Underwater Image Enhancement based on Residual and Adversarial Network

Authors

Md Tosadduk Rahman
180041129

Md. Tawratur Rashid Tanha
180041141

Ishrak Hossain
180041218

Supervised by

Dr. Md. Hasanul Kabir
Professor

Shahriar Ivan
Lecturer

Md. Zahidul Islam
Lecturer

Department of Computer Science and Engineering
Islamic University of Technology
Organization of the Islamic Cooperation (OIC)

Dhaka, Bangladesh

May, 2023

Declaration of Authorship

This is to certify that the work presented in this thesis, titled, **"Underwater Image Enhancement based on Residual and Adversarial Network"**, is the outcome of the investigation and research carried out by Md Tosadduk Rahman, Md. Tawratur Rashid Tanha, Ishrak Hossain, under the supervision of Professor Dr. Md. Hasanul Kabir, Lecturer Shahriar Ivan and Lecturer Md. Zahidul Islam. The submission of this thesis or any portion of it elsewhere for the granting of a degree, diploma, or other qualification is hereby expressly prohibited. A list of references is provided, and information taken from other people's published or unpublished work has been recognized in the text.

Authors:

Md Tosadduk Rahman	Md. Tawratur Rashid Tanha	Ishrak Hossain
Student ID: 180041129	Student ID: 180041141	Student ID: 180041218

Supervisor:

Dr. Md. Hasanul Kabir
Professor
Department of Computer Science and Engineering
Islamic University of Technology

Shahriar Ivan
Lecturer
Department of Computer Science and Engineering
Islamic University of Technology

Md. Zahidul Islam
Lecturer
Department of Computer Science and Engineering
Islamic University of Technology

Dedication

Dedicated to our parents, who have provided their constant encouragement, been with us during the years with their love, guidance, and inspiration.

Acknowledgement

We would like to express our gratitude and appreciation to everyone who contributed significantly to the completion of our thesis by providing motivation, inspiration, and helpful advice.

Our supervisor, **Prof. Md. Hasanul Kabir**, PhD, Dept. of Computer Science and Engineering, Islamic University of Technology (IUT), has provided us with invaluable supervision, expertise, and encouragement. We extend our gratitude towards him for his constant guidance.

We are appreciative of our co-supervisors, **Shahriar Ivan**, Lecturer, Dept. of Computer Science and Engineering, Islamic University of Technology, **Md. Zahidul Islam**, Lecturer, Dept. of Computer Science and Engineering, Islamic University of Technology and their excellent guidance and suggestions during the course of this project.

Contents

List of Figures

List of Tables

Abstract

Due to the complexity and constraints of the underwater environment, underwater picture enhancement is a difficult task. In order to improve underwater images that have problems with low contrast, blurriness, and color mistakes, this research suggests a deep learning-based technique. Residual Networks (ResNet) and Super-Resolution Generative Adversarial Networks (SRGANs) are combined in the suggested method. In order to restore fine details and improve overall contrast and sharpness, ResNet extracts residual information. SRGANs produce enhanced underwater picture versions at high resolution, enhancing visual integrity.

Extensive testing on several underwater picture datasets reveals the suggested method's superior performance. Comparing it to cutting-edge methods, objective quality indicators such as contrast augmentation, image sharpness, and color accuracy confirm its efficiency. Qualitative evaluations show that the underwater photographs have significantly improved in terms of contrast, blurriness, and color reproduction. This increases their ability to be analyzed and interpreted as well as their visual appeal. Marine research, underwater robots, and inspection systems can all benefit from better underwater image quality. Improved visual quality is beneficial for accurate underwater object identification, biodiversity measurement, and extending our understanding of underwater ecosystems. In conclusion, this study provides a deep learning-based technique for enhancing underwater image quality that combines ResNet and SRGANs. The method addresses low contrast, blurriness, and color mistakes to produce notable improvements. Its effectiveness is supported by the experimental findings and qualitative evaluations, emphasizing its potential to advance underwater photography methods and applications.

Chapter 1

Introduction

We provide a succinct summary of the theory that we have discussed so far in this part. We first provide a quick overview of underwater image improvement. An introduction to picture enhancement follows this. Based on our research and the research gaps we identified in the challenges section, we then present the problem statement. We'll wrap off the chapter by outlining the goals, anticipated contributions, and structure of the thesis.

1.1 Image enhancement

A potent method for improving an image's visual appeal and communication potency is image augmentation. Image augmentation enhances the visual quality of an image by changing its colors, contrast, and sharpness, for example, or makes the intended message more powerful. This method can be applied manually using specialized picture editing software or automatically utilizing specialized algorithms that enhance photographs.

Image augmentation has a wide range of uses in numerous industries. Image augmentation is essential to improving the aesthetic appeal of photos in the realm of photography since it helps photographers better express their artistic vision. Image augmentation gives photographs life, increases their visual impact, and brings out features that may have been muted during the original capture by changing colors, contrast, and sharpness.

Similar to photography, image augmentation techniques are used in videography to improve video frames, guaranteeing viewers an engaging visual experience. Videos become more bright, interesting, and intriguing when the color balance, contrast, and sharpness are optimized.

Additionally, picture augmentation has useful uses in medical imaging. By enhancing medical pictures like X-rays or MRI scans with this approach, medical personnel can better see important anatomical features or abnormalities. Medical image augmentation improves patient care by changing picture features to help with accurate diagnosis and better interprofessional communication.

The development of deep learning algorithms in recent years has completely changed picture augmentation methods. Deep learning-based methods for producing realistic and high-quality augmented images, like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have proven to be remarkably effective. These methods can accurately imitate different artistic styles and learn from massive datasets, enhancing photos in a more intelligent and context-aware way.

Last but not least, image augmentation is a flexible approach used in photography, videography, and medical imaging to improve the aesthetic appeal and communicative effect of images. It offers a way to improve colors, contrast, and sharpness, enabling photographers, videographers, and medical professionals to get their desired visual results. This is done by utilizing sophisticated algorithms and image editing software. Image augmentation techniques are constantly evolving thanks to deep learning developments, enabling more creative expression and enhancing visual comprehension across a variety of areas.

1.2 Underwater Image Processing

Underwater image processing can be divided into some categories. The following diagram will give a clear idea.

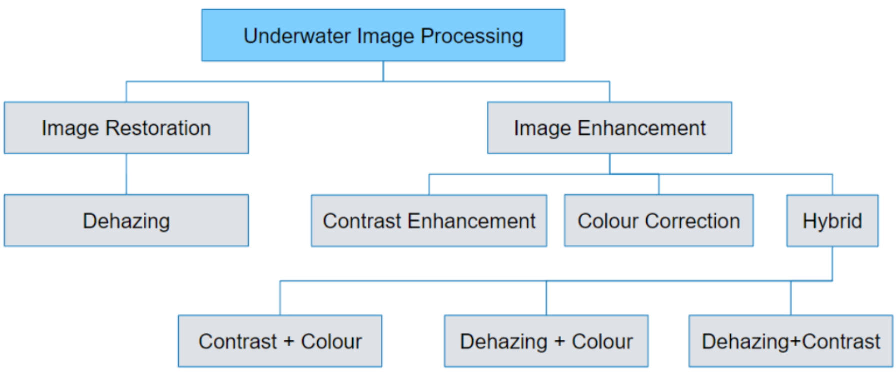

Figure 1.1: Underwater Image Processing categories

1.3 Underwater Image enhancement

Our focus is mainly on strategies for underwater image improvement among the different types of image processing techniques. Enhancing underwater photos entails enhancing the visual appeal and interpretability of photographs taken in aquatic settings. Underwater photographs frequently experience issues including blurriness, distortion, and low contrast because of the peculiar characteristics of water, including light absorption and dispersion. A variety of strategies can be used to solve these problems and improve the visual quality of underwater photos, including:

- **Color correction:** Images are colored differently when taken underwater due to changes in light absorption at various wavelengths. The natural hues of underwater photos can be adjusted and restored using color correction procedures, improving their visual accuracy and appeal. These methods account for the specific color biases introduced by the aqueous medium by adjusting the color balance.

- **Brightness and contrast adjustment:** Because of the light's scattering properties in water, underwater photographs may appear low contrast and have a small dynamic range. Details in the image can be improved by modifying brightness and contrast levels, making the image more aesthetically appealing and increasing visibility in general. Contrast adjustments aid in sharpening the distinction between various objects and elements in the image, while brightness adjustments improve visibility of darker areas.

- **Sharpening:** The sharpness and visibility of minute details in underwater photos can be diminished by the refraction of light as it travels through the water. To improve the image's sharpness overall and bring out minute details, sharpening procedures can be used. Underwater photographs can recover the clarity that was lost by using algorithms that highlight edges and high-frequency components, revealing the fine details of the scenes that were photographed.

- **Noise reduction:** Due to the difficult lighting circumstances and the presence of particles in the water, underwater photographs frequently have high degrees of noise and grain. These artifacts are suppressed using noise reduction techniques, producing images that are clearer and more appealing to the eye. The quality of underwater photos can be greatly enhanced by using denoising algorithms that effectively decrease noise while preserving crucial image information.

The visual quality of underwater photographs can be considerably improved by using these image enhancement techniques, allowing for better interpretation and analysis in a variety of underwater imaging applications. Enhancing underwater photographs and making them appropriate for scientific research, marine exploration, and other underwater imaging domains requires a mix of color correction, brightness and contrast improvement, sharpening, and noise reduction. These methods aid in restoring the scene's natural appearance, enhancing image details, and enhancing underwater photographs' overall visual appeal and clarity.

1.4 Problem Statement

Autonomous underwater vehicles (AUVs) have become effective instruments for a variety of tasks recently, such as mapping the seafloor, analyzing underwater scenes, collaborating with humans and robots, and monitoring marine organisms. For essential functions like object tracking and scene comprehension, these vehicles significantly rely on high-quality photos. However, because of things like light scattering, absorption, and low visibility, underwater imaging is extremely difficult. These problems frequently cause the quality of acquired photographs to degrade, leading to the loss of crucial details, decreased contrast, and distorted colors. The usefulness of AUVs in underwater environments can be hampered by the hazy and noisy appearance of underwater photographs, even when using high-tech cameras.

A complete image enhancement solution is urgently needed to address these issues and allow AUVs to function optimally in the underwater environment. In order to meet this need, this research suggests a cutting-edge and reliable deep learning-based method for improving underwater image quality. The suggested model seeks to address important challenges such color distortion correction, contrast enhancement, noise reduction, and resolution improvement by utilizing the power of deep neural networks.

Many computer vision tasks, such as object detection, segmentation, and image recognition, have been successfully completed using deep learning. Deep neural networks are well-suited for addressing the particular difficulties of underwater picture enhancement because of their capacity to learn sophisticated image representations and capture fine details.

The development of a sophisticated deep learning model specifically suited for underwater image enhancement is the main goal of this study. We seek to develop a strong framework capable of successfully repairing and enhancing underwater photographs by training the model on substantial datasets of underwater image data and fine-tuning its parameters. In addition to improving visual quality, the suggested approach will also make underwater photographs easier to understand and use for a variety of purposes in AUVs and similar underwater imaging systems.

The results of this study have enormous potential for the underwater imaging industry. The suggested deep learning-based approach has the potential to revolutionize AUV operations by greatly enhancing the quality and authenticity of underwater photographs. This would enable more accurate data analysis, effective marine research, and accurate environmental monitoring. The knowledge gathered from this study can also help us comprehend the undersea environment better and help us realize the enormous potential that lies beneath the waters. Based on our overall objective and purpose, we name our problem as follows-

"Designing a residual and generative adversarial network based underwater image enhancement method that can improve images by reducing bluish and greenish color and boosting resolution."

1.5 Challenges

To increase the overall quality and visual integrity of the photos that are being taken, it is necessary to solve a number of issues that underwater image processing brings. These difficulties result from the undersea environment's inherent features, such as light refraction, absorption, and scattering effects. These problems lead to decreased image quality, which makes it challenging to extract useful information. The following difficulties must be overcome-

- The distortion brought on by light refraction, absorption, and scattering in water results in poor image quality for underwater images. These effects cause color changes, which lessen how natural-looking underwater photographs appear. To address these problems and improve the appearance of underwater photos' colors, color-correcting techniques are frequently used.
- Absorption of Red Wavelengths: In deep water, red wavelengths are absorbed more quickly, creating some difficulties:
  - Low Contrast: It is difficult to detect objects and details in underwater photographs due to the absorption of red wavelengths.
  - Blurred Images: The blurring that results from light dispersion in water lessens the sharpness and clarity of underwater photographs.

