# DEVELOPMENT OF DEEP LEARNING BASED PORTABLE DEHAZING SYSTEM

THESIS SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

**Master of Engineering**
**in**
**Eectronics and Telecommunication Engineering**

Electronics and Telecommunication Engineering Department
Jadavpur University

June,2023

by

**MD SOHEL AKHTAR**
**Class Roll No: 002110702013**
**Examination Roll No: M4ETC23015**
**Registration No: 160210 of 2021-2022**

Under the guidance of
**Prof. Sheli Sinha Chaudhuri**

Department of Electronics and Telecommunication Engineering
Jadavpur University
Kolkata - 700032
West Bengal, India

# DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this thesis contains previous work and original work by the undersigned candidate, as part of my Master of Engineering in Electronics and Telecommunication Engineering in the Department of Electronics and Telecommunication Engineering. All information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that I have thoroughly cited and referenced all material and findings which are not original to this research, as provided by these rules and conduct.

Name : Md Sohel Akhtar
Exam Roll No : M4ETC23015
Thesis Title : DEVELOPMENT OF DEEP LEARNING BASED PORTABLE DEHAZING SYSTEM

Md. Sohel Akhtar

Signature of Candidate

FACULTY OF ENGINEERING & TECHNOLOGY
JADAVPUR UNIVERSITY

# CERTIFICATE

This is to certify that the thesis entitled — **"DEVELOPMENT OF DEEP LEARNING BASED PORTABLE DEHAZING SYSTEM"** has been carried out by **Md Sohel Akhtar** bearing Class Roll No: **002110702013**, Examination Roll No.: **M4ETC23015** and Registration No: **160210** of **2021-2022**, under my guidance and supervision and be accepted in partial fulfillment of the requirement for the degree of **Master of Engineering** in **Electronics and Telecommunication Engineering** in the Department of Electronics and Telecommunication Engineering.

S. S. Chaudhuri 13/6/23

Dr. Sheli Sinha Chaudhuri
*Professor*
Dept. of Electronics & Tele-Comm. Engg.
JADAVPUR UNIVERSITY
Kolkata-700 032

**Prof. Sheli Sinha Chaudhuri**
Supervisor
Department of Electronics and
Telecommunication Engineering
Jadavpur University

Manotosh Biswas 14/06/23

**Prof. Manotosh Biswas**
Head of the Department
Department of Electronics and
Telecommunication Engineering
Jadavpur University

MANOTOSH BISWAS
Professor and Head
Electronics and Telecommunication Engineering
Jadavpur University, Kolkata - 32

Ardhendu Ghoshal 13/06/23

**Prof. Ardhendu Ghoshal**
Dean
Faculty of Engineering & Technology
Jadavpur University
Kolkata - 700032

DEAN
Faculty of Engineering & Technology
JADAVPUR UNIVERSITY
KOLKATA-700 032

# FACULTY OF ENGINEERING & TECHNOLOGY
# JADAVPUR UNIVERSITY

## CERTIFICATE OF APPROVAL

The thesis titled **"DEVELOPMENT OF DEEP LEARNING BASED PORTABLE DEHAZING SYSTEM"** is hereby approved as a creditworthy study of an engineering subject conducted and presented satisfactorily to warrant its acceptance as a precondition to the degree for which it was submitted. It is understood that the undersigned does not automatically support or accept any argument made, opinion expressed, or inference drawn in it by this approval, but only approves the thesis for the reason it was submitted.

**Committee on Final**
**Examination for Evaluation of**
**the Thesis**

---

**Signature of External Examiner**

---

**Signature of Supervisor**

# ACKNOWLEDGEMENTS

This thesis entitled "**DEVELOPMENT OF DEEP LEARNING BASED PORTABLE DEHAZING SYSTEM**" is the result of the work whereby I have been accompanied and supported by many people, including my guide, my friends, and lab seniors. It is a pleasant aspect that now I can express my gratitude to all of them.

First and foremost, I would like to express my sincere gratitude to my thesis supervisor **Dr. Sheli Sinha Chaudhuri**, Professor of the Department of Electronics and Telecommunication Engineering, at Jadavpur University for her valuable guidance, insightful suggestions, and support while conducting this thesis work as well as in the writing of this thesis. I have been very fortunate to have a guide like her. Her positivity, confidence, and ideas helped me to complete my thesis work successfully and she guided me as a guardian. I am also very grateful to our **HOD Prof. Manotosh Biswas** for his continuous help and support and for letting me use all the available resources for my thesis work.

I would like to acknowledge the help of **Mr. Asfak Ali**, who is a research scholar in this department for helping me with the idea, implementation, and analysis. I am also thankful to my fellow project mates, friends, and technical and non-technical staff of Jadavpur University who have helped me directly or indirectly during the tenure of my thesis work.

I want to express my gratitude to my parents and family also, as, without their sacrifices, I can't do anything. And their invaluable love, encouragement, and support make me. whatever I am today.

Md. Sohel Akhtar

**Md Sohel Akhtar**
**Electronics & Telecommunication Engineering**
**Electronics & Telecommunication Engineering Department**
**Jadavpur University**
**Kolkata-32, West Bengal, India**

# Abstract

Dehazing is a crucial task in computer vision that aims to remove the effect of atmospheric haze from images. Generative Adversarial Networks (GANs) have recently emerged as a powerful solution for image restoration, including dehazing. This thesis proposes a novel approach to address the challenges of dehazing by introducing a GAN-based dehazing model. The core of the proposed model lies in the generator and discriminator components which work in an adversarial manner, constantly improving and refining the dehazing process. The generator aims to generate more realistic and visually appealing haze-free images, while the discriminator strives to become more adept at distinguishing real and generated images. The generator utilizes different architectures, such as U-Net, PSPNet, MANet, and FPN, known for their effectiveness in capturing spatial and contextual information. The goal is to assess the impact of these architectures on dehazing performance and identify the most suitable architecture for the model. Additionally, the influence of various encoder blocks, including MobileNet, EfficientNet, ResNet, and Vision Transformer, is investigated to improve feature extraction and representation learning.To train and evaluate the model, the RESIDE (Realistic Single Image Dehazing) dataset is utilized which enables training on various hazy scenarios, enhancing the model's ability to generalize to unseen hazy images. The model's performance is evaluated on benchmark datasets, including iHAZE, NHaze, and Dense Haze, and compared against state-of-the-art dehazing methods. Objective metrics, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), are employed to assess image quality in a quantitative manner. This comprehensive evaluation allows for an analysis of the model's effectiveness and the identification of the most suitable generator architecture and encoder block for dehazing tasks.Furthermore, the application of the proposed model is extended by implementing the U-Net model with MobileNet architecture on a Raspberry Pi device. This implementation allows for real-time dehazing by capturing frames from a live camera feed. This research contributes to the advancement of dehazing techniques by harnessing the power of GANs. The proposed model aims to surpass the limitations of traditional methods and enhance image quality and visibility in hazy conditions. Through the exploration of different generator architectures and encoder blocks, valuable insights are gained regarding their impact on dehazing performance. The evaluation on benchmark datasets provides a thorough analysis of the model's effectiveness. Ultimately, this research strives to improve the performance of dehazing algorithms in real-world applications where image quality and visibility are critical factors.

**Keywords**: Dehazing, Computer Vision, Generative Adversarial Networks,Image Restoration,Raspberry Pi

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Haze[2] is a natural atmospheric occurrence characterized by the presence of small particles, such as dust, smoke, and pollutants, which reduces clarity, contrast, and color saturation of the images. It happens when there is a high concentration of airborne particles in the atmosphere, which scatters and absorbs light, leading to a degradation in image quality characterized by reduced visibility, diminished contrast, and altered color fidelity. It poses significant challenges in the field of computer vision and image processing, as it introduces unwanted artifacts and affects the overall visual appeal of images.The presence of haze in images is a prevalent issue that significantly degrades their visual quality and poses challenges in computer vision applications.

The effects of haze can be observed in several ways. Firstly, haze attenuates the intensity of light, resulting in a loss of contrast and making objects located at a distance appear faded and less distinct compared to those in clear conditions. Secondly, haze introduces a color shift, often manifesting as a bluish or grayish tint, which distorts the true color representation of scenes. Lastly, haze reduces the visibility of fine details and textures, making it challenging to perceive important image features accurately.

In computer vision applications, such as object detection, recognition, and tracking, the presence of haze can severely impact the performance and accuracy of algorithms. Haze reduces the clarity and sharpness of objects, making it difficult to distinguish fine details and identify objects accurately. This can lead to errors in object localization and recognition, compromising the reliability and robustness of computer vision systems.In remote sensing applications, such as aerial and satellite imagery, haze can obstruct the observation and interpretation of important features and phenomena on the Earth's surface. Hazy conditions limit the visibility of geographical landmarks, vegetation, water bodies, and other crucial environmental elements, making it challenging to extract meaningful information and derive accurate insights from the captured data. This can impede tasks such as land cover classification, environmental monitoring, and disaster assessment.In image processing applications, such as photography, cinematography, and

video surveillance, haze can have detrimental effects on the overall visual quality and aesthetic appeal. Hazy images exhibit low contrast, washed-out colors, and reduced sharpness, diminishing their visual impact and communicative power. Moreover, haze can introduce unwanted artifacts and distortions, affecting the fidelity and authenticity of the captured scenes.Addressing the problem of haze in images has been a topic of significant research and development in recent years.

