Unsupervised Cycle GAN based Homogeneous and Non-homogeneous Image Dehazing

by

Jaymin Vekariya
202111015

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

July, 2023

Declaration

I hereby declare that

i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,

ii) due acknowledgment has been made in the text to all the reference material used.

Jaymin Vekariya

Certificate

This is to certify that the thesis work entitled Unsupervised Cycle GAN based Homogeneous and Non-homogeneous Image Dehazing has been carried out by Jaymin Vekariya for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my supervision.

Dr. Srimanta Mandal
Thesis Supervisor

Scanned with CamScanner

Acknowledgments

I am extremely thankful to my supervisor, Dr. Srimanta Mandal, for their invaluable guidance, consistent advise and support, and patience throughout my MTech Thesis. Their vast knowledge and extensive experience have helps me during my academic research.

I want to thank my peers and friends Yash Joshi, Harsh Savaliya, Gaurav Jha, Vinay Sheth, Vyom Shah, and Rohit Mishra who were always there in my ups and downs. Their continuous help and support have made my study and life at DA-IICT a wonderful and memorable experience.

Finally, I would like to express my gratitude to my parents and family for their love, care, and invaluable support throughout my life.

Contents

Abstract

Atmospheric phenomena like haze, fog, and smoke degraded image visibility. As a result, there is less contrast, colour distortion, etc. in the obtained image. In remote sensing, computer vision, and photography, haze reduction is greatly desired. in photography, dehazing can increase the visibility and quality of outdoor images and landscapes, making them more vibrant and appealing. In computer vision, dehazing can improve the quality of object detection task, recognition, and tracking algorithms, especially in outdoor and low-light environments. In remote sensing, dehazing can improve the quality of satellite and aerial images, making them more useful for environmental monitoring, disaster management, and urban planning. Dehazing removes haze, improves scene vision, and adjusts the airlight's colour change. There are two types of haze, homogeneous and non-homogeneous. To increase the dehaze quality for both homogeneous and nonhomogeneous haze, several techniques were used. Methods —classical, Deep learning-based, and GAN-based. For non-homogeneous haze, it is challenging to estimate the spread of haze. Due to the less availability of real world ground-truth images, many recent methods focus on the unsupervised approach to solve this issue. GAN, and cycle-GAN based unsupervised methods are highly used in this technique. But still, there is not any prominent unsupervised technique for non-homogeneous haze removal. This paper proposed the unsupervised cycle-GAN based approach, which has worked on both homogeneous and non-homogeneous haze. Specifically, we use cycle-GAN with non-homogeneous and homogeneous haze removal generator. Generator use modified Unet with pixel, channel attention and pretrained resnet as haze removal. Overall proposed architecture gives better results for both non-homogenous and homogeneous images compared to the existing unsupervised methods.

Index Terms: *Unsupervised, Non-homogeneous haze, Homogeneous haze, Image Dehazing, GAN, Attention mechanism*

List of Principal Symbols and Acronyms

CA Channel Attention
PA Pixel Attention
D4 Dehazing via Density and Depth Decomposition
NH Non-homogeneous Haze
H Homogeneous Haze
GAN Generative Adversarial Networks
ASM Atmospheric Scattering Model
PSNR Peak Signal-to-Noise Ratio
SSIM Structural Similarity Index

List of Tables

List of Figures

CHAPTER 1
Introduction

Atmospheric particles, such as haze and fog, can impede the visibility of a scene by attenuating the reflected light along its path and scattering the atmospheric light. This scattering effect, as shown in Fig. 1.1, adds a portion of atmospheric light (known as airlight) to the direct light received by a camera. As a result, the captured image exhibits lower contrast, color distortion, and other undesirable effects, rendering it less visually appealing and useful for various applications. These atmospheric phenomena can have a significant impact on the quality of outdoor and other types of images, underscoring the importance of dehazing techniques for enhancing their visibility and usefulness.

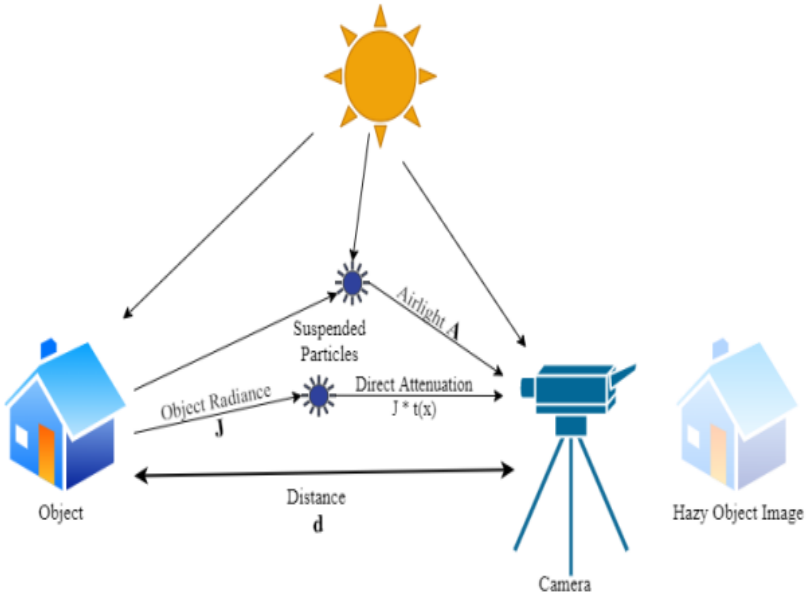


Figure 1.1: Atmospheric Scattering Model (ASM)

This type of degraded image is not desirable for vision-related applications. As a result, the impact of haze and other atmospheric phenomena must be minimised. For example, we must create a picture similar to Fig.1.3(b) from 1.3(a), which is degraded because of haze.

(a) Hazy Image

(b) Haze-free Image

Figure 1.2: Example of homogeneous image dehazing

(a) Hazy Image

(b) Haze-free Image

Figure 1.3: Example of non-homogeneous image dehazing

There are two types of Haze.: i) homogeneous, and ii) non-homogeneous. In homogeneous hazy images, haze is evenly distributed at the same depth level. As illustrated in Fig.1.2(a), the haze is evenly spread throughout the entire scene. Haze presence is unequally distributed in non-homogeneous hazy images, which means that some part of the scene is highly affected by haze or has a high density of haze compared to other regions. As shown in Fig.1.3(a), the haze is distributed unevenly across the entire scene, with some areas having high-density haze and others having very low-density haze.

The atmospheric scattering model (ASM) [24, 25, 26] can approximate haze degradation. Mccartney initially proposed this model [24], and later Narasimhan and Nayar [25] improved it.

The atmospheric scattering model:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1.1}$$

In this equation, the hazy image captured by the camera, the haze-free image, and the transmission map are represented as $I(x)$, $J(x)$ and $t(x)$ respectively. The global atmospheric light is denoted as A, and x represents the pixel location in the image. The term $J(x)t(x)$ represents the direct attenuation, while the term $A(1 - t(x))$ is known as the airlight. In this method, we aim to find the unknowns $t(x)$ and A from the input hazy image $I(x)$. The transmission map $d(x)$ is directly related to the distance d(x) of the scene point from the camera.

$$t(x) = e^{-\beta d(x)} \tag{1.2}$$

This equation demonstrates that the intensity of the scene decreases exponentially with depth. If we can recover the transmission, we can estimate the depth up to an unknown scale.

Here β is the scattering coefficient. Several dehazing methods follow this model to find out a haze-free image. However, the ASM can not reflect the nonhomogeneous nature of haze. As a result, methods developed based on ASM find it difficult to produce an optimal dehazing result on non-homogeneous hazy image.

1.1 Objective

The objectives of the thesis can be summarized as follows:

1. Developing an architecture that can handle both non-homogeneous and homogenous haze.
2. Developing an architecture that can overcome the challenges posed by limited data and the non-homogeneous nature of haze, using an unsupervised approach due to the limitation of real-world ground truth images.
3. Even when a dataset is available, designing an architecture that can handle the various density levels seen in non-homogeneous haze images..
4. Including a module in the architecture to address the issue of color distortion and ensure the preservation of accurate colors and structure in the dehazed images.