    - Color Degradation: The overall aesthetic quality of underwater photos can be impacted by the absorption of red wavelengths, which can also cause color deterioration.

- Diversity of Underwater Conditions: Diverse degradation variables are presented by the vastly different underwater environments:

    - Shallow Coastal Waters: Further picture deterioration occurs in shallow coastal waters due to other factors including suspended particles and vegetation.
    - Deep Oceanic Waters: The difficulties in deep oceanic waters are mostly caused by increasing water pressure and a lack of light.
    - Muddy Waters: Suspended sediments and extreme turbidity in murky or muddy waters significantly degrade visibility and image quality.

Choosing the right architecture becomes essential when using deep learning techniques to improve underwater image quality. Datasets of underwater picture training have been used to train a variety of deep learning architectures. To meet the above-mentioned special issues, it is crucial to take into account architectures that produce effective outcomes quickly. The requirement for a comprehensive design that can handle both difficulties at once is shown by recent architectures that either concentrate on color distortion reduction or resolution upscaling separately.

In this research, we offer a brand-new deep learning architecture created especially for the enhancement of underwater images. By efficiently resolving color distortion and using cutting-edge techniques for resolution enhancement, our suggested design attempts to overcome the difficulties associated with underwater photography. Our method offers a complete solution for improving the visual quality of underwater images by concurrently addressing these important problems. We illustrate the effectiveness and potential of our architecture to fully realize underwater imaging's potential and improve a variety of underwater domains through comprehensive experimentation and evaluation.

1.6 Objectives of the Thesis

The objectives of this thesis are as follows:

1. Create a sophisticated image enhancement model that efficiently adjusts the contrast, gets rid of the haze, fixes color divergence, and lessens the effect of various

water kinds. The model promises to improve the visual appeal of underwater photos dramatically.

2. Implement a resolution enhancement component within the proposed model to improve the image resolution. This component should effectively enhance the fine details and improve the overall sharpness of the underwater images.

3. Analyze the effectiveness of the suggested architecture both with and without changes. We want to evaluate how the changes will affect the model's overall performance using comparison analysis. The performance of the model will be optimized with the aid of this analysis, which will offer insights into the efficacy of various components. The quality of the augmented images will be assessed using quantitative metrics like peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and subjective evaluations.

4. Perform a complete performance analysis of the suggested approach. In order to evaluate our architecture, results from previous studies using other architectures will be compared to those from our architecture. The proposed method will be thoroughly examined in this comparative analysis, which will also highlight its contributions to and developments in the field of underwater image enhancement.

5. Describe the drawbacks and difficulties of the suggested strategy and make suggestions for future research trajectories. Determine what needs further work so that underwater image enhancing systems can perform better and be more reliable.

These objectives will enable this thesis to advance cutting-edge techniques for underwater image enhancement. The proposed model and its evaluations will provide beneficial insights that will aid researchers and industry professionals working in the field in improving the visual quality of underwater images for a variety of applications, including marine research, underwater exploration, and underwater robotics.

Chapter 2

Literature Review

We will undertake a thorough and in-depth analysis of recent developments in the field of underwater image enhancement in this section. The goal is to examine and evaluate the previous research that has been done to solve the difficulties unique to improving underwater photographs. The emphasis will be on examining publications that use deep learning techniques, which are consistent with the methodology outlined in our model.

Image enhancement is only one of the many computer vision jobs where deep learning has proven to be a potent tool. It is especially well suited for addressing the distinctive properties of underwater photos due to its capacity to learn complicated features and patterns from large-scale data sets. We intend to discover cutting-edge strategies and methodologies that have been successfully applied for underwater image improvement by thoroughly reviewing the literature on deep learning approaches.

We will examine and assess the major advancements made in the field of underwater picture enhancement as a result of these studies during our evaluation. We will look at the approaches, architectures, and algorithms put forth in these works, paying close attention to how well they deal with issues like color distortion, poor contrast, and blurriness that are frequently present in underwater pictures.

We will also take into account how well these methods work in various underwater circumstances and with various types of water (such as dirty water, deep ocean water, and shallow coastal water). We can decide whether deep learning-based techniques are best for our proposed model by analyzing their advantages and disadvantages in distinct undersea situations.

We want to get a thorough grasp of the present state-of-the-art deep learning algorithms for underwater image enhancement through a thorough assessment of the existing literature. Our proposed model will be developed using this knowledge, enabling us to take use of the most efficient methods and strategies to improve the visual quality of underwater photos. Our ultimate goal is to enable a wide range of applications in industries including marine research, underwater exploration, and underwater robots while also helping to enhance underwater imaging technology.

Jun Ao et al. introduced an original technique called Adaptive Linear Stretch (ALS) [8] in the field of improving underwater images. By improving the objective quality of the stretched image and maintaining computing efficiency, ALS strives to go beyond the constraints of conventional linear stretching techniques. The essential characteristic of ALS is its adaptability, which is attained by using a threshold that can be adjusted and obtained from the histogram of the image.

The main goal of the ALS approach is to restore color in underwater photographs. The adaptable threshold is established by studying the image's histogram, which enables ALS to precisely alter the color distribution and resolve color cast problems. Due to ALS's adaptive nature, which enables it to dynamically adjust to each image's unique properties, the contrast and color appearance of the images are improved.

The usefulness of the ALS approach for improving underwater photographs has been experimentally evaluated. According to the results, there has been a noticeable improvement in image contrast, a decrease in color cast, and an overall improvement in the subjective quality of the photographs. Additionally, ALS maintains a low computing cost while achieving these improvements, making it an effective option for real-time applications or situations with limited resources.

The ALS method's contribution to underwater image color correction rests in its effectiveness and simplicity. ALS offers the path for enhanced visual quality and greater interpretability of underwater photographs by effectively addressing the issues related to color distortion. Numerous applications, such as marine research, underwater exploration, and underwater robots, are significantly impacted by these developments.

In conclusion, the ALS method developed by Jun Ao et al. offers a fresh strategy for improving underwater images. As a result of ALS' adaptability and excellent computational implementation, color aberrations in underwater photographs are successfully corrected, resulting in better image contrast and subjective quality. The successful outputs of this research lead to the creation of straightforward yet effective algorithms for correcting color aberrations in underwater images, enabling a variety of real-world uses and facilitating further developments in the field.

The goal of this work (ALS method) is to create a simple and efficient algorithm for correcting the color of underwater images. The results -

	Def	SSIM	MES	PSNR	H	PT(s)
Original	21.64	1.00	–	–	6.79	–
HE	137.93	0.72	1680.44	15.88	7.95	8.05
Gani's Method	58.34	0.92	326.42	22.99	7.47	5.08
ALS	113.00	0.77	1343.65	16.85	7.90	0.56

Table 2.1: Quantitative comparison between ALS and various methods

Figure 2.1: Comparison of ALS method and original picture

Smitha Raveendran et al. have proposed a revolutionary way for improving the quality of underwater photos by integrating a revised image creation model into an existing deep learning-based methodology. A backscatter estimating module and a direct-transmission estimate module, both built using convolutional neural networks, make up the two modules that make up the suggested technique. The enhanced underwater image is created by further processing these modules and the input image with a reconstruction module.

The parametric rectified linear unit (PRelu) [9] and dilated convolution [10] techniques are used to enhance the neural network's performance. Dilated convolution gives the network a bigger receptive field, allowing for improved contextual information acquisition, while PRelu increases the network's fitting capability by adding learnable parameters to the rectified linear unit.

The effectiveness of the suggested strategy is assessed using the UIEB and URPC-2020 benchmark datasets. The experimental findings show how the technique can considerably improve the quality of underwater photos. The suggested method successfully decreases the negative impacts brought on by underwater imaging conditions, such as scattering and absorption, by accurately measuring the back-scatter and direct transmission components.

The performance of the deep learning-based system for underwater picture enhancement has improved overall thanks to the adoption of the revised image creation model and the use of PRelu and dilated convolution techniques. The testing results show that the suggested method is effective and that it is suitable for a variety of underwater imaging applications and has the potential to improve the quality of underwater images.

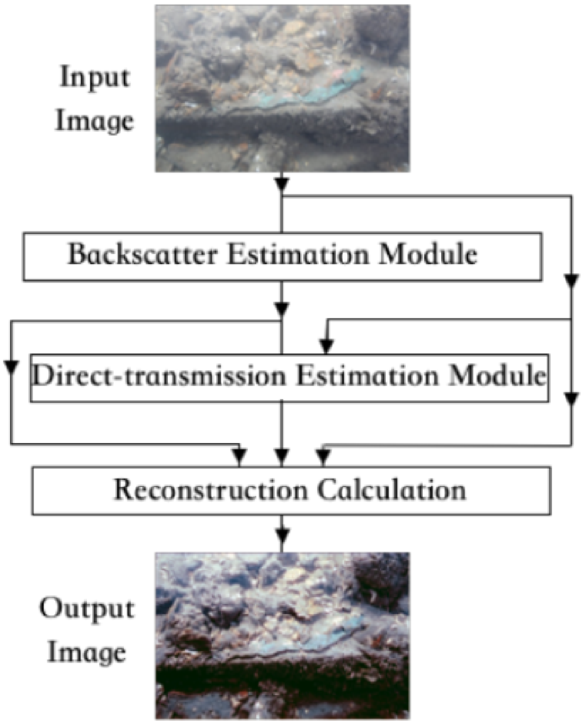

Figure 2.2: Framework of proposed method

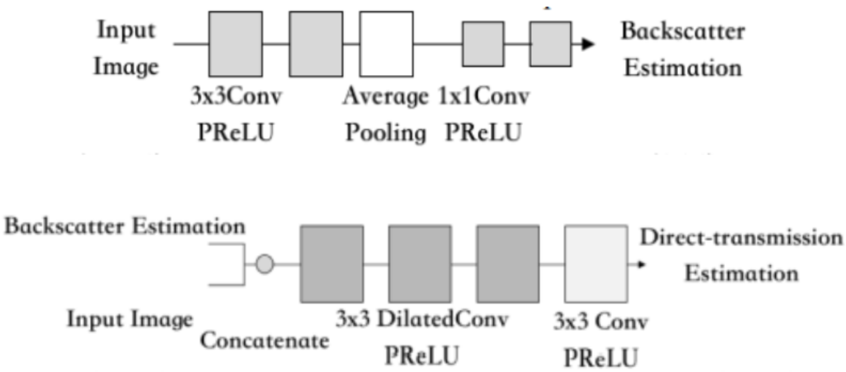

Figure 2.3: Backscatter and direct-transmission estimation

Peng Liu et al. proposed cycle-consistent adversarial networks (CycleGAN) [11] to generate synthetic underwater images for training convolutional neural network models. The very-deep super-resolution reconstruction model (VDSR) [12] is also introduced for use in underwater image resolution applications, leading to the development of the Underwater Resnet [13] model, a residual learning model for underwater image enhancement tasks. The loss function and training mode are also improved by introducing a multi-term loss function incorporating mean squared error loss and a proposed edge difference loss, and by implementing an asynchronous training mode. These modifications are intended to enhance the performance of the multi-term loss function. Many image-to-image translation models use per-pixel difference loss functions such as the MSE or L1 loss function. These loss functions attempt to minimize the difference between two images at the pixel level. However, using the MSE loss function, as the original VDSR model did, can lead to higher peak signal-to-noise ratio (PSNR) [14] scores but poor visual results. This is because the MSE loss function averages the differences at the pixel level and does not consider higher-level information, such as overall structure. As a result, the MSE loss function can average the solution and make image details too smooth, which is not conducive to enhancing high-frequency information.

A residual learning model is the UResnet that has been proposed. It is made up of ResBlocks [13], which combine the result of one convolution layer with the input of another. By using ResBlocks, the information from the top layer can be fully communicated to the layers below ResBlock [13] stacking enables the training of deeper networks. The super-resolution reconstruction models EDSR [15] and SRResnet [16] served as inspiration for UResnet. The head, torso, and tail make up the three primary parts of the suggested UResnet model.