The presence of haze in images presents challenges for various computer vision tasks, including object detection, recognition, and scene understanding. Hazy images can hinder the performance of algorithms that rely on accurate color information, precise edge detection, or reliable texture analysis. Therefore, the development of effective dehazing techniques is crucial to restore the visual quality and improve the usability of hazy images.In addition to enhancing the visual quality of hazy images, dehazing techniques find applications in various domains. For example, in outdoor surveillance, haze often impairs the effectiveness of surveillance systems, limiting their ability to detect objects or individuals at a distance. Dehazing techniques can significantly improve the visibility and accuracy of surveillance cameras, thereby enhancing security and surveillance operations. In the context of autonomous driving, haze poses significant challenges as it reduces the visibility of the road and surrounding objects. Dehazing techniques can aid autonomous driving systems by improving the clarity of images captured by onboard cameras, enabling better object detection and obstacle avoidance.Remote sensing applications also benefit from dehazing methods. Hazy conditions can impact the quality of satellite imagery and aerial photographs used in remote sensing, affecting the accuracy of analysis and interpretation. Dehazing techniques play a crucial role in improving the visibility and interpretability of these images, facilitating environmental monitoring, disaster management, and urban planning efforts. Additionally, the removal of haze effect techniques are valuable in the field of image and video editing, allowing photographers and videographers to enhance the visual appeal of their captured content by eliminating haze and ensuring vibrant colors, sharp details, and overall improved image quality.

By addressing the challenges posed by haze and developing robust dehazing techniques, the visual quality of hazy images can be enhanced, improve the performance of image analysis algorithms, and enable better understanding of scenes captured under hazy conditions.This thesis aims to contribute to the field of image dehazing by proposing a GAN-based dehazing model that leverages different generator architectures and encoder blocks to enhance dehazing performance. By exploring the effectiveness of various architectural choices, the proposed model aims to provide improved image quality, enhanced visibility, and better preservation of important details and structures. The evaluation of the model on different datasets and comparison with state-of-the-art dehazing methods will help assess its capabilities and highlight its advantages in terms of objective metrics and visual quality.

## 1.1 Haze Imaging Model

The haze imaging model is based on the understanding that light undergoes scattering and absorption when it passes through a hazy medium. These interactions with atmospheric particles lead to a degradation in image quality, including reduced contrast, color fidelity, and overall visibility. The purpose of the haze imaging model is to mathematically describe this degradation process.

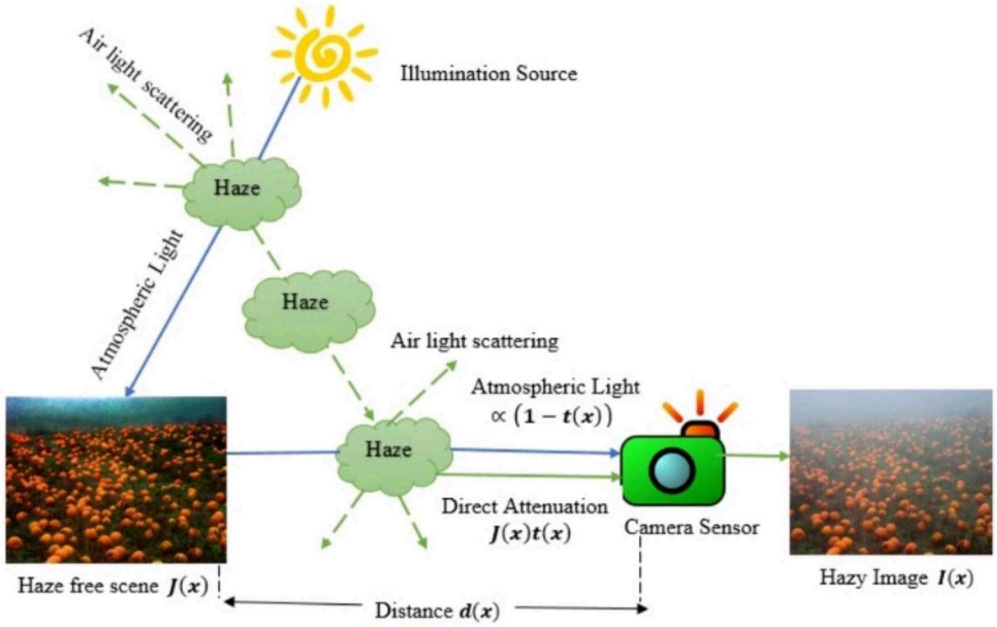


FIGURE 1.1: Image formation under hazy weather conditions and atmospheric scattering model[1].

In the haze imaging model, the observed hazy image $I(x)$ at a specific pixel $x$ can be represented as a combination of the scene radiance $J(x)$ and the atmospheric light $A$, adjusted by the transmission map $t(x)$:

$$I(x) = J(x)t(x) + A(1 - t(x))$$

The scene radiance $J(x)$ corresponds to the actual radiance values of the scene at pixel $x$ in the absence of haze. It serves as the ideal image that is aimed to recover through dehazing algorithms. The atmospheric light $A$ represents the global illumination resulting from light scattering and attenuation by atmospheric particles. It is assumed to be constant across the image.The transmission map $t(x)$ determines the fraction of scene radiance reaching the camera sensor at pixel $x$. It indicates the degree of haziness or opacity in the scene, with a value of 1 indicating no haze and a value of 0 representing complete haze. The transmission map accounts for local variations in haze density within the image.

According to the haze imaging model, the observed hazy image is a combination of the scene radiance attenuated by the haze ($J(x)t(x)$) and mixed with the atmospheric light ($A(1-t(x))$). By estimating the transmission map and separating the haze component, the clear scene radiance can be restored, leading to enhanced visibility and improved image quality.

Various dehazing techniques utilize the haze imaging model to estimate the transmission map and atmospheric light, which are essential for effective haze effect removal from hazy images. By accurately modeling the degradation process, these techniques aim to recover the true scene radiance and enhance the visual quality of hazy images.It is important to acknowledge that the haze imaging model involves simplifications and has certain limitations. It assumes a single scattering model and does not consider complex light interactions or multiple scattering effects. Nonetheless, despite these simplifications, the haze imaging model provides a valuable framework for comprehending and addressing the challenges associated with haze effect removal in different dehazing methods.

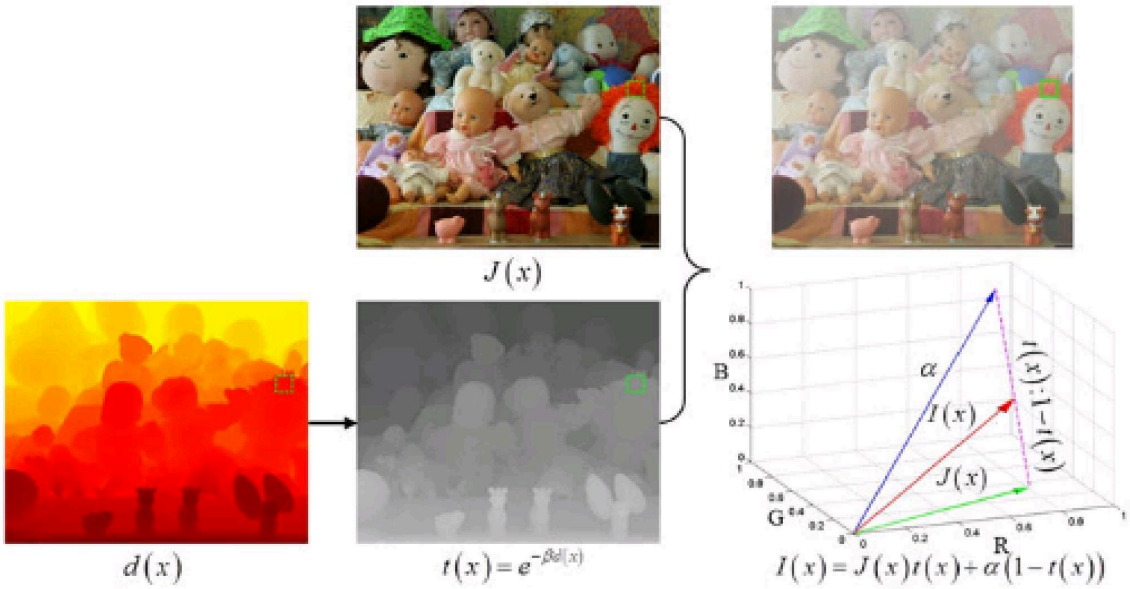


FIGURE 1.2: Different components in Atmospheric Scattering Model[1].

## 1.2 Importance of Dehazing

Dehazing techniques are of great importance in enhancing image quality and visibility by effectively removing the degrading effects of haze. Haze, caused by the scattering and absorption of light by atmospheric particles, significantly reduces the clarity, contrast, and color fidelity of images. This reduction in image quality can have a negative impact on various applications across different fields.

One crucial aspect of dehazing techniques is their ability to improve visibility. Haze obscures important details and reduces the visibility of objects, scenes, and structures. Dehazing techniques aim to remove or mitigate the effect of haze, thereby revealing hidden details and enhancing the visibility of objects and features. This is particularly crucial in applications such as surveillance, remote sensing, and aerial photography, where clear and detailed images are necessary for accurate analysis and interpretation.

Another significant aspect is the enhancement of image quality. Hazy images often suffer from low contrast, washed-out colors, and reduced sharpness. Dehazing techniques restore the original visual quality by recovering contrast, restoring colors, and sharpening details. This leads to visually appealing and more engaging images, making them suitable for various applications such as photography, cinematography, and advertising.

Dehazing techniques also play a vital role in accurate computer vision. Haze poses challenges for computer vision tasks such as object detection, recognition, and tracking. Hazy conditions reduce the clarity and sharpness of objects, making it difficult for algorithms to accurately identify and localize them. Dehazing techniques improve the performance of computer vision algorithms by enhancing the visibility of objects and improving their appearance, leading to more reliable and accurate results.

In remote sensing applications, such as satellite imagery and aerial surveys, haze can hinder the extraction of meaningful information from the captured data. Dehazing techniques enable the identification and analysis of important features on the Earth's surface, such as land cover, vegetation, water bodies, and infrastructure. This is crucial for tasks such as environmental monitoring, disaster assessment, and urban planning.

Furthermore, dehazing techniques contribute to enhanced visual communication. Haze can significantly impact the visual communication of images, videos, and multimedia content. Dehazing techniques restore the clarity and vibrancy of the content, improving its visual impact and communicative power. This is important in various domains, including media production, advertising, and virtual reality, where visually compelling and immersive experiences are desired.

