

TRANSFORMER BASED DIFFUSION MODELS FOR UNDERWATER IMAGE RECONSTRUCTION

BY

M. HASHEM M. KENANA SHULLAR

A Thesis Presented to the
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

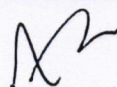
JULY, 2024

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **M. HASHEM M. KENANA SHULLAR** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**.

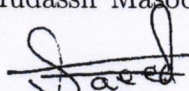
Thesis Committee



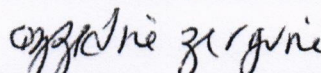
Dr. Ali Nasir (Adviser)



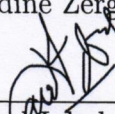
Dr. Mudassir Masood (Co-Adviser)



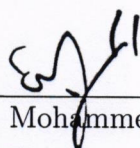
Dr. Saeed Anwar (Member)



Dr. Azzedine Zerguine (Member)



Dr. Naveed Iqbal (Member)



Dr. Mohammed Mustafa
Al-Muhaini

Department Chairman



Dr. Suliman Saleh AlHomidan

Dean of Graduate Studies

Date



©M. Hashem M. Kenana Shullar
2024

To my family.

ACKNOWLEDGMENTS

I would like to dedicate my thanks first and foremost to Allah Almighty. Secondly, I want to express my sincere gratitude to my dear father and mother. I would also like to thank my brother, who provided significant help during the final months of my Master's journey. Without his support, this work would not have been completed. Lastly, this work could not have been accomplished without the constant help and guidance of my advisors, Dr. Ali Nasser, Dr. Mudassir Masood, and Dr. Saeed Anwar. I am deeply grateful for their patience and the support they have shown me, and for giving me the opportunity to finish my Master's. Therefore, I would like to express my heartfelt appreciation to them.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
ABSTRACT (ENGLISH)	xii
ABSTRACT (ARABIC)	xiv
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Thesis Outcomes	2
1.3 Thesis organization	3
CHAPTER 2 LITERATURE REVIEW	4
2.1 Non-physical model methods	4
2.2 Physical model methods	5
2.3 Deep Learning methods	6
2.3.1 Transformers	7
2.3.2 Diffusion Models	8
2.4 Summery	10

CHAPTER 3 BACKGROUND	11
3.1 Underwater Light Propagation	11
3.2 Deep Learning	12
3.2.1 Convolutional Neural Networks (CNNs)	13
3.2.2 Transformers	15
3.2.3 Encoder-Decoder (U-Net)	19
3.2.4 Diffusion Models	20
3.3 Summary	23
CHAPTER 4 RESULTS OF APPLYING THE DIFFUSION MODEL ON THE DENOISING AND DERAINING TASKS	25
4.1 Baseline Details	26
4.2 Datasets	26
4.3 Results and Discussion	27
4.4 Summary	30
CHAPTER 5 DIFFUSION MODELS FOR IMAGE RECONSTRUC- TION	31
5.1 Diffusion Models for Underwater Image Reconstruction	31
5.2 Results and Analysis	33
5.3 Summary	35
CHAPTER 6 CONCLUSION	36
6.1 Summary	36
6.2 Future Work	37
REFERENCES	38
VITAE	50

LIST OF TABLES

- 4.1 Comparing our model against the PromptIR baseline 27
- 4.2 Comparing our model against state of the art on Denoising and Deraining 27
- 5.1 Comparing our approach against different underwater reconstruction
methods 33

LIST OF FIGURES

3.1	The three components of underwater imaging: direct component, forward component, and backward scatter component	13
3.2	Overview of the Basic CNN Architecture.	15
3.3	Overview of ViT Architecture	16
3.4	Overview of the U-Net Architecture.	20
3.5	A general overview of the Diffusion Model Architecture.	21
3.6	A comparison between Diffusion Models and other generative models. .	23
4.1	Overview of our Approach	26
4.2	Visual comparisons between our model and the literature on the Denoising task along with corresponding PSNR/SSIM values	28
4.3	Visual comparisons between our model and the literature on the De-raining task along with corresponding PSNR/SSIM values	29
5.1	Visual comparisons between our model and state of the art on the UIEB dataset along with corresponding PSNR/SSIM values	34

LIST OF ABBREVIATIONS

AUVs	Autonomous Underwater Vehicles
CNN	Convolutional Neural Network
CV	Computer Vision
GAN	Generative Adversarial Network
HSV	Hue-Saturation-Value Color Space
IFM	Image Formation Model
Lab	lightness-red/green coordinate-yellow/blue, a Color space defined by the International Commission on Illumination
LCH	Hue-Chroma-Luminance Color Space
NLP	Natural Language Processing
PSNR	Peak Signal-to-Noise Ratio
RGB	Red-Green-Blue Color Space
SOTA	State of the Art

SSIM	Structural Similarity Index Measure
UCIQE	Underwater Color Image Quality Evaluation
UIEB	Underwater Image Enhancement Benchmark
UIQM	Underwater Image Quality Measure
ViT	Vision Transformer

THESIS ABSTRACT

NAME: M. Hashem M. Kenana Shullar

TITLE OF STUDY: Transformer based Diffusion Models for Underwater Image Reconstruction

MAJOR FIELD: Electrical Engineering

DATE OF DEGREE: July, 2024

Images taken underwater are affected by multiple types of degradation due to the nature of these environments. Such effects include color cast, non-uniform lighting and light scattering due to suspended particles. These effects severely lower the quality of the images which in turn would make them not applicable for use in downstream tasks such as underwater mapping and detection. In this work, a diffusion model was trained to specifically tackle the underwater image reconstruction task. This model was compared against existing deep learning methods, across multiple quality metrics. Our findings reveal that the diffusion model outperforms the existing methods in terms of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Underwater Image Quality Measure (UIQM). These results underscore the effectiveness of diffusion models in enhancing the quality of underwater imagery,

thereby extending their applicability in downstream tasks that require high-fidelity visual data.

ملخص الرسالة

الاسم: هاشم كنانة شلار

عنوان الدراسة: نماذج الانتشار القائمة على المحولات لمهام إعادة بناء الصور الملتقطة في البيئات المائية

التخصص: قسم الهندسة الكهربائية

تاريخ الدرجة العلمية: يوليو، 2024

تتأثر الصور الملتقطة في البيئات المائية بأنواع متعددة من عوامل التشويه بسبب طبيعة هذه البيئات. تشمل هذه التأثيرات تشوه لون الصور، والإضاءة غير المتساوية، وتشتت الضوء. تؤدي هذه التأثيرات إلى خفض جودة الصور بشكل كبير مما يجعلها غير مناسبة للاستخدام في مهام أخرى مثل رسم الخرائط واستكشاف البيئات تحت الماء. في هذا العمل، تم تدريب نموذج انتشار (Diffusion Model) خصيصاً لمهمة إعادة بناء الصور تحت الماء. تمت مقارنة هذا النموذج بالطرق الحالية للتعلم العميق عبر مقاييس جودة متعددة. أظهرت نتائجنا أن نموذج الانتشار يتفوق على الطرق القائمة من حيث مؤشر نسبة قوة الإشارة إلى الضوضاء (PSNR)، ومقياس التشابه الهيكلي (SSIM)، ومقياس جودة الصورة تحت الماء (UIQM). تؤكد هذه النتائج فعالية نماذج الانتشار في تحسين جودة الصور تحت الماء، مما يوسع تطبيقاتها في المهام اللاحقة (Downstream tasks) التي تتطلب بيانات بصرية عالية الدقة.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Image restoration is a fundamental problem in the field of computer vision and image processing, focusing on the recovery of an original, clean image from a corrupted or degraded version. Corruption can arise from various sources, such as noise, blur, and compression artifacts. The task of image restoration is crucial in enhancing image quality for diverse applications, including medical imaging, satellite imaging, surveillance, and photography. By reconstructing images that have been compromised, restoration techniques enable better interpretation, analysis, and decision-making across various fields [1], [2].

In medical imaging, for example, the precision and clarity of images are vital for accurate diagnosis and treatment planning [3]. Noise or artifacts in medical images can obscure critical details, potentially leading to misdiagnoses or overlooked conditions. Similarly, in satellite imaging, the ability to restore clear and precise images is essential for monitoring environmental changes, managing natural disasters, and supporting

agricultural practices. High-quality restored images provide more reliable data for researchers and policymakers, facilitating informed decisions based on accurate visual information.

Furthermore, in surveillance and security, image restoration enhances the ability to identify and analyze objects, faces, and activities captured by cameras, particularly in low-light or adverse weather conditions. This capability is crucial for ensuring public safety and effective law enforcement. Overall, the advancement of image restoration methodologies continues to expand the potential and reliability of visual data across an array of critical applications.

Underwater image reconstruction refers to a subset of image restoration tasks where the problem involves processing an input image taken underwater to either enhance the quality of the image (i.e., in terms of the ability to distinguish different objects present in that image) or restore some of the details that might be missed due to certain circumstances (Such as color cast) [4], [5], [6].

