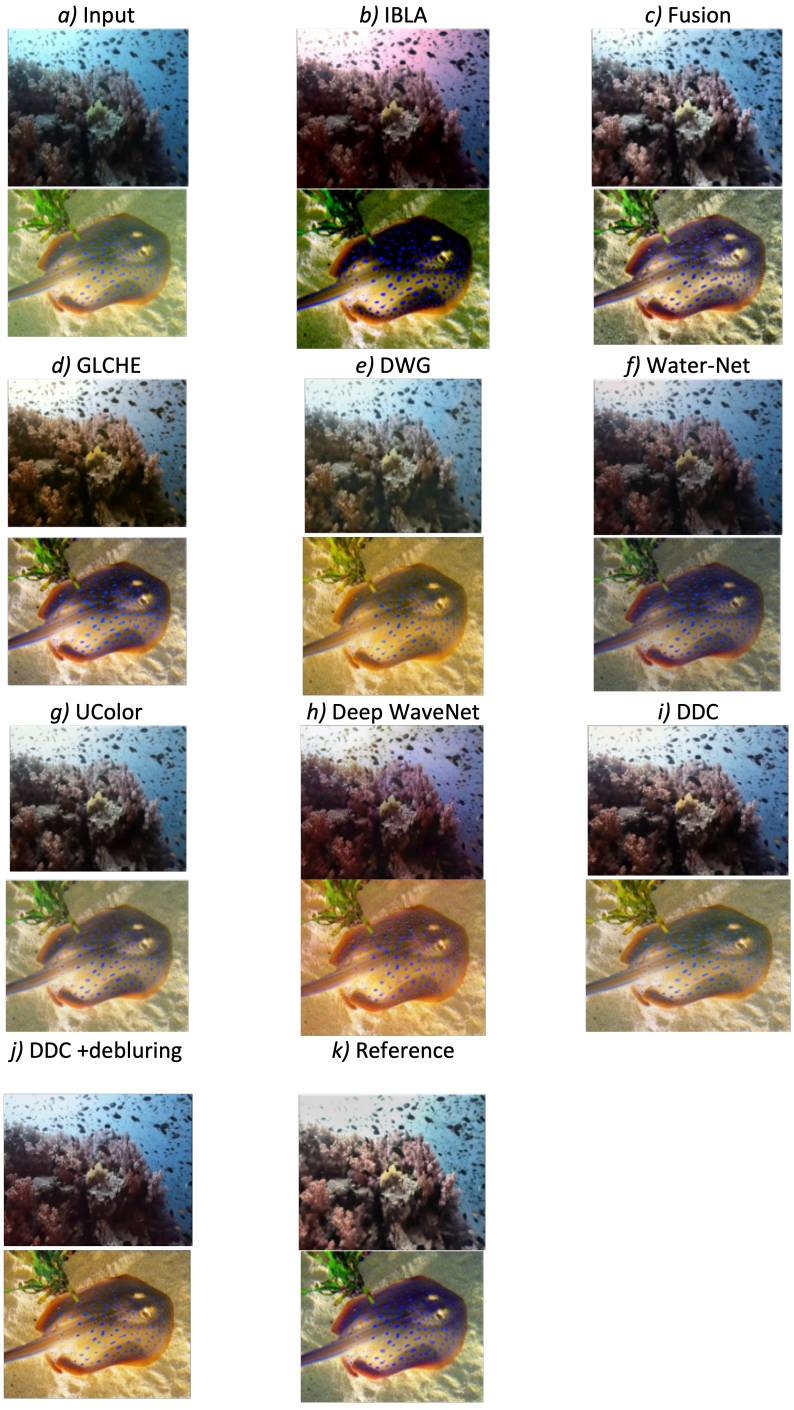

Figure 3.4: Results of different models for UIEBD dataset for qualitative comparison. These same models are used during the quantitative analysis.

Method	**PSNR↑**	**SSIM↑**	**Inference Time↓ (s)**	**Type**
IBLA [18]	14.3856	0.4299	38.71 (E)	Classic
Fusion [11]	21.1849	0.8222	6.58 (P)	Classic
GLCHE [19]	21.027	0.8487	<u>0.05</u> (E)	Mix
DWG [39]	19.6727	0.8614	0.4487 (E)	Deep
Water-Net [2]	19.3134	0.8303	0.61 (E)	Deep
Ucolor [20]	20.63	0.77	2.75 (P)	Deep
SCNet [15]	**22.08**	0.8625	0.4495 (E)	Deep
Deep WaveNet [21]	<u>21.57</u>	0.8	1.16 (P)	Deep
DDC	20.8694	<u>0.8703</u>	**0.025**	Deep
DDC + deblurring branch	20.9226	**0.8802**	0.3266	Deep