There is only one convolution layer in the head. In light of how long training takes, the body portion stacks 16 ResBlocks in the following sequential order: [Conv-BN-ReLU-Conv-BN]. One convolution can be found in the tail. There are 34 convolution layers in total. The network uses a 33 convolution with a 1 pixel stride and a 1 pixel zero-padding to preserve the geometry of feature maps, enabling UResnet to receive inputs of any shape. This work achieved good result in PSNR [14] and SSIM [17] metrics.

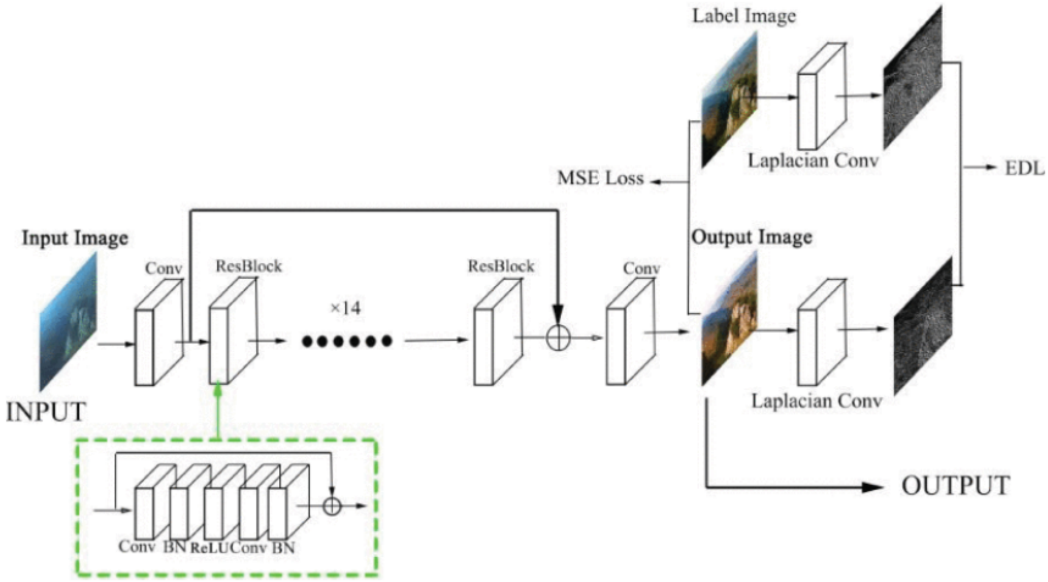

Figure 2.4: UResnet model with EDL

Hema Krishnan et al. presented a method for improving the perceptual quality of underwater images. To do this, synthetic underwater images are generated using UwGAN. The UResNet [7] model, which incorporates various loss functions such as edge difference loss (EDL) and mean square edge (MSE) loss, is then used for underwater image enhancement. An asynchronous training mode is also implemented to enhance the performance of the multi-term loss function. This proposed UnderwaterGAN achieved a better score than CycleGAN [11] in UICM, UISM, UIconM and UIQM [18] metrics but could not have better results in PSNR [14] and SSIM [17] metrics.

Figure 2.5: Input images and corresponding enhanced images

Alireza Aghelan et al. improved Real-ESRGAN by using a high-order degradation process to generate low-resolution (LR) degraded images. Instead of training the Real-ESRGAN model from scratch, transfer learning is used by fine-tuning the model on underwater image datasets after it has been trained on natural image datasets. The model is fine-tuned using Google Colab Pro, with a batch size of 10 per GPU, Adam optimizer with a learning rate of 0.0001, and exponential moving average. A combination of L1 loss, perceptual loss, and GAN loss functions is used for fine-tuning. The model is fine-tuned for 10300 iterations (approximately 8 epochs) and saved, and then fine-tuned for an additional 26000 iterations (approximately 20 epochs) and evaluated for performance.

Md Jahidul Islam et al [6] present a model for enhancing the quality of underwater images in real-time using a generative adversarial network. They use an objective function to evaluate the perceptual quality of the images based on various factors such as color, texture, and style. The authors also introduce a dataset of underwater images with varying levels of quality, captured using different cameras under different visibility conditions. The model is trained using both paired and unpaired data from this dataset. The enhanced images are shown to improve the performance of standard models for tasks such as underwater object detection, human pose estimation, and saliency prediction. The authors suggest that the model could be used in the autonomy pipeline of visually-guided underwater robots.

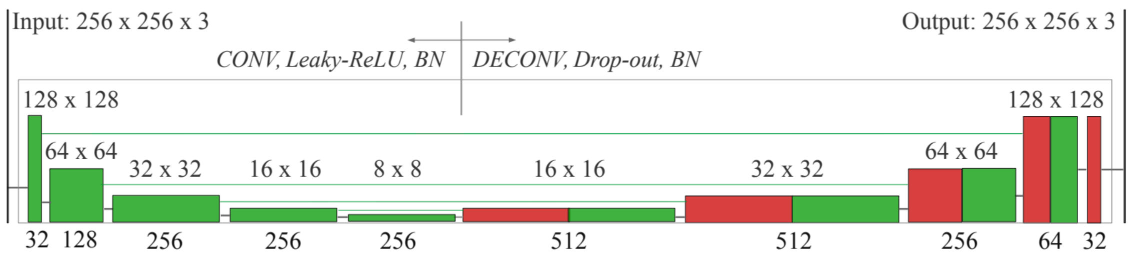

Figure 2.6: Generator

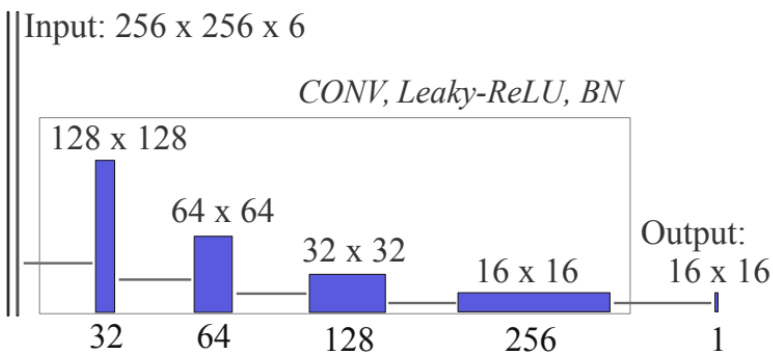

Figure 2.7: Discriminator: A Markovian PatchGAN

Pritish Uplavikar et al [19] propose a model for enhancing underwater images by learning to disentangle the features of the images from the effects of different types of water (viewed as different domains). They train the model on a dataset of images of 10 different types of water, and show that it performs better in terms of SSIM and PSNR scores compared to previous methods for most of the water types, and also generalizes well to real-world datasets. The enhanced images produced by the model also improve the performance of a high-level vision task (object detection).

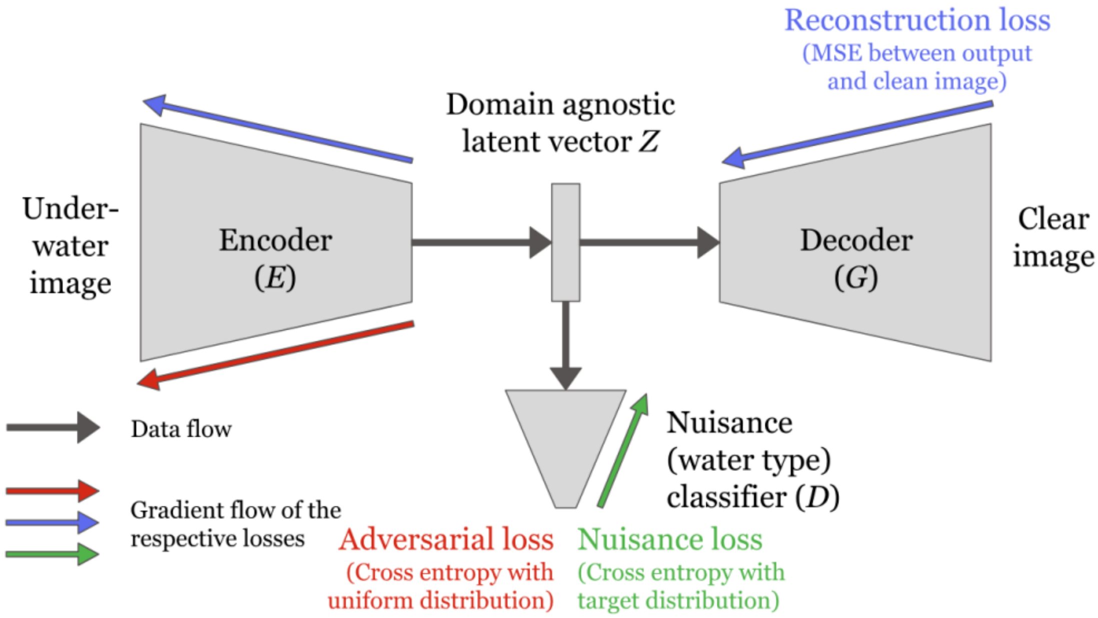

Figure 2.8: CNN based encoder decoder

Zhenqi Fu et al. [1] address the difficulty of boosting underwater photos that suffer from various degradation induced by different water types by introducing a potent approach called SCNet. The authors provide a number of normalization approaches within SCNet to accomplish their goal of learning features that are unaffected by the particular water conditions.

To create multi-scale representations, SCNet uses a U-Net architecture and normalization methods at each size. Whitening and standardization with re-injection of the first two seconds are the two essential steps in the normalization procedure [1]. The method effectively decorrelates activations across spatial dimensions by using whitening, and channel-wise correlation is eliminated by standardizing and re-injecting the initial two moments. SCNet may learn resilient and water-type invariant features thanks to these normalizing techniques, producing improved underwater photos.

The authors ran tests on two real-world underwater image datasets to gauge SCNet's performance [1]. The outcomes show that SCNet is quite good at improving photos of a variety of water kinds. The technique performs admirably in terms of visual quality improvement, suggesting that it could be an effective tool for enhancing underwater images.

In conclusion, Fu et al. suggest SCNet [1], a technique that solves the various degradation in underwater photos brought on by various types of water. To successfully enhance underwater photos, SCNet learns water-type invariant features and incorporates normalization techniques within a U-Net architecture. The experimental outcomes on real-world datasets confirm SCNet's capability to manage various types of water and generate notable visual quality enhancements. This study advances the field by outlining a possible method for improving underwater images.

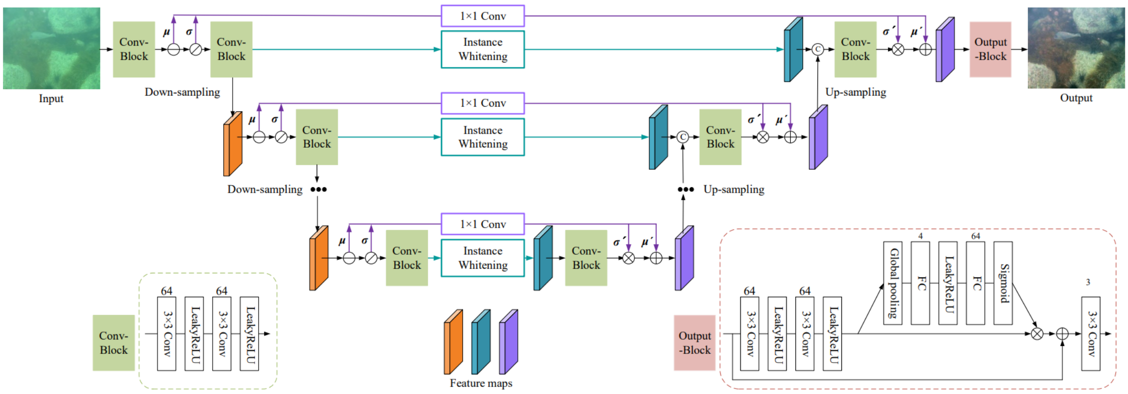

Figure 2.9: Normalization of the U-Net's spatial and channel dimensions [1]

Chongyi Li et al. make a substantial addition to the field of underwater image improvement through the development of a brand-new dataset, the Underwater Image Improvement Benchmark (UIEB) dataset [2]. This dataset was created primarily to assess how well image-enhancing techniques work in aquatic settings. It has a sizable collection of actual underwater photographs that have been painstakingly tagged with factual data, making it an invaluable tool for furthering this field's state-of-the-art.