1.2 Thesis Outcomes

- A literature survey on image restoration methods in general and underwater image reconstruction (UIR) specifically is conducted, providing an understanding of the current state-of-the-art techniques.
- The use of transformers for image restoration is investigated.
- The use of transformer-based diffusion models for image restoration is explored and evaluated, showing how these models can effectively handle various degra-

dation types such as noise, rain, and underwater distortions.

- Achieved significant enhancements over existing methods in the literature by demonstrating superior performance in terms of PSNR, SSIM, and UIQM across various image restoration tasks, specifically, denoising, deraining, and underwater image reconstruction.

1.3 Thesis organization

This thesis is organized into six chapters. Chapter 1 introduces the topic, outlines the thesis outcomes, and describes the organization of the thesis. Chapter 2 provides a literature review, discussing the different approaches existing in the literature to solve image restoration and UIR in general. Chapter 3 covers the needed background of image restoration, deep learning techniques such as Transformers, encoder-decoder architectures (U-Net), and diffusion models. Chapter 4 presents the results of applying the diffusion model to denoising and deraining tasks, detailing the baseline model, datasets used, and discussing the results and their implications. Chapter 5 focuses on the application of diffusion models for underwater image reconstruction, presenting the methodology, results, and analysis. Finally, Chapter 6 concludes the thesis with a summary of the findings and suggestions for future work.

CHAPTER 2

LITERATURE REVIEW

Generally, the underwater reconstruction methods investigated in the literature could be categorized into three main categories: 1) Non-physical model methods, 2) Physical model methods, and 3) Deep Learning methods.

2.1 Non-physical model methods

Non-physical model underwater image reconstruction techniques improve the visual properties of the degraded image by working on the pixel values of the input image and modifying them. Such techniques do not assume a physical model that simulates the imaging process as in other techniques. An example of a technique falling under this category is the histogram equalization method [7]. Histogram equalizers improve image contrast by non-uniform scaling of pixel values of the image, and since images captured in underwater environments suffer from color shift issues due to the different attenuation factors, using histogram equalization was investigated for the underwater reconstruction problem by many researchers [8], [9], [10]. Another color

correction method used for underwater images is the White balance technique [11]. White balance works under the observation that in non-uniformly lit environments, color deviations are significant. White balancing aims at correcting those color deviations and thus, it is somewhat suitable for underwater image reconstruction [12]. In [13], authors operated on both the Red-Green-Blue (RGB) and Hue-Saturation-Value (HSV) color spaces by scaling the dynamic pixel range of the image. Authors in [14], [15] propose retinex-based approach where they attempt to enhance image quality by decomposing the image into reflectance and illumination components, correcting color distortions. Although these methods are simple and computationally efficient [16], [17], [18], they are naive and do not take into consideration the varying types of underwater environments and the effects they introduce to the image. Such techniques rarely produce satisfying results.

2.2 Physical model methods

Physical model techniques go through a simple process: first researchers start by choosing or designing an appropriate image degradation model of the underwater environment. Then, they either estimate the model parameters. Finally, the model is used to solve an inverse problem utilizing whatever assumptions or priors they make about the environment. Image models, especially complex ones, require the use of special equipment to measure some of the parameters used in them. Priors are usually used to estimate the model parameters, instead of using measuring equipment [19], [20], [21], [22]. The Dark Channel Prior [23], for example, inspired by its use in the

image dehazing problem, was adapted for use in underwater restoration [24]. In [25], the authors tested a combined wavelength-dependent compensation and Dark Channel Prior approach for underwater reconstruction. The authors in [26] note that the dark channel of underwater images is effectively a zero map. They solve an optimization problem where they minimize an objective function based on that assumption. Their optimal transmission map proves to be able to produce images with high contrast. In [27], a generalized Dark Channel Prior for image reconstruction is proposed. The method combines adaptive color correction with an image formation model. Using priors for parameter estimation limits the use of such reconstruction techniques to cases where these priors hold. For the dark channel prior case, for example, results are extremely poor under nonuniform light.

2.3 Deep Learning methods

Deep learning for low-level vision problems has shown significant improvements in the last couple of years. Such techniques are trained using degraded/non-degraded image pairs. The use of CNNs and GANs showed impressive results [28], [29], [30], [31]. In [30], the authors proposed WaterGAN, a Generative Adversarial Network based method. Synthetic underwater images are generated from air-depth map pairs. Then, the synthetic dataset is used to train a network consisting of two stages for color shift correction. Due to the difficulty of having real clear/unclear image pairs, researchers usually use synthetic data sets to train their models such as in [28], where they use their synthetic dataset to train a CNN network on underwater images of different

types. [32] proposes Water CycleGAN. CycleGAN-based approaches ease degraded-clear underwater image pairs constraint during training. This method, however, may output inauthentic images due to having multiple possible outputs.

2.3.1 Transformers

Inspired by their success in the Natural Language Processing (NLP) field, the use of Transformers for Computer Vision has been investigated and recent Transformers achieved SOTA results, in many high-level vision tasks. However, using those architectures is still not fully addressed for low-level vision tasks such as dehazing, deblurring, denoising, and underwater reconstruction [33]. In [34], the authors pre-trained a Transformer for low-level vision. The model accommodates different tasks, including super-resolution, denoising, and deraining. This is done by using multiple pairs of head and tail corresponding to different tasks and a single shared body. A Transformer architecture inspired by UNet is built for image restoration [33], where the Transformer is applied to local windows to overcome their quadratic complexity. Specific to the underwater image reconstruction problem, there are multiple works that incorporate Transformers for [35], [36], [37]. The authors in [38], integrated a U-shape Transformer with a channel-wise and spatial-wise Transformers. The network is trained on a loss function that takes into account the RGB, Lab, and LCH color spaces.

An adaptive group attention was proposed in [39], which they claim that it can dynamically select visually complementary channels that would help guide the network in the image enhancement process. In [40], the authors used a model based on

residual dense attention and convolution blocks. The majority of these works utilize the attention mechanism without using the complete Transformer based architecture despite the fact that it showed SOTA results in high-level vision tasks.

2.3.2 Diffusion Models

In the context of computer vision, non-generative methods showed good results when tackling an array of high-level vision tasks. However, their efficacy when confronted with the challenges inherent in image reconstruction, such as underwater image reconstruction, is not comparable to their results on high-level vision tasks. Low-level vision tasks are characterized by their complex forms of image degradation, which can significantly hinder these models' capacity for precise and high-quality visual restoration.

Generative models emerge as particularly adept in this context, given their proficiency in generating images from inputs that have been substantially degraded. Their effectiveness is largely attributable to their capacity to navigate multimodal distributions. This attribute is crucial for image reconstruction tasks, where the desired outcome may correspond to several possible degraded states, and implementations applying such methods have shown promising results.

Diffusion Models, as a class of generative models, have demonstrated promising results in image generation tasks. Unlike traditional approaches, which may not fully capture the diversity of degradation modes present in underwater imagery, Diffusion Models thrive in navigating these complex distributions. This enables a more nuanced restoration of underwater images that closely aligns with the real-world variability of

these challenging environments.

The work in [41] introduced an enhancement approach for degraded low-light images by proposing a degradation-aware learning scheme utilizing diffusion models. The Retinex-conditioned diffusion model, ReCo-Diff, aims to separate the target image into two components similar to the human eye perspective: the illumination and the reflectance images. This method involves the usage of a network to estimate the initial separation. Then, another network works to clean these separated components to produce less noisy images. It was shown that this model surpasses its peers like LIME [42] and DRBN [43] in several matrices including the PSNR and SSIM. The work in [44] explored the ability of diffusion models to enhance the quality of images taken underwater. Their main idea was to facilitate the two UNets to perform denoising followed by applying several transformations to change the distribution. The work in [45] introduces a framework that consists of a commonly used image enhancement technique like histogram equalization and color correction and a pre-trained diffusion model. The proposed pipeline starts by applying the image enhancement techniques on the image. From there, the diffusion model will use its knowledge from pretraining to denoise the image. In addition to its ability to produce more natural images, this approach also provides better adaptability (i.e., allowing the user to select the intended image enhancement technique) and variation resilience related to the image conditions.

2.4 Summery

To sum up, the methods for solving the underwater image reconstruction problem can be categorized into three categories, which are non-physical model methods, physical model methods, and deep learning-based methods. While non-physical model methods are computationally efficient, they are naive and do not account for the underlying nature of images, this makes them impractical for the underwater image reconstruction task. On the other hand, physical models assume an image model that captures the degradation process the taken images have gone through, however, these techniques usually do not generalize enough to different underwater environments and require the use of handcrafted priors. Deep learning-based methods depend on the availability of large and diverse enough dataset and do not require any handcrafted assumptions before their operation. Among the deep learning methods, Transformers and Diffusion models, have achieved SOTA results for image restoration tasks and show great potential for the underwater image reconstruction problem.