Table 3.2: Quantitative results on the UIEDB [1] test set. We show state-of-the-art results in SSIM and inference speed while having competitive results in Peak Signal to Noise Ration (PSNR). Best scores shown in bold, second best underlined. In the inference time column, the letter E indicates the experiment was run with the same training conditions, while the letter P indicates it is a published result. DDC makes reference to our base proposed approach (Debluring, Dehazing, and Color Correction)

ration, only achieving 20.8694 and 20.9226 with our minimal and deblurring branch configurations respectively. We argue this is not a significant blemishing, because PSNR is not as good of a tool for image quality especially in this domain of underwater image improvement and on this dataset [55][56][2]. The goal of this dataset and our method and this dataset was qualitative improvements on images, even the 'ground truths' in this dataset are based purely on subjective studies [2]. For this qualitative goal, SSIM is a much better metric than PSNR as it is based on luminance, contrast, and structure as a human perception analog rather than an unnormalized absolute divergence metric like PSNR [55][56].

It is possible to observe that prior-based methods (classic) are computationally expensive and do not achieve the best scores in the evaluation metrics. There are several causes for this. For instance, the classic methods generally require more complicated pipelines to achieve strong performance as they are required to generalize for many possible underwater image degradation cases. In general, as statistical deep learning methods use CNNs to

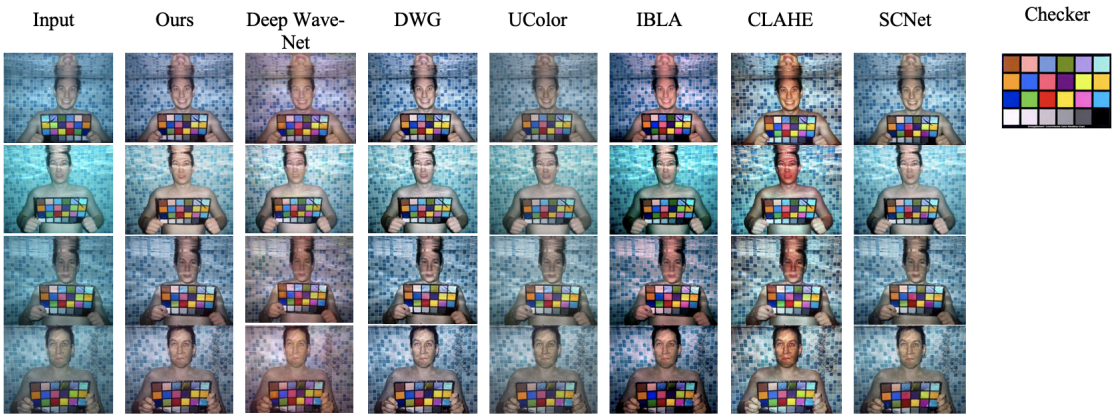

Figure 3.5: Results of different models for the ColorChecker dataset for qualitative comparison. These same models are used during the quantitative analysis.

process the underwater images they are able to leverage GPU parallelization to be faster than existing classic methods . The fastest method outside of the one we propose for image enhancement is a hybrid method which preformed local and global statistical methods, synthesizing them together to increase their efficiency.

We present the fastest high performance method for underwater image enhancement on this dataset. Our proposed method (minimal configuration) is able to process images at a rate of 40/s on low-end consumer hardware. This is two times faster than any comparable previous method in this domain, and the only to achieve the real-time performance necessary for being deployed in applications like on controlled Unmanned Underwater Vehicles (UUV)s. We achieve this speed through a minimalist design philosophy starting simply with a well tested encoder-decoder structure and only adding efficient modifications which address specific challenges in underwater images. Our deblurring branch configuration does not perform at real time speeds but still shows competitive results compared to other methods with an inference speed of 0.3266s/image.

3.3.5 Qualitative Results

Figure 1.2 shows qualitative results of the proposed model in different types of water. It is possible to appreciate that our method recovers color efficiently by addressing the red channel absorption effect. Also, it retrieves information close to the truthiness colors of objects, especially in the red channel. On the other side, our model still creates brightness environment artifacts on images with objects in different depths when the green channel is more representative, visible through some examples presented in Figure 1.2. This might be related to the nature of the dehazing design. Nevertheless, details are retrieved avoiding unnatural looks with sharp results.

Additionally, Figure 3.4 presents a qualitative subjective comparison of the results achieved for each model. We appreciate that classical methods create unnatural color artifacts while our results show a closer real recovery of colors, maintaining a balance between red, green, and blue channels to create natural looking images with minimal brightness artifacts. The coral example in Figure 3.4 clearly shows that our model handles the red channel absorption effect efficiently and still performs well on recovering the blue, white and yellow objects. This means that our model effectively learns the weighting of color channels, justifying that attention (CBAM) is a fundamental component of the architecture. Furthermore, we might see a reduction of the brightness artifacts of our model when we add the deblurring branch, which leads to an increase in the image sharpness as shown in the second sample of Figure 3.4.