Overall, dehazing techniques are essential for improving image quality and visibility. They enable the extraction of meaningful information, improve the accuracy of algorithms and systems, and enhance the overall visual experience for various applications. The development of effective dehazing techniques continues to be a critical research area, enabling advancements in computer vision, remote sensing, image processing, and other domains reliant on visual data.

There are many kinds of dehazing techniques. Image Enhancement-Based Dehazing [3] is one of the simplest dehazing techniques and involves increasing the hazy image's contrast and brightness. Histogram equalisation[4], gamma correction[5], and colour balance adjustments [6]are some of the methods that may be used to accomplish this. However, for


FIGURE 1.3: An example of hazy image and it's clear version (a) Indoor (b) Outdoor

images with a lot of haze, this approach might not be very effective. The physical model-based dehazing technique is based on how light travels through the atmosphere. The atmospheric scattering model[7] is the most commonly used physical model,which explains how airborne particles scatter light.The hazy image can be dehazed by measuring the air transmission and scene radiance. However, this method can be computationally expensive and may not work well for images with complex lighting conditions. In most natural images, there exists a dark channel that has low intensity values in haze-free regions. Dark Channel Prior-Based Dehazing[8] involves estimating the transmission pattern from the dark channel . However, this method may not work well for images with low contrast or complex scenes. Deep learning-based dehazing[9, 10, 11] is the most recent advancement in the area of computer vision, which has shown promising results in the removal of haze from images. In order to understand the mapping between hazy and clear images, these techniques involve training a deep convolutional neural network (CNN).Utilising neural network feature extraction capabilities, it analyses the characteristics of the haze to achieve the dehazing effect.

## 1.3 Potential of GAN based Dehazing

Traditional dehazing methods typically rely on assumptions and prior knowledge about the degradation process caused by haze.Algorithms based on image enhancement may result in loss of image details[6], whereas physical model-based algorithms require manual feature extraction and are not universally applicable. These methods, such as the Dark Channel Prior (DCP) and Color Attenuation Prior (CAP), exploit statistical properties and priors related to haze and scene characteristics. While these methods have shown promising results in specific scenarios, they often struggle with complex scenes, varying haze conditions, and the presence of non-uniform haze distributions. As a result, there is a growing need for more advanced and adaptive dehazing techniques. Despite the positive outcomes of deep learning-based dehazing techniques, the majority of them still

evaluate ambient light and transmittance values by using a physical scattering model through a convolutional neural network (CNN)[6]. Additionally, these methods often use mean square loss, which is calculated on a per-pixel basis, leading to image smoothing, loss of detail, and reduced image clarity.

To address these limitations, there is potential to develop more effective learning algorithms and loss functions. Some promising approaches include the use of Generative Adversarial Networks (GANs) and conditional GANs, which have demonstrated success in various image translation and reconstruction tasks. Perceptual loss functions can also be used to improve the visual quality of pictures. Based on these considerations, I have developed an trainable patch GAN based network that can eliminate haze from a single image of a scene.

GAN-based models have emerged as a promising approach in the field of dehazing, offering significant potential for improving the quality and visibility of hazy images. The use of Generative Adversarial Networks (GANs) in dehazing has gained attention due to their ability to capture complex image patterns and generate visually realistic outputs. GANs leverage the power of adversarial training, where a generator network competes against a discriminator network to produce dehazed images that closely resemble real clear images.

One of the key advantages of GAN-based models in dehazing is their ability to learn from data and generate realistic and visually pleasing results. By training on large datasets of paired hazy and clear images, GANs can capture the intricate relationships between haze and scene properties. The generator network learns to transform hazy inputs into dehazed outputs, effectively removing the haze and enhancing visibility. This learning-based approach allows GANs to handle various types and levels of haze, adapt to different scenes and lighting conditions, and generate dehazed images that exhibit improved contrast, color fidelity, and sharpness.

Comparative analysis is crucial when evaluating the effectiveness of GAN-based models in dehazing. As the field of GAN-based dehazing is relatively new and rapidly evolving, it is essential to assess and compare different architectural choices, loss functions, and training strategies. Comparative analysis helps in understanding the strengths and limitations of different GAN models and their components. It enables researchers to identify the most effective configurations and techniques for dehazing, leading to advancements in the field.

By conducting comparative analysis, researchers can evaluate the performance of different GAN architectures, such as U-Net, PSPNet, MANet, and FPN, in terms of their ability to generate high-quality dehazed images. They can investigate the impact of using different encoder blocks, such as MobileNet, EfficientNet, ResNet, and Vision

Transformer, on the dehazing performance. Comparative analysis allows for the identification of the best-performing models, which can guide future research and application development.

Furthermore, comparative analysis helps in assessing the performance of GAN-based dehazing models on different datasets. Testing the models on diverse datasets, including benchmark datasets like RESIDE, iHAZE, NHaze, and Dense Haze, provides insights into their generalization capabilities and robustness. It allows for the evaluation of their performance under varying hazy conditions, scene types, and imaging scenarios. Comparative analysis also facilitates the comparison of dehazing outcomes in terms of metrics such as PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index), providing quantitative measures of the dehazing quality.

In summary, GAN-based models hold significant potential in dehazing tasks due to their ability to generate realistic and visually appealing dehazed images. However, to fully harness this potential, comparative analysis is essential. By systematically evaluating different architectural choices, encoder blocks, loss functions, and datasets, researchers can identify the most effective configurations and techniques for GAN-based dehazing. Comparative analysis enables advancements in the field and paves the way for the development of robust and high-performance dehazing models.

## 1.4 Objective of the thesis

The objectives of this research paper are as follows:

1. To propose a GAN-based model for dehazing that effectively removes haze effect from images. The model will leverage different architectures such as U-Net, PSPNet, MANet, and FPN as generators. The objective is to explore the performance of these architectures in capturing and removing haze, and to evaluate their ability to enhance image quality and visibility.

2. To investigate the impact of different encoder blocks on the performance of the dehazing model. The encoder blocks under consideration include MobileNet, EfficientNet, ResNet, and Vision Transformer. This analysis aims to identify the encoder block that achieves the best results in terms of PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) values, indicating the quality of the dehazed images.

3. To train and evaluate the proposed GAN model using the RESIDE dataset, which consists of a diverse range of hazy images captured in various real-world scenarios. By using this dataset, the model can be trained on a comprehensive set of haze conditions and better generalize to different environments.

4. To test the proposed GAN model on other benchmark datasets such as iHAZE, NHaze, and DenseHaze, which represent different levels of haze and challenging scenarios. This evaluation will assess the model's performance and robustness across different datasets, providing a comprehensive analysis of its effectiveness.

5. To compare the outcomes of the different models based on objective metrics, including PSNR and SSIM values. These metrics will be used to quantitatively evaluate the performance of the proposed GAN model and compare it with other state-of-the-art dehazing techniques. This analysis will help identify the strengths and weaknesses of each model and determine the most effective approach for haze effect removal.

6. To conduct a qualitative analysis by visually comparing the dehazed images generated by the proposed GAN model with those obtained from other techniques. Visual assessments will be made to evaluate the clarity, contrast, and overall visual quality of the dehazed outputs. This analysis will provide insights into the realism and perceptual quality of the dehazed images.

7. To evaluate the generalization capability of the proposed GAN model by testing it on unseen or test images. This analysis will assess the model's ability to effectively remove haze effect from images that were not part of the training dataset. It will demonstrate the model's potential for real-world applications where it encounters new and unseen hazy conditions.

8. The final objective is to implement the most efficient and lightweight architecture on a Raspberry Pi device. The performance and real-time capabilities of the model will be evaluated using live camera input. This research aims to contribute to practical and effective haze effect removal solutions for applications like surveillance systems and autonomous devices.

By achieving these objectives, this research paper aims to contribute to the field of dehazing by proposing a novel GAN-based approach. The findings will provide insights into the impact of different architectures and encoder blocks on dehazing performance, and help researchers and practitioners make informed decisions in selecting suitable models for their specific dehazing tasks. Additionally, the comparative analysis with other techniques will contribute to the understanding of GAN-based dehazing techniques and their potential in enhancing image quality and visibility.

## 1.5 Overview of the thesis work

In this thesis, a GAN-based dehazing model is proposed to enhance image quality and visibility in hazy conditions. The model utilizes GANs to learn the mapping between hazy and haze-free images, enabling the generation of high-quality dehazed outputs.

Different generator architectures, namely U-Net, PSPNet, MANet, and FPN, are employed in this research due to their effectiveness in capturing spatial and contextual information. The goal is to investigate the influence of these generator architectures on the dehazing performance and identify the most suitable architecture for the proposed model.Additionally, the impact of different encoder blocks on the dehazing process is studied. Encoder blocks play a crucial role in feature extraction and representation learning. Popular encoder blocks such as MobileNet, EfficientNet, ResNet, and Vision Transformer are considered in this research to extract meaningful features and representations from the hazy input images. By evaluating the effectiveness of different encoder blocks, the aim is to assess their ability to preserve important details and structures in the dehazed outputs. This analysis provides insights into how the choice of encoder block influences the performance of the proposed GAN-based dehazing model.

To train and evaluate the proposed model, the RESIDE (Realistic Single Image Dehazing) dataset is utilized. The RESIDE dataset provides a diverse range of hazy and corresponding haze-free images, which allows for training the model on various hazy scenarios and enhances its generalization to unseen hazy images. During the training process, a combination of loss functions is employed to guide the model. These include Mean Squared Error (MSE) loss, perceptual loss, and adversarial loss. By utilizing these loss functions, the generator is encouraged to generate dehazed outputs that are both visually accurate and realistic.In addition to training on the RESIDE dataset, the performance of the model is evaluated on several benchmark datasets, including iHAZE, NHaze, and Dense Haze. This evaluation enables the assessment of the model's generalization capabilities across different hazy scenarios and facilitates a comparison with other state-of-the-art dehazing methods. The comparison is based on quantitative metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), which provide objective measures of image quality and allow for a comprehensive analysis of the model's performance.