The authors also suggest the Water-Net architecture, a gated fusion network made to improve underwater photos [2], in addition to the dataset. In comparison to earlier state-of-the-art methods, the Water-Net architecture achieves superior improvement outcomes by using projected confidence maps and performing controlled fusion. Comprehensive experimental assessments are used to show the success of the suggested strategy and emphasize its potential for overcoming the difficulties associated with underwater photography circumstances.

Li et al. underline the need of using specialized datasets and customized algorithms for the goal of underwater picture enhancement by introducing the UIEB dataset and the

Water-Net architecture. The report emphasizes the potential for additional advancements in this sector while also highlighting the necessity for specialized tools and methods to address the distinctive features of underwater imaging.

Future research and development in underwater image enhancement will have a strong basis thanks to the combination of the UIEB dataset and the Water-Net architecture. This research increases our understanding of underwater imaging difficulties while also encouraging the investigation of novel strategies to raise the bar for excellence in this field.

Method	MSE↓	PSNR↑	SSIM↑
Fusion-Based [20]	1.1280	17.6077	.07721
retinex-based [21]	1.2924	17.0168	0.6071
GDCP [22]	4.0160	12.0929	0.5121
histogram prior [23]	1.7019	15.8215	0.5396
blurriness-based [24]	1.9111	15.3180	.6029
Water CycleGAN [25]	1.7298	15.7508	0.5210
Dense GAN [26]	1.2152	17.2843	0.4426
Water-Net [2]	0.7976	19.1130	0.7971

Table 2.2: Full-reference image quality assessment in terms of MSE, PSNR and SSIM on testing set

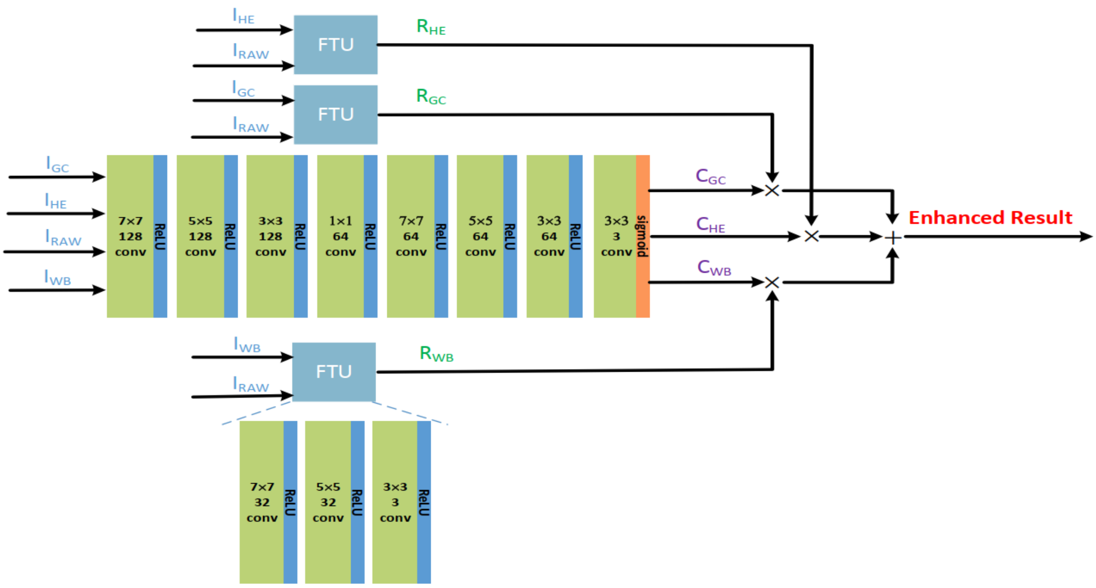

Figure 2.10: An overview of the Water-Net architecture that has been proposed [2]. A gated fusion network called Water-Net fuses the inputs with the projected confidence maps to produce the improved outcome.

Chongyi Li et al. describes a novel strategy for improving the visibility of underwater photographs using a multi-color space embedding method [3]. Their suggested technique entails converting underwater photographs into several color spaces before merging the generated images to get an improved output. The authors use a dataset of actual underwater images to show the efficacy of their method and contrast it with a number of cutting-edge picture enhancing techniques.

Li et al. provide the Ucolor architecture together with an attention mechanism in order to handle the difficulties posed by underwater environments such as turbidity and scattering [3]. To adaptively improve the visibility of underwater photographs, the Ucolor architecture integrates several color spaces and automatically chooses essential elements. The suggested method significantly enhances image clarity by fusing the properties of many color spaces.

The authors also suggest a decoder network that makes use of underwater image medium transmission information in addition to the Ucolor architecture. This guidance mechanism aids the network in prioritizing the areas with the worst quality degradation, improving picture enhancement performance. This method significantly raises the quality of underwater photographs by making use of deep neural network technology and the understanding of underwater imaging.

The suggested method beats numerous cutting-edge image enhancement algorithms across a variety of assessment parameters, according to experimental results [3]. In situations where there is extreme degradation, it successfully improves the visibility and quality of underwater photographs. This innovative method has a lot of potential for real-world use and makes a significant addition to the field of underwater image improvement.

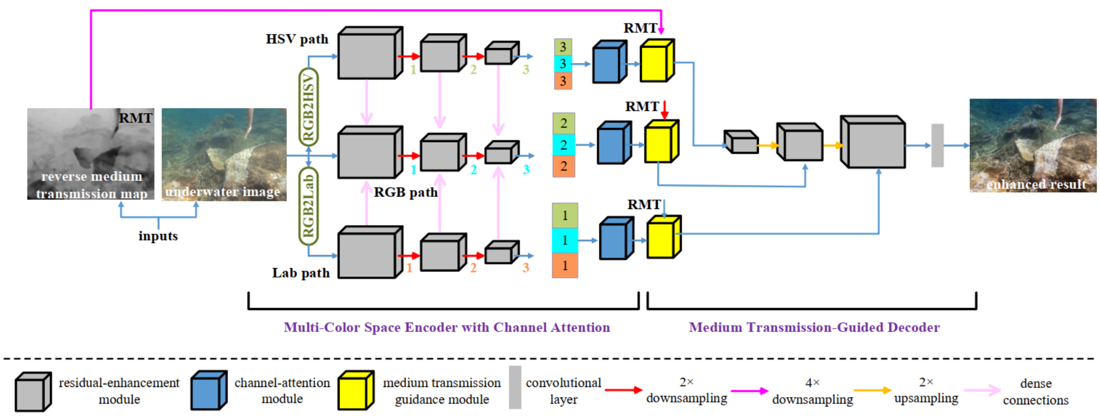

Figure 2.11: A description of Ucolor's architecture [3].

Christian Ledig et al. proposed SRGAN, a generative adversarial network (GAN) For image super resolution (SR). They suggest an adversarial loss and a content loss make up a perceptual loss function. The use of SRGAN results in highly significant improvements in perceptual quality, according to a thorough mean-opinion-score (MOS) test. In comparison to any state-of-the-art method, the MOS scores generated with SRGAN are closer to those of the original high-resolution photos. The main contribution of the paper is -

- With the 16 blocks deep ResNet (SRResNet) that is tuned for MSE, they achieved a new state of the art for image SR with high upscaling factors (4), as evaluated by PSNR and structural similarity (SSIM).
- They suggested SRGAN, a GAN-based network that has been tailored for a novel perceptual loss. Here, instead of using the MSE-based content loss, they calculated a loss using the VGG network's feature maps, which are more resistant to changes in pixel space.
- With the help of a thorough mean opinion score (MOS) test performed on images from three publicly available benchmark datasets, they were able to establish that SRGAN is now, by a wide margin, the state of the art for the estimate of photo-realistic SR images with high upscaling factors (4).

The result comparison of their work is mentioned in the table-

	SRResNet-			SRGAN-	
Set5	MSE	VGG22	MSE	VGG22	VGG54
PSNR	32.05	30.51	30.64	29.84	29.40
SSIM	0.9019	0.8803	0.8701	0.8468	0.8472
MOS	3.37	3.46	3.77	3.78	3.58
Set14					
PSNR	28.49	27.19	26.92	26.44	26.02
SSIM	0.8184	0.7807	0.7611	0.7518	0.7397
MOS	2.98	3.15*	3.43	3.57	3.72*

Figure 2.12: Performance of different loss functions for SRResNet and the adversarial networks on Set5 and Set14 benchmark data. MOS score significantly higher $p < 0.05$ than with other losses in that category

Zhiming Zhang at al. The authors proposed an approach based on SRGAN (Super-Resolution Generative Adversarial Network) to enhance the quality of underwater images. In their method, they modified the standard SRGAN pipeline. Initially, the training images were preprocessed using the Dark Channel Prior (DCP) technique. Subsequently, the SRGAN model was applied to increase the resolution of the photos. In this process, a low-resolution image was initially upscaled to the desired target resolution using bicubic interpolation. Then, a three-layer convolution network performed nonlinear mapping to generate high-resolution images. They made several improvement in the vanilla srgan -

- **Generator improvement:** The authors optimized the SRGAN model by replacing a single 9*9 convolution layer with two 5*5 convolution layers. They found that both configurations have similar feature extraction capabilities, but using two 5*5 convolutions significantly reduces computational complexity. Additionally, the traditional SRGAN algorithm faced challenges related to inadequate reconstruction of image details and unstable training. To address these issues, the authors removed the Batch Normalization layer in the residual block and reduced the number of parameters, aiming to achieve stable training.

- **Discrimonator improvement:** The authors employed the VGG19 model as a discriminator in their approach. They randomly combined the generated super-resolution (SR) image and the high-resolution (HR) image and fed it into the discriminator model for evaluation. The final layer of the network incorporated the Sigmoid activation function after applying the convolution-BN (Batch Normalization) layer and Leaky PReLu. The output of the discriminator was expected to be a discriminative probability within the range of [0, 1]. Ideally, the HR image should have a value close to 1, indicating a clear underwater image, while the SR image should approach 0, indicating an underwater image generated by the generator. The training of the model was concluded when the output of the discriminator approached 0.5.

- **Improvement of loss function:** In their approach, the authors incorporated both L1 content loss and VGG19 perceptual loss in addition to utilizing the loss mechanism provided by the GAN (Generative Adversarial Network). Due to the presence of outliers in underwater photos, L1 loss was chosen as it is more robust and less affected by these outliers compared to L2 loss. Although applying only the content loss function can achieve a higher peak signal-to-noise ratio (PSNR), it leads to a loss of high-frequency information and the texture details of the image, resulting in suboptimal overall image quality. To preserve sharper edge information in the reconstructed high-resolution image, the authors employed a pre-trained 19-layer VGG network to extract high-level features from the image. These features were then added as a perceptual loss term to the model's overall loss, enhancing the final image quality.