Furthermore, the ColorChecker dataset [54] was used for additional qualitative analysis. Some of the results are presented in Figure 3.5. We appreciate that classical models do not provide the best color recovery and introduce unnatural brightness artifacts. DWG and SCNet provide high quality results, however, some colors such as blue and pink are not correctly recovered. Our model seems to avoid unnatural looking artifacts and recover colors close to the true ones.

3.4 Outcome

In this chapter, we proposed an efficient real-time method for underwater image enhancement that achieves state-of-the-art results. Our quantitative analysis is primarily focused on the SSIM index since it is based on luminance, contrast, and structure informed by human perception. We achieved fast results in inference time and SSIM on the UIEB dataset. Additionally to setting comparison points with other models, the UIEB dataset is ideal to demonstrate the model behavior in real underwater images and not only in synthetic data. Our proposed method effectively builds up from a minimal U-Net based encoder-decoder architecture with DWT skip connections, CBAM blocks for better color casting, domain-specific training loss for GAN training, and an optional deblurring branch configuration. Our method debuts the only real-time (40frames/second) method on this dataset while presenting state-of-the-art results on SSIM of 0.8703. Our model variant increases this improvement over other methods to an SSIM of 0.8802, outperforming state-of-the-art results.

For future work, it will be important to address the underwater image enhancement requirements from an unsupervised learning perspective, since results will not depend on visual qualitative analysis for groundtruth construction. Also, this will lead to more realistic models that could act accurately in different types of water.

Chapter 4

Unsupervised Deep-Learning Approach for Underwater Image Enhancement

This chapter presents an unsupervised deep-learning approach to enhance underwater images. Our methodology was inspired by the YOLY approach [5] previously described in section 2.3, with a focus on underwater disturbances. The purpose of this chapter is to suggest an alternative way to train deep-learning models without relying on ground truth data. In this study, we use the base network described in chapter 3. The content of this chapter was recently submitted for publication at The International Symposium on Visual Computing (ISVC) conference. At the moment of writing this chapter, we have not received a notification from ISVC reviewers yet. Minimal modifications were made such as updating references and rephrasing some statements for clarity.

4.1 Introduction

On underwater image datasets it is usually common that ground truth paired data is not available. This is due to the environment and physical difficulties of obtaining images free of disturbances. Therefore, most of the available ground truths used up to date for under-

water image enhancement approaches are generated synthetically as described in chapter 1. To avoid dealing with this lack of information, we propose an unsupervised deep learning approach based on the mathematical image formation model to enhance underwater images. Our approach reaches state-of-the-art results in the structural similarity index and is performed in speeds around 10 frames per second. We primarily concentrate our attention on formulating a methodology to improve the images, rather than designing a new CNN architecture. Specifically, our main contributions could be listed as:

- To the best of our knowledge, this is one of the first unsupervised deep-learning approaches proposed for underwater image enhancement.
- Introduction of two novel loss functions to regulate contrast and the transmission map.
- An efficient approach capable of real-time data processing once trained.
- A methodology that learns and generalizes the necessary transformations for enhancing underwater images in various water types and disturbances.

4.2 Proposed Approach

We describe an unsupervised approach capable of improving underwater images and operating efficiently based on the mathematical formulation of hazy images.

4.2.1 RGB channels

In underwater images, all pixels tend to distribute nearby a specific plane in the RGB space [13]. Liu et al. proposed a linear transformation into the UV space based on a universal observation that projecting the pixels will not lead to a severe color shift.

Instead of reducing the dimensionality of color space, we propose to split images into three single channels (RGB) and treat them independently. This will help us to estimate the appropriate distribution of pixels across each dimension. This will allow the network to create a haze-free representation of the image channel-wise while being able to do a color shifting.

4.2.2 Image Haze mathematical equation

Equation 4.1 presents the general underwater optical image formation model, where I_c is the corrupted image, J_c is the haze-free enhanced image, T_c is the transmission map and A_c is the global atmospheric light.

$$I_c(x) = J_c(x) * T_C(x) + A_c * (1 - T_c(x)), c\epsilon(R, G, B) \tag{4.1}$$

Inspired by [5], our approach is based on the previous mathematical model. We use three networks to individually estimate the transmission map, the ambient light, and the enhanced image. Figure 4.1 illustrates this idea. Then, we could construct a predicted hazed image using the model outputs in equation 4.1 and compare it directly to our input.

4.2.3 Networks

In order to maximize our possible output through this approach, we use the light model described by [4] as our J network. Their key components, like channel and attention blocks, plus the use of the Discrete Wavelength Transform, make it ideal to recover color and enhance shapes on underwater images. Also, the network showed to be efficient by being able to process in real-time, being practical for diverse applications.