1.2 Contribution

The contributions of the thesis can be summarized as follows:

The thesis makes significant contributions by developing a novel architecture that utilizes unsupervised learning based on Cycle-GAN to overcome the unavailability of real-world ground truth images. To address color distortion, the proposed method incorporates channel and pixel attention mechanisms. Additionally, the proposed model effectively handles varying density levels of haze using the U-Net architecture, enabling the removal of both homogeneous and non-homogeneous haze.

Extensive evaluations comparing the proposed architecture against state-of-the-art methods demonstrate its superior effectiveness in haze removal tasks, as measured by objective metrics such as PSNR and SSIM. The adaptability and generalizability of the architecture make it suitable for various applications in fields such as photography, computer vision, and remote sensing.

1.3 Organization of Thesis

Chapter 2 provides an overview of existing techniques for image dehazing, including classical methods, supervised learning-based methods, and unsupervised learning-based methods.

Chapter 3 presents the first proposed method, which introduces a cycle-GAN network architecture. It explains how the two-branch Generator network utilizes pretrained models and attention mechanisms to handle both non-homogeneous and homogeneous haze.

Chapter 4 focuses on the modification of the second branch of the generator architecture by replacing the pretrained ResNet with a pretrained ConvNet model.

Chapter 5 introduces a single-branch generator architecture that incorporates a UNet with channel attention (CA) and pixel attention (PA), pretrained ResNet, and an enhancer network.

Chapter 6 discusses the definition of cycle loss, GAN loss functions, and the total loss function.

Chapter 7 presents the results of the experiments conducted in the study. This chapter compares the obtained results with existing state-of-the-art methods. The model is directly tested on the I-haze 2018 dataset and also evaluated on real-world images.

Chapter 8 concludes the thesis, summarizing and contributions.

Chapter 9 outlines potential areas for future research and development.

CHAPTER 2
Literature Survey

Haze is a phenomenon that occurs when light passes through a medium, such as air, that is filled with small particles of haze, fog, smoke etc. These particles scatter the light, making the image appear hazy. The amount of haze depends on the particles' density, the distance between the camera and the object, and the atmospheric light.

Image dehazing is the process of removing haze from an image. This is a challenging task because the hazy image does not contain any haze-related information. To deal with the density of haze in hazy images, the distance between the object and the camera is difficult to measure, Because hazy images do not carry this distance information.

There are a number of different methods that have been proposed for image dehazing. These can be generally divided into Three categories:

1. Classical methods
2. Supervised learning-based methods
3. Unsupervised learning based methods

2.1 Classical methods

According to equation 1.1, the process of getting haze-free images depends on determining two unknowns: the atmospheric light A and the transmission map $t(x)$. The transmission map is dependent on the depth or distance of each scene point from the camera. However, since this depth information is not directly available in the image, estimating the distance values can be challenging. This limitation has led to the development of classical methods [7, 13, 23, 25, 29] that focus on

addressing this issue.

Classical methods determine priors on haze-free images based on statical observations. Most approaches compute A the atmospheric light and $t(x)$ the transmission map in the ASM model.

He et al. introduced the dark channel method [16], which is based on the observation that the haze-free images mostly have at least one color channel with a noticeably lower pixel intensity. This approach, known as the dark channel prior, is derived from statistical analysis of haze-free outdoor photographs. The study discovered that in most local regions excluding the sky, certain pixels referred to as dark pixels have very low intensity values in at least one of the RGB channels.

$$D(x) = min_{y \in \Omega_r(x)}(min_{c \in r,g,b} I^c(y)) \tag{2.1}$$

Here, $\Omega_r(x)$ represents a local patch of size r×r entered at pixel x and I^c denotes the intensity value of the RGB color channel c of image I. he dark channel method, as indicated by this equation, provides an estimate of the amount of haze present in the image. It is directly related to the transmission map $t(x)$. implying that a higher value of $D(x)$ suggests a higher level of haze in the corresponding region.

$$t(x) \propto 1 - D(x) \tag{2.2}$$

The dark channel method is a type of statistics, it not work for all images. The dark channel prior is invalid when the scene objects are similar to the atmospheric light, and no shadow is cast on them. The scene's radiance dark channel contains bright values near such objects. As a result, the DCP [16] method will overestimate the haze layer while underestimating the transmission of these objects.

The color attenuation prior [36]. This method discover that the brightness and saturation of pixels in a hazy image change dramatically as the haze concentration changes. They argue that In a linear model, the difference between the value and saturation of pixels should be positively correlated to the scene's depth.

$$D(x) \propto C(x) \propto V(x) - S(x) \tag{2.3}$$

Where $D(x)$ is the scene depth, $C(x)$ is the concentration of the haze, $V(x)$ is the

brightness of the scene and $S(x)$ is the saturation

Classical methods methods are based on certain assumptions about the characteristics of hazy images and the atmospheric scattering process.

These assumptions may not always hold true for real-world hazy images, leading to inaccurate results. Real-world hazy scenes might have complex lighting conditions, different atmospheric properties, and variations in picture depth, which can challenge these classical methodologies into question. As a result, the performance of these methods may be limited in handling the wide range of hazy image scenarios encountered in practical applications.

To overcome these limitations, more advanced and data-driven approaches, such as deep learning-based methods, have been proposed. These methods can learn complex mappings between hazy and haze-free images from large-scale datasets, allowing them to capture and model the characteristics of real-world haze more effectively.

2.2 Supervised learning-based methods

Instead of relying on hand-crafted features like classical approaches, deep learning methods learn features automatically from training data. Hence, it has become accurate for many vision applications, including image dehazing. Cai et al. have developed DehazeNet [8] as a deep-learning method for determining medium transmission. DehazeNet takes a hazy image as input and produces a medium transmission map as output. After computing t(x), it employs the ASM to generate a dehaze image.

Unlike DehazeNet, AOD-net [19] calculates the atmospheric light and transmission map simultaneously and then uses ASM to produce a haze-free image. Several deep learning approaches calculate haze-free images directly without using ASM by end-to-end mapping between haze-free and hazy images.

Qin et al. suggested FFA-net [27], an attention-based neural network that can directly generate haze-free images without calculating atmospheric light or transmission map. These approaches perform well on homogeneous hazy images but fail to produce satisfactory results on non-homogeneous datasets.

Jing Liu suggested the Trident Dehazing method [21] for dense and non-homogeneous haze. The Trident Dehazing method learns the hazy to hazy-free image mapping with automatic haze density recognition using unet style architecture. TDN comprises three sub-nets: the Encoder-Decoder Net (EDN), which is the main net of TDN and helps reconstruct the coarse hazy-free feature. The Detail Refinement sub-Net (DRN) helps refine the high-frequency details that were easily lost in the max pooling layers in the encoder, and the Haze Density Map Generation sub-Net (HDMGN), which can automatically distinguish the thick haze region from the thin one.

Supervised learning methods for haze removal, such as Dehazenet [3], AOD-net [4], Trident [5], DMPHN [10],and KTDN [30], have been proposed in the literature. These methods rely on paired training data, where pairs of hazy and corresponding haze-free images are available for training the models.

However, these supervised approaches face certain limitations. One of the challenges is the availability of real-world ground truth images, which are required for training the models effectively. Obtaining such ground truth data is often difficult and time-consuming, which restricts the applicability of these methods in practical scenarios.

Another limitation is the risk of overfitting, where the trained models may become too specialized and perform poorly on new or unseen data. Overfitting occurs when the models excessively adapt to the training data and fail to generalize well to diverse real-world hazy images.

To address these limitations, alternative approaches such as unsupervised learning methods, which do not require paired training data, have been explored. These methods aim to overcome the reliance on ground truth images and mitigate the risk of overfitting, offering more flexibility and applicability in real-world scenarios.