CHAPTER 3

BACKGROUND

3.1 Underwater Light Propagation

There are two main interactions the underwater light goes through when traveling through water:

1. **Absorption:** This refers to light's attenuated power as it goes through the medium.
2. **Scattering:** This is the process of deflecting light from its straight-line propagation path.

The factors affecting the light traveling through water could be described mathematically, as the light travels through the medium, the radiance E at a position r can be described by [46]:

$$E(r) = E(0)e^c \tag{3.1}$$

Where c is the attenuation coefficient of the water.

The foundations for the widely used image formation model were laid initially in [47]. Then, [48] extended the model and tested it on different sub-sea image-capturing systems. According to this model, underwater images could be treated as a combination of three components (See Fig. 3.1):

- Direct light E_d : light coming from the object directly without being scattered by the water,
- Forward-scattered light E_f : Object reflected light, scattered at a small angle.
- Backscatter component E_b : Light was not in the target scene but entered the camera due to floating particles.

The total irradiance E_T :

$$E_T = E_d + E_f + E_b \quad (3.2)$$

This model captures the main effects affecting underwater images and is the main model used in the literature for model-based restoration or synthetic data generation. While this model is not used in this work, it is worth mentioning here since the dataset in this work utilizes the imaging model described here in some ways.

3.2 Deep Learning

Deep learning is a subfield of machine learning. The primary goal of machine learning is to create systems capable of learning the relationships between data points to

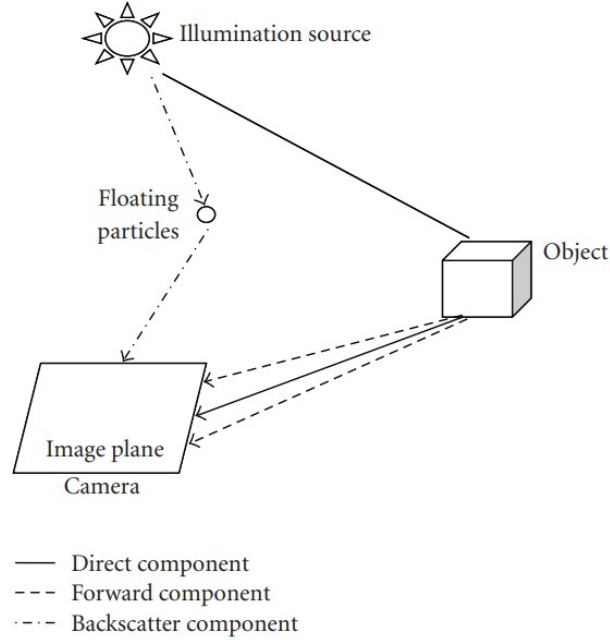


Figure 3.1: The three components of underwater imaging: direct component, forward component, and backward scatter component

determine an underlying function, which can then be used to make predictions based on input data. Deep learning, a more advanced form of machine learning, excels in learning complex functions without the need for prior feature engineering, leading to superior accuracy compared to traditional methods and machine learning models. This capability makes deep learning particularly effective for data with numerous features, such as images. To address various tasks, researchers have developed different deep learning architectures, which vary in their building blocks, depth, and details [49], [50].

3.2.1 Convolutional Neural Networks (CNNs)

CNN is one category of deep learning algorithms that is primarily used for processing structured grid data (i.e., data containing important spatial information) like images.

It operates by convolving stacked filters over input images to automatically learn different abstractions of spatial features, such a mechanism enables the CNN to detect patterns like shapes, surfaces, and edges necessary for the recognition and classification tasks. Though CNNs were introduced in the late 90s in [51], they showed their significant importance in many CV tasks, especially after presenting the new variants like ResNet introduced in [52] and MobileNet presented in [53].. The vanilla architecture of CNNs is shown in Fig. 3.2. Other variants build upon this model with some additions to enhance certain aspects. For example, the MobileNet model utilizes a pointwise convolution to combine information across the various channels.

CNNs (Convolutional Neural Networks) are deep learning algorithms primarily used for processing structured grid data, such as images. They work by applying filters over input images to automatically learn spatial features like shapes, surfaces, and edges, which are crucial for recognition and classification tasks. Introduced in the late 90s [51], CNNs have become essential in computer vision, with notable advancements such as ResNet [52] and MobileNet [53]. The basic CNN architecture (For image classification) is shown in Fig. 3.2. CNNs are also highly effective for image restoration tasks such as denoising, dehazing, and deraining. By learning to identify and remove noise, haze, or rain from images, CNNs can significantly improve image quality. Specialized CNN architectures can be trained to enhance images by reducing artifacts and recovering lost details, making them invaluable in applications like photography, surveillance, and autonomous driving.

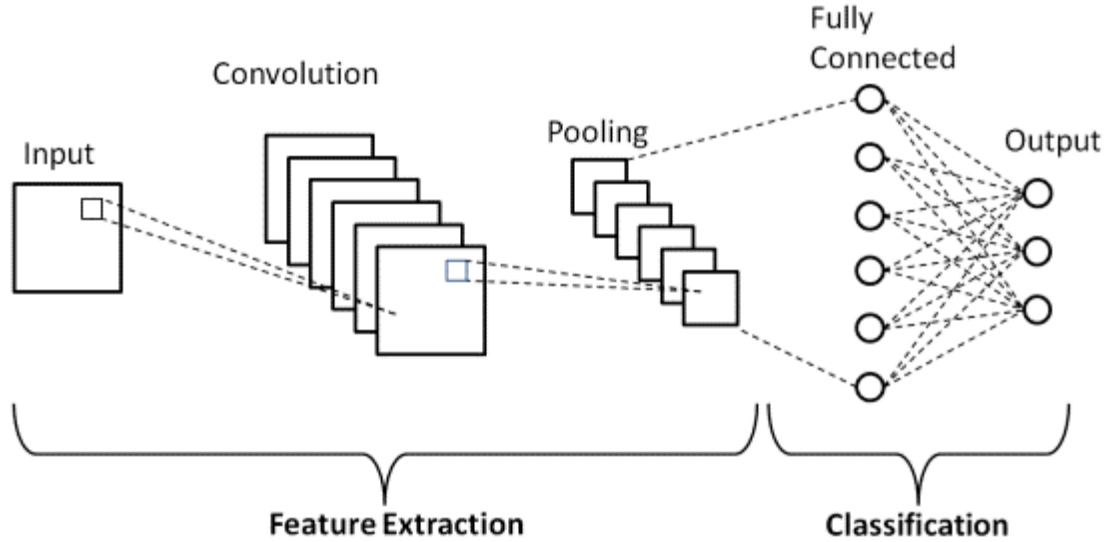


Figure 3.2: Overview of the Basic CNN Architecture.

3.2.2 Transformers

Architectures based on the Attention mechanism such as Transformers, have given SOTA results across different natural language processing (NLP) tasks. Typically, a model is first pre-trained on a large corpus and then fine-tuned on smaller task-specific datasets. In the UIR problem and computer vision in general, convolutional neural network architectures are still the model of choice. Initial works investigating using the Transformer architecture in computer vision first tried using hybrid CNN-Self attention architectures, Others considered removing convolutions entirely. These models, however, either did not scale well or were not as good as CNNs. In [54], the authors experimented with using the original Transformer architecture directly on images. To transform images to a form applicable to the Transformer, the authors split images into square patches. Then, each patch is transformed into a one-dimensional vector. The sequence of those embedded patches is fed to the Transformer as an input.

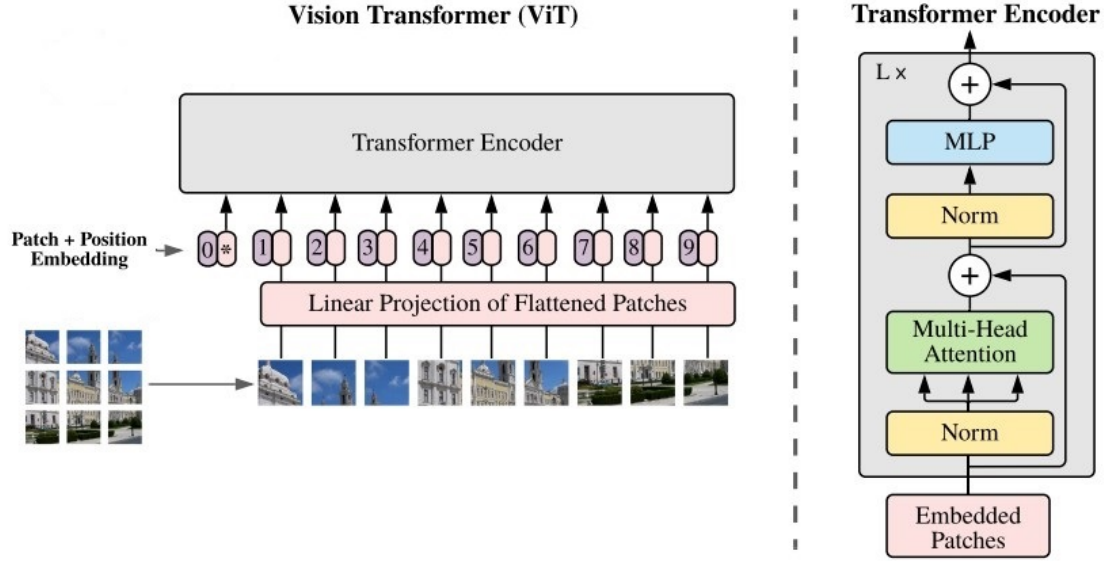


Figure 3.3: Overview of ViT Architecture

Each patch is treated in similar manner to tokens (words) in NLP tasks.