On the other hand, it is possible to assume in A network that the atmospheric light is global and independent from the image content [5]. Hence, we decided to use a ResNet

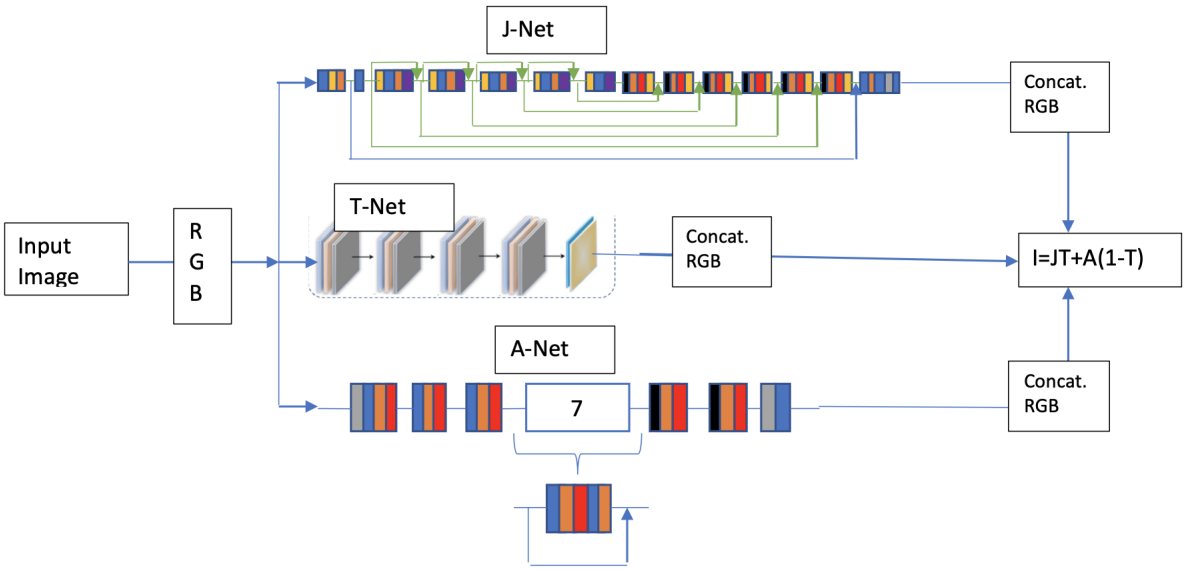

Figure 4.1: Approach schematic using DWT branch of [4] as J net, Res-Net with 7 residual blocks as A-Net and [5] light network as T-Net.

with 7 residual blocks and a global skip connection, commonly used to enhance the image from blurring and integrate the residual image knowledge on aerial and underwater images [27, 4, 51]. Due to the nature of this architecture, we can estimate the light conditions in a general way for different types of underwater images. Therefore, this network would help retain the common pattern allowing training in multiple images at once instead of individually doing it. Using this network increases computational complexity compared with the lither model of [5]. However, this will only be determinant during the training process since inference enhanced image will, after all, only require the J network.

For the T network, we use the structure defined by [5, 57], which is a non-degenerative light architecture since it avoids down-sampling. Then, it is designed to retain the details and perform fast.

4.2.4 Loss Function

Since our approach is purely based on the mathematical formulation of an underwater hazy image, the loss function plays a key role in the output results generated by the model.

Mean Square Error (MSE) loss

Let I_h be the input image, I_{hp} be the hazed image obtained through equation 4.1 with the J, T, and A networks outputs. The MSE loss is defined by equation 4.2 where n is the total number of pixels in the image.

$$MSE = \frac{1}{n} * \sum_{i=1}^{n} (I_h(i) - I_{hp}(i))^2 \tag{4.2}$$

Variational Inference Atmospheric Error (VAE) and Blur Losses

As proposed in the work of [5], these losses make the A model learns the latent Gaussian distribution of the atmospheric light. The blur loss compares the disentangled atmospheric light ($F * A(x)$) and the original output prediction of the A network ($A(x)$). Equation 4.3 presents the formulation of this loss, where F is a mean (average) filter of 5 by 5, MSE is the Mean Squared Error and $*$ denotes convolution operation. On the other side, the variational inference (VAE) aims to minimize the difference between the latent code and the predicted one [5]. Equation 4.4 shows how this loss is defined, where μ is the mean of the input and σ is the variance.

$$Blur = mse(A(x) * F, A(x)) \tag{4.3}$$

$$VAE = \frac{1}{2} \sum ((\mu)^2 + (\sigma)^2 - 1 - exp(\sigma)^2) \tag{4.4}$$

CAP loss

This loss is based in [58] observations and formulated by [5] which states that the deep of a clean image is positively correlated to the difference between the brightness and the saturation.

$$CAP = \|V(J(x)) - S(J(x))\|_p \tag{4.5}$$

LOG loss

We propose an additional loss to control the transmission map estimation. We want to minimize the distance L1 between the second-order derivative of the transmission map (T) and the Laplacian of Gaussian of the recovered image. This avoids blurriness and preserves the edges in the recovered image (J). It is valid since the medium of transmission describes the portion of the light that is not scattered and reaches the camera [32].

$$LOG = \sum \sqrt{(LoG(T(x)) - LoG(J(x)))^2} \tag{4.6}$$

In equation 4.6, LoG defines the convolution of the image with the Laplacian of Gaussians filter [0 1 0; 1 -4 1; 0 1 0].