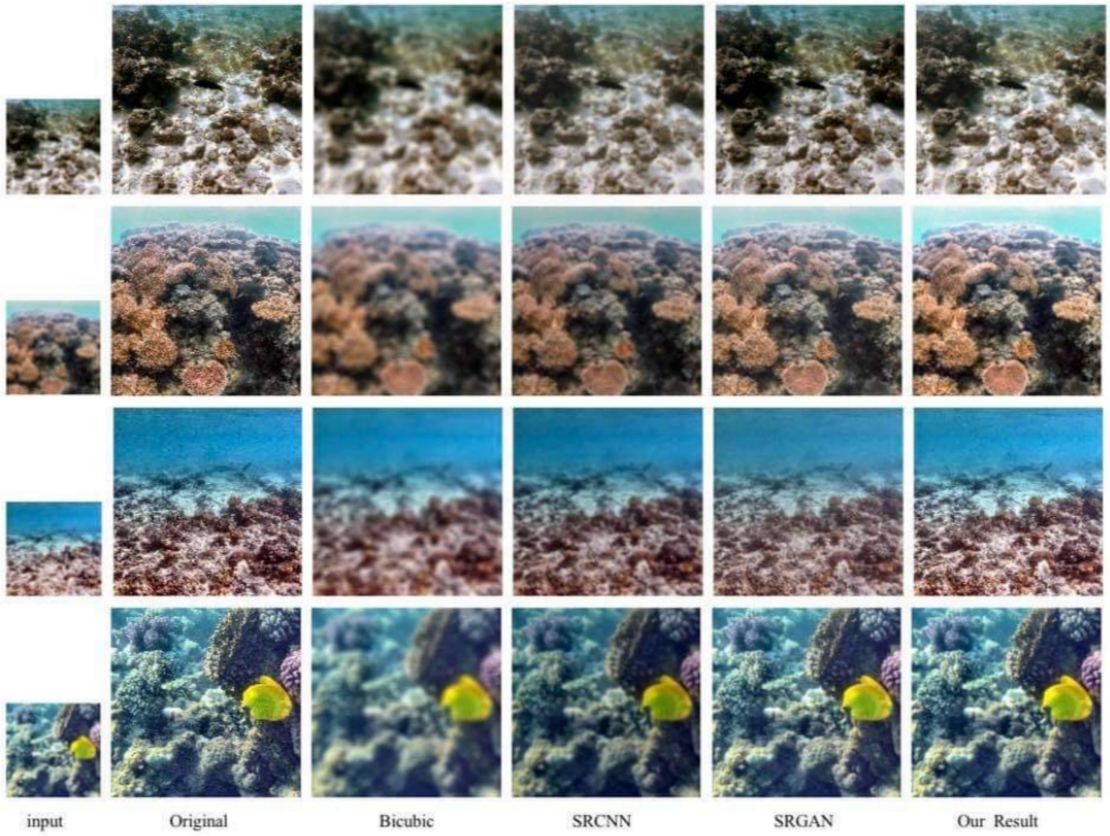

Figure 2.13: Improved SRGAN image reconstruction result

Chapter 3

Datasets

Image datasets that are taken underwater are important for the advancement of image processing methods.

3.1 UIEB Dataset

A benchmark dataset for improving underwater photographs, consisting of 950 genuine underwater shots, was generated by Li et al. [2] in 2019. Out of these, 890 images had appropriate reference pictures, while the remaining 60 underwater images were deemed difficult since relevant reference pictures weren't available. This dataset was used to conduct a thorough qualitative and quantitative analysis of current underwater image enhancing techniques. Additionally, the suggested underwater image enhancement benchmark for training Convolutional Neural Networks (CNNs) was applied to the development and training of a baseline underwater image enhancement network known as Water-Net. The benchmark analyses and the introduction of Water-Net provided important insights for future research in the area of underwater picture enhancement by shedding light on the capabilities and constraints of cutting-edge algorithms. Some sample images from this dataset are shown below-

Figure 3.1: Constructed Image from the UIEB dataset. Top row: raw underwater image taken in diversed underwater scene, Bottom row: corresponding reference results

3.2 MBARI dataset

The Monterey Bay Aquarium Research Institute provided 666 photos of fish for the MBARI underwater image dataset, which was assembled by Yang et al. (2019).

3.3 RUIE Dataset

The Underwater Image Quality Set (UIQS), Underwater Color Cast Set (UCCS), and Underwater Task-driven Testset (UHTS) are the three distinct subsets that make up the Real-world Underwater Image Enhancement (RUIE) dataset, which was first introduced by Liu et al. (2020). Each subgroup focuses on solving a particular difficulty in underwater picture enhancement, such as enhancing visibility quality, removing color casts, and simplifying more complex detection and classification tasks.

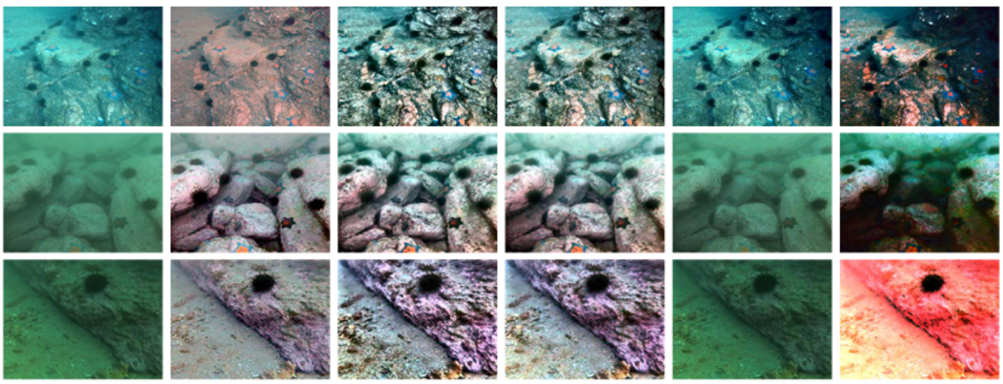

Figure 3.2: Sample images from RUEI dataset [4] and improvement through different method

3.4 SQUID Dataset

The collection includes a variety of file types, including distance maps, RAW photos, TIF files, and camera calibration files. It comprises of 57 stereo pairs that were recorded at four separate locations in Israel: two in the Mediterranean Sea and two in the Red Sea, which reflect temperate and tropical water conditions, respectively. One location in the Red Sea is a coral reef known as "Katzaa," which has 15 stereo pairs and is situated at a depth of 10-15 meters. The second location is a shipwreck called "Satil," which has eight stereo pairs and is located at a depth of 20 to 30 meters. Two rocky reef ecosystems can be found in the Mediterranean Sea, 30 kilometers apart. The first location is Nachsholim, which has 13 stereo pairs with a depth range of 3-6 meters. The second site is Mikhmoret, which has 21 stereo pairs and is located at a depth of 10 to 12 meters.

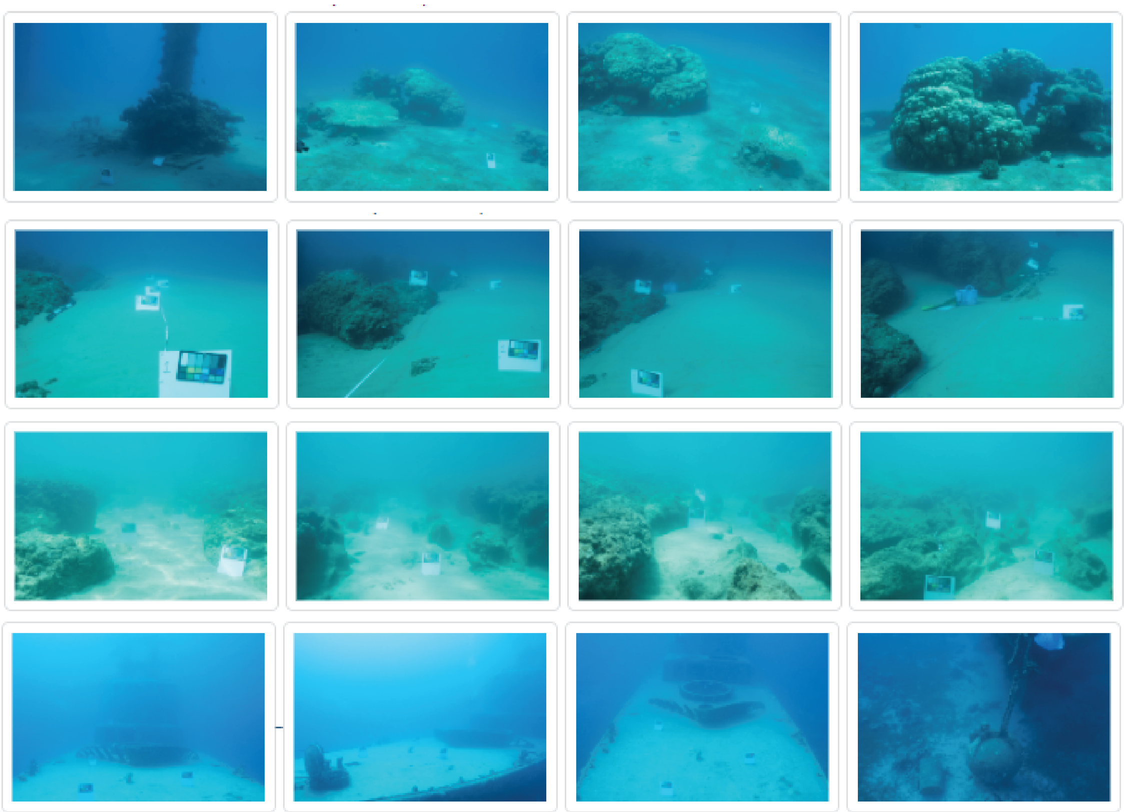

Figure 3.3: Sample images from SQUID Dataset [5]

3.5 DIV2K Dataset

This study analyses the state of the art based on the outcomes of the NTIRE 2017 challenge and proposes a new dataset (DIV2K) [27] for example-based single image super-resolution. This was the first of its kind challenge, which had six competitions, attracted hundreds of participants, and had many different solutions submitted.

3.6 EUVP dataset

Islam et al.'s work introduces the EUVP dataset [6] which aims to enhance underwater visual perception. The photos in the EUVP collection were captured by seven distinct cameras during deep-sea excursions and research involving humans and robots in a range of visibility situations. Almost 12,000 pictures are divided into 3 groups:

- **Underwater Dark:** It has total 5,550 images
- **Underwater ImageNet:** It has total 3,700 images
- **Underwater Scenes:** It has total 2,185 images

Each image has a companion ground truth image. For adversarial training, this dataset can be used. We have selected the EUVP dataset to work with our model because of its variety and size.

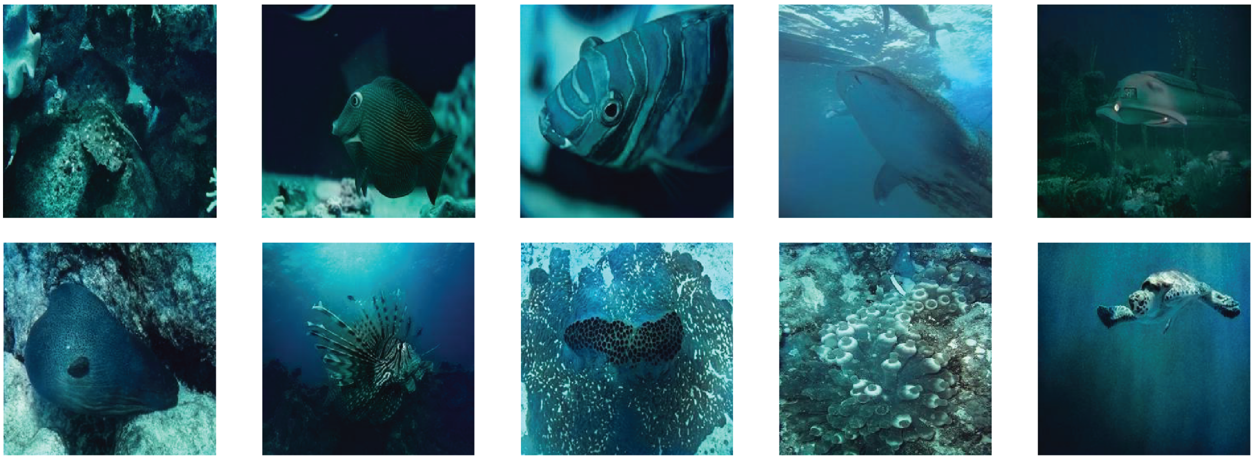

Figure 3.4: Sample images from EUVP dataset(Underwater Dark)