In addition to the contributions mentioned earlier, this research also includes the implementation of the U-Net model with MobileNet architecture on a Raspberry Pi device for real-time dehazing. The choice of using MobileNet architecture is motivated by its lightweight nature and higher performance, making it suitable for deployment on edge devices with limited computational resources.The implemented model takes input from a live camera feed, allowing for on-the-fly dehazing of captured frames. This real-time capability is particularly valuable in applications such as surveillance cameras, drones, or any scenario where immediate visibility enhancement is required. By executing the dehazing process directly on the Raspberry Pi, the need for transmitting data to a remote server for processing is eliminated, thereby reducing latency and enabling rapid decision-making based on dehazed images.The integration of the U-Net model with MobileNet architecture on the Raspberry Pi platform adds a practical dimension to this research.

It showcases the potential for real-time haze effect removal in resource-constrained environments, where traditional cloud-based solutions may not be feasible or practical. The implementation on Raspberry Pi not only validates the effectiveness of the proposed model but also opens up opportunities for real-world deployment in various scenarios that demand immediate and efficient dehazing capabilities.

## 1.6 Outline of the thesis

The thesis is organized into several chapters, each focusing on a specific aspect of the research on GAN-based dehazing techniques. The outline of the thesis is as follows:

1. Introduction
    - Overview of haze effect removal and its significance in various applications
    - Explanation of the hazing model and its impact on image quality and visibility
    - Importance of dehazing techniques in enhancing image quality and restoring visibility
    - Potential of GAN-based approaches in addressing the challenges of haze effect removal
    - Clear statement of objectives for the thesis
2. Motivation and Contribution
    - Discussion of the motivation behind the research on haze effect removal techniques
    - Explanation of the specific contributions of the thesis
3. Related Work
    - Review of prior-based, learning-based, deep learning-based, and GAN-based approaches for haze effect removal
    - Explanation of the principles, techniques, and contributions of each approach
    - Comparison of strengths and weaknesses of different methods
    - Elaboration of Generative Adversarial Network
4. Proposed Method
    - Description of the GAN-based method for dehazing
    - Explanation of the architecture of the generator and discriminator networks
    - Elaboration on the different architectures employed for the generator (UNet, PSPNet, FPN, MA-Net)

- Discussion of the encoder blocks used (MobileNet, EfficientNet, ResNet, Vision Transformer)
- Detailed explanation of the loss function employed

5. Real Time Dehazing on Raspberry pi
    - Implementation of the lightweight U-Net model with MobileNet architecture.
    - Description of the hardware setup and software configuration
    - Performance analysis and evaluation of real-time dehazing on Raspberry Pi
6. Experimental Setup
    - Overview of the datasets used for training and evaluation (RESIDE, iHAZE, NHaze, Dense Haze)
    - Description of preprocessing steps applied to the datasets
    - Discussion of training details (optimizer, learning rate schedule, number of epochs)
    - Specification of performance metrics used for evaluation (PSNR, SSIM)
7. Evaluation
    - Presentation of quantitative results using performance metrics
    - Comparison of outcomes among different models and encoders
    - Qualitative evaluation through visual comparisons with original hazy images and state-of-the-art techniques
8. Conclusion
    - Summary of contributions and key findings of the research
    - Suggestion of future research directions

# Chapter 2

# Motivation and Contribution

This chapter presents the motivation behind the research conducted in this thesis and outlines the specific contributions made in the field of dehazing using GAN-based models.

## 2.1 Motivation

Haze is a common atmospheric phenomenon that significantly impacts the quality and visibility of images. It occurs due to the scattering and absorption of light by particles and pollutants in the atmosphere, resulting in reduced contrast, color distortion, and loss of fine details. Hazy images pose significant challenges in various domains, including computer vision, remote sensing, and autonomous navigation systems.

In computer vision and image processing, haze obstructs object detection, recognition, and tracking algorithms, leading to degraded performance and compromised accuracy. In remote sensing and aerial imagery, haze reduces the visibility of important features, such as terrain, buildings, and landmarks, hindering applications like environmental monitoring, disaster management, and map generation. In autonomous driving systems, haze impairs the detection of obstacles, road signs, and pedestrians, posing safety risks. Even in everyday photography, haze can ruin the aesthetic appeal and visual impact of images, affecting both amateur and professional photographers.

The importance of dehazing techniques lies in their ability to restore image quality, enhance visibility, and improve the interpretability of visual data captured in hazy conditions. By effectively removing haze, these techniques enable better analysis, decision-making, and understanding of visual information. They help reveal hidden details, restore natural colors, and improve the overall perceptual quality of images, making them invaluable in a wide range of applications.

The motivation behind this research stems from the critical need to develop effective dehazing techniques that can restore the true visual appearance of hazy images.Traditional

dehazing methods have made significant contributions to mitigating the effects of haze. However, they often rely on handcrafted priors or assumptions about the scene and haze properties, which limit their applicability to diverse and complex hazy scenes. Additionally, they may struggle to preserve fine details, result in color artifacts, or exhibit limited robustness to varying haze conditions. As a result, these methods may produce suboptimal dehazing results, limiting their practical applicability.Deep learning approaches, particularly GAN-based models, have shown great promise in various image synthesis tasks, including dehazing. GANs offer a powerful framework for learning complex mappings between hazy and clear images by leveraging adversarial training. They can capture the underlying distribution of haze-free images and generate visually realistic dehazed outputs. The motivation for this research is to explore the potential of GAN-based models in addressing the challenges of haze effect removal and enhancing image quality and visibility.

The motivation for exploring GAN-based dehazing techniques stems from the potential advantages they offer. GANs are capable of learning complex mappings between hazy and clear images without explicitly modeling the degradation process. They can capture intricate image details, textures, and structures, leading to visually realistic and high-quality dehazed outputs. GANs also have the ability to generalize well to unseen or test images, making them suitable for real-world applications.

By proposing a GAN-based dehazing model in this thesis, the goal is to contribute to the development of advanced techniques that overcome the limitations of traditional methods and provide superior dehazing results. The use of different generator architectures, such as U-Net, PSPNet, MANet, and FPN, along with various encoder blocks like MobileNet, EfficientNet, ResNet, and Vision Transformer, allows for the exploration of different model configurations and their impact on dehazing performance.

The motivation extends to the evaluation and comparison of the proposed model and different encoder configurations using multiple datasets. By assessing the dehazing outcomes on datasets such as RESIDE, iHAZE, NHaze, and Dense Haze, the thesis aims to provide a comprehensive understanding of the model's performance in various hazy scenarios. Performance metrics, including peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), will be utilized to objectively measure the quality and fidelity of the dehazed images.

The motivation behind this research stems from the increasing demand for effective haze effect removal techniques in various applications, coupled with the need for efficient and real-time implementations on resource-constrained devices. Haze in images significantly degrades their quality and visibility, impacting tasks such as image analysis, object detection, and scene understanding. Therefore, there is a strong motivation to develop

advanced dehazing methods that can enhance image quality and visibility in hazy conditions.Furthermore, the deployment of dehazing models on edge devices, such as Raspberry Pi, has gained significant attention. Edge devices offer the advantage of processing data locally, reducing latency and dependence on cloud-based solutions. However, the limited computational resources of edge devices pose challenges in implementing complex models. Therefore, there is a need for lightweight and high-performance architectures that can be effectively deployed on edge devices for real-time dehazing tasks.

In line with this motivation, this research explores the implementation of the U-Net model with MobileNet architecture on a Raspberry Pi device. MobileNet is known for its lightweight characteristics and higher performance, making it a suitable choice for resource-constrained devices. By taking input directly from a live camera feed, the implemented model enables real-time dehazing with a processing speed of 1 frame per second on the Raspberry Pi. This capability opens up possibilities for various real-time applications, such as surveillance systems, robotics, and augmented reality, where immediate and efficient haze effect removal is crucial.

By addressing the need for both effective dehazing techniques and their efficient deployment on edge devices, this research aims to contribute to the field of haze effect removal and extend its practical applications. The combination of the U-Net model with MobileNet architecture on a Raspberry Pi showcases the potential of lightweight and high-performance models for real-time dehazing, providing valuable insights and paving the way for future advancements in image enhancement and edge computing.The findings of this research can have practical implications in various fields, including computer vision, remote sensing, autonomous systems, and photography, leading to improved performance, safety, and visual experiences.

## 2.2 Contribution

The primary contribution of this thesis is the proposal of a GAN-based model for dehazing that leverages different generator architectures and encoder blocks. The following are the specific contributions made in this research:

1. Proposed GAN-based Dehazing Model: The thesis introduces a novel GAN-based model that is specifically designed for dehazing tasks. This model harnesses the power of Generative Adversarial Networks to effectively remove haze effect from images and enhance their visual quality. The proposed model comprises a generator network, responsible for generating dehazed images, and a discriminator network, which distinguishes between real clear images and the generated dehazed images. By leveraging the adversarial learning framework of GANs, the proposed model learns to produce dehazed outputs that closely resemble real clear images. This contribution provides a new and promising approach to address the challenging problem of haze effect removal.

2. Investigation of Different Generator Architectures: The thesis conducts an in-depth exploration and comparison of different generator architectures in the context of dehazing. Specifically, the UNet, PSPNet, MANet, and FPN architectures are evaluated. Each architecture is carefully analyzed in terms of its structure, capacity to capture and retain relevant image information, and ability to generate visually pleasing dehazed outputs. The performance of each architecture is assessed on various benchmark datasets, allowing for a comprehensive understanding of their strengths and weaknesses in dehazing tasks. This contribution provides valuable insights into the suitability of different generator architectures for dehazing and assists researchers and practitioners in selecting the most appropriate architecture for their specific applications.

3. Evaluation of Different Encoder Blocks: The thesis investigates the impact of different encoder blocks on the performance of the dehazing model. Various encoder blocks, including MobileNet, EfficientNet, ResNet, and Vision Transformer, are examined and compared. The encoder blocks play a crucial role in capturing and representing image features, which directly affect the quality of the dehazed outputs. By evaluating different encoder blocks, the thesis provides insights into their capabilities in handling haze and their effectiveness in preserving important image details. This contribution aids in understanding the importance of encoder block selection in improving the overall dehazing performance.