2.3 Unsupervised learning based methods

Unlike the supervised methods, unsupervised learning methods do not required paired dataset for training. Some unsupervised methods can be directly trained

on hazy images, which disentangles the hazy image into the clean image and other components.

Hodges et al.[17] introduced a model consisting of two separate networks: the dehazing and the discriminator networks. The dehazing network estimates the transmission map to obtain an initial haze-free image using the dehazing formula. The second network, the discriminator, helps fine-tune the earlier weights obtained.

The primary objective of the structure is to create an abstract measure of the dehazing model's performance, which helps to enhance the weights of the input parameters via the feedback mechanism.

Engin et al. proposed an enhanced cycle-GAN called Cycle-Dehaze [12] by combining cycle consistency and perceptual losses to preserve the textural information of the dehaze images.

Most unpaired dehazing methods are based on cycle-GAN with additional specific designs. However, these methods usually ignore the density and the spread of haze when generating hazy images. Neglecting these factors leads to unrealistic haze generation, which will further affect the dehazing performance. To deal with these problems, our proposed framework focuses on exploring haze density and the spread of haze in non-homogeneous hazy images.

2.4 Chapter Summary

In the field of image dehazing, researchers have shifted towards unsupervised approaches due to the lack of real-world ground truth data for hazy images. Unsupervised methods aim to learn the underlying characteristics of hazy images without relying on pre-existing paired datasets.

For homogeneous hazy images, where the haze is uniformly distributed, several methods have been proposed, such as Enhanced cycle-GAN and CycleDehaze. These methods leverage the power of generative adversarial networks (GANs) and cyclic training to generate haze-free images from hazy counterparts.

When it comes to non-homogeneous haze removal, where the haze varying den-

sities and spatial distributions, developing effective unsupervised methods still needs improvement. Non-homogeneous haze poses additional challenges due to its complex nature, making it difficult to capture and model the diverse characteristics of such hazy scenes.

In brief, these methods usually ignore the density and the spread of haze when generating dehaze images. The absence of these factors leads to unrealistic haze generation, which will further affect the dehazing performance. To deal with these issues, our proposed Unsupervised Cycle-gan based framework focuses on exploring haze density and the spread of haze in non-homogeneous hazy images.

CHAPTER 3
Proposed Method-I

In our proposed method, we adopt a cycle-GAN based architecture to address the problem of non-homogeneous haze removal. This architecture consists of two generator networks and two discriminator networks, which are trained on an unpaired dataset.

The first generator network is designed to remove haze from a given hazy image. It takes a hazy image as input and generates a corresponding haze-free image as output. By learning the mapping from hazy images to haze-free images, this generator aims to effectively restore the underlying scene information and eliminate the adverse effects of haze.

The second generator network in our architecture is responsible for generating haze. Unlike traditional methods that generate haze using simple heuristics or predefined models, our generator takes into account factors such as haze density and haze spreading. By considering these factors, the generator can produce more realistic and diverse hazy images that closely resemble real-world hazy conditions.

The discriminator networks play a crucial role in training the generators. They are responsible for distinguishing between real haze-free images and generated haze-free images, as well as between real hazy images and generated hazy images. By providing feedback to the generators based on the discriminators' prediction, the generators can continuously improve their performance and generate more accurate results.

Overall, our proposed method combines the power of cycle-GAN with the consideration of haze density and spreading, enabling effective non-homogeneous haze removal using an unsupervised approach. By leveraging unpaired data and

training two generators with specific objectives, our method aims to achieve more accurate and realistic haze removal results.

3.1 Proposed Method Architecture

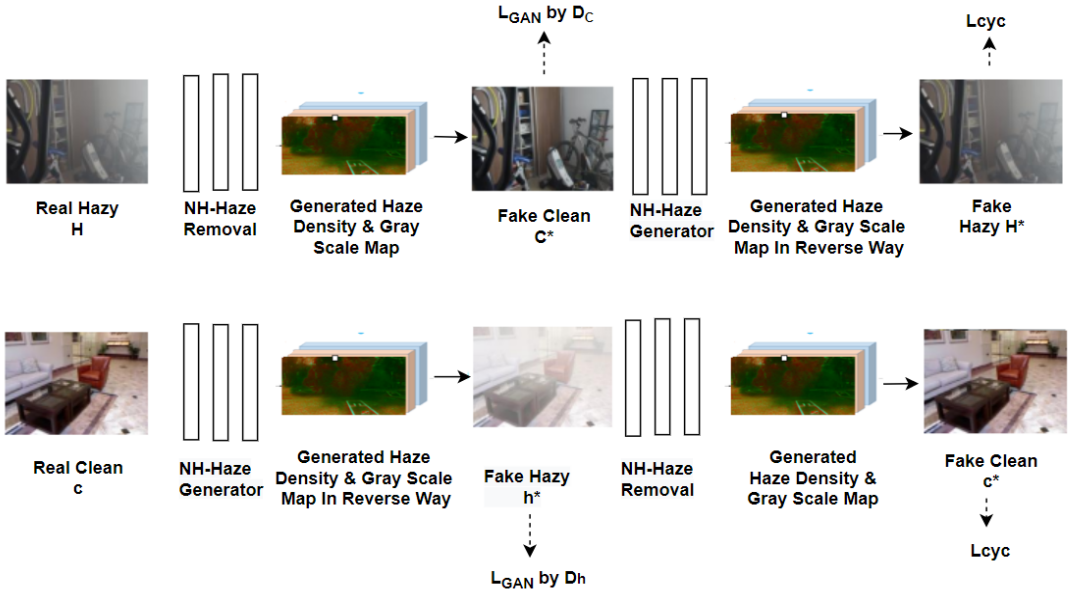


Figure 3.1: cycle-Dehaze network with G and F as Generators and Dx and Dy as Discriminators

Our network architecture incorporates two distinct branches to enable bidirectional conversion between hazy and clean images. The first branch is responsible to converting hazy images to clean images and then hazy images, while the second branch focuses on the reverse conversion from clean to hazy images and then clean images again. This two-branch approach ensures that the network learns a consistent mapping between the hazy and clear image domains, which is important for generating high-quality dehazed images.

The primary objective of the first branch is to remove haze from hazy images and produce clean versions. Through a series of learnable transformations and feature extraction operations, this branch effectively learns to recover the underlying scene details and restore the original appearance of the image. By training the network to perform hazy to clean image conversion, it becomes adept at recognizing and mitigating the effects of haze, resulting in visually pleasing dehazed images.

The second branch of the network serves as a complementary component by focusing on the clean to hazy and hazy to clean images. This branch aims to capture the essence of the hazy appearance and generate realistic hazy versions of clean images. By incorporating this branch, the network gains a deeper understanding of the characteristics of hazy images, including factors such as haze density, haze spreading, and atmospheric effects. This knowledge enables the network to generate accurate and convincing hazy counterparts of clean images.

The two-branch approach plays a important role in ensuring a consistent mapping between the hazy and clear image domains. By training the network to handle both hazy to clean and clean to hazy conversions, it learns to model the complex relationship between these domains effectively. This consistency in mapping is crucial for generating high-quality dehazed images, as it enables the network to capture the details and visual characteristics of the scene while removing haze.

During the testing phase, only the hazy to clean generator network is utilized. This means that the network is specifically designed to generate clean images from hazy inputs. The training process of the network focuses on learning the transformation from hazy to clean images, which is the primary task of image dehazing. By specializing in this direction, the network becomes proficient in producing visually pleasing and realistic dehazed images during testing.