The model proposed by the authors in [54], known as ViT, is shown in Fig.3.3. The original Transformer expects a sequence of 1D tokens as input. To address this, a 2D image, $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$, is reshaped to a sequence of 1D patches, $\mathbf{x}_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$, where (H, W) is the size of the input image, C is the number of channels, (P, P) is the size of each patch, and $N = HW/P^2$ is the resulting number of patches, which also represents the input sequence length for the Transformer. After the patchification step, the patches are transformed into a latent representation using a linear layer. Since the Transformer processes all input tokens (patches) at the same time, position information is lost. The authors add learnable what is called a position embedding to each patch to keep position information, so this position embedding works as a position identifier. After this step, the data is ready to flow through the network layers. The Transformer architecture is based mainly on the self-attention mechanism

which tries to understand how the different tokens in the input sequence are related to each other.

To calculate the attention layer output, each patch from the input image (1D vector) is projected into three different vectors: Query, Key, and Value. Those vectors are abstractions used for simulating attention. For example, for the patch $\mathbf{x}_p$, we have:

$$\mathbf{q}_p = \mathbf{x}_p \times \mathbf{W}^q \quad (3.3)$$

$$\mathbf{k}_p = \mathbf{x}_p \times \mathbf{W}^k \quad (3.4)$$

$$\mathbf{v}_p = \mathbf{x}_p \times \mathbf{W}^v \quad (3.5)$$

Where $\mathbf{W}^q$, $\mathbf{W}^k$, and $\mathbf{W}^v$ are learned matrices that are used to calculate the corresponding query, key, and value vectors for each patch. Essentially, the attention for the patch $\mathbf{x}_p$ is calculated by the weighted sum of all the value vectors generated from the image patches (Including its own value vector). To generate the weights scaling the value vectors, scores are first calculated by the dot product of the query vector of the patch for which the attention is being calculated and the key vectors of all the other patches (Including its own key vector), then those scores are fed into a soft-max layer to ensure they all sum up to one. So for the patch $\mathbf{x}_1$ for example, scores are first calculated as follows (Assuming we have three patches only):

$$\mathbf{s}_{11} = \mathbf{q}_1 \cdot \mathbf{k}_1 \quad (3.6)$$

$$\mathbf{s}_{12} = \mathbf{q}_1 \cdot \mathbf{k}_2 \quad (3.7)$$

$$\mathbf{s}_{13} = \mathbf{q}_1 \cdot \mathbf{k}_3 \quad (3.8)$$

Then, the corresponding weight for each score is calculated through:

$$\mathbf{a}_{1i} = \frac{\exp(\mathbf{s}_{1i})}{\sum_{j=1}^3 \exp(\mathbf{s}_{1j})} \quad (3.9)$$

Where $\mathbf{a}_{1i}$ represents the attention weight for the i -th key with respect to the query $\mathbf{q}_1$. These weights are then used to compute the weighted sum of the value vectors to obtain the attention output $\mathbf{z}_1$, for the patch $\mathbf{x}_1$:

$$\mathbf{z}_1 = \sum_{i=1}^3 \mathbf{a}_{1i} \mathbf{v}_i \quad (3.10)$$

In this equation, $\mathbf{v}_i$ represents the value vectors corresponding to the patches, and the weighted sum of these value vectors, scaled by their respective attention weights $\mathbf{a}_{1i}$, gives the final attention output $\mathbf{z}_1$ for the patch $\mathbf{x}_1$.

Transformers use multi-head self-attention, which means each layer will learn multiple attention relationships between the different tokens in the input. In practice, attention is computed on a set of queries simultaneously, packed into a matrix $\mathbf{Q}$. The keys and values are also packed together into matrices $\mathbf{K}$ and $\mathbf{V}$:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \quad (3.11)$$

In this equation:

- $\mathbf{Q}$ (Query matrix): Contains the query vectors for all the patches in the input.
- $\mathbf{K}$ (Key matrix): Contains the key vectors for all the patches in the input.

- $\mathbf{V}$ (Value matrix): Contains the value vectors for all the patches in the input.
- $\sqrt{d_k}$: The square root of the dimensionality of the key vectors. This scaling factor is used to stabilize the gradients during training [55].

This compact form efficiently computes the attention for multiple tokens in parallel, enabling the transformer to capture complex relationships in the input data. Since there is no element-wise non-linearity, self-attention is simply performing a re-averaging of the value vectors. To address this, the attention output is fed through a feed-forward layer, thus providing non-linear activation.

3.2.3 Encoder-Decoder (U-Net)

The U-shaped architecture, commonly referred to as U-Net, was introduced in [56] as an advanced version of the encoder-decoder networks [57]. The concept behind the encoder-decoder architecture involves two main parts: the encoder and the decoder. The encoder compresses the input image into a lower-dimensional representation by progressively reducing the spatial dimensions while increasing the feature depth. This compressed representation captures the essential features of the input image. The decoder then reconstructs the image from this representation by progressively up-sampling and combining it with features from the corresponding encoder layers via skip connections. These skip connections, introduced by U-Net, transfer important details directly from the encoder to the decoder, preserving spatial hierarchies and significantly improving performance in tasks like image segmentation. The basic U-Net architecture, shown in Fig. 3.4, does not include attention mechanisms found in some other encoder-decoder models, but it remains highly effective for various com-

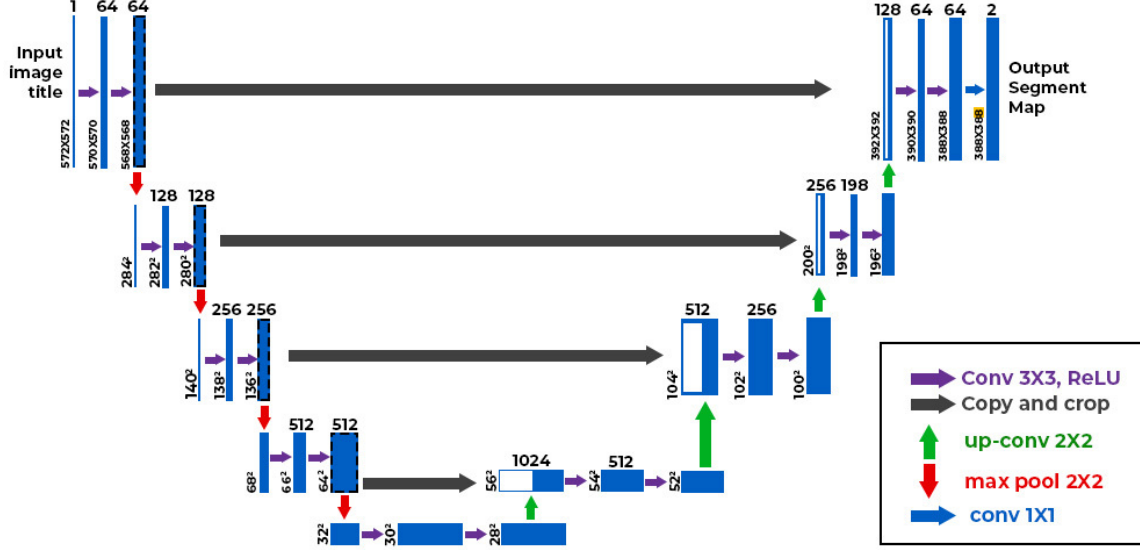


Figure 3.4: Overview of the U-Net Architecture.

puter vision applications.

3.2.4 Diffusion Models

Diffusion Models, displayed in Fig. 3.5, transform standard Gaussian distribution samples into empirical data distribution samples via a recurrent denoising process. Conditional variants of these models condition the denoising process on the input so that the output is not a randomly sampled image from the learned distribution and rather, depending on the task, a transformed version of the input -Usually an un-degraded restored image.