LC loss

Finally we defined a loss to control the brightness and enhance the contrast on the output image. The contrast could be defined as the difference between the pixel with maximum intensity and the pixel with minimum intensity as shown in equation 4.7. Although, this do not guarantee that the image will enhance its contrast uniformly. Then, we decide to compare the contrast of the global image with the contrast of random patches as defined by equation 4.9, where $G_C(x)$ is the global contrast of the image, $Np_C(x)$ is the contrast of the

n-th patch and n is the number of patches. We used 10 patches with size 7x7.

$$Contrast(x) = max(x) - min(x) \tag{4.7}$$

$$LC = \sqrt{(G_C(x) - \frac{1}{n}\sum_{i=1}^{n} Np_C)^2} \tag{4.8}$$

Total Loss

Note that all the losses depend only on the input image or on the predicted output images. Therefore, there is no actual need to have a ground truth to train using this approach. The total loss can be computed as:

$$TotalLoss = a * MSE + b * VAE + c * CAP + d * Blur + e * LOG + f * LC \tag{4.9}$$

Where a=1.5,b=0.7,c=0.7,d=0.01,e=0.3 and f=0.7 are constants determined through hyperparameter tuning.

4.3 Results and Discussion

We present state-of-the-art results in the UIEBD [2] while training without ground truth and operating efficiently in nearly real-time.

4.3.1 Dataset

We use a combination of two different datasets OceanDark [3] and UIEBD [2] (including the additional 60 challenging images), to train our model. Note that we decide to use only real-world data in our training since we want a model to be able to work in real-time

applications, and we do not have limitations for ground truth. OceanDark dataset presents low-light underwater images while the UIEBD includes degradation related to color cast and blurriness. Then, a combination of both makes our models to improve in generalization and reduce the probability of bias. Additionally, UIEBD contains 890 real underwater images with their corresponding ground truths, being one of the most common underwater datasets to evaluate models performances. Therefore, working with it allows us to directly compare our approach with multiple deep-learning models.

4.3.2 Training

We train our model on a GPU NVIDIA GeForce RTX 3060 with 12GB of memory and used a batch size of 1. We use the same augmentation described by [4], but instead of cropping the patches, we re-scale the images to 256x256. We train using the Adam optimizer for 300 epochs reducing the learning rate by half every 100 epochs.

4.3.3 Quantitative results

We evaluate the performance of our proposed approach in the UIEBD dataset [2], which contains real-world underwater images. We use standard splits, evaluating the performance of the approach on 190 images for direct comparison [15, 4]. Ground truths were only used to compute the score of the structural similarity index (SSIM) of our output by comparing it to the expected result and the peak signal-to-noise ratio (PSNR). Table 4.1 presents a comparison with multiple methods in these two metrics of performance.

Compared to other methods, our approach achieves better SSIM scores of 0.8313 and an inference time of 0.0996 seconds, faster than the traditional computer vision approaches in the field. Also, quantitative results show improvements in the images near to the deep learning methods training in a supervised way, showing that our approach is comparable with current state of the art. Notably our method still have deficiencies that could be im-

Method	PSNR↑	SSIM↑	Inference Time↓ (s)	Type
IBLA [18]	14.3856	0.4299	38.71	Classic
Fusion [59]	21.1849	0.8222	6.58	Classic
GLCHE [19]	21.027	0.8487	0.05	Mix
DWG [39]	19.6727	0.8614	0.4487	Deep
Water-Net [2]	19.3134	0.8303	0.61	Deep
Ucolor [20]	20.63	0.77	2.75	Deep
SCNet [15]	**22.08**	0.8625	0.4495	Deep
Deep WaveNet [21]	21.57	0.8	1.16	Deep
DDC [26]	20.8694	**0.8703**	**0.025**	Deep
UnUn	18.3508	0.8313	0.0996	Deep-Unsupervised

Table 4.1: Quantitative results on the UIEDB [2] test set. We show state-of-the-art results in SSIM. UnUn refers to the unsupervised underwater approach proposed.

prove since we reached low state-of the art results on the PSNR. However, different authors have shown that small differences are not significant since PSNR is not as good of tool in underwater image quality enhancement [55, 56, 2, 4].