4. Real-time Dehazing on Raspberry Pi: To extend the application of the proposed model, an implementation of the U-Net model with MobileNet architecture is carried out on a Raspberry Pi device. This implementation enables real-time dehazing by taking input from a live camera feed. The achieved processing speed of 1 frame per second demonstrates the feasibility of deploying the proposed model on resource-constrained edge devices, such as Raspberry Pi.

5. Dataset Selection and Evaluation: To ensure a comprehensive evaluation of the proposed dehazing model, the thesis utilizes multiple datasets, including RESIDE, iHAZE, NHaze, and DenseHaze. These datasets cover a wide range of hazy conditions and scene characteristics, allowing for a thorough assessment of the model's performance in various real-world scenarios. By evaluating the model on diverse datasets, the thesis demonstrates its robustness and generalizability across different hazy conditions. This contribution ensures that the proposed approach can effectively handle a variety of hazy environments and provides a reliable basis for comparing the performance of different models and encoder blocks.

6. Performance Metrics and Comparative Analysis: The thesis employs a comprehensive set of performance metrics, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), to quantitatively evaluate and compare the outcomes of different models and encoder blocks. These metrics provide objective measures of image quality and similarity between the dehazed outputs and ground truth clear images. By

conducting a thorough comparative analysis of the dehazing results, the thesis highlights the strengths and weaknesses of each model and encoder combination. This comparative analysis allows for a comprehensive understanding of the effectiveness of the proposed approach and facilitates informed decision-making in selecting the best model and encoder for dehazing tasks.

In summary, the thesis makes significant contributions to the field of dehazing based on GANs by proposing a novel model, exploring different generator architectures and encoder blocks, evaluating the performance on diverse datasets, and providing a comprehensive comparative analysis of the results. These contributions advance the knowledge and understanding of GAN-based dehazing techniques, offering insights into their effectiveness and applicability in enhancing image quality and visibility in hazy conditions. The findings of this thesis pave the way for further advancements in the field and contribute to the development of more effective dehazing algorithms.

# Chapter 3

# Literature Review

## 3.1 Existing Dehazing Techniques

haze effect removal from hazy images has emerged as a vibrant and active research domain within the field of computer vision and image processing. Numerous techniques have been proposed, including prior-based methods[3, 7, 8], learning-based methods [9, 10, 11], and the utilization of generative adversarial networks (GANs) [12, 13, 14, 15, 16]. Here, we present a comprehensive overview of these approaches and highlight their contributions to the field.

### 3.1.1 Prior-Based Dehazing Techniques

Prior-based dehazing techniques belong to a category of methods that utilize statistical assumptions and prior knowledge about the haze and scene properties to estimate the transmission map and atmospheric light for haze effect removal. These techniques aim to recover the original scene radiance by modeling the degradation caused by haze.

The Dark Channel Prior (DCP) is a widely used prior-based method that assumes the intensity of the dark channel, which represents the minimum value across color channels, is close to zero in haze-free regions. DCP analyzes the dark channel of a hazy image to estimate the lower bound of the transmission map. This lower bound indicates the maximum haze density within local patches of the image. DCP assumes a uniform haze distribution within each patch and employs a soft matting operation that considers local haze density and patch intensity range to estimate the transmission map.Another popular prior-based technique is the Color Attenuation Prior (CAP), which capitalizes on the observation that haze primarily affects the color channels of an image, leading to color distortion. CAP assumes that the ratio between hazy and haze-free image intensities remains constant for pixels in the same scene region. By analyzing the color

attenuation caused by haze, CAP estimates the transmission map by comparing the color distortion between hazy and haze-free regions. It uses the ratio of color channel intensities to estimate the transmission map, assuming a uniform effect of haze on each color channel.Multi-scale fusion methods extend the prior-based approach by considering information from multiple scales to improve transmission map estimation. These methods perform haze effect removal at various scales by using a pyramid or multi-scale representation of the image. The transmission maps estimated at different scales are then fused to obtain a refined transmission map. By incorporating different scales, multi-scale fusion methods can handle variations in haze density across different image regions. They combine the coarse-scale transmission map estimation with fine-scale details to achieve enhanced dehazing performance.

Dark Channel Prior(DCP)[17], introduced by He et al. , leverages the dark channel prior to estimate the transmission map. DCP-based methods have demonstrated strong performance in haze effect removal, providing a foundation for subsequent research in the field.FFA-Net[9] (Zhang et al., 2019) utilized feature fusion modules and attention blocks to capture contextual information and emphasize relevant features. AECR-Net[10] developed by Zhang et al., incorporated adaptive enhancement and context reasoning modules to improve dehazing performance. PFF-Net( proposed by Cai et al.[11], uses a pyramid feature fusion module to extract multi-level features from the ground truth image for estimatin the transmission pattern and then generating the dehazed image. Liu et al.[18] presents a method based on channel attention and spatial transform networks,which uses channel attention to extract the most informative features.

Researchers are actively exploring advanced prior-based methods by incorporating additional priors, refining existing assumptions, and investigating hybrid approaches that integrate prior knowledge with machine learning techniques. By leveraging statistical priors and exploiting the inherent characteristics of haze, prior-based dehazing techniques contribute to the development of effective algorithms for haze effect removal and restoration of hazy images.

### 3.1.2 Learning-Based Dehazing Techniques

Learning-based methods in dehazing involve using machine learning algorithms, particularly convolutional neural networks (CNNs), to learn the relationship between hazy and clear images. These methods have gained attention due to their ability to capture complex patterns and effectively remove haze effect.

The basic concept of learning-based dehazing is to train a model using a dataset of hazy images and their corresponding clear versions. During training, the model's parameters are optimized to minimize the difference between the predicted clear image and the

ground truth clear image. By learning from diverse hazy-clear image pairs, the model can generalize and dehaze unseen or test images.

Regression-based modeling is one approach in learning-based dehazing. It involves learning a mapping function that predicts the clear image from the hazy input. Regression models like support vector regression (SVR) or random forest regression (RFR) can be used to estimate this mapping. These models consider image features, statistical properties, or handcrafted priors to capture the relationship between hazy and clear images. The models are trained using input-output pairs and then applied to unseen hazy images to generate their clear counterparts.Convolutional neural networks (CNNs) are another powerful approach in learning-based dehazing. Designed for visual data, CNNs excel in dehazing tasks. Their architecture consists of interconnected layers that learn hierarchical representations of input images. In dehazing, CNNs are trained to directly estimate the clear image from the hazy input, eliminating the need for explicit modeling of intermediate variables. The network learns the transformations required to remove haze effect by optimizing its parameters through a training process.

Notable learning-based dehazing techniques include DehazeNet, AOD-Net (Atmospheric Optical Depth Network), and DCPDN (Deep Convolutional Prior and Dehazing Network). These methods showcase the effectiveness of learning-based approaches in enhancing visibility and image quality, making them valuable for computer vision, remote sensing, and image processing applications. Li et al. presented AOD-Net[19], a convolutional neural network (CNN) architecture designed for single-image dehazing. The method estimates the transmission and illumination properties using a joint optimization framework. AOD-Net achieved notable performance by leveraging the power of CNNs in learning haze-related features. Zhang et al. introduced a multi-scale fully convolutional network(MS-CNN)[20] for dehazing. The MS-CNN utilizes the scale-invariant features of CNNs to improve the performance of dehazing in different atmospheric conditions. This approach is capable of handling non-uniform atmospheric light and outperforms Li et al.'s [19] method on benchmark datasets. DehazeNet[21] employed a CNN architecture that employs a transmission-guided layer-wise propagation strategy to eliminate the haze. Gated Fusion Network(GFN) [22], proposed by Ren et al., aimed to enhance dehazing performance by effectively fusing multi-scale information. GFN utilized a gated fusion module to selectively integrate features from different scales, improving the model's ability to handle variations in haze density across the image. These learning-based approaches harnessed the power of machine learning and CNNs, demonstrating significant progress in generating haze-free image.

### 3.1.3 GAN-based Dehazing Techniques

GAN-based dehazing techniques utilize Generative Adversarial Networks (GANs) to address the problem of haze effect removal in images. GANs consist of a generator network

and a discriminator network. The generator takes a hazy image as input and aims to generate a dehazed output that closely resembles a clear image. The discriminator, on the other hand, is trained to differentiate between the generated dehazed images and real clear images.

During the training process of GAN-based dehazing methods, an adversarial learning scheme is employed. The generator and discriminator networks compete against each other in a min-max game. The generator's objective is to produce dehazed images that can deceive the discriminator into classifying them as real clear images. Conversely, the discriminator's goal is to accurately distinguish between real and generated images. This adversarial training process encourages the generator to generate dehazed images that are visually indistinguishable from real clear images.To train a GAN, a large dataset of paired hazy and clear images is required. This dataset can be obtained by either capturing real-world hazy images or generating synthetic hazy images using physically-based haze models. The paired images are utilized to compute pixel-wise differences between the hazy and clear images, serving as training targets for the generator network. The generator learns to minimize these differences and generate dehazed images that are perceptually similar to the clear images.In addition to the adversarial loss, GAN-based dehazing methods often incorporate additional loss functions to guide the training process and ensure the production of visually pleasing dehazed images. Commonly used loss functions include content loss, which measures the similarity between features extracted from the generated and real clear images, and perceptual loss, which captures high-level image characteristics. These additional losses aid in preserving important details and structures in the dehazed images.During the inference stage, the trained generator network is applied to unseen hazy images to produce dehazed outputs. The generator processes the hazy input through its learned layers and applies the acquired transformations to remove haze effect and enhance visibility. The resulting dehazed images exhibit improved contrast, reduced haze effects, and enhanced details compared to the original hazy inputs.