Diffusion Models consist of a forward diffusion process and a backward denoising stage. The forward operation progressively incorporates Gaussian noise into the input $y_0 = y$ over T iterations [58]:

$$q(y_t|y_{t-1}) = \mathcal{N}(y_t; \sqrt{1 - \beta_t}y_{t-1}, \beta_t\mathbf{I}) \quad (3.12)$$

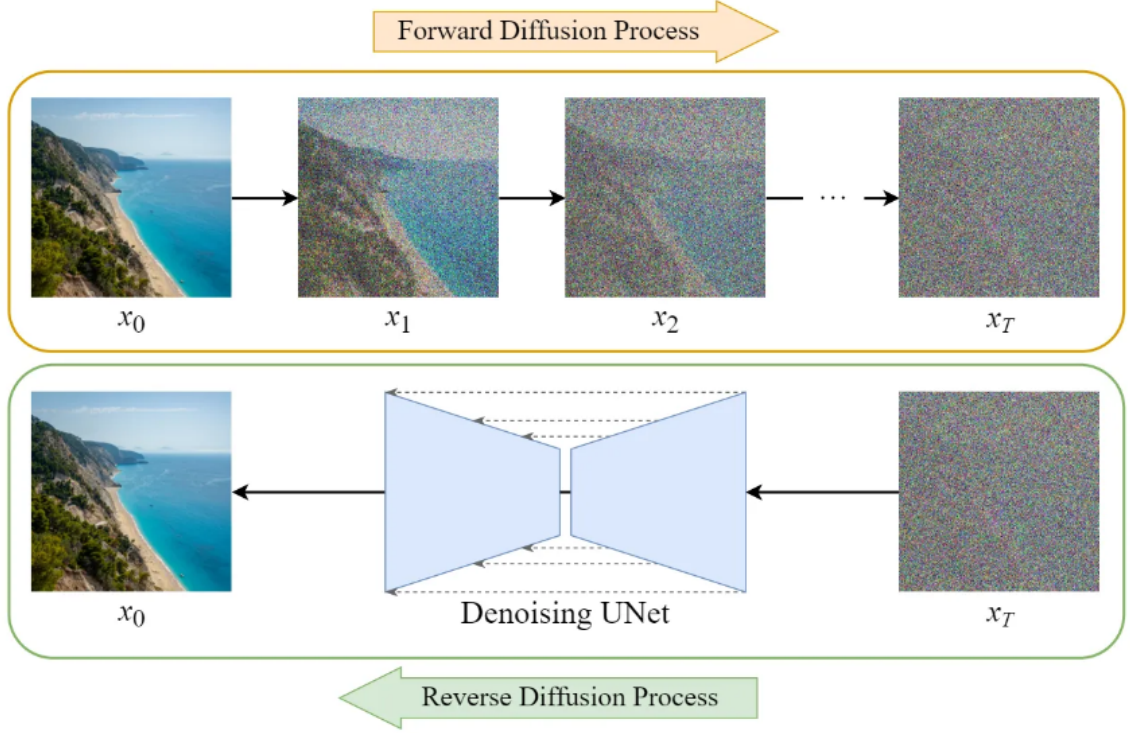


Figure 3.5: A general overview of the Diffusion Model Architecture.

where β_t represents the noise scheduler at each timestep. By the final step T , y_T resembles Gaussian noise. We can directly find y_t without going over all the iterations:

$$\alpha_t = 1 - \beta_t \quad (3.13)$$

$$\overline{\alpha}_t = \prod_{s=1}^t \alpha_s \quad (3.14)$$

$$q(y_t|y_0) = \mathcal{N}(y_t, \sqrt{\overline{\alpha}_t}y_0, (1 - \overline{\alpha}_t)I) \quad (3.15)$$

As for the backward denoising process, a denoising network learns to go from the Gaussian distribution to the empirical distribution of the dataset on which the network

is being trained.

Fig. 3.6 shows how diffusion models differ from other generative approaches.

Generative Adversarial Networks (GANs) consist of a generator and a discriminator trained in an adversarial manner. The generator creates data samples from random noise, while the discriminator attempts to distinguish between real and generated data. The objective is for the generator to produce data that the discriminator cannot differentiate from real data. While GANs are capable of generating high-quality images, they often suffer from instability during training and can be difficult to optimize due to issues like mode collapse, where the generator produces a limited variety of samples.

Variational Autoencoders (VAEs) use an encoder to compress input images into a latent space and a decoder to reconstruct the images from this compressed representation. The encoder maps the input to a probability distribution, and the decoder samples from this distribution to generate new images. This approach ensures that the latent space follows a standard distribution, making it easier to generate new, similar images.

Diffusion Models present a different approach by gradually adding noise to data through a forward process and learning a reverse process to denoise and reconstruct the data. This method is robust and effective, leading to high-quality image generation and restoration. Diffusion models address some of the shortcomings of the previous methods, such as the training instability in GANs and the blurriness in VAEs. They also do not require the complex invertibility constraints of flow-based models, making them more straightforward to implement while still achieving impressive results in

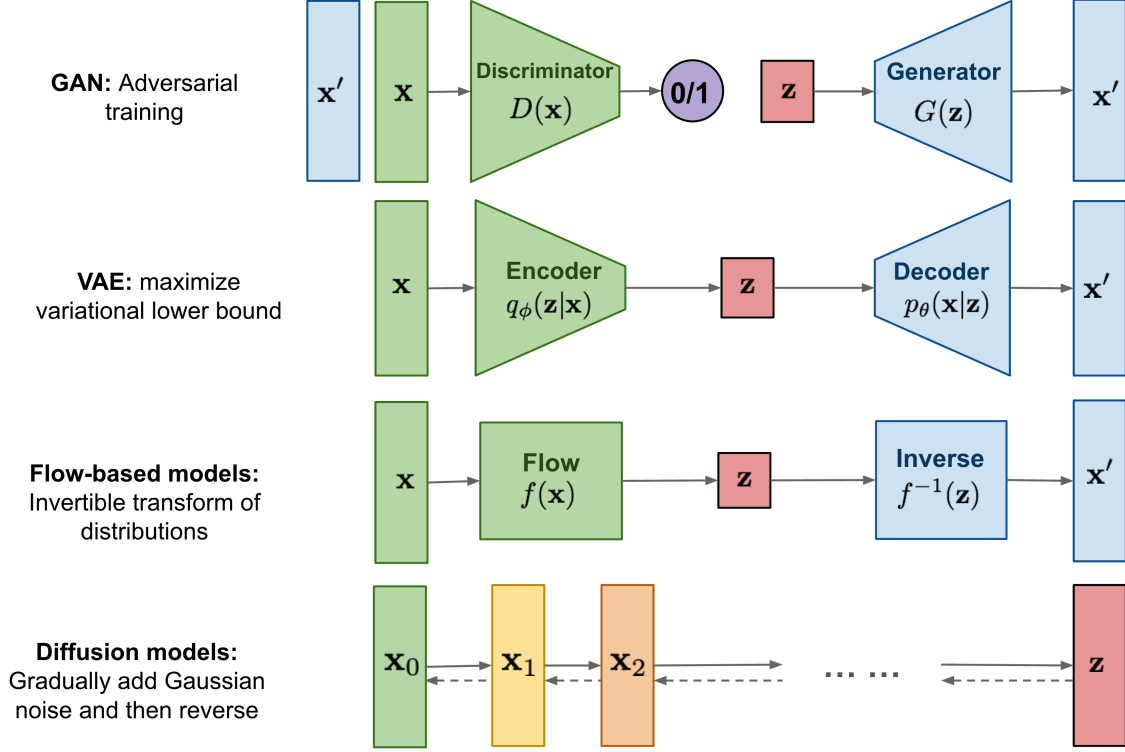


Figure 3.6: A comparison between Diffusion Models and other generative models.

terms of image quality and diversity.

3.3 Summary

This section provided an overview of key concepts and methodologies relevant to image restoration. It begins with an introduction to image restoration, highlighting its importance in fields like medical imaging, satellite imaging, and surveillance. The section then explains underwater light propagation, describing the roles of absorption and scattering. Next, the section covers deep learning, focusing on Transformers, and detailing their use in computer vision through mechanisms like self-attention. The encoder-decoder architecture, specifically U-Net. Finally, Diffusion Models are

introduced, emphasizing how they compare to other generative methods.

CHAPTER 4

RESULTS OF APPLYING THE DIFFUSION MODEL ON THE DENOISING AND DERAINING TASKS

In [59], the authors introduce PromptIR, a Transformer based U-Net architecture for image restoration, shown in Fig.4.1, the model shows promising results for handling various degradation types without requiring specific models for each case. It leverages prompt modules to dynamically adapt to different degradations and achieve good results in denoising, deraining, and dehazing tasks, this model is used in this work as a denoiser. In this chapter, we aim to test the efficacy of diffusion models on the denoising and dehazing tasks.