3.1.1 Generator Architecture

Generator Architecture

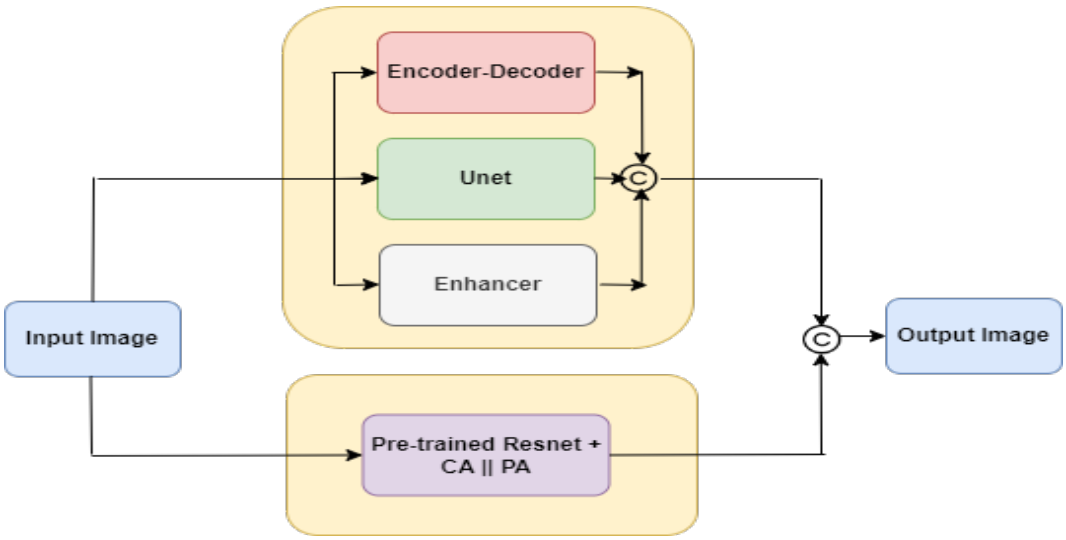


Figure 3.2: Generator Architecture - I

The **Generator** contains two branches. The first **haze removal branch** contains three subnets, and the second **attention branch** [14] contains a pre-trained resnet with channel and pixel attention parallelly.

Haze removal branch architecture

1. Encode-Decoder
2. Unet
3. Enhancer

Attention branch architecture

1. Pre-trained Resnet + CA || PA

3.1.2 Encode-Decoder

The haze-free feature is achieved using. Downsampling factor of encoder is 2 and upsampling factor of decoder is also same. This architecture allows the network to capture and encode features at different scales, which can be beneficial for haze removal tasks.

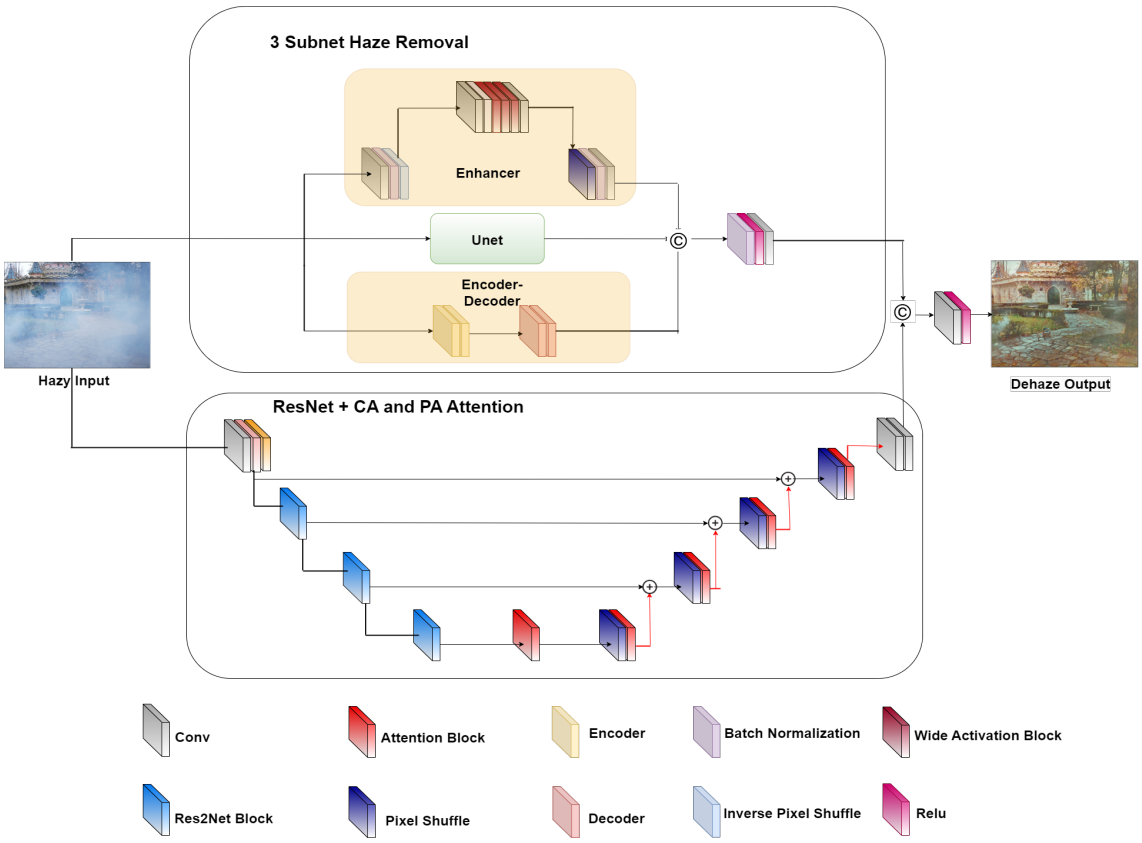


Figure 3.3: Generator Detailed Architecture

3.1.3 Unet

This sub-network incorporates a U-Net [28]. The U-Net architecture is a generally used in various image processing tasks due to its effectiveness in capturing both local and global features of an image. It consists of an encoder-decoder structure with skip connections that allow for the transfer of information between different resolution levels.

In this particular case, a standard U-Net architecture is utilized, which consists of six downsampling (encoding) layers followed by six upsampling (decoding) layers.

3.1.4 Enhancer network

The Enhancer sub-network plays a important role in refining the image generated by the previous stages. It operates on the downsampled image, reducing its size by a factor of 4. The network applies non-linear feature mapping to enhance the image quality further.

To achieve this, the Enhancer network utilizes inverse pixel shuffle layers, which transform the feature maps from a spatial dimension to a depth dimension. This transformation allows the network to process the image more compactly and efficiently. Subsequently, pixel shuffle layers are used to reverse the process, converting the feature maps from the depth dimension back to the spatial dimension.

The Wide Activation Blocks (WAB) [32] are employed in the Enhancer network to provide non-linear feature mapping. These blocks consist of two 3x3 convolutional layers and a wide activation laye. The wide activation layer helps capture complex patterns and details in the image, enhancing the overall output quality of the network.

Finally, a 3x3 convolutional layer is used as the final concatenation operation. This layer combines the features obtained from the different modules within the network, enabling the integration of valuable information from each stage.

The Enhancer network is essential in refining the image generated by the previous stages. By leveraging inverse pixel shuffle layers, Wide Activation Blocks, and concatenation operations, the network is able to produce sharper and more precise image outputs, resulting in improved image quality and visual appeal.

3.1.5 Pre-trained Resnet + CA || PA network

The pre-trained ResNet-based network employed in this approach takes advantage of transfer learning [11, 33] to enhance the dehazing process. By using prior information gained from image classification tasks and applies it to the task of dehazing. Using a pre-trained Res2Net [15], a variant of ResNet architecture that improves the representation power, as the backbone of the encoder, the network benefits from the learned features and hierarchical representations from the ImageNet [9] pretraining.

In the decoder module, pixel-shuffle layers are utilized for upsampling. This technique helps in reducing computational overhead by avoiding the use of traditional upsampling methods such as transposed convolutions. Pixel-shuffle layers rearrange the channel-wise information to progressively recover the size of feature maps to the original resolution. This progressive recovery helps in maintaining the spatial details and fine-grained information, contributing to the generation

of high-quality dehazed images.

To further improve performance, channel attention and pixel attention mechanisms are incorporated in parallel with the pre-trained ResNet in each decoder. These attention mechanisms allow the model to focus on relevant features and enhance the dehazing process, leading to visually appealing and accurate dehazed images.

CHAPTER 4
Proposed Method-II

4.0.1 Generator Architecture

We have changes generator architecture in second method. Generator architecture is given below.