Recently, the use of GAN-based feature extraction methods has yielded promising outcomes for different image processing tasks. For example, the GAN-based DehazeGAN [15], introduced by Hu et al., addressed haze effect removal by utilizing a generator-discriminator framework. DehazeGAN employed a conditional GAN to learn the mapping between hazy and ground truth images, generating dehazed outputs with improved visual quality. The discriminator network facilitated the generation of more realistic dehazed images, enhancing the overall dehazing performance. Furthermore DCPDN[12] utilized a GAN-based architecture to dehaze images by learning the underlying context prior. The network employed an encoder-decoder structure combined with an adversarial loss to generate visually pleasing haze-free results. By effectively capturing the context prior information, DCPDN significantly improved dehazing performance. Following this idea, Wang et al.[13] proposed a GAN-based technique for dehazing that exploits a

feature extractor network to obtain high-level features from the hazy image. The feature extractor network is trained within a GAN-based framework, where the generator network generates high-quality feature maps and the discriminator network verifies the quality of the synthesized features. The method attained state-of-the-art results on benchmark datasets and exhibited resilience to various types of haze. DehazeFormer [14] proposed by Liu et. al., adopted a transformer-based architecture, leveraging self-attention mechanisms to capture global contextual information and exploit dependencies between image patches. These deep learning-based approaches have pushed the boundaries of haze effect removal and showcased the potential of deep learning techniques in handling complex haze conditions.

GAN-based dehazing techniques offer several advantages over traditional methods. They are capable of learning complex mappings between hazy and clear images, enabling the removal of various types and levels of haze. GANs excel at capturing fine details and structures, resulting in visually realistic dehazed results. Moreover, GAN-based methods demonstrate good generalization to unseen or test images, making them applicable in real-world scenarios.

Nevertheless, GAN-based dehazing techniques also face challenges and limitations. Training GANs requires a substantial amount of data and computational resources, as well as careful hyperparameter tuning. GAN training can be unstable and susceptible to issues such as mode collapse, where the generator fails to produce diverse outputs. Generating a diverse and representative training dataset that covers a wide range of hazy conditions can be challenging. Additionally, interpreting and controlling GAN-based dehazing models may be less straightforward compared to traditional methods based on explicit mathematical models.Researchers continue to explore and enhance GAN-based dehazing methods by developing novel architectures, loss functions, and training strategies. Adapting GANs to specific application domains and addressing the challenges associated with GAN training remain active areas of research. GAN-based dehazing techniques hold great potential for improving image quality, enhancing visibility, and advancing various applications in computer vision, remote sensing, and image processing.

## 3.2 Strengths and limitations of Existing techniques

The limitation of prior-based dehazing methods is their reliance on specific assumptions about haze and scene properties. These assumptions may not hold true in all scenarios, leading to inaccurate transmission map estimation and compromised dehazing results, especially in complex scenes where the assumptions about haze distribution or color distortion may not be valid. Nonetheless, prior-based techniques have demonstrated promising outcomes in numerous practical applications and provide valuable

insights into the dehazing process.Learning-based methods have demonstrated superior performance compared to traditional approaches in terms of accuracy and visual quality. They capture complex relationships between hazy and clear images, adapt to various haze types, and handle complex scenes. However, learning-based methods require substantial training data and computational resources. Availability of suitable datasets and computational complexity are important considerations when applying learning-based dehazing.

GAN-based dehazing techniques offer several key advantages over prior-based and learning-based methods, making them an exciting and promising approach for enhancing image quality and visibility in the context of dehazing

Firstly, GAN-based methods have the ability to capture complex mappings between hazy and clear images. Traditional prior-based methods often rely on simplistic assumptions about haze and scene properties, which may not hold in all scenarios. Learning-based methods, on the other hand, require a large amount of training data and may struggle with generalization to unseen images. GANs, with their adversarial training process, can learn the underlying distribution of clear images directly from the data, enabling them to capture intricate relationships and handle various types and levels of haze.

Secondly, GANs are capable of preserving important details and structures in dehazed images. Prior-based methods typically rely on explicit modeling of intermediate variables like transmission maps, which may introduce artifacts and lead to information loss. Learning-based methods, although effective in some cases, may struggle with accurately representing fine details. GANs, with their ability to generate high-resolution and visually realistic outputs, can retain and enhance the fine-grained information in the dehazed images, resulting in improved image quality and visibility.

Furthermore, GAN-based dehazing methods can generalize well to unseen or test images. Prior-based methods and some learning-based methods are often tailored to specific assumptions or training data, limiting their applicability to new and diverse scenarios. GANs, by learning the underlying data distribution, can generate dehazed images that exhibit realistic visual characteristics, making them suitable for a wide range of real-world applications.

Another advantage of GAN-based dehazing is their flexibility in terms of architecture and loss functions. Researchers can explore and design various network architectures, such as different generator and discriminator configurations, to improve the performance of the dehazing models. Additionally, different loss functions can be utilized, including content loss, perceptual loss, or adversarial loss, to guide the training process and ensure the generation of visually pleasing dehazed images.

It is important to note that GAN-based dehazing techniques are not without their challenges. Training GANs requires a large amount of data and computational resources, as

well as careful tuning of hyperparameters. GAN training can be unstable and prone to issues like mode collapse or lack of diversity in the generated outputs. Additionally, generating a diverse and representative training dataset that covers a wide range of hazy conditions can be challenging. However, researchers continue to explore and develop novel architectures, loss functions, and training strategies to overcome these challenges and improve the effectiveness of GAN-based dehazing methods.

Deep learning-based methods have gained significant success in removing haze from hazy images. CNN-based methods have demonstrated favorable outcomes in single image dehazing, whereas GAN-based methods have obtained state-of-the-art results by generating high-quality haze-free images. Researchers have also proposed several variations of these techniques to address their limitations and achieve improved performance. GAN-based feature extraction methods have shown great potential for dehazing.However, a common issue with GAN-based dehazing methods is the underutilization of feature information during the process of generating a haze-free image. Typically, the generator extracts features from the hazy image through a multi-layer encoder, uses a feature converter to convert between the hazy and clear image features, and then employs a decoder to reconstruct the haze-free image. Motivated by the ideas presented in [14, 23, 24, 15], this paper introduces a novel single-image dehazing algorithm that utilizes a generative adversarial network based on a Unet network. Theproposed method improves the feature utilization within the generator, resulting in enhanced image dehazing quality and efficiency.

## 3.3 Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a class of deep learning models that consist of two neural networks: a generator and a discriminator. GANs were introduced by Ian Goodfellow and his colleagues in 2014 and have since gained significant attention and popularity in the field of machine learning.

The generator network takes random noise or a latent vector as input and transforms it into a complex output sample, typically an image. Its objective is to learn the underlying data distribution and generate synthetic samples that resemble real data from a target distribution. The generator consists of several layers, such as fully connected layers, convolutional layers, or a combination of both, to progressively transform the input noise into a higher-dimensional output. The architecture of the generator can vary depending on the specific task and dataset.The generator's goal is to generate samples that are convincing enough to fool the discriminator into classifying them as real. Through the training process, the generator learns to produce samples that capture the patterns, structures, and statistical properties of the training data. By optimizing its parameters, the generator aims to minimize the discernibility between the generated samples and

real data.On the other hand, the discriminator network, sometimes referred to as the critic, is responsible for distinguishing between real and fake samples. It takes as input either a real sample from the target distribution or a synthetic sample generated by the generator. The discriminator's role is to classify whether the input sample is real or fake. Typically, the discriminator is a binary classifier that outputs a probability score indicating the likelihood of the input being real or fake.

The discriminator is trained on a combination of real and generated samples. Its objective is to accurately discriminate between real and fake samples, thus learning to capture the distinguishing characteristics and statistical properties of the real data. As training progresses, the discriminator's ability to differentiate between real and fake samples improves.The training process involves an adversarial game between the generator and discriminator. The generator aims to generate samples that the discriminator cannot distinguish from real data, while the discriminator aims to accurately classify the samples as real or fake. The generator and discriminator are updated iteratively, with each update pushing them closer to their respective objectives. This adversarial learning process encourages the generator to generate more realistic samples, while the discriminator becomes more adept at distinguishing between real and fake samples.The interplay between the generator and discriminator during training leads to a competitive relationship, where the generator learns to produce samples that are increasingly similar to real data, and the discriminator becomes more skilled at identifying generated samples. This iterative process continues until a desired level of realism is achieved, or convergence is reached.

Together, the generator and discriminator form a powerful framework for generative modeling. The generator learns to capture the underlying data distribution and generate samples, while the discriminator acts as a critical evaluator, providing feedback to the generator to improve the quality of generated samples. This adversarial interplay drives the GAN towards generating high-quality and diverse samples that resemble the real data distribution.The training process of GANs involves an adversarial game between the generator and the discriminator. The generator aims to generate synthetic samples that are indistinguishable from real samples, while the discriminator tries to improve its ability to differentiate between real and synthetic samples. The generator and discriminator are trained iteratively, with each update pushing them closer to achieving their respective goals. This adversarial learning process drives the generator to improve its ability to generate more realistic samples, while the discriminator becomes more adept at discriminating between real and fake samples.

One of the key advantages of GANs is their ability to generate highly realistic and diverse samples. The generator network learns to capture the underlying data distribution and generate samples that exhibit the same characteristics as the training data. GANs have been successfully applied to various domains, including image generation, text generation, music synthesis, and more. GANs have also been extended to conditional GANs

(cGANs), where additional information or constraints are provided to both the generator and discriminator. This allows for more controlled and targeted generation, such as generating images conditioned on specific attributes or generating images based on textual descriptions.GANs have revolutionized many areas of research and have led to significant advancements in image generation, data augmentation, and other related tasks. They have also been used in combination with other techniques, such as in GAN-based dehazing, where GANs are employed to remove haze effect and enhance visibility in images. However, training GANs can be challenging and often requires careful hyperparameter tuning, large amounts of training data, and extensive computational resources. GAN training can be unstable, and issues like mode collapse, where the generator fails to produce diverse outputs, can occur. Researchers are actively exploring techniques to address these challenges and improve the stability and performance of GANs.

Overall, GANs are a powerful framework for generative modeling, enabling the generation of realistic and diverse samples. Their ability to capture complex data distributions and generate high-quality samples has made them a popular choice in various domains, including image generation, style transfer, super-resolution, and dehazing, among others.