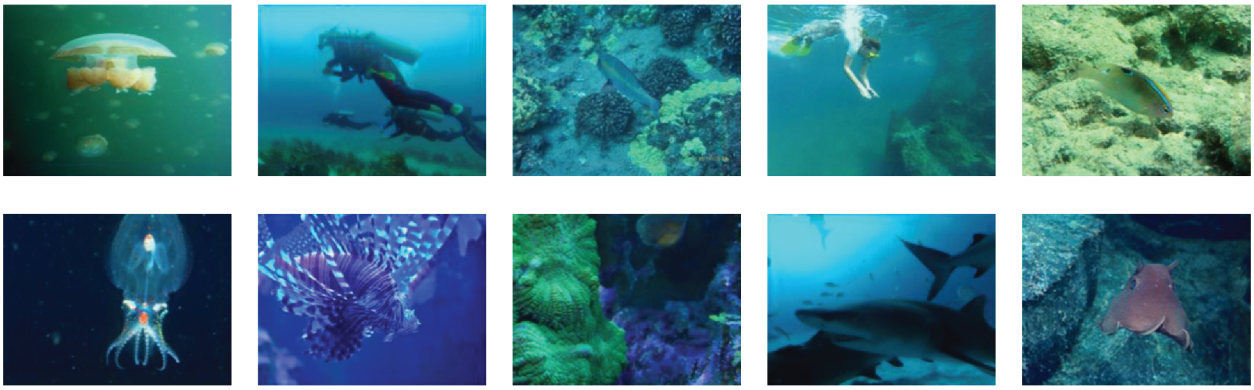

Figure 3.5: Sample images from EUVP dataset(Underwater ImageNet)

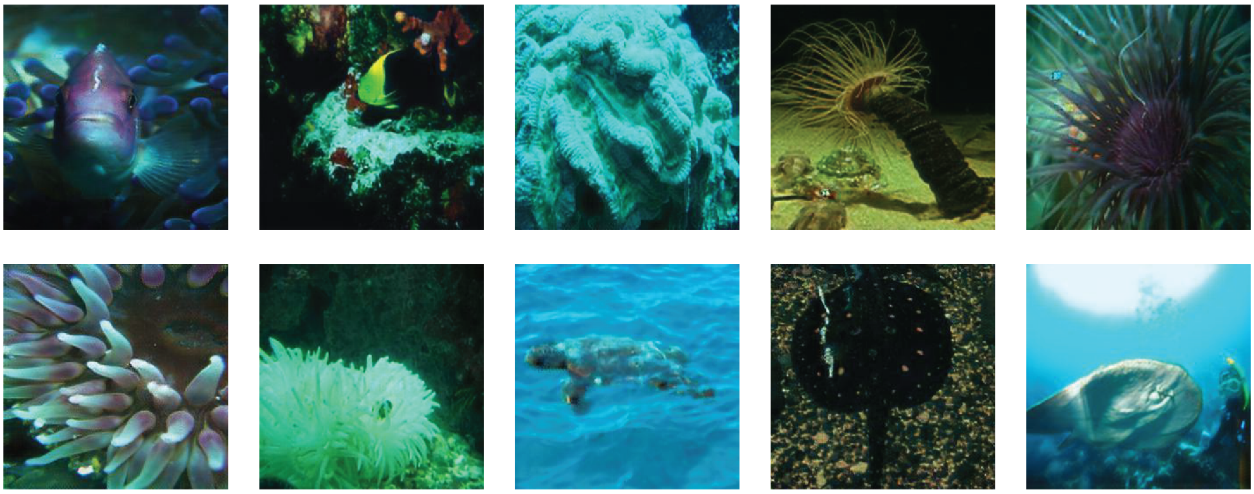

Figure 3.6: Sample images from EUVP dataset(Underwater Scene)

Chapter 4

Proposed Architecture

The objective of our proposed method is to enhance underwater image along with increasing the resolution in a minimal amount of time. To do so, we have proposed a model that consists of Residual Network (ResNet) and Super-Resolution Generative Adversarial Networks (SRGANs). We will first enhance the image by removing blue and greenish hue and then increase the resolution by SRGAN. But we will not use the default SRGAN model. We will bring several changes [28] in the SRGAN model. We will discuss them one by one.

4.1 ResNet

ResNet is a neural network architecture that was introduced in 2015 in a paper called ”Deep Residual Learning for Image Recognition” [13]. It was developed by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. The name ”ResNet” [16] is short for ”Residual Network.”

It has been observed that adding more layers to a neural network can actually decrease its performance. This phenomenon may be due to various factors such as the optimization function, the initialization of the network, and the vanishing gradient problem. The vanishing gradient problem refers to the difficulty that the network has in learning from the back-propagated error signal when the gradients are very small, which can occur when the network is very deep.

To solve this, skip connection [29] is introduced through Resnet. Skip connection is the core of residual blocks. In a neural network with skip connections, the output of a layer is computed differently than in a traditional neural network. In a traditional network, the output of a layer is obtained by multiplying the input by the weights of the layer and adding a bias term. However, in a network with skip connections, the output is obtained by adding the input to the output of the previous layer using a skip connection. This helps to preserve the gradient signal as it flows through the network, making it easier for the model to learn. So Skip connections in ResNets help to alleviate the vanishing gradient problem [30] in deep neural networks by providing an alternative path for the gradient to flow through. They also allow the model to learn identity functions, which means that the performance of the higher layers is at least as good as that of the lower layers, and not worse.

Key features of Resnet -

1. ResNet incorporates Batch Normalization [31], which is a technique used to improve the performance of neural networks. Batch Normalization adjusts the input layer in a way that reduces the issue of covariate shift, which refers to a change in the distribution of the input data that occurs during training. By normalizing the inputs, the network is able to learn more efficiently and achieve better performance.

2. ResNet uses Identity Connections to help prevent the vanishing gradient problem [30]. This can occur when the network is very deep, and can make it difficult for the network to learn effectively. Identity connections work by connecting the output of a layer to the input of a higher layer, allowing the higher layer to access both the output and the input of the lower layer. This helps to preserve the gradient signal as it flows through the network, making it easier for the model to learn.

3. ResNet uses a type of block called a bottleneck residual block to improve the performance of the network. A bottleneck residual block is a design element that is used to increase the capacity of the network while also reducing the number of parameters that the network needs to learn. This is done by using a smaller number of filters in the block, which allows the network to learn more complex features while also making it more computationally efficient. The use of bottleneck residual blocks can help to improve the overall performance of the network.

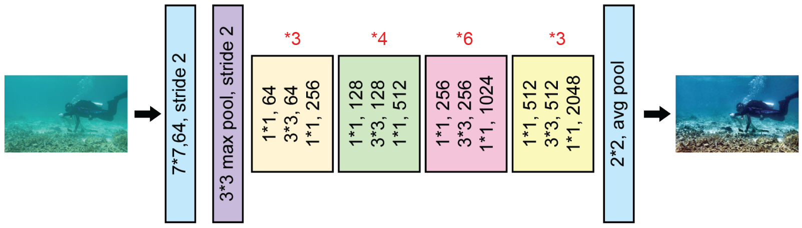

Figure 4.1: ResNet-50 architecture

4.2 SRGAN

SRGAN [16] is a technique that enables models to significantly increase the resolution of most images by a factor of almost 4x. It is a challenging task to generate a high-resolution image from a low-resolution image, and CNNs have been used to produce high-resolution images that are quick to train and have high accuracy. However, they may struggle to recover fine details and can produce blurry images. The SRGAN architecture is designed to address these issues and produce high-quality, state-of-the-art images.

Many supervised algorithms for super-resolution use mean squared error loss to compare the acquired high-resolution image to the ground truth image. This approach is convenient because minimizing mean squared error automatically maximizes peak signal-to-noise ratio (PSNR) [14], which is a commonly used evaluation metric for super-resolution images. However, these metrics focus on individual pixel features rather than visually perceptible attributes such as high texture detail in the image.

Any GAN [32] architecture consists of two main things-

1. Generator
2. Discriminator

4.2.1 Generator

This method uses a fully convolutional SRRESNET model as the generator architecture to produce excellent super-resolution images. To improve the overall architecture and guarantee high-quality photos, a discriminator model that functions as an image classifier is also incorporated. The SRGAN architecture produces natural-looking, perceptually high-quality images.

A low-resolution input is handled by a first convolutional layer with 9x9 kernels and 64 feature maps before being sent on to a parametric ReLU layer in the SRRESNET

[16] generating network. Because parametric ReLU is a powerful non-linear activation function for the job of converting low-resolution images to high-resolution images, it is used across the entire generator design.

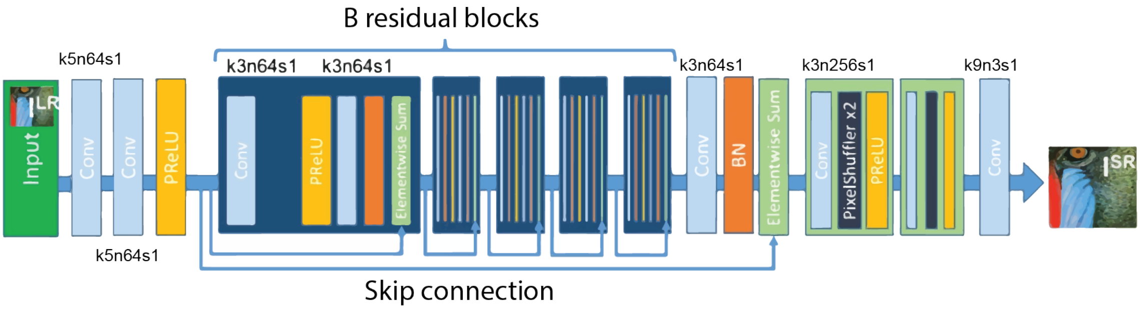

Figure 4.2: Improved generator

Although values less than zero can be simply mapped to zero when using an activation function like ReLU, this can result in dead neurons. Leaky ReLU, which maps values less than zero to a user-defined value, is an alternate choice. In this case, parametric ReLU is a superior option because it enables the neural network to select the best value on its own. This is due to parametric ReLU's ability to allow the network to change the value used to map negative inputs, which prevents the problem of dead neurons.

A number of residual blocks are included in the feed-forward, fully convolutional SR-RESNET model's design. Each residual block is made up of a batch normalization layer, a parametric ReLU activation function, a convolutional layer with batch normalization, a layer with 3x3 kernels and 64 feature mappings, and a final element wise sum method. The feed-forward output and the skip connection output are combined to create the final output using the element wise sum technique.

The remainder of the generator model is built as illustrated in the accompanying image once the residual blocks are formed. In order to produce super-resolution images, the generator model design incorporates a pixel shuffler after the convolutional layer has been 4x upsampled. The pixel shuffler shuffles the height and width dimensions with values from the channel dimension. In this instance, the channel is half while the height and width are both doubled.

4.2.2 Discriminator

The generator and discriminator compete with one another and improve at the same time in the standard GAN process, which is supported by the discriminator architecture. While the generator tries to create realistic images to evade detection by the discriminator, the discriminator network seeks to recognize fraudulent images. The differentiable discriminator D, trained to discriminate between genuine images and super-resolved images, is what the generative model G in SRGANs attempts to trick.

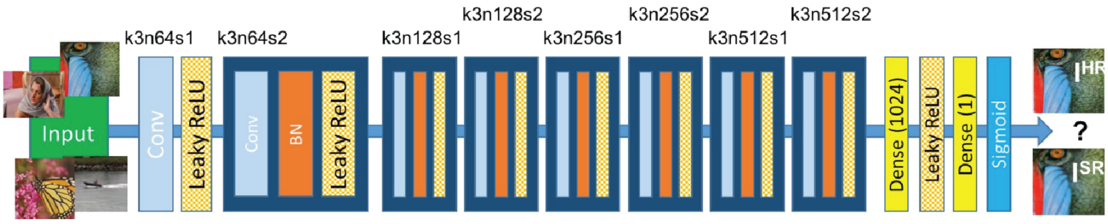

Figure 4.3: Discriminator

The discriminator architecture is uncomplicated and simple to comprehend. It is made up of an initial convolutional layer, followed by a 0.2 alpha Leaky ReLU activation function. A batch normalization layer, a Leaky ReLU activation function, and a sequence of repeated blocks of convolutional layers are also included in the architecture. After five of these repeating blocks, there are thick layers and then a classification-focused sigmoid activation function. The initial convolutional size is 64x64 and doubles every two full blocks until it reaches the 512x512 8x upscaled size. The generator learns more effectively and generates better outcomes thanks to this discriminator model.

- Generator Improvement

  The first modification we will do is to switch from one 9*9 convolution layer to two 5*5 convolution kernels. Although the execution time will be optimized, the outcome will be about the same. The discriminator's Batch Normalization layer will then be removed in order to eliminate undesirable artifacts and stabilize the training.

- Improvement of loss function

  We will use content loss instead of MSE loss. The following image gives the same MSE loss value though they are not similar.