4.3.4 Qualitative Results

The UIEDB dataset was built for subjective visual improvement of images not accurate reconstruction of the underlying hazy and discoloured objects, based on the feedback of a set of human judges. For that reason often other methods supervised methods [2, 20, 15] generate these same kinds of images improved but towards a specific subjective goal, often including exaggerated contrast and vibrance of colour. Our method however is built directly from the hazy image formulation and therefore does not seek to create images with extremely high contrast and color vibrance but rather more subtle improvements as seen in Figure 4.2. From the zoomed regions of the image, the quantitative performance of the method becomes more clear, with our dehazing and sharpening part of the coral as well as improving the contrast and lighting vibrance slightly. These somewhat more measured improvements still increase the scores on the target dataset as we do improve

both the contrast and the colour vibrance. However, our method is based on the hazy image model and not the subjective goals of the dataset ground truths. Then, our model's image enhancements look more subdued. This subdued improvement is generally a welcome feature for fields such as underwater biological applications where large shifts in colour vibrance or over-zealous sharpening can lead to incorrect species identification.

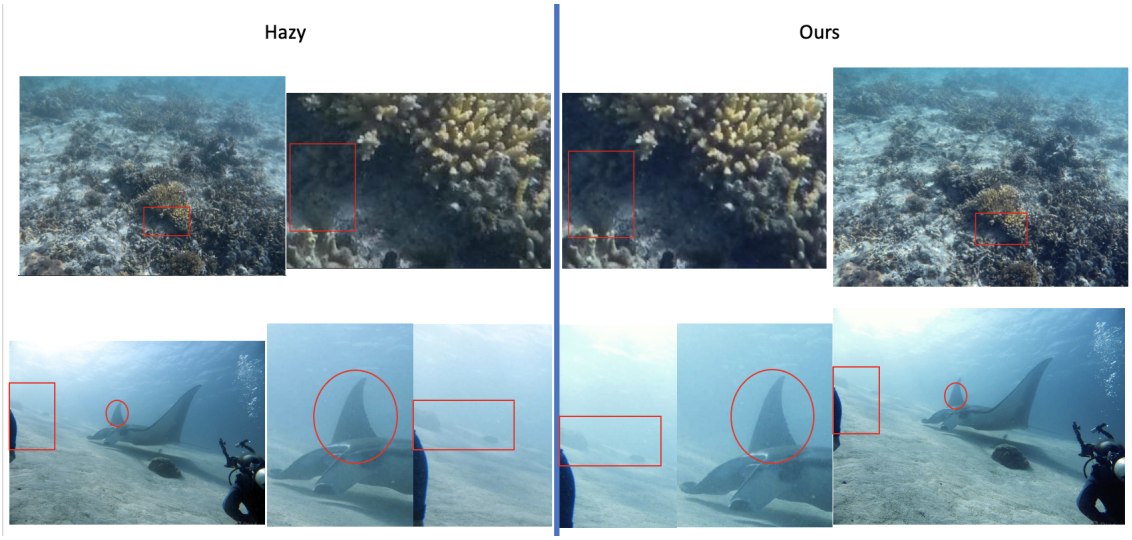

Figure 4.2: Qualitative results with scaled samples showing improvement in light and shadows, colors, and edges

4.4 Outcome

We presented an efficient unsupervised approach for underwater image enhancement which achieves state-of-the-art results with a score of 0.8313 in the SSIM index. Our approach is primarily based on the mathematical definition of a haze image, training independent networks to estimate the atmospheric light, the transmission map and the enhanced image. We defined two new loss functions to control the image contrast and ensure the transmission map preserves the details. Our approach showed fast processing since the architectures used were light. Finally, we run our approach in different types of water showing that our suggested method could generalize well in the enhancement task. Future

approaches include comparing the performance in different ways since other configurations of our method achieved higher human perception enhancement but lower structural similarity index scores.

Chapter 5

Conclusions and Future Work

The objective of this thesis was to thoroughly research the current state of the art in underwater image enhancement and propose an approach to address distortions. In order to achieve this, we approached the problem from two deep-learning computer vision perspectives.

In Chapter 3, a supervised approach was discussed, involving the use of two parallel networks to restore the quality of underwater images by addressing color casting, blurriness, and scattering. Additionally, the chapter examined the impact of employing a single branch for real-time data processing, achieving faster speeds while still reaching significant improvement in image distortions. Our proposed approach demonstrates superior performance compared to state-of-the-art methods, as indicated by SSIM metrics, particularly when incorporating key components such as the DWT algorithm, CBAM attention blocks, adversarial network training with gradient penalty, and skip connections.

Chapter 4 presents an underwater image enhancement approach using an unsupervised learning. We employ three different networks to estimate each term of the mathematical equation for a hazy image. We also propose the inclusion of new expressions for the compound loss function, independent channel processing, and the selection of network

architectures. Our experiments highlight the potential of our approach in achieving state-of-the-art scores in SSIM.

Our results presented in this thesis lead to two broad observations. First, computer vision deep-learning approaches can provide powerful solutions for image enhancement, although they are limited by the availability of data. Second, unsupervised approaches could be a promising alternative in this domain and requires further exploration in the context of image enhancement.