# Chapter 4

# Methodology

The proposed GAN model for dehazing in this thesis utilizes a variety of architectures for the generator and explores different encoder blocks to enhance its performance. The generator, responsible for transforming hazy images into dehazed ones, incorporates different architectures, including U-Net, PSPNet, MANet, and FPN, to leverage their unique capabilities in capturing and reconstructing image details.Each generator architecture is coupled with different encoder blocks, such as MobileNet, EfficientNet, ResNet, and Vision Transformer, which serve as feature extractors to capture the underlying characteristics of hazy images effectively. These encoder blocks play a crucial role in learning discriminative representations and facilitating the generation of high-quality dehazed outputs.

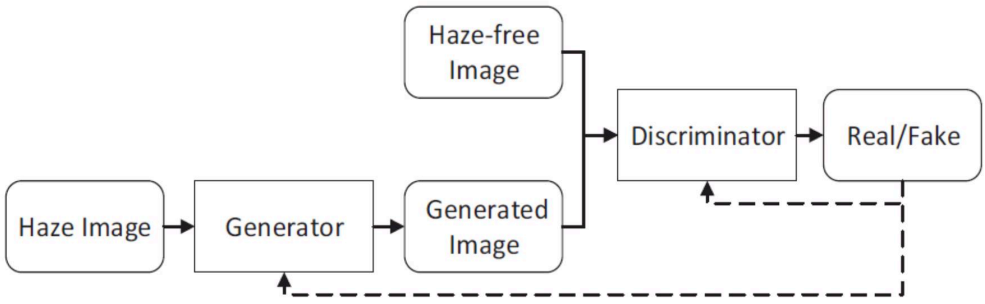


FIGURE 4.1: The Architecture of dehazing algorith using GAN

During the training process, a well-curated dataset, such as the RESIDE dataset, is employed to train the proposed GAN model. The RESIDE dataset consists of a diverse range of hazy images, enabling the model to learn from various hazy conditions and generalize its dehazing capabilities to different scenarios.To guide the training and optimize the performance of the generator, multiple loss functions are utilized. The losses employed in the generator include MS-SSIM loss, L1 loss, and BCEWithLogits loss. The MS-SSIM loss focuses on preserving structural information and perceptual

quality, while the L1 loss emphasizes pixel-wise similarity between the generated and ground truth images. The BCEWithLogits loss aids in generating dehazed outputs that are visually similar to real clear images.Furthermore, the discriminator, which plays a crucial role in the adversarial learning process, is trained using the BCEWithLogits loss. By distinguishing between real and generated images, the discriminator provides feedback to the generator, facilitating the generation of more realistic and visually pleasing dehazed outputs.

Overall, the proposed GAN model for dehazing combines different generator architectures, encoder blocks, loss functions, and utilizes the RESIDE dataset for training. This comprehensive approach aims to enhance the model's ability to effectively remove haze effect from images, improve visibility, and produce visually realistic dehazed outputs.

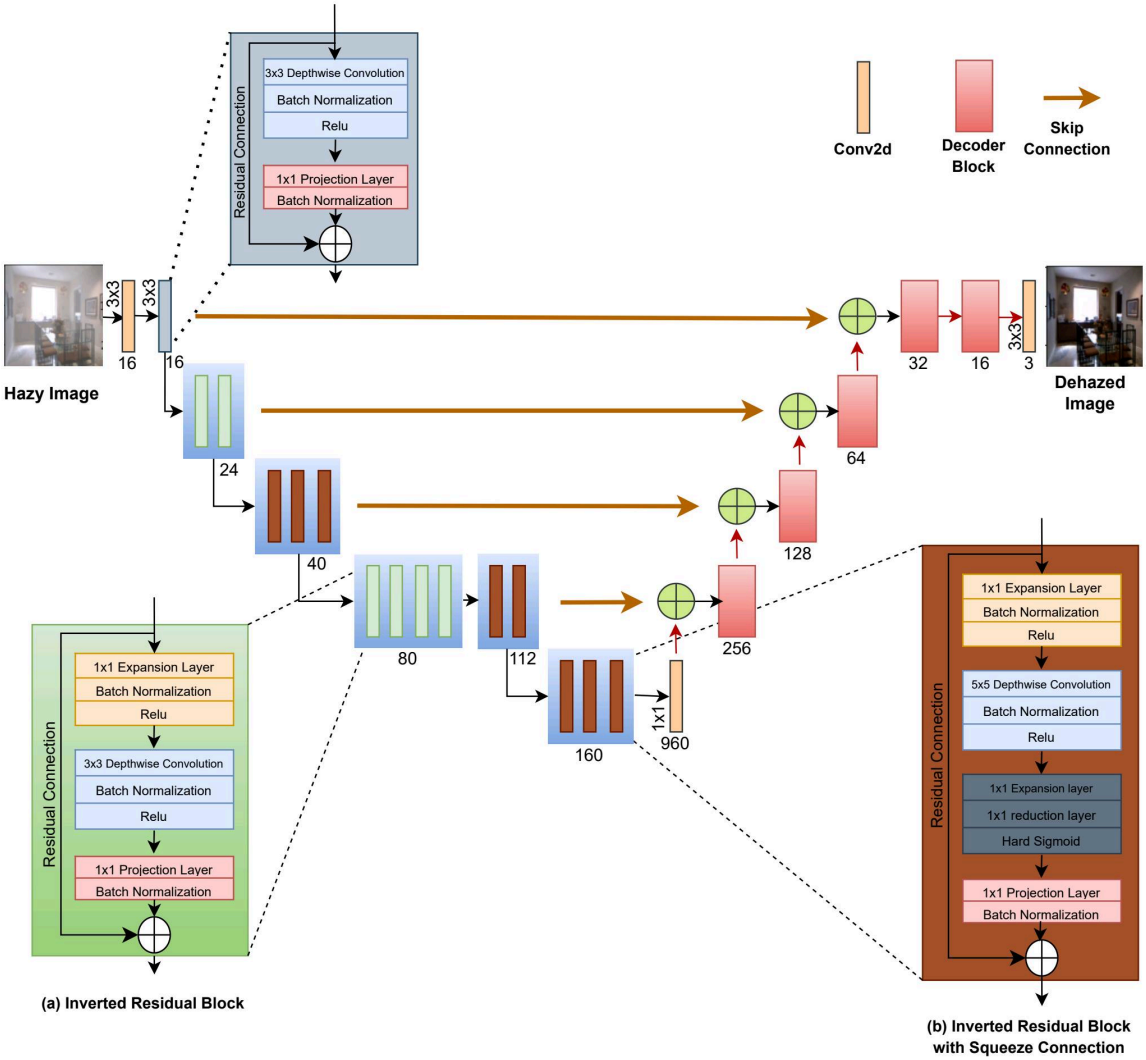


FIGURE 4.2: Illustration of Proposed Generator Architecture-(a)Inverted Residual Block. (b) Decode Block. (c)Inverted Residual Block with Squeeze Connection.

## 4.1 Generator

The generator in the proposed GAN model for dehazing is responsible for transforming hazy input images into clear and visually appealing dehazed outputs. It follows an encoder-decoder architecture with different generator blocks and encoder blocks.

The generator begins with an encoder block that encodes the hazy input image, extracting high-level features that capture the underlying information about the haze. Various encoder blocks such as MobileNet, EfficientNet, ResNet, and Vision Transformer can be employed in this stage. These encoder blocks have different architectural configurations and are designed to capture and represent image features effectively.The encoded features from the encoder block are then passed to the decoder block, which is responsible for reconstructing the dehazed output. The decoder block employs architectures like U-Net, PSPNet, MANet, or FPN. These architectures are specifically designed to capture and integrate contextual information, enabling accurate and detailed reconstruction of the dehazed image.The decoder block gradually upsamples the encoded features, applying a series of upsampling layers or transposed convolutions. This process enlarges the spatial dimensions of the encoded features, allowing the generator to generate a dehazed output that matches the size of the original input image.During the upsampling process, skip connections are often incorporated. These connections establish direct connections between corresponding encoder and decoder layers. By utilizing skip connections, the generator can leverage low-level and high-level features simultaneously, capturing fine details and contextual information for better dehazing results.

To optimize the generator's performance and guide its training process, various loss functions can be used. Commonly employed loss functions include Mean Squared Error (MSE) loss, Structural Similarity Index (SSIM) loss, and adversarial loss. These loss functions measure the difference between the generated dehazed image and the ground truth clear image, encouraging the generator to produce outputs that are perceptually similar to real clear images.The generator's objective is to learn the mapping between hazy and clear images, capturing the complex relationships and patterns within the training data. Through an iterative training process, the generator's parameters are optimized to minimize the difference between the generated dehazed images and the ground truth clear images. This enables the generator to produce high-quality dehazed outputs when applied to unseen hazy images during the inference phase.

By incorporating different generator blocks, encoder blocks, and loss functions, the proposed GAN model can effectively dehaze images and enhance their visibility. The generator's architecture and training process allow it to learn and capture the underlying characteristics of haze, leading to the generation of visually pleasing dehazed outputs.

The proposed GAN-based dehazing model utilizes various architectures, namely U-Net, PSPNet, MANet, and FPN, as the generator component. These architectures are carefully chosen due to their distinct characteristics and strengths in capturing and reconstructing image details during the dehazing process. Each architecture offers a different approach to feature extraction and image reconstruction, allowing for a comprehensive exploration of their performance in haze effect removal as described below:

### 4.1.1 Unet

U-Net is a deep learning architecture that has been widely used in various image-to-image translation tasks, including dehazing. It was originally introduced for biomedical image segmentation but has shown great effectiveness in the domain of dehazing.