4.3 Underwater ResNet (U-ResNet)

We have already announced that we want to use Underwater ResNet to improve images before sending them on to SRGAN to boost their resolution. We'll talk about the Underwater ResNet architecture right now.

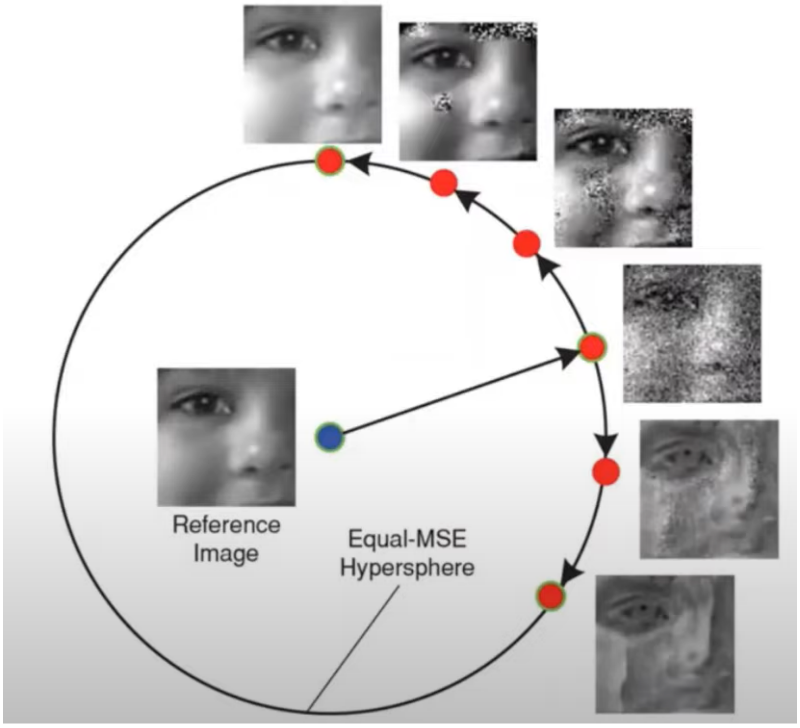

Figure 4.4: Complication of MSE loss

A residual learning model, UResnet is what we've suggested. ResBlocks, which are the building blocks of this system, merge the output of one convolution layer with the input of another. The data from the top layer can be fully transmitted to the layers below by using ResBlocks. ResBlock stacking makes it possible to train deeper networks. The three main components of the proposed UResnet model are the head, torso, and tail. Through a long-distance skip connection, the outputs from the head and body sections are joined. The body's output layer receives feature information from the input layer via the long-distance skip link, necessitating the use of ResBlock modules to discriminate between input images and label images. In the head, there is only one convolution layer. The body component stacks 16 ResBlocks in the following order to account for the time needed for training: [Conv-BN-ReLU-Conv-BN]. The tail contains one convolution layer. There are a total of 34 convolution layers. UResnet can accept inputs of any shape since the network utilizes a 33 convolution with a 1 pixel stride and a 1 pixel zero-padding to retain the geometry of feature maps.

In our underwater resnet, Edge Difference Loss (EDL) along with asynchronous training mode has been included. These topics are discussed next.

4.3.1 Edge Difference Loss (EDL)

The MSE or L1 loss function, which seeks to educate the model how the two images differ at the pixel level, is used by the majority of image-to-image translation models. Because MSE Loss averages differences at the pixel level and ignores higher-level information like an overall structure, the resulting images do not have strong visual effects, but the model can get a higher peak signal-to-noise ratio (PSNR) score using this approach. This means that the MSE Loss function is not the best choice for boosting high-frequency information because it tends to average the solution and smooth out visual features.

A punishment phrase called edge difference loss (EDL) is suggested due to the underwater photos' considerable feature loss, particularly with regard to edge information. The level of the resulting image's details is raised by punishing the models with EDL. A Laplacian operator $\begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ is used as a sensitive edge detection operator to calculate the EDL. The output of the final layer of the model are subjected to the convolution operation using the Laplacian template as the convolution kernel. The label image is simultaneously subjected to the convolution operation using the Laplacian template. The EDL is then computed using the MSE Loss of these two feature maps.

$$Loss = MSE\ Loss + k * EDL \tag{4.1}$$

The coefficient *k* is a hyperparameter that determines the balance between the two components of the loss function. It adjusts the proportion of each part in the overall loss calculation. The specific value of *k* in equation (4.1) is determined through a greedy search method, which involves conducting numerous experiments to find the optimal value. Finding of *K* value leads us to asynchronous training mode.

4.3.2 Asynchronous Training Mode

Finding the ideal value of "k" in equation (4.1) to optimize the output can be difficult in real-world situations. This problem arises because the Laplacian operator used in the equation is susceptible to noise and sensitive to edge information. The output image's quality can be negatively impacted by noise amplification if the value of "k" is improperly selected. An asynchronous training strategy is suggested as a solution to this problem. This method of training can be used with other deep learning models that use a multi-term loss function. The following actions are included in the asynchronous training mode

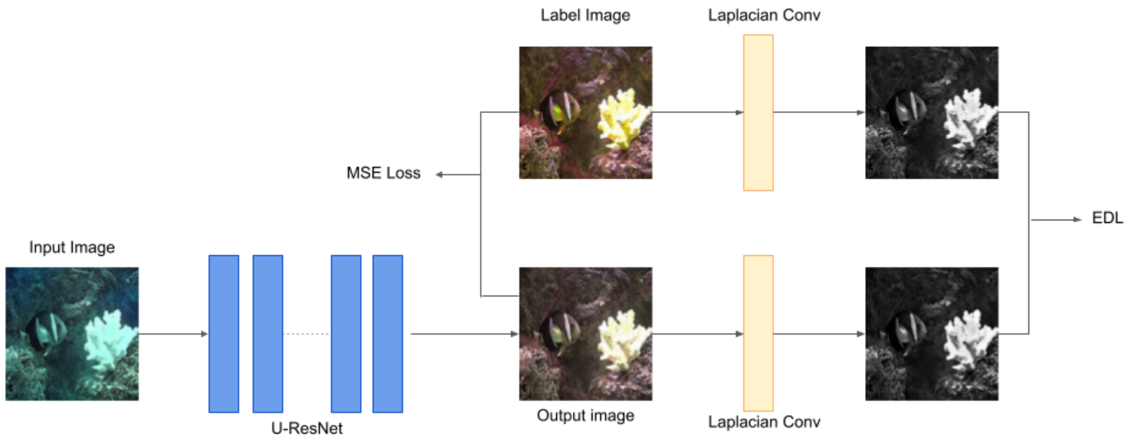

Figure 4.5: U-ResNet with EDL

steps:

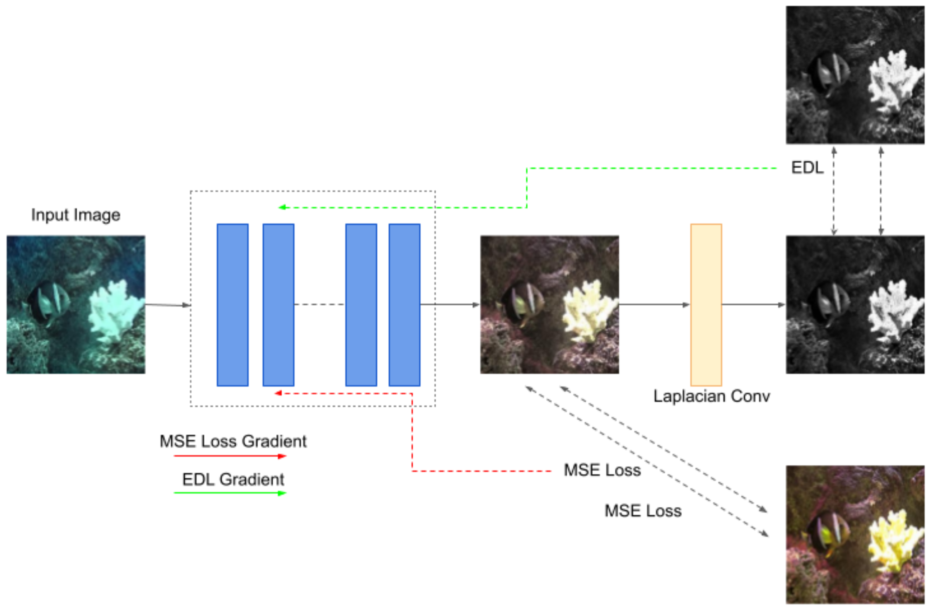

Figure 4.6: Asynchronous Training Mode

The network undergoes a two-round training process for each batch. In the first round, the gradients are calculated and back propagation is performed using the EDL (Edge-Enhancing Loss), leading to updates in the network weights. In the second round, the gradients are computed using MSE loss, and they are propagated back to update the network weights once again. As a result, every batch undergoes two rounds of training, and the network weights are updated twice for each batch.

The asynchronous training mode is employed, where the first training round utilizes EDL (Edge-Enhancing Loss). This approach leverages the capability of EDL to preserve edge information and aid the network in recovering fine details and edges. However, the influence of EDL on the network is constrained by the second training round, which emphasizes the pixel-level differences between the output and label images. Consequently, this limitation prevents the amplification effect of the Laplacian operator on noise, ensuring that noise is suppressed during the training process.

Furthermore, if the two components of the loss function are trained with different weights, as done in traditional multi-loss training models, determining the appropriate allocation of these weights becomes a challenge that requires conducting numerous experiments. Finding the optimal weights becomes crucial, as their determination is typically fixed and can potentially compromise the robustness of the model.

Chapter 5

Results

We will compare our result from two perspective - from **Quantitative** perspective and also from **Qualitative** view.

5.1 Quantitative Analysis

For Quantitative analysis, we have evaluated PSNR, UICM, UISM, UIConM and UIQM metrics. We have compared out model with other benchmark model on EUVP [6] datasets based on these metrics and also compared the performance of vanilla U-ResNet model on the EUVP dataset. Before diving into the value comparison, we are discussing about the evaulation metrics first.

- **PSNR:** A popular metric for rating the quality of improved underwater photos is the Peak Signal-to-Noise Ratio (PSNR). It gives a numerical evaluation of the degree of distortion or noise present in the improved image by comparing the differences between the original and enhanced images' pixel values. An improved image's quality and fidelity are suggested by a greater PSNR value, which shows a stronger similarity between the enhanced and original image. PSNR is frequently used as one of the quantitative measurements to evaluate the efficacy of different algorithms and strategies in raising the quality of underwater images. However, it is important to note that while PSNR provides a numerical assessment, it may not capture the perceptual quality or visual appeal of the enhanced underwater images. Therefore, it is usually complemented with qualitative analysis and other metrics to obtain a comprehensive evaluation.
- **Underwater Image Contrast Measure (UICM):** The improvement in foreground

and background contrast in underwater photographs is evaluated by UICM. It measures the improvement in visual details as well as the decrease in haze or color cast.

- **Underwater Image Sharpness Measure (UISM):** The improvement in image sharpness or clarity is measured by UISM. It assesses how well tiny features are preserved and how much blur is reduced as a result of underwater lighting.

- **Underwater Image Content Measure (UIConM):** UIConM evaluates the preservation of important image content, such as the visibility of underwater objects, marine life, or underwater structures. It quantifies the improvement in image content visibility and detail.

- **Underwater Image Quality Measure (UIQM):** UIQM provides an overall assessment of the quality of underwater images by considering multiple factors, including color reproduction, contrast, sharpness, and content visibility. It combines various image attributes to generate a comprehensive quality score.

To add on these metrics, by quantifying particular elements relating to contrast, sharpness, content visibility, and overall image quality, UICM, UISM, UIConM and UIQM metrics aid in the objective evaluation of the performance of underwater image enhancing systems. Researchers can examine the performance of various enhancement algorithms and gain a deeper knowledge of their usefulness by combining these measurements.

Now, let's see the table. *Table* 5.1 gives us a comparison between our model with the existing architectures:

- **PSNR:** ”LAFFNet” has the greatest PSNR score (28.42), followed in third by our model (25.46). With a 26.78, ”FUnIE-GAN” also performs admirably. The PSNR ratings for the ”DCP” and ”ILBA” approaches are the lowest.