As discussed, we have been limited by the amount of available data in this domain. Consequently, most of the current approaches have only been tested on a limited number of datasets, which can introduce bias towards specific types of images. A research direction of generating diverse datasets would be of interest for future optimization and the development of models that can handle a wider range of environments. It is recommended that datasets are built using different types of water in real conditions and multiple locations to guarantee the diversity of the samples. Additionally, various color checkers should be used at different depths and distances in the image, so that higher accuracy ground truths could potentially be established. This will also lead to a better qualitative analysis.

In relation to the previous statement, another research direction would be the evaluation metrics. While multiple non-reference metrics have been suggested to assess model performance (UIQM, PCQI, UCIQE), reference metrics appear to provide better descriptors of the results. I believe that more research should be conducted to accurately unify the comparison of underwater image enhancement. For instance, during our research with the unsupervised approach, we encountered configurations that resulted in qualitatively improved images from a human perspective, despite having lower scores compared to the state-of-the-art. A well-defined metric should be normalized and percentage-based to enable comparison across models effectively. In the specific underwater image enhancement task, it is suggested that the metric be independent of any ground truth. Additionally, it

should highly focus on the recovery of the red channel, sharpness of edges and corners, and overall scattering reduction.

Additionally, a potential research direction could involve an analysis of videos. Our supervised model demonstrates the ability to process videos quickly enough for real-time frame-by-frame processing. However, it would be necessary to apply additional post-processing to smooth object movement and average the differences between enhanced frames, which could potentially result in an unnatural appearance of the video. Moreover, suggesting different methodologies for training the proposed architectures with an additional dimension (time) could lead to an interesting research path.

Finally, further exploration of unsupervised learning techniques should be conducted, as they offer a solution to the challenges associated with the availability of ground truth data. Currently, there is limited information and research in this area, despite its potential as a promising path to explore. Our study was based on the classical formulation of a hazy image model; exploring different definitions of this could result in better enhancement. For example, [60] proposed a reformulated mathematical model that takes into account the wavelength dependent attenuation in underwater images. Their model suggest a correction for the potential instabilities of underwater image color reconstruction. In a similar way to our presented approach in Chapter 4, future work could start by studding deep-learning architectures that estimates the individual terms of [60] formulation. For instance my future research will be concentrated in finding a better mathematical descriptor of the underwater images model and suggest newer improvements for the unsupervised training approach, since I believe that this is a path that should be explored.

Bibliography

[1] Muhammad Aldila Syariz, Chao-Hung Lin, Manh Van Nguyen, Lalu Muhamad Jaelani, and Ariel C. Blanco. Waternet: A convolutional neural network for chlorophyll-a concentration retrieval. *Remote Sensing*, 12(12), 2020.

[2] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020.

[3] Tunai Porto Marques and Alexandra Branzan Albu. L2uwe: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 538–539, 2020.

[4] Alejandro Rico Espinosa, Declan McIntosh, and Alexandra Branzan Albu. An efficient approach for underwater image improvement: Deblurring, dehazing, and color correction. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, pages 206–215, 2023.

[5] Boyun Li, Yuanbiao Gou, Shuhang Gu, Jerry Zitao Liu, Joey Tianyi Zhou, and Xi Peng. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129, 2021.

[6] Delphine Mallet and Dominique Pelletier. Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). *Fisheries Research*, 154:44–62, 2014.

[7] Martin Heesemann, Tania L Insua, Martin Scherwath, Kim S Juniper, and Kate Moran. Ocean networks canada: from geohazards research laboratories to smart ocean systems. *Oceanography*, 27(2):151–153, 2014.

[8] Declan GD McIntosh, Tunai Porto Marques, Alexandra Branzan Albu, Rodney Rountree, and Fabio De Leo Cabrera. Movement tracks for the automatic detection of fish behavior in videos. In *NeurIPS 2020 Workshop on Tackling Climate Change with Machine Learning*, 2020.

[9] Jonathan A Bergshoeff, Nicola Zargarpour, George Legge, and Brett Favaro. How to build a low-cost underwater camera housing for aquatic research. *Facets*, 2(1):150–159, 2017.

[10] Tunai Porto Marques, Alexandra Branzan Albu, and Maia Hoeberechts. A contrast-guided approach for the enhancement of low-lighting underwater images. *Journal of Imaging*, 5(10):79, 2019.

[11] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 81–88, 2012.

[12] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1682–1691, 2019.

[13] Yongbin Liu, Shenghui Rong, Xueting Cao, Tengyue Li, and Bo He. Underwater image dehazing using the color space dimensionality reduction prior. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1013–1017, 2020.

[14] Jaulin Luc Bazeille Stéphane, Quidu Isabelle. Color-based underwater object recognition using water light attenuation. *Intelligent Service Robotics*, 5, 2012.

[15] Zhenqi Fu, Xiaopeng Lin, Wu Wang, Yue Huang, and Xinghao Ding. Underwater image enhancement via learning water type desensitized representations. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2764–2768, 2022.