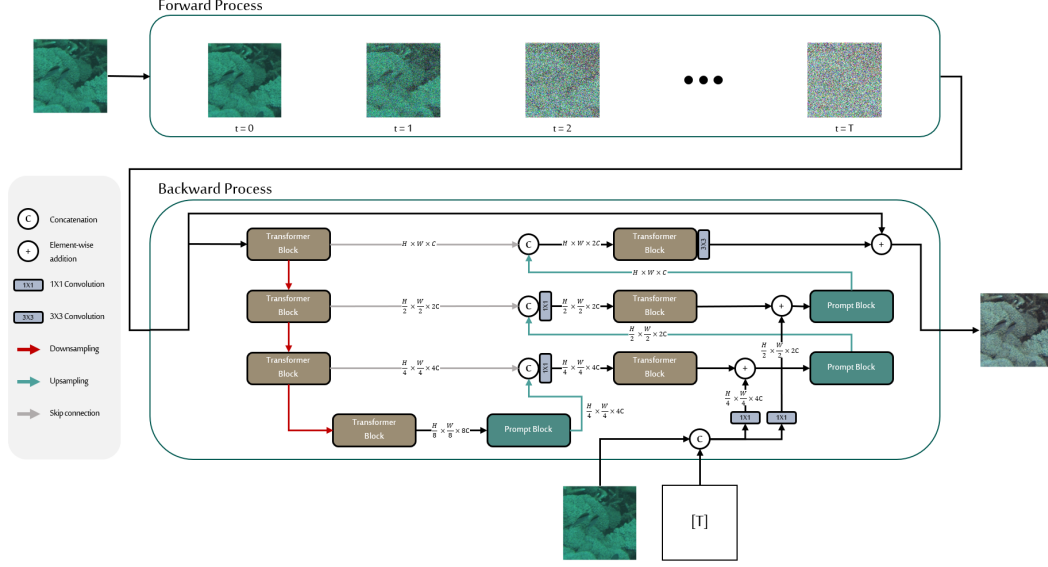


Figure 4.1: Overview of our Approach

4.1 Baseline Details

The PromptIR model comprises 4 levels. The prompt blocks are inserted after each decoder stage, adding up to a total of 3 prompt blocks. The baseline is trained with a batch size of 32. The model was trained on an L1 loss, with an Adam optimizer.

4.2 Datasets

For training the model on the image denoising task two datasets were used, BSD400 and WED. From these two datasets, images were degraded by introducing Gaussian noise with different noise levels $\sigma \in \{15, 25, 50\}$, where σ refers to the standard deviation of the Gaussian noise that is introduced to degrade the images. Testing is performed on the BSD68 dataset. As for the deraining task, the Rain100L [60] dataset was used to train and test the diffusion model, the same dataset used in the PromptIR paper. This dataset contains 300 clean-rainy image pairs.

Table 4.1: Comparing our model against the PromptIR baseline

Methods	Denoising (BSD68)			Deraining (Rain100L)
	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	
PromptIR (Trained only on denoising)	34.34 /0.938	31.71 /0.898	28.49/ 0.813	-
PromptIR (Trained only on deraining)	-	-	-	37.03/ 0.9786
All in one PromptIR (Baseline)	33.98/0.933	31.31/0.888	28.06/0.799	36.37/0.972
Our model	34.12/ 0.942	31.53/ 0.911	28.62 /0.653	37.43 /0.947

Table 4.2: Comparing our model against state of the art on Denoising and Deraining

Methods	Denoising (BSD68)			Deraining (Rain100L)
	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	
BRDNet [61]	32.26/0.898	29.76/0.836	26.34/0.836	27.42/0.895
LPNet [62]	26.47/0.7782	24.77/0.748	21.26/0.552	24.88/0.784
FDGAN [63]	30.25/0.910	28.81/0.868	26.43/0.776	29.89/0.933
MPRNet [64]	33.54/0.927	30.89/0.880	27.56/0.779	33.57/0.954
DL [65]	33.05/0.914	30.41/0.861	26.90/0.740	32.62/0.931
AirNet [66]	33.92/0.933	31.26/0.888	28.00/0.797	34.90/0.967
PromptIR [59]	33.98/0.933	31.31/0.888	28.06/0.799	36.37/0.972
Ours	34.12/0.942	31.53/0.911	28.62 /0.653	37.43 /0.947

4.3 Results and Discussion

As can be seen in Table 4.1, our PromptIR Diffusion Model outperforms the baseline in all of the noise levels tested in the original PromptIR paper. This result might be expected as the way we train the Diffusion Model, where we expose it to different levels of noise, will expose the model to a larger dataset, and thus the model will be better at the task. This result suggests that using PromptIR in diffusion models enhances the performance of such models on reconstruction tasks and thus motivates the further investigation of this approach on the underwater reconstruction task. One other interesting observation is when comparing the results of the diffusion model against a model that was trained only on the denoising task. As seen in the table, our model is outperformed by the model trained on a single task, this also makes sense

Groundtruth	Input	AirNet	PromptIR	Ours
				
	23.24/0.63	24.43/0.79	35.17/0.95	39.17/0.98
				
	21.35/0.58	25.81/0.89	33.50/0.82	38.9/0.95
				
	22.41/0.69	25.01/0.76	33.95/0.86	39.85/0.97
				
	24.32/0.7	26.45/0.83	37.70/0.95	38.95/0.98

Figure 4.2: Visual comparisons between our model and the literature on the Denoising task along with corresponding PSNR/SSIM values

as the denoiser of our diffusion model was trained on three different tasks including denoising, which makes it a more complex training setup, and results in a solution that is further away from an optimal solution for denoising. Surprisingly though, our generalist diffusion model still managed to outperform the specialist PromptIR model in the $\sigma = 50$ noise level, qualitative results are presented in 4.2. From the figure, it can be seen how the Diffusion model gives good results where non-Diffusion model solutions fail. Notice how in the second and third images, all other models let some visible noise remain whereas the Diffusion Model solution is better at removing them. This pattern is consistent across various types of images, including animals, landscapes, and underwater scenes, demonstrating the robustness of our model.

The same can be said for the deraining task, notice how the model outperforms

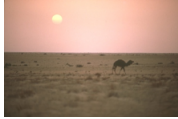
Groundtruth	Input	AirNet	PromptIR	Ours
				
	22.02/0.35	34.83/0.81	39.45/0.90	42.67/0.86
				
	20.72/0.27	33.00/0.83	34.81/0.84	37.61/0.82
				
	20.8/0.31	31.21/0.84	32.17/0.85	34.45/0.86
				
	23.8/0.37	31.09/0.78	32.39/0.79	33.89/0.82

Figure 4.3: Visual comparisons between our model and the literature on the Deraining task along with corresponding PSNR/SSIM values

both the baseline and the model trained on deraining alone, which is a result that was not expected. This further proves the efficacy of diffusion models on image restoration tasks. Qualitative results on the deraining task are shown in 4.3. notice again how the other techniques fail to retrieve the entirety of the true scene they let some rain streaks pass through, especially the second and third images. We have also benchmarked our model against other models available in the literature as can be seen in Table 4.2. Our model outperforms all the benchmarked methods in both denoising and deraining.

4.4 Summary

In this chapter, we investigate the effectiveness of diffusion models on the tasks of denoising and deraining. Using the PromptIR Transformer-based U-Net architecture as our baseline, we trained our model on datasets such as BSD400 and WED for denoising, and Rain100L for deraining. Our results demonstrate that our PromptIR Diffusion Model outperforms the baseline model in all tested noise levels for the denoising task and achieves superior performance compared to other state-of-the-art methods in both denoising and deraining. Notably, our model shows robustness across various image types, effectively removing noise and rain streaks where other models fail. These findings highlight the potential of diffusion models for enhancing image restoration tasks and justify further exploration of this approach for underwater image reconstruction.

CHAPTER 5

DIFFUSION MODELS FOR IMAGE RECONSTRUCTION

5.1 Diffusion Models for Underwater Image Reconstruction

As mentioned earlier, a diffusion model is employed in this work to address the underwater image reconstruction task. For the denoising process of the training, given a clear image (Groundtruth), $\mathbf{y}$, a degraded version of the image, $\tilde{\mathbf{y}}$, is generated following the forward noising process described in the previous section. Then a Transformer based UNet model, f_{θ} , is trained on the loss function:

$$\|f_{\theta}(\tilde{\mathbf{y}}, \gamma) - \epsilon\|_1^1, \quad (5.1)$$

where ϵ is the noise affecting the degraded image.

Since the transformer is reported to be data hungry, a pre-trained encoder is used

as a backbone of the model, the encoder was trained on 4 different image restoration tasks [59]. Hence, only part of the decoder is trained in this work, while the encoder parameters are kept constant with no change. The model was trained for 70 epochs.

The UIEB (Underwater Image Enhancement Benchmark) [67] dataset was used in this work. The UIEB contains 890 underwater images with their corresponding "reference" images. The reference images were constructed using the following methodology: For each underwater image, a number of proposed reference images are produced using different underwater image reconstruction methods. Then, several subjects were asked to choose the best image, out of the proposed images, that better captures the underwater scene while removing the color shift effects. While training on this dataset might mean that the model is only as good as the techniques used to produce the references in the UIEB, it is used here because multiple works have used the dataset and published the models that were trained on it.