- **UICM:** With a UICM score of 10.606, our model tops the list, followed by ”ILBA” (7.892) and ”FUnIE-GAN” (7.040). The score of ”DCP” has the lowest UICM.

- **UISM:** The two programs with the highest UISM scores are ”LAFFNet” (6.724) and ”FUnIE-GAN” (5.606). In terms of UISM, alternative techniques range from 4.005 to 5.292. Our model falls behind on this metric but is still not the lowest-valued model for this.

- **UIConM:** With the lowest UIConM score (0.258), ”LAFFNet” exhibits superior content preservation. The UIConM values for our model beats the score of all other models with a score of 0.262

- **UIQM:** The highest value is achieved by the ”LAFFNet” model(3.092) while followed by the ”Deep SESR”(2.638) with the second position. Our model achieves a comparatively lower score here.

In conclusion, ”LAFFNet” routinely outperforms the competition in terms of a variety of measures, earning excellent marks in the PSNR, UISM, and UIQM. Our model also performs well in terms of UICM, UIConM and has a PSNR score that is quite high. The performance of ”FUnIE-GAN” is competitive in terms of PSNR, UICM, and UISM. The scores for ”DCP” and ”U-GAN” are generally lower across all criteria.

Method	PSNR	UICM	UISM	UIConM	UIQM
DCP	17.55	6.781	4.005	0.056	1.575
ILBA	18.83	7.892	4.389	0.123	1.958
U-GAN	23.67	6.052	5.120	0.224	2.483
WaterNet	20.14	6.736	5.292	0.212	2.511
FUnIE-GAN	26.78	7.040	5.606	0.185	2.514
Deep SESR	25.25	5.975	5.211	0.260	2.638
LAFFNet	28.42	6.502	6.724	0.258	3.092
Ours	25.46	10.606	4.431	0.262	2.544

Table 5.1: Quantitative comparison on the EUVP dataset [6]. The best results are RedFaced and the second best ones are BlueFaced

Now the following table in Table 5.2 gives us the comparison ont the superior EUVP dataset trained and tested on both our model and UResNet model:

Method	PSNR	UICM	UISM	UIConM	UIQM
UResNet	24.87	8.54	5.896	0.203	2.707
Ours	25.45	10.606	4.431	0.262	2.544

Table 5.2: Quantitative comparison on the EUVP dataset [6] between vanilla UresNet [7] and our Model. The best results are RedFaced.

- **PSNR:** Our model outperforms UResNet, which only receives a score of 24.87, with a PSNR score of 25.45.
- **UICM:** Compared to UResNet’s score of 8.54, our model’s UICM score of 10.606 is noticeably higher.
- **UISM:** Our model receives a UISM score of 4.431, while UResNet receives a better score of 5.896.
- **UIConM:** In comparison to UResNet’s value of 0.203, our model obtains a slightly higher UIConM score of 0.262.

- **UIQM:** Our model receives a UIQM score of 2.544, but UResNet scores 2.707, which is a little higher.

In conclusion, when compared to UResNet, our model performs better in terms of PSNR, UIConM and UICM. In terms of UISM, UResNet performs better than our model. Our model performs somewhat better in UIConM and UResNet performs slightly better in UIQM, according to the values for UIConM and UIQM.

When compared to UResNet on the EUVP dataset, our model generally shows promise in terms of enhancing image quality, notably in terms of PSNR and UICM. To reach more thorough conclusions, additional research and comparisons on more datasets would be helpful.

5.2 Qualitative Analysis

In order to evaluate the quality, clarity, and overall appearance of underwater photos, numerous visual aspects must be subjectively evaluated. It requires looking at elements including subject recognition, image visibility, color reproduction, contrast and dynamic range, image sharpness, noise, and artifacts. To evaluate the efficacy of underwater image enhancing techniques and algorithms, experts or observers carefully examine these elements. The analysis aids in understanding how underwater images are viewed differently by different people and serves as a roadmap for future developments in underwater imaging technologies. We should also keep these information in mind for qualitative analysis-

- **Image Visibility:** Analyzing the clarity of objects, amount of detail, and overall visibility of underwater scenes under various lighting situations.
- **Color Reproduction:** Assessing the accuracy and richness of colors in underwater photos, with a focus on preserving natural tones and removing any color casts or distortions brought on by the water.
- **Contrast and Dynamic Range:** Analyzing the distribution of tonal values and contrast in underwater photographs to make sure there is a fair mix of dark and light areas and to capture a variety of features in both shadows and highlights.
- **Image Sharpness:** Evaluating the degree of clarity and sharpness in underwater photos, including the retention of minute details and the avoidance of blurring or softening brought on by water turbulence or photo editing.

- **Noise and Artifacts:** Locating and assessing any visual noise, distortions, or undesired abnormalities, such as graininess, chromatic aberrations, or compression artifacts, that may be present in the underwater photos.

- **Overall Aesthetics:** Taking into account composition, lighting, and the ability to capture the beauty and distinctiveness of underwater surroundings, as well as the overall visual appeal and aesthetic aspects of underwater photos.

After achieving notable quantitative improvements in some metrics in enhancing underwater images through our research, we will now present a set of images for qualitative analysis.

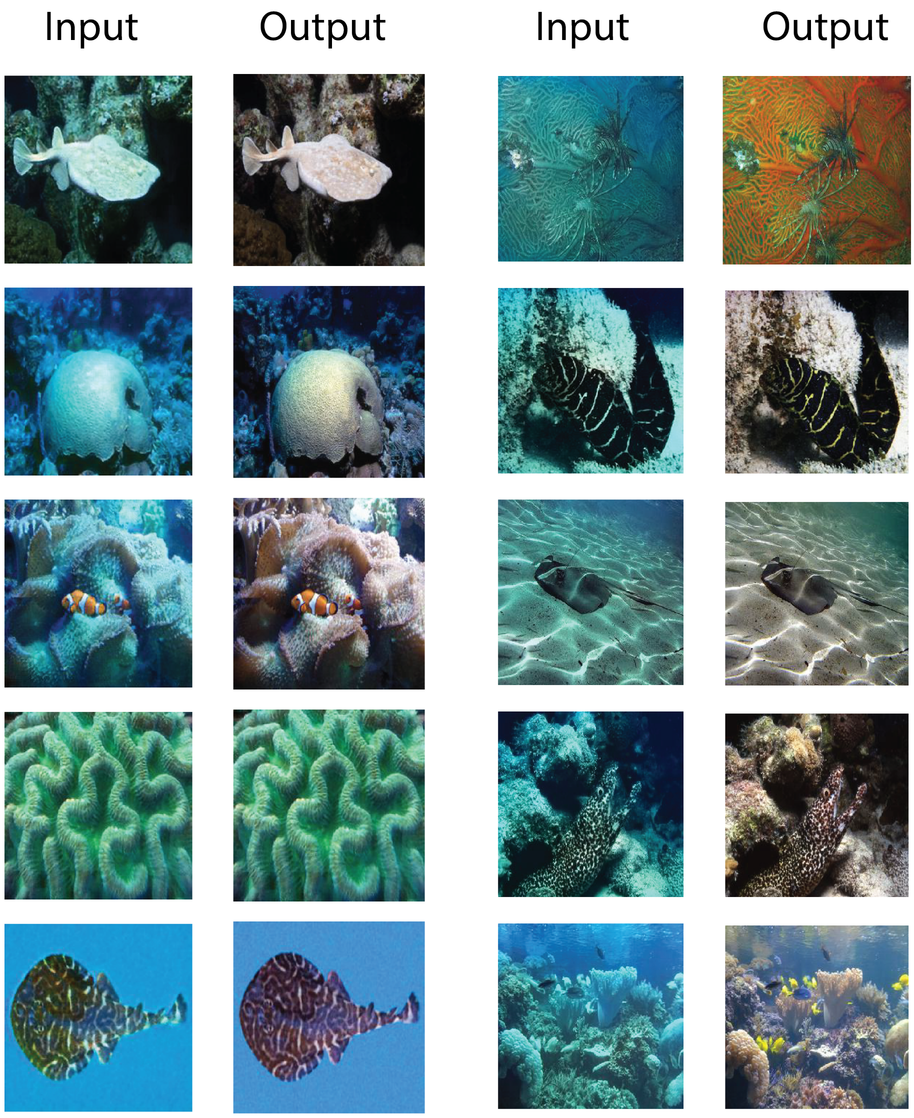

Figure 5.1: Comparison between the raw underwater image and enhanced image by our model

Chapter 6

Conclusion and Future Works

6.1 Conclusion

In this study, we investigated convolutional neural networks, generative adversarial networks, and residual networks as potential deep learning methods for improving the quality of underwater photographs. Residual networks have been the most cutting-edge method among these.

To increase the resolution of underwater photos, we have nonetheless suggested a technique that blends residual networks with generative adversarial networks. Our proposed model has proven to perform better through thorough quantitative study and comparison with UResNet, a standard residual network.

Our model demonstrates its potential to improve the quality of underwater photos by attaining the greatest PSNR and UICM scores. Our model's competitive performance in UIConM and UIQM further proves its efficacy, even though UResNet exceeds it in terms of UISM.

We admit that the evaluation is based on the EUVP dataset, and that other varied and large datasets should be used to train our model to increase its generalizability. This will increase the model's accuracy and provide it the opportunity to learn a wider variety of underwater image features.

Overall, our suggested approach has shown its ability to considerably improve the resolution of underwater photographs by fusing residual networks with generative adversarial networks. To improve the model's performance and confirm its efficacy in real-world circumstances, future research should concentrate on extending the dataset utilized for training and testing the model.

6.2 Future Works

By doing more training iterations, we want to improve our current model's performance by boosting accuracy and streamlining the output. We also intend to investigate more architectures than the first ResNet implementation. We anticipate improving the overall performance of our system by training the model utilizing various architectures. Additionally, we'll look into using different Generative Adversarial Networks (GANs) instead of the SRGAN used in the final stages of our suggested model. This investigation of several GANs has the potential to improve our outcomes even further. We also understand how crucial it is to use stronger and more effective computing resources for training. In order to speed up the training process and ultimately produce better results, we plan to switch to a machine that is more capable. We will be able to get more specialized and superior results because to this improvement in computational power.

References

[1] Z. Fu, X. Lin, W. Wang, Y. Huang, and X. Ding, "Underwater image enhancement via learning water type desensitized representations," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2764–2768, IEEE, 2022.

[2] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[3] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Transactions on Image Processing*, vol. 30, pp. 4985–5000, 2021.

[4] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4861–4875, 2020.

[5] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 8, pp. 2822–2837, 2020.

[6] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.

[7] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu, "Underwater image enhancement with a deep residual framework," *IEEE Access*, vol. 7, pp. 94614–94629, 2019.

[8] J. Ao and C. Ma, "Adaptive stretching method for underwater image color correction," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 02, p. 1854001, 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

[10] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *2018 IEEE winter conference on applications of computer vision (WACV)*, pp. 1451–1460, Ieee, 2018.

[11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

[12] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[14] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic imaging*, vol. 11, no. 2, pp. 206–223, 2002.

[15] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, 2017.

[16] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.

[17] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th international conference on pattern recognition*, pp. 2366–2369, IEEE, 2010.

[18] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2015.

[19] P. M. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning.," in *CVPR workshops*, pp. 1–8, 2019.

[20] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 81–88, IEEE, 2012.

[21] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *2014 IEEE international conference on image processing (ICIP)*, pp. 4572–4576, IEEE, 2014.

[22] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2856–2868, 2018.

[23] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.

[24] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE transactions on image processing*, vol. 26, no. 4, pp. 1579–1594, 2017.

[25] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal processing letters*, vol. 25, no. 3, pp. 323–327, 2018.

[26] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multi-scale dense generative adversarial network," *IEEE Journal of Oceanic Engineering*, vol. 45, no. 3, pp. 862–870, 2019.

[27] R. Timofte, S. Gu, J. Wu, L. Van Gool, L. Zhang, M.-H. Yang, M. Haris, *et al.*, "Ntire 2018 challenge on single image super-resolution: Methods and results," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[28] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, pp. 630–645, Springer, 2016.

[30] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[31] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?," *Advances in neural information processing systems*, vol. 31, 2018.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.