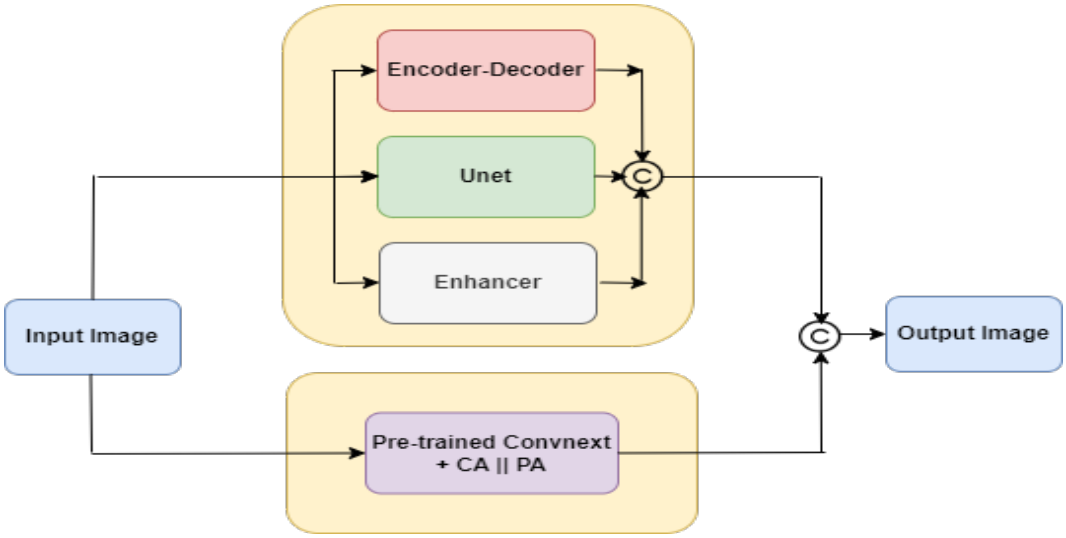


Figure 4.1: Generator Architecture - II

Haze removal branch architecture

1. Encode-Decoder
2. Unet
3. Enhancer

Attention branch architecture

1. Pre-trained Convnext + CA || PA

4.0.2 Convnext + CA || PA network

In our second method, we opted to replace the pre-trained ResNet with a pre-trained Convnext model [22] in the generator architecture. Convnext is similar to ResNet. It is composed of multiple layers of convolutional and pooling operations. However, Convnext offers certain advantages over ResNet that make it suitable for our task.

One key benefit of Convnext is its ability to handle non-homogeneous datasets more effectively. The Convnext architecture is specifically designed to capture and model complex spatial dependencies within images. This makes it well-suited for addressing the challenges posed by non-homogeneous haze, where the haze distribution and density vary across different regions of the image.

Additionally, Convnext has demonstrated better performance in tasks that require capturing fine-grained details and intricate patterns. Its architecture enables the extraction of more localized and specific features, which can be advantageous when dealing with images affected by non-uniform haze.

By replacing ResNet with Convnext in our generator architecture, we aimed to leverage these benefits and explore whether Convnext can offer improved performance for non-homogeneous haze removal compared to the ResNet architecture that we have studied in Chapter I. This experimental comparison allowed us to gain insights into the strengths and weaknesses of different pre-trained feature extraction model and their impact on haze removal results.

It's important to note that while ResNet is known for its ability to extract deep features and perform well in various tasks, Convnext can still be effective in certain scenarios. By training the Convnext on dataset and evaluating its performance, we aimed to assess its suitability for the specific task of dehazing.

By conducting experiments and comparing the results obtained with the Convnext-based method to those achieved with the pre-trained ResNet-based method, we can gain insights into the effectiveness and limitations of both approaches. This comparison allows us to make informed decisions about which model architecture to use for dehazing based on their performance on the dataset.

All other architectures are the same as we have seen in the first method.

CHAPTER 5
Proposed Method-III

5.0.1 Generator Architecture

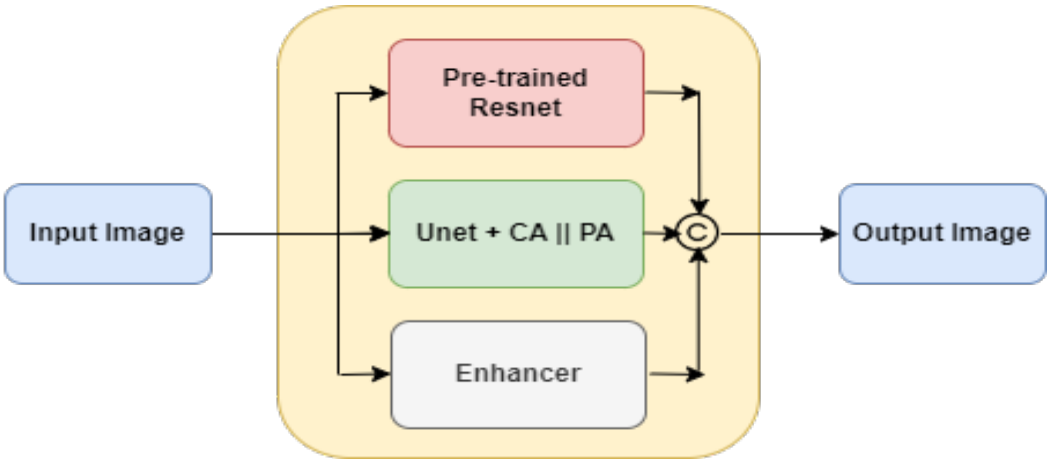


Figure 5.1: Generator Architecture - III

Our final method utilizes a single branch in the generator network, as we found extra branch channel and pixel attention branch increase overhead in the model. Instead, we merged the channel and pixel attention mechanisms with the U-Net architecture. We made this decision based on the observation that U-Net has superior capabilities in identifying the distribution of haze in images.

In our approach, we removed the encoder-decoder module used in the previous method because the pre-trained ResNet already provide a powerful encoder-decoder architecture. ResNet has the ability to capture haze-free features from hazy images, making it suitable for the dehazing task. One of the key advantages of ResNet is its effectiveness in training deep networks, even with hundreds of layers, while maintaining good optimization and performance. This is achieved through skip connections, which facilitate the direct flow of gradients, overcoming the vanishing gradient problem and enabling better information propagation

throughout the network.

By combining the U-Net architecture with channel attention and pixel attention, we observed improved results in addressing haze. U-Net's capability to identify the distribution of haze, coupled with the integration of channel attention and pixel attention, led to more effective haze removal and enhanced visual effects in the non-homogeneous dehazed images.

Overall, our final method leverages the powerful encoder-decoder architecture of ResNet while taking advantage of U-Net's capabilities in haze distribution identification. The combination of channel attention and pixel attention with U-Net enhances the model's ability to address haze and improve the visual quality of the dehazed images.

Generator architecture

1. Unet + CA || PA
2. Pre-trained resnet
3. Enhancer

5.0.2 Unet + CA || PA network

This sub-network incorporates a U-Net [28] structure along with Channel Attention (CA) and Pixel Attention (PA) modules. The U-Net architecture is a generally used in various image processing tasks due to its effectiveness in capturing both local and global features of an image. It consists of an encoder-decoder structure with skip connections that allow for the transfer of information between different resolution levels.

In this particular case, a standard U-Net architecture is utilized, which consists of six downsampling (encoding) layers followed by six upsampling (decoding) layers. Each upsampling module is augmented with both the Channel Attention (CA) and Pixel Attention (PA) modules. This means that at each stage of upsampling, the network incorporates the CA and PA modules in parallel.

The Channel Attention (CA) module helps the network to selectively focus on important channels or feature maps, enabling it to capture and highlight significant image features during the upsampling process. On the other hand, the Pixel

Attention (PA) module allows the network to attend to specific pixels within their local contexts, preserving fine-grained details and enhancing the overall spatial information during the upsampling operation.