[16] Jian Chen, Hao-Tian Wu, Lu Lu, Xiangyang Luo, and Jiankun Hu. Single underwater image haze removal with a learning-based approach to blurriness estimation. *Journal of Visual Communication and Image Representation*, 89:103656, 2022.

[17] Simon Emberton, Lars Chittka, and Andrea Cavallaro. Underwater image and video dehazing with pure haze region segmentation. *Computer Vision and Image Understanding*, 168:145–156, 2018. Special Issue on Vision and Computational Photography and Graphics.

[18] Yan-Tsung Peng and Pamela Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing*, PP:1–1, 02 2017.

[19] Xueyang Fu and Xiangyong Cao. Underwater image enhancement with global–local networks and compressed-histogram equalization. *Signal Processing: Image Communication*, 86:115892, 2020.

[20] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30:4985–5000, 2021.

[21] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters (RA-L)*, 5(2):3227–3234, 2020.

[22] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[23] Ziyin Ma and Changjae Oh. A wavelet-based dual-stream network for underwater image enhancement. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2769–2773, 2022.

[24] Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98:107038, 2020.

[25] Stefanie James, Chris Harbron, Janice Branson, and Mimmi Sundler. Synthetic data use: exploring use cases to optimise data utility. *Discover Artificial Intelligence*, 1(1):15, Dec 2021.

[26] Alejandro Rico Espinosa, Declan McIntosh, and Alexandra Branzan Albu. An efficient approach for underwater image improvement: Deblurring, dehazing, and color correction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 206–215, January 2023.

[27] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks.

In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8183–8192, 2018.

[28] Woong Bae, Jaejun Yoo, and Jong Chul Ye. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1141–1149, 2017.

[29] Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, Jul 2002.

[30] S.G. Narasimhan and S.K. Nayar. Chromatic framework for vision in bad weather. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, volume 1, pages 598–605 vol.1, 2000.

[31] Robby Tan. Visibility in bad weather from a single image. *Computer Vision and Pattern Recognition, 2008. CVPR 2008.IEEE Conference on*, 06 2008.

[32] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, 2009.

[33] A. Preetham, Peter Shirley, and Brian Smits. A practical analytic model for daylight. *Proceedings of ACM SIGGRAPH*, 99:91–100, 01 1999.

[34] R. Sathya, M. Bharathi, and G. Dhivyasri. Underwater image enhancement by dark channel prior. In *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, pages 1119–1123, 2015.

[35] Sheezan Fayaz, Shabir A. Parah, and G. J. Qureshi. Efficient underwater image restoration utilizing modified dark channel prior. *Multimedia Tools and Applications*, 82(10):14731–14753, Apr 2023.

[36] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015.

[37] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2016.

[38] Chau Yi Li, Riccardo Mazzon, and Andrea Cavallaro. Underwater image filtering: methods, datasets and evaluation, Dec. 2020.

[39] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform GAN for nonhomogeneous dehazing. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 203–212, 2021.

[40] Yi Li, Yang Sun, and Syed Naqvi. U-shaped transformer with frequency-band aware attention for speech enhancement, 12 2021.

[41] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[42] Alain Horé and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010.

[43] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[44] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M.

Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[45] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. June 2014.

[46] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. *Computer Vision – ECCV 2018*, pages 3–19, 2018.

[47] Fayaz Ali Dharejo, Yuanchun Zhou, Farah Deeba, Munsif Ali Jatoi, Muhammad Ashfaq Khan, Ghulam Ali Mallah, Abdul Ghaffar, Muhammad Chhattal, Yi Du, and Xuezhi Wang. A deep hybrid neural network for single image dehazing via wavelet transform. *Optik*, 231:166462, 2021.

[48] Zhengxuan Zhao, Kaixu Chen, and Satoshi Yamane. Cbam-unet++:easier to find the target with the attention module "cbam". In *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pages 655–657, 2021.

[49] Wenmei Li, Jiaqi Wu, Huaihuai Chen, Yu Wang, Yan Jia, and Guan Gui. Unet combined with attention mechanism method for extracting flood submerged range. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:6588–6597, 2022.

[50] Jinjiang Li, Guihui Li, and Hui Fan. Image dehazing using residual-based deep cnn. *IEEE Access*, 6:26831–26842, 2018.

[51] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 2016.

[52] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017. cite arxiv:1701.07875.

[53] Z. Wang, Eero Simoncelli, and Alan Bovik. Multiscale structural similarity for image quality assessment. *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, 2:1398 – 1402 Vol.2, 12 2003.

[54] Codruta O. Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing*, 27(1):379–393, 2018.

[55] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019.

[56] De Rosal Igantius Moses Setiadi. PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications*, 80(6):8423–8444, 2021.

[57] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[58] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015.

[59] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 81–88. IEEE, 2012.

[60] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.