In our approach, we subject each input image—sourced from the degraded, underwater domain—to an initial color tone analysis to find the dominant hues, which are predominantly bluish or greenish owing to the characteristic light absorption and scattering underwater. The outcome of this analysis directs the application of a suitable color shift to our target images, instituting a blue or green shift aligned with the detected dominant hue. Subsequently, we introduce a vertical gradient to mimic the depth-based light behavior in aquatic environments. We proceed to subtly reduce contrast and saturation to mirror the inherent optical properties of underwater scenes more accurately. A blur effect is also applied, emulating the visual interference caused by particulate matter in the water—such as silt and plankton—that typically leads to

Table 5.1: Comparing our approach against different underwater reconstruction methods

Methods	Datasets			
	UIEB		Test-60	
	PSNR	SSIM	UIQM	UCIQE
ACDC [68]	18.04	0.8212	1.03	0.38
Rank1 [69]	18.16	0.8443	1.36	0.40
MMLE [70]	15.48	0.8242	1.53	0.38
UWCNN [71]	13.43	0.6253	0.63	0.29
WaterNet [67]	18.35	0.8397	1.63	0.41
Funie [72]	21.53	0.8512	1.61	0.38
Ucolor [35]	20.63	0.8495	1.63	0.34
TACL [73]	20.41	0.8478	1.58	0.41
MetaUE [74]	22.01	0.8607	1.65	0.42
PromptIR [59]	21.37	0.7827	1.61	0.37
Our model	22.90	0.8724	2.73	0.41

light scattering and absorption.

The aforementioned modifications effectively serve as a form of data augmentation, adding layers of complexity and underwater-like qualities to the synthetic degradations. Further blending these intentionally degraded images with actual degraded photographs from the Underwater Image Enhancement Benchmark (UIEB) dataset allows us to refine our augmentation process. This method not only enhances the training data but also potentially increases the robustness of the denoising model against various real-world underwater degradations.

5.2 Results and Analysis

Multiple methods are used to benchmark the model used in this work, including the PromptIR baseline. The models are benchmarked on two test sets, Test60 and UIEB.

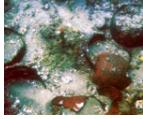
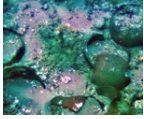
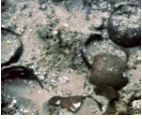
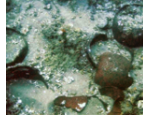
Groundtruth	Input	Rank1	TACL	MMLE	PromptIR	Ours
						
	13.91/0.42	14.22/0.54	21.83/0.87	15.47/0.86	17.63/0.79	22.87/0.88
						
	14.53/0.46	13.85/0.58	18.61/0.79	14.96/0.72	17.03/0.83	22.39/0.85
						
	14.21/0.39	14.3/0.55	18.93/0.85	15.51/0.88	17.83/0.82	22.32/0.85
						
	13.88/0.31	14.06/0.45	19.26/0.78	14.89/0.75	20.77/0.79	22.46/0.82

Figure 5.1: Visual comparisons between our model and state of the art on the UIEB dataset along with corresponding PSNR/SSIM values

The difference is that UIEB has reference images while Test60 does not have reference images. For the UIEB images, the average PSNR and SSIM are calculated and are shown in Table 5.1. As can be seen from the table, in terms of both metrics, the Diffusion Model outperforms the other techniques. The margin by which it outperforms them is not very high though. This could be attributed to the fact that the denoiser used in the Diffusion model is a Transformer based which processes images as patches. This introduces blocking artifacts that reduce SSIM and PSNR between the image and its reference.

As for Test60, since there are no reference images, full reference measures can not be used to assess the performance of each method on the test set. Two no-reference quality measures for the underwater image enhancement problem are used.

The underwater color image quality evaluation (UCIQE) [75] metric extracts the most relevant CIELab space statistical features that are representative of underwater image degradations, such as color shift and noise caused by suspended particles. The other metric is the underwater image quality measure (UIQM) [76]. This measure takes into consideration three attributes: colorfulness measure, sharpness, and contrast. These two measures are usually used in the literature to assess underwater reconstruction methods. As shown in Table 5.1, the diffusion model scores higher than the other methods in UIQM, while for UCIQE, the model outperforms all the other methods except the work in [74], which is still not that better than ours. A set of reconstructed images using the three techniques is visualized in Fig. 5.1. For the most part, the model does well in removing underwater image degradations. However, as can be seen from the first picture, it seems to struggle with images that lack depth, such as scenes with a subject in front of a close backdrop.

5.3 Summary

This chapter explores the application of diffusion models for UIR tasks. Using a transformer-based U-Net model trained, the study demonstrates the effectiveness of diffusion models in handling underwater image degradations. The model shows superior performance compared to other methods. Benchmarking against various techniques on the UIEB dataset, the diffusion model achieves higher PSNR, SSIM, and UIQM scores, proving its robustness and efficacy. Visual comparisons highlight the model’s ability, although it shows some limitations with images lacking depth.

CHAPTER 6

CONCLUSION

6.1 Summary

In this work, a transformer-based Diffusion Model was trained and tested on three different image restoration tasks: image denoising, image deraining, and underwater image reconstruction. Our model demonstrated significant improvements over existing methods. For underwater image reconstruction, our model achieved the highest PSNR and SSIM on the UIEB dataset and the best UIQM on the Test-60 dataset, indicating superior fidelity and perceptual quality. In image denoising, our model outperformed others across various noise levels ($\sigma = 15, 25, 50$) on PSNR. For image deraining, our model surpassed other state-of-the-art methods on the Rain100L dataset in PSNR and SSIM. These results demonstrate the model’s effectiveness in restoring image quality across different scenarios.

6.2 Future Work

For future work, one potential direction is to develop a synthetic dataset where the same scene is used to generate multiple degraded images simulating different water environments. This approach would allow us to fully utilize the capabilities of diffusion models by exposing them to a wide range of degradation scenarios. Additionally, exploring the use of a pre-trained model trained on a pre-text task more relevant to the underwater image reconstruction (UIR) problem, such as image colorization, could provide a strong initialization and potentially lead to better performance.

REFERENCES

- [1] P. Tang, L. Li, Y. Xue, M. Lv, Z. Jia, and H. Ma, “Real-world underwater image enhancement based on attention u-net,” *Journal of Marine Science and Engineering*, vol. 11, no. 3, 2023.
- [2] Y. Guan, X. Liu, Z. Yu, Y. Wang, X. Zheng, S. Zhang, and B. Zheng, “Fast underwater image enhancement based on a generative adversarial framework,” *Frontiers in Marine Science*, vol. 9, p. 964600, 2023.
- [3] V. P and J. M, “A survey on recent approaches in image reconstruction,” in *2020 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Thrissur, India, 2020, pp. 1–5.
- [4] X. Chu, Z. Fu, S. Yu, X. Tu, Y. Huang, and X. Ding, “Underwater image enhancement and super-resolution using implicit neural networks,” in *2023 IEEE International Conference on Image Processing (ICIP)*, 2023, pp. 1295–1299.
- [5] Z. Wang, D. Zhou, Z. Li, Z. Yuan, and C. Yang, “Underwater image enhancement via adaptive color correction and stationary wavelet detail enhancement,” *IEEE Access*, vol. 12, pp. 11 066–11 082, 2024.

- [6] M. Yang, J. Hu, C. Li, G. Rohde, Y. Du, and K. Hu, “An in-depth survey of underwater image enhancement and restoration,” *IEEE Access*, vol. 7, pp. 123 638–123 657, 2019.
- [7] R. Hummel, “Image enhancement by histogram transformation,” *Computer Graphics and Image Processing*, vol. 6, no. 2, pp. 184–195, 1977. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0146664X77800117>
- [8] G. Karam, Z. Abood, and R. Saleh, “Enhancement of underwater image using fuzzy histogram equalization,” *International Journal of Applied Information Systems*, vol. 6, pp. 1–6, 12 2013.
- [9] R. Singh and M. Biswas, “Adaptive histogram equalization based fusion technique for hazy underwater image enhancement,” in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2016, pp. 1–5.
- [10] M. S. Hitam, E. A. Awalludin, W. N. J. H. W. Yussof, and Z. Bachok, “Mixture contrast limited adaptive histogram equalization for underwater image enhancement,” in *2013 International conference on computer applications technology (IC-CAT)*. IEEE, 2013, pp. 1–5.
- [11] Y.-C. Liu, W.-H. Chan, and Y.-Q. Chen, “Automatic white balance for digital still camera,” *IEEE Transactions on Consumer Electronics*, vol. 41, no. 3, pp. 460–466, 1995.
- [12] C. Li and X. Zhang, “Underwater image restoration based on improved background light estimation and automatic white balance,” in *2018 11th Interna-*

- tional Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2018, pp. 1–5.
- [13] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib, “Enhancing the low quality images using unsupervised colour correction method,” in *2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 1703–1709.
- [14] 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)*, 2014, pp. 4572–4576.
- [15] S. Zhang, T. Wang, J. Dong, and H. Yu, “Underwater image enhancement via extended multi-scale retinex,” *Neurocomputing*, vol. 245, pp. 1–9, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231217305246>
- [16] A. Mahiddine, J. Seinturier, D. P. J. M. Boï, P. Drap, D. Merad, and L. Long, “Underwater image preprocessing for automated photogrammetry in high turbidity water: An application on the arles-rhone xiii roman wreck in the rhodano river, france,” in *2012 18th International Conference on Virtual Systems and Multimedia*, 2012, pp. 189–194.
- [17] B. Yin, H. Lei, Z. Wei, and L. Gui, “Reconstruction with underwater stereo vision,” in *Proceedings of the International Conference on Logistics, Engineering, Management and Computer Science*, 2014.