By combining the U-Net architecture with the CA and PA modules in a parallel fashion, the sub-network can effectively leverage the benefits of both local and global feature extraction, as well as channel-wise and pixel-wise attention mechanisms. This enables the network to generate high-quality dehazed images with enhanced color, structure, and spatial details.

Pre-trained Resnet [15] and Enhancer subnet are similar to those we used in our first method.

CHAPTER 6
Loss Function

In this section, we discuss the loss used to train our model.

6.1 Cycle Loss

Cycle Loss imposes that an intermediate image transferred from one domain to another should be able to transfer back. The cycle loss in our method can be written as follows

$$L_{cyc} = \|\mathbf{c} - \hat{\mathbf{c}}\|_1 + \|\mathbf{H} - \hat{\mathbf{H}}\|_1 \tag{6.1}$$

6.2 GAN Loss

GAN Loss It checks whether a generated image belongs to a specific domain. It means that the output of the Dehaze network belongs to a clean set and the same for the Rehaze network.
For the Dehazing network G_d and corresponding discriminator D_c, the GAN loss can be expressed as follow:

$$L_{adv}(\mathcal{D}_c) = \mathbb{E}[(\mathcal{D}_c(\boldsymbol{c}) - 1)^2] + \mathbb{E}[(\mathcal{D}_c(\hat{\boldsymbol{C}}))^2] \tag{6.2}$$

$$L_{adv}(\mathcal{G}_D) = \mathbb{E}[(\mathcal{D}_c(\hat{\boldsymbol{C}}) - 1)^2]$$

The total loss function is defined as

$$L = \lambda_1 L_{cyc} + \lambda_2 L_{adv} \tag{6.3}$$

We have set values λ_1, λ_2 are 1, 0.2 respectively.

CHAPTER 7
Experiments and Results

In this section we will discuss regarding the datasets, training details, and evaluation metrics. Then, we compare our method quantitatively and qualitatively with state-of-the-art dehazing methods. Finally, to further understand the effects of the various modules, we undertake ablation studies.

7.0.1 Dataset for Training

In our study, we trained our model on two datasets separately: NH 2021 [6] and Reside [20]. For the NH 2021 dataset, we selected 20 images for training. To effectively handle these images, we divided them into smaller patches with a size of 256x256 pixels, considering a stride value of 118 pixels. Similarly, for the Reside dataset, we utilized 45 images and used the same patch size of 256x256, with a stride value of 50 pixels.

For testing our model, we chose 5 images from each dataset.

During the model training phase, we employed the Adam optimizer [18] with $\beta_1 = 0.9$, $\beta_1 = 0.999$. A batch size of 1 was used for training. We set 0.0001 as the initial learning rate. In order to increase the training data's diversity, we used horizontal flip augmentation.

We calculated two generally used metrics to evaluate the performance of our model: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). These metrics provide insights into the quality of the dehazed images generated by our model. We compared our results with state-of-the-art methods and the base paper [31] to assess the effectiveness of our approach.

In our experiments, we followed the official train and test split for the NH-Haze 2021 dataset. The computed PSNR and SSIM values, along with the visual results,

are presented in the next section, providing a comprehensive assessment of our model's performance.

We tested our model directly on the whole I-haze 2018 [5] dataset without any prior training on that dataset. This approach allowed us to assess the adaption capability of our model and its performance on unseen data.

By testing our model on the I-haze 2018 [5] dataset and comparing the results with other methods, we aimed to demonstrate the effectiveness and competitive performance of our model in the field of image dehazing.

7.0.2 Comparison with state-of-the-art Methods

We compared our proposed method with several state-of-the-art methods, including DCP [16], AOD [19], FFA-Net [27], TDN [21], Non-homogeneous Dehazing of Images by Attention Mechanism in Deep Framework [3], Homogeneous and Non-homogeneous Image Dehazing [2], Cycle-GAN [35], CycleDehaze [12], and D4 [31]. It is worth noting that some of these methods, such as DCP, Cycle-GAN, CycleDehaze, and D4, do not require paired data for training and follow an unsupervised learning approach. On the other hand, methods like DCP, AOD, FFA-Net, and TDN follow a supervised learning method, which relies on paired data for training.

Our proposed method also follows an unsupervised learning approach and achieves better results compared to other unsupervised methods. The results are presented in Figure 7.2.

7.0.3 Non-homogeneous, Homogeneous PSNR and SSIM comparison with state-of-the-art methods

The results presented in Table 6.1 demonstrate the effectiveness of our method in addressing both types of haze removal. Our approach does not rely on any atmospheric formula for homogeneous haze removal, making it versatile in various environmental conditions. Moreover, it is capable of identifying and effectively removing the spread of non-homogeneous haze.

Figure 7.1: Non-homogeneous, Homogeneous result comparison with state-of-the-art methods

Training Method	Method	NH 21		SOTS-Indoor		I-Haze	
		PSNR	SSIM	PSNR	PSNR	PSNR	SSIM
Supervised	AOD [19]	13.3	0.469	19.06	0.852	15.61	0.712
	FFA [27]	20.45	0.804	36.36	0.993	12.00	0.592
	TDN [21]	20.23	0.762	34.59	0.975	19.33	0.828
	NH-Haze Dehazing [3]	22.87	0.893	-	-	-	-
	NH-Haze and H-Haze Dehazing [2]	16.93	0.695	-	-	16.43	0.698
Unsupervised	DCP [16]	11.68	0.709	13.10	0.699	13.10	0.699
	CycleGAN [35]	-	-	21.34	0.898	15.29	0.756
	CycleDehaze [12]	-	-	20.11	0.854	14.69	0.751
	D4 [31]	10.83	0.699	25.42	0.932	15.61	0.78
	Proposed Method-I	**14.5**	**0.521**	**25.4**	**0.948**	**16.13**	**0.668**
	Proposed Method-II	**15.52**	**0.641**	**21.96**	**0.861**	**14.48**	**0.515**
	Proposed Method-III	**18.04**	**0.78**	**-**	**-**	**-**	**-**

Table 7.1: PSNR and SSIM comparison

For the homogeneous dataset, our method achieves a PSNR value of **25.4** and an SSIM value of **0.948**, indicating high-quality dehazing results. Similarly, for the non-homogeneous dataset, our method attains a PSNR value of **18.04** and an SSIM value of **0.78**. These results highlight the robustness and adaptability of our method in handling different types of haze, leading to significant improvements in image quality.