- [18] Z. M. Abood, G. S. Karam, and R. N. Saleh, “Enhancement of underwater image using fuzzy histogram equalization,” *International Journal of Applied Information Systems*, vol. 6, pp. 1–6, 2013.
- [19] J. Y. Chiang and Y.-C. Chen, “Underwater image enhancement by wavelength compensation and dehazing,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.
- [20] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, “Automatic red-channel underwater image restoration,” *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1047320314001874>
- [21] P. L. Drews, E. R. Nascimento, S. S. Botelho, and M. F. Montenegro Campos, “Underwater depth estimation and image restoration based on single images,” *IEEE Computer Graphics and Applications*, vol. 36, no. 2, pp. 24–35, 2016.
- [22] 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.
- [23] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1956–1963.

- [24] P. Drews Jr, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, “Transmission estimation in underwater single images,” in *2013 IEEE International Conference on Computer Vision Workshops*, 2013, pp. 825–830.
- [25] J. Y. Chiang and Y.-C. Chen, “Underwater image enhancement by wavelength compensation and dehazing,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.
- [26] H. Liu and L.-P. Chau, “Underwater image restoration based on contrast enhancement,” in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, 2016, pp. 584–588.
- [27] 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.
- [28] C. Li, S. Anwar, and F. Porikli, “Underwater scene prior inspired deep underwater image and video enhancement,” *Pattern Recognition*, vol. 98, p. 107038, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320319303401>
- [29] 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, 2020.
- [30] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, “Watergan: Un-supervised generative network to enable real-time color correction of monocular

- underwater images,” *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 387–394, 2018.
- [31] S. Anwar and C. Li, “Diving deeper into underwater image enhancement: A survey,” *arXiv preprint arXiv:[insert arXiv ID here]*, 2019.
- [32] 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.
- [33] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, “Uformer: A general u-shaped transformer for image restoration,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 17 683–17 693.
- [34] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, “Pre-trained image processing transformer,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 12 294–12 305.
- [35] 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.
- [36] Z. Guo, D. Guo, Z. Gu, H. Zheng, B. Zheng, and G. Wang, “Unsupervised underwater image clearness via transformer,” in *OCEANS 2022 - Chennai*, 2022, pp. 1–4.

- [37] T. Ren, H. Xu, G. Jiang, M. Yu, X. Zhang, B. Wang, and T. Luo, “Reinforced swin-convs transformer for simultaneous underwater sensing scene image enhancement and super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.
- [38] L. Peng, C. Zhu, and L. Bian, “U-shape transformer for underwater image enhancement,” 2021. [Online]. Available: <https://arxiv.org/abs/2111.11843>
- [39] Z. Huang, J. Li, Z. Hua, and L. Fan, “Underwater image enhancement via adaptive group attention-based multiscale cascade transformer,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–18, 2022.
- [40] K. Sun, F. Meng, and Y. Tian, “Multi-level wavelet-based network embedded with edge enhancement information for underwater image enhancement,” *Journal of Marine Science and Engineering*, vol. 10, no. 7, 2022. [Online]. Available: <https://www.mdpi.com/2077-1312/10/7/884>
- [41] Y. Wu, G. Wang, Z. Wang, Y. Yang, T. Li, P. Wang, C. Li, and H. T. Shen, “Reco-diff: Explore retinex-based condition strategy in diffusion model for low-light image enhancement,” 2023.
- [42] X. Guo, Y. Li, and H. Ling, “Lime: Low-light image enhancement via illumination map estimation,” *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.
- [43] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, “From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement,” in *2020*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3060–3069.
- [44] S. Lu, F. Guan, H. Zhang, and H. Lai, “Underwater image enhancement method based on denoising diffusion probabilistic model,” *Journal of Visual Communication and Image Representation*, vol. 96, p. 103926, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1047320323001761>
- [45] D. Du, E. Li, L. Si, F. Xu, J. Niu, and F. Sun, “Uiedp:underwater image enhancement with diffusion prior,” 2023.
- [46] R. Schettini and S. Corchs, “Underwater image processing: State of the art of restoration and image enhancement methods,” *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 746052, Apr 2010. [Online]. Available: <https://doi.org/10.1155/2010/746052>
- [47] B. McGlamery, “A computer model for underwater camera systems,” in *Ocean Optics VI*, vol. 208. SPIE, 1980, pp. 221–231.
- [48] J. S. Jaffe, “Computer modeling and the design of optimal underwater imaging systems,” *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101–111, 1990.
- [49] L. Alzubaidi, J. Zhang, A. J. Humaidi, and et al., “Review of deep learning: concepts, cnn architectures, challenges, applications, future directions,” *J Big Data*, vol. 8, no. 1, p. 53, 2021.

- [50] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, p. 85–117, Jan. 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.neunet.2014.09.003>
- [51] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [52] 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*, 2016, pp. 770–778.
- [53] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” 2017.
- [54] A. Kolesnikov, A. Dosovitskiy, D. Weissenborn, G. Heigold, J. Uszkoreit, L. Beyer, M. Minderer, M. Dehghani, N. Houlsby, S. Gelly, T. Unterthiner, and X. Zhai, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021.
- [55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023.
- [56] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015.

- [57] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” 2016.
- [58] C. Saharia, W. Chan, H. Chang, C. A. Lee, J. Ho, T. Salimans, D. J. Fleet, and M. Norouzi, “Palette: Image-to-image diffusion models,” 2022.
- [59] V. Potlapalli, S. W. Zamir, S. H. Khan, and F. Shahbaz Khan, “Promptir: Prompting for all-in-one image restoration,” in *Advances in Neural Information Processing Systems*, A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 71 275–71 293. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/e187897ed7780a579a0d76fd4a35d107-Paper-Conference.pdf
- [60] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, “Learning texture transformer network for image super-resolution,” 2020.
- [61] C. Tian, Y. Xu, and W. Zuo, “Image denoising using deep cnn with batch renormalization,” *Neural Networks*, vol. 121, pp. 461–473, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608019302394>
- [62] H. Gao, X. Tao, X. Shen, and J. Jia, “Dynamic scene deblurring with parameter selective sharing and nested skip connections,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3843–3851.
- [63] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, “Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing,” *Proceedings*

- of the *AAAI Conference on Artificial Intelligence*, vol. 34, pp. 10 729–10 736, 04 2020.
- [64] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, “Multi-stage progressive image restoration,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14 816–14 826.
- [65] Q. Fan, D. Chen, L. Yuan, G. Hua, N. Yu, and B. Chen, “A general decoupled learning framework for parameterized image operators,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 33–47, 2021.
- [66] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, “All-in-one image restoration for unknown corruption,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 17 431–17 441.
- [67] 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, 2020.
- [68] W. Zhang, Y. Wang, and C. Li, “Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement,” *IEEE Journal of Oceanic Engineering*, vol. 47, no. 3, pp. 718–735, 2022.
- [69] J. Liu, R. W. Liu, J. Sun, and T. Zeng, “Rank-one prior: Real-time scene recovery,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8845–8860, 2023.

- [70] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, “Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement,” *IEEE Transactions on Image Processing*, vol. 31, pp. 3997–4010, 2022.
- [71] C. Li, S. Anwar, and F. Porikli, “Underwater scene prior inspired deep underwater image and video enhancement,” *Pattern Recognition*, vol. 98, p. 107038, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320319303401>
- [72] 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.
- [73] R. Liu, Z. Jiang, S. Yang, and X. Fan, “Twin adversarial contrastive learning for underwater image enhancement and beyond,” *IEEE Transactions on Image Processing*, vol. 31, pp. 4922–4936, 2022.
- [74] Z. Zhang, H. Yan, K. Tang, and Y. Duan, “Metaue: Model-based meta-learning for underwater image enhancement,” 2023.
- [75] M. Yang and A. Sowmya, “An underwater color image quality evaluation metric,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.
- [76] 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, 2016.

VITAE

- Name: Hashem Shullar
- Nationality: Syrian
- Date of Birth: 09/09/1998
- Email: *hashem.shullar@gmail.com*
- Permenant Address: 6920 Prince Salman Ibn Muhammad, 13321 Sahafah, Riyadh
- Research: Interested in the areas of artificial intelligence, machine learning, and optimization. Throughout my Bachelor's and Master's studies, I have worked on a variety of projects, including Energy-Efficient Active Federated Learning on Non-IID Data, Monocular-Based Collision Avoidance System for Unmanned Aerial Vehicles, Optimal Real-Time Pricing-Based Scheduling in Home Energy Management Systems Using Genetic Algorithms, and Object-Aware Adaptive Image Retargeting via Importance Map Fusion. Those projects were published in IEEE conferences and journals.