7.0.4 Non-homogeneous result comparison with state-of-the-art methods

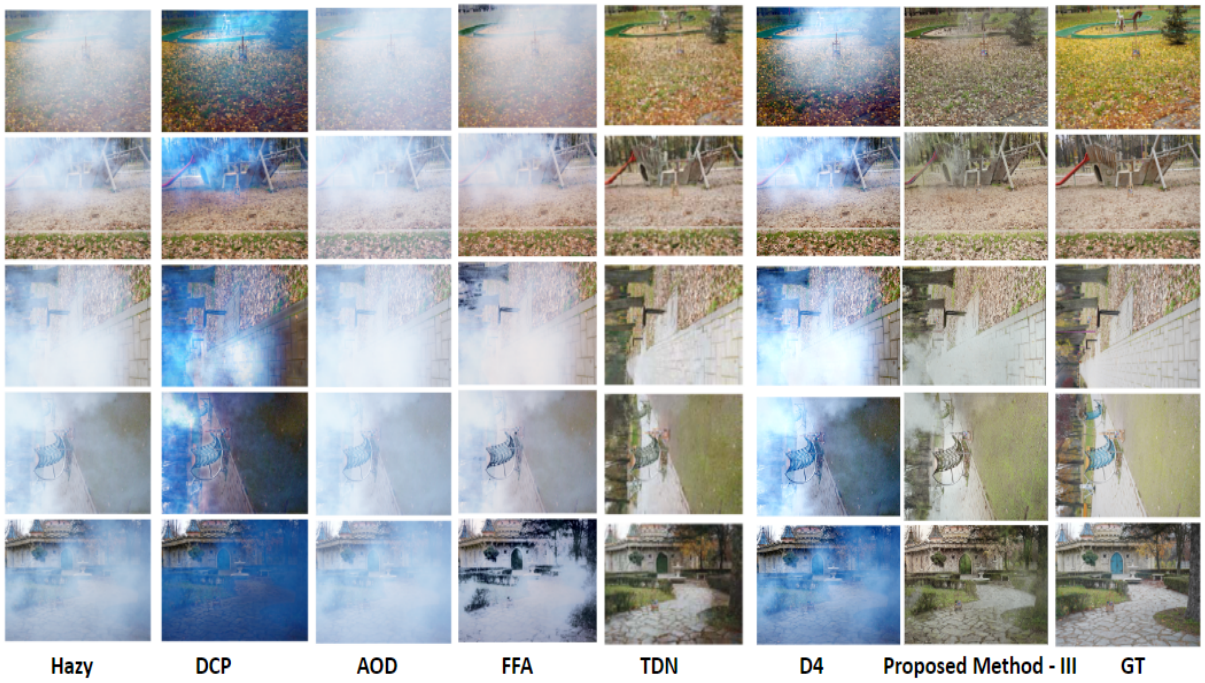


Figure 7.2: Non-homogeneous result comparison with state-of-the-art methods

The analysis of Figure 7.2 for a non-homogeneous dataset reveals several important observations regarding different image dehazing methods. The Dark Channel Prior (DCP) method generates dehazed results with noticeable color distortion. This can be attributed to its reliance on an atmospheric scattering model and a prior-based approach, which struggles to effectively remove haze with varying density. Both DCP and D4 exhibit a tendency to produce bluish results, further indicating their limitations in effectively removing haze.

D4, a unsupervised learning approach, encounters difficulties in removing non-homogeneous haze due to its sequential process of predicting a transmission map and then generating a dehazed image based on atmospheric scattering. While this approach works reasonably well for homogeneous haze removal, it fails to accurately identify and address varying density haze.

In our proposed method, we leverage a U-Net style sub-module that demonstrates efficient identification of varying density haze. This sub-module plays a crucial role in accurately removing haze. The encoder-decoder sub-module helps to produce haze-free features from hazy images, while the enhancer network is

utilized to restore sharpness to the image that may be lost during the encoding and decoding procedure.

While FFAnet and TDN methods are capable of properly removing haze, they heavily rely on a supervised approach, which necessitates access to paired datasets. This reliance on paired data restricts their applicability in real-world scenarios where acquiring such datasets is challenging.

Our proposed method distinguishes itself by combining the benefits of unsupervised learning while effectively addressing the limitations of other unsupervised methods, including cycle-GAN, cycle-Dehaze, and D4. CycleDehaze, an improved version of cycle-GAN for image dehazing, employs a perceptual loss inspired by EnhanceNet to enhance visual quality metrics such as PSNR and SSIM. However, it still fails to handle varying haze density and color distortion. In contrast, our model uses the power of U-Net and channel and pixel attention mechanisms parallelly to efficiently tackle these challenges. Consequently, our proposed method demonstrates superior performance in haze removal, producing visually pleasing results.

Overall, our approach showcases a compelling solution by combining the advantages of unsupervised learning and effectively addressing the limitations encountered by other unsupervised methods. By effectively handling varying haze density and color distortion, our method outperforms existing techniques in the task of image dehazing.

7.0.5 Method-I,II,III comparison

We evaluated three different methods for non-homogeneous haze removal on the NH21 dataset. The first method utilized a two-branch generator network, where the first branch consisted of an encoder-decoder subnet, Unet, and an enhancer, while the second branch incorporated a pre-trained ResNet with channel attention (CA) and pixel attention (PA). However, this method resulted in dehazed images with over-saturated color information, indicating a limitation in effectively removing haze from non-homogeneous hazy images.

In the second method, we replaced the pre-trained ResNet with a pre-trained ConvNet architecture (Convnext) in the second branch of the generator network.

The third and final method involved a single-branch generator network that combined three different subnets: a pre-trained ResNet, Unet with channel attention (CA), and pixel attention (PA), along with an enhancer. This method proved to be highly effective in accurately removing haze from non-homogeneous hazy images. The combination of Unet with CA and PA allowed for better identification of the spread of haze and its removal without causing color saturation artifacts. This method achieved a PSNR value of 18.04 and an SSIM value of 0.78, which were higher compared to the results obtained from method-I and method-II, as shown in Table 6.1. The visual results displayed in Figure 6.3.

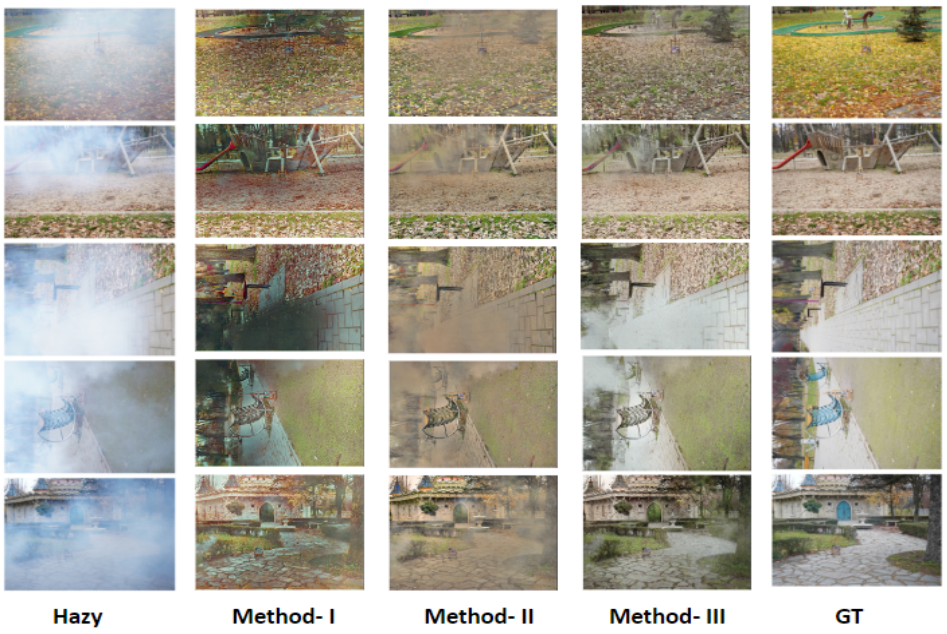


Figure 7.3: Method-I,II,III comparison on Non-homogeneous dataset

7.0.6 Homogeneous result comparison with state-of-the-art methods

Figure 7.4 show the homogeneous result comparison with state-of-the-art-methods

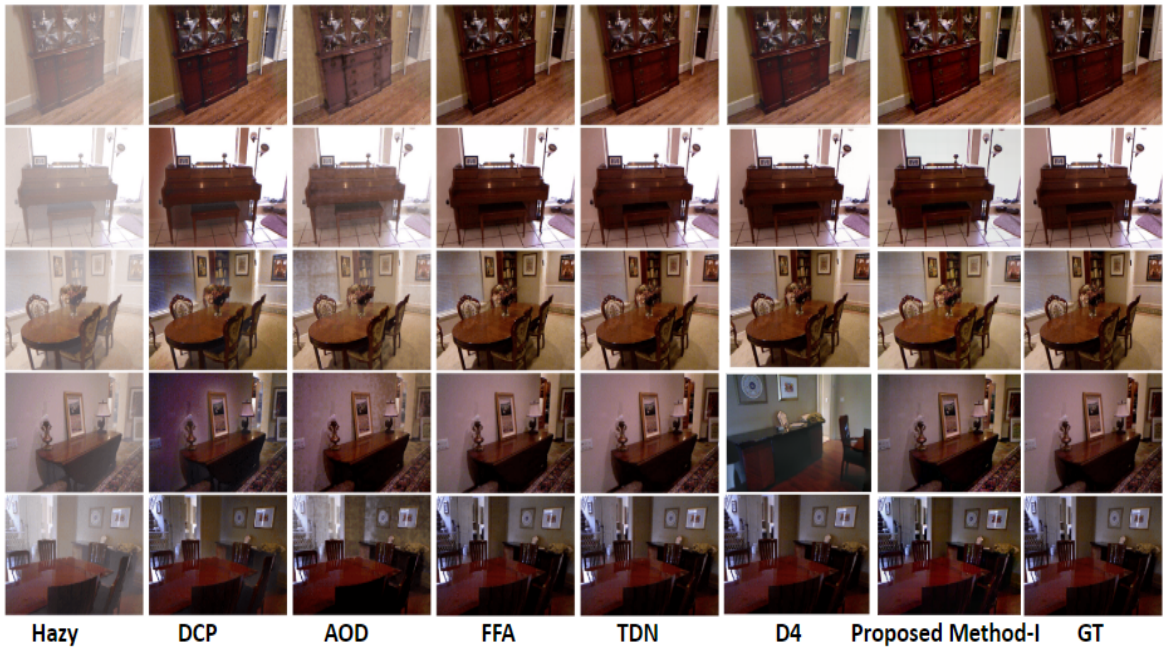


Figure 7.4: Homogeneous result comparison with state-of-the-art methods

7.0.7 I-haze 2018 results

Figure 7.5: I-haze 2018 results

We conducted experiments on the entire I-haze 2018 dataset without any prior training specifically on that dataset. We achieved a PSNR value of 16.13, which was the highest among the unsupervised approaches evaluated. Additionally, we obtained an SSIM value of 0.668 on this dataset.

These results demonstrate the effectiveness of our model in handling different datasets without the need for dataset-specific training. The ability to achieve competitive performance on the I-haze 2018 dataset without prior training highlights the robustness and adaptability of our proposed method. It suggests that our model has learned meaningful representations of haze and can effectively remove it from images across various datasets.

7.0.8 Testing on real world images

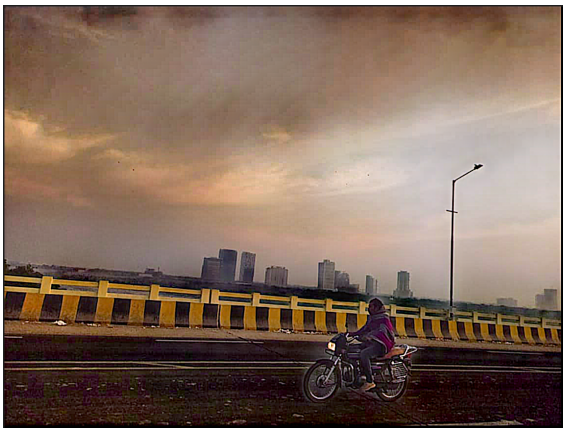
Figure 7.6 and 7.7 present the output results of our method on real-world images that lack ground truth image. The photographs used in our experiment was taken at GIFT City, located in Gandhinagar, India.

(a) Hazy Image

(b) Output

Figure 7.6: Real world image testing

(a) Hazy Image



(b) Output

Figure 7.7: Real world image testing

CHAPTER 8
Conclusions

In our study, we addressed the problem of image dehazing, aiming to improve visibility and image quality under hazy conditions. We introduced a novel two-branch architecture based on cycle-GAN, which effectively removes both homogeneous and non-homogeneous haze from images in an unsupervised manner. Our approach incorporates haze density map generation for a better way to address the spread of the haze and attention mechanisms to preserve color and structure information in the dehazed images.

Our proposed model is the first to address the unsupervised approach for removal of non-homogeneous haze. We successfully integrated pixel and channel attention mechanisms into our model. Pixel attention is employed to preserve structural details, while channel attention helps maintain accurate color information.

Comprehensive experiments on multiple datasets, we demonstrated that our method outperforms state-of-the-art techniques in terms of PSNR and SSIM within an unsupervised framework. Our model exhibits robustness and generalizability, effectively handling various types of haze with different densities and spread patterns. This contributes significantly to the field of image restoration, offering a novel and effective solution for image dehazing that can benefit applications in photography, computer vision, and remote sensing. Overall, our proposed method demonstrates better performance for both homogeneous and non-homogeneous datasets within an unsupervised approach.

CHAPTER 9
Future Work

Incorporating self-attention GAN to enhance further the feature extraction and domain adaptation capabilities of our model. Self-attention GAN [34] can capture long-range dependencies and global context information, which could improve the quality and realism of the generated images.

Extending our method to other image restoration tasks, such as denoising and super-resolution [1, 4]. Our approach can be easily adapted to different domains by changing the input and output data. It would be interesting to explore how our method performs on other image degradation and enhancement types.

References

[1] J. johnson, a. alahi, and l. fei-fei. perceptual losses for real-time style transfer and super-resolution. in european conference on computer vision, pages 694– 711. springer, 2016.

[2] Manan gajjar and srimanta mandal. "homogeneous and nonhomogeneous image dehazing using deep neural network." in computer vision and image processing: 6th international conference, cvip 2021, rupnagar, india, december 3–5, 2021, revised selected papers, part i, pages 375–386. springer, 2022.

[3] Srimanta mandal, a. dhedhi, rajib lochan das. "non-homogeneous dehazing of images by attention mechanism in deep framework" mtech thesis, daiict, 2022.

[4] Y. zhang, k. li, k. li, l. wang, b. zhong, and y. fu. image super-resolution using very deep residual channel attention networks. in proceedings of the european conference on computer vision (eccv), pages 286–301, 2018.

[5] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *Advanced Concepts for Intelligent Vision Systems: 19th International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings 19*, pages 620–631. Springer, 2018.

[6] C. O. Ancuti, C. Ancuti, F.-A. Vasluianu, and R. Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 627–646, 2021.

[7] D. Berman, S. Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016.

[8] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[9] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 1375–1383. IEEE, 2019.

[10] S. D. Das and S. Dutta. Fast deep multi-patch hierarchical network for non-homogeneous image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 482–483, 2020.

[11] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655. PMLR, 2014.

[12] D. Engin, A. Genç, and H. Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 825–833, 2018.

[13] R. Fattal. Single image dehazing. In *ACM SIGGRAPH 2008 Papers, SIG?GRAPH '08*, page 72:1–72:9. ACM, New York, NY, USA, 2008.

[14] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 203–212, 2021.

[15] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019.

[16] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[17] C. Hodges, M. Bennamoun, and H. Rahmani. Single image dehazing using deep neural networks. *Pattern Recognition Letters*, 128:70–77, 2019.

[18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[19] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017.

[20] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Reside: A benchmark for single image dehazing. *arXiv preprint arXiv:1712.04143*, 1, 2017.

[21] J. Liu, H. Wu, Y. Xie, Y. Qu, and L. Ma. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 430–431, 2020.

[22] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022.

[23] S. Mandal and A. Rajagopalan. Local proximity for enhanced visibility in haze. *IEEE Transactions on Image Processing*, 29:2478–2491, 2019.

[24] E. J. McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York*, 1976.

[25] S. G. Narasimhan and S. K. Nayar. Contrast restoration of weather degraded images. *IEEE transactions on pattern analysis and machine intelligence*, 25(6):713–724, 2003.

[26] S. K. Nayar and S. G. Narasimhan. Vision in bad weather. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 820–827. IEEE, 1999.

[27] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11908–11915, 2020.

[28] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[29] R. T. Tan. Visibility in bad weather from a single image. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008.

[30] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma. Knowledge transfer dehazing network for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 478–479, 2020.

[31] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings*

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2037–2046, 2022.

[32] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. Huang. Wide activation for efficient and accurate image super-resolution. *arXiv preprint arXiv:1808.08718*, 2018.

[33] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pages 818–833. Springer, 2014.

[34] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena. Self-attention generative adversarial networks. In *International conference on machine learning*, pages 7354–7363. PMLR, 2019.

[35] 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*, pages 2223–2232, 2017.

[36] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing*, 24(11):3522–3533, 2015.