The U-Net architecture is characterized by its symmetric encoder-decoder structure with skip connections. The encoder portion of U-Net consists of multiple convolutional layers followed by downsampling operations, such as max pooling or strided convolutions. This downsampling process reduces the spatial resolution of the feature maps while increasing their depth or number of channels. This allows the network to capture hierarchical and abstract representations of the input image.The decoder part of U-Net, which is symmetric to the encoder, aims to recover the spatial resolution of the feature maps. It consists of upsampling operations, such as transposed convolutions or bilinear interpolation, to increase the spatial dimensions. These upsampling operations are typically followed by convolutional layers to refine the features.The key innovation of U-Net is the use of skip connections that connect corresponding encoder and decoder layers. These skip connections provide a shortcut for low-level features from the encoder to directly reach the decoder layers. By combining low-level details with high-level context information, U-Net can preserve fine-grained information during the dehazing process.The skip connections enable U-Net to effectively capture and leverage both local and global features. The low-level details captured in the early layers of the network can help in recovering fine structures, while the high-level context information in the later layers can aid in understanding the overall scene and haze distribution. This combination of local and global information contributes to the accurate dehazing performance of U-Net.

Overall, the U-Net architecture with its encoder-decoder structure and skip connections has proven to be effective in capturing fine details and preserving global context, making it a popular choice for dehazing tasks.

### 4.1.2 MANet

MaNet, short for Multi-scale Attention Network, is a deep learning architecture that has been designed specifically for the task of image dehazing. It is known for its ability

to capture multi-scale features and effectively model the complex haze distribution in images.

The MaNet architecture incorporates a series of convolutional layers, attention modules, and skip connections to enhance the performance of dehazing. One of the key components of MaNet is the multi-scale feature extraction module, which extracts features at different scales to capture both local and global information. This allows the network to handle haze of varying densities and sizes.In MaNet, attention modules are used to selectively focus on relevant features and suppress irrelevant ones. These attention mechanisms help the network allocate more resources to important regions in the image and enhance the dehazing process. By adaptively weighting the features based on their importance, the attention modules improve the network's ability to distinguish between haze and underlying scene structures.Skip connections, similar to those used in U-Net, are incorporated in MaNet to facilitate the flow of information between different layers. These connections enable the network to preserve low-level details and effectively reconstruct fine structures during the dehazing process.One notable aspect of MaNet is its ability to handle complex scenes with varying haze distributions and lighting conditions. The multi-scale feature extraction and attention mechanisms enable the network to adaptively process different regions of the image based on their characteristics. This makes MaNet robust and capable of handling challenging dehazing scenarios.

### 4.1.3 PSPNet

PSPNet, which stands for Pyramid Scene Parsing Network, is a deep learning architecture originally designed for semantic segmentation tasks. However, it has also been successfully applied to the task of image dehazing. PSPNet focuses on capturing multi-scale contextual information by leveraging the power of pyramid pooling modules.

The core idea behind PSPNet is to utilize pyramid pooling modules to aggregate features at multiple scales and resolutions. This enables the network to have a global understanding of the image context while also capturing fine-grained details. By incorporating pyramid pooling, PSPNet effectively captures both local and global information, which is essential for accurately modeling and removing haze from images.The architecture of PSPNet consists of a feature extraction backbone, followed by a pyramid pooling module and a final convolutional layer for prediction. The feature extraction backbone, often based on a pre-trained network like ResNet or VGGNet, extracts high-level features from the input image. These features are then fed into the pyramid pooling module.The pyramid pooling module divides the feature maps into non-overlapping regions and applies adaptive average pooling to each region. By doing so, it captures the context at different scales, ranging from fine details to global information. The pooled features are then concatenated and further processed through convolutional layers to

generate a comprehensive representation of the image. Finally, the output of the pyramid pooling module is passed through a convolutional layer for pixel-wise prediction. In the case of image dehazing, the final prediction aims to estimate the transmission map or the dehazed image itself. PSPNet's ability to capture multi-scale contextual information makes it suitable for dehazing tasks. By considering both local and global features, PSPNet can effectively model the haze distribution and capture the underlying scene structures. This enables it to produce high-quality dehazed images with enhanced visibility and reduced haze effects.

In summary, PSPNet is a deep learning architecture that utilizes pyramid pooling modules to capture multi-scale contextual information. It has demonstrated promising results in image dehazing by effectively modeling the haze distribution and preserving scene details. Its ability to leverage both local and global information makes it a valuable tool for improving image quality and visibility in hazy conditions.

### 4.1.4 FPN

FPN, which stands for Feature Pyramid Network, is a popular deep learning architecture that addresses the challenge of object detection and segmentation at different scales. It has also been applied to the task of image dehazing to effectively capture and utilize features from different levels of abstraction.

The main goal of FPN is to enable robust feature representation by constructing a feature pyramid that incorporates features from different spatial resolutions. The architecture of FPN consists of a backbone network, typically a pre-trained convolutional neural network (CNN) such as ResNet or VGGNet, and a top-down pathway with lateral connections.The backbone network extracts features from the input image at various levels of abstraction. These features typically capture different levels of detail, with higher-level features capturing more semantic information and lower-level features capturing finer details. However, the spatial resolutions of these features decrease as the network goes deeper.To address the resolution discrepancy, the top-down pathway in FPN enhances the features from the backbone network by upsampling them to match the resolution of higher-level features. This process is achieved through lateral connections, which perform 1x1 convolutions on the higher-level features to reduce their dimensionality. The resulting features are then combined with the upsampled features from the previous level to create a fused representation.By integrating features from different scales, FPN creates a feature pyramid that maintains both semantic richness and spatial resolution. The top-down pathway allows the network to capture and propagate information from higher-level features to lower-level features, enabling the model to have a holistic understanding of the image context while preserving fine-grained details.

In the context of image dehazing, FPN can effectively leverage the multi-scale information to estimate and remove haze effect. The fused features from different levels of the feature pyramid provide a comprehensive representation of the hazy image, allowing the model to capture both the global scene structure and the fine details affected by haze. This leads to accurate dehazing results with improved visibility and reduced haze effects.

## 4.2 Encoders

In this work, different encoder block architectures, namely MobileNet, EfficientNet, Vision Transformer, and ResNet, have been employed to capture and encode the information at various levels of abstraction from the input hazy image. These encoder blocks, play a crucial role in extracting meaningful features that serve as input for the subsequent processing stages of the generator network, are described below:

### 4.2.1 Mobilenet

In the proposed GAN-based dehazing model, the MobileNet architecture is utilized as the backbone encoder for feature extraction. MobileNet is a lightweight convolutional neural network architecture known for its efficiency and compact size. It is particularly suitable for resource-constrained environments, making it a favorable choice for the dehazing task.

The MobileNet encoder block consists of a series of depth-wise separable convolutions, followed by point-wise convolutions. This design allows for efficient feature extraction while reducing the computational complexity and the number of parameters compared to traditional convolutional networks.In each depth-wise separable convolution layer, the input is convolved with a set of filters that operate on a single channel independently, capturing spatial information. This is followed by a point-wise convolution, where a 1x1 convolution is applied to combine the features across channels. The point-wise convolution enables cross-channel interaction and the learning of complex representations.By utilizing the MobileNet encoder as the backbone, the dehazing model can effectively extract hierarchical features from hazy images. The encoder captures relevant information about the image content, including haze distribution, textures, and edges, which are crucial for generating high-quality dehazed outputs.Furthermore, the lightweight nature of the MobileNet architecture allows for faster training and inference times, making it feasible for real-time or near real-time dehazing applications. The reduced computational complexity also enables the model to be deployed on resource-constrained devices without compromising performance.

The integration of the MobileNet encoder as the backbone in the GAN-based dehazing model combines efficient feature extraction with computational effectiveness, contributing to the model's ability to effectively remove haze effect and enhance the visibility of images.

### 4.2.2 EfficientNet

EfficientNet is a family of convolutional neural network architectures that have been widely used in various computer vision tasks, including image classification, object detection, and segmentation. In the context of the GAN generator for dehazing, EfficientNet is employed as an encoder to extract and encode features from the input hazy image.

EfficientNet models are known for their excellent trade-off between accuracy and efficiency. They achieve state-of-the-art performance by carefully balancing the model's depth, width, and resolution using a compound scaling method. This approach ensures that the model is both powerful and computationally efficient.The EfficientNet encoder consists of a series of blocks that progressively downsample the input image while increasing the number of channels. Each block typically includes convolutional layers, activation functions, and normalization layers to extract and transform the image features. The depth and width of the EfficientNet model can be controlled by adjusting the scaling coefficients.One notable characteristic of EfficientNet is the use of mobile inverted bottleneck convolution, which combines depthwise separable convolutions with pointwise convolutions to reduce computational complexity. This technique significantly reduces the number of parameters and computations while maintaining good representation capacity.

As an encoder in the GAN generator for dehazing, EfficientNet extracts hierarchical features from the hazy image, capturing both low-level and high-level information. The depth and width of the EfficientNet model influence the level of abstraction and complexity of the extracted features. By employing EfficientNet as an encoder, the generator can leverage its powerful feature extraction capabilities to capture important haze patterns and scene structures, enhancing the dehazing performance.

The choice of a specific variant of EfficientNet (e.g., EfficientNet-B0, EfficientNet-B4) depends on the specific requirements of the dehazing task, such as the desired balance between accuracy and computational efficiency. It is important to consider the available computational resources when selecting the EfficientNet variant, as larger models may require more memory and computational power.

### 4.2.3 Resnet

ResNet, short for Residual Network, is a widely used convolutional neural network architecture known for its ability to effectively train very deep networks. In the context of the GAN generator for dehazing, ResNet is employed as an encoder to extract and encode features from the input hazy image.

The key innovation of ResNet is the introduction of residual blocks, which allow for the training of deep networks by addressing the degradation problem. Each residual block consists of two main components: a skip connection and a set of convolutional layers. The skip connection directly passes the input of the block to the output, enabling the network to learn residual mappings. This bypasses the need for the network to learn identity mappings and facilitates the training of deeper networks.ResNet architectures come in different variants, such as ResNet-18, ResNet-50, ResNet-101, etc., which differ in the number of layers and parameters. Deeper variants have more residual blocks and can capture more complex features but may also be more computationally intensive.

As an encoder in the GAN generator for dehazing, ResNet is responsible for extracting hierarchical features from the hazy image. Each residual block in the encoder processes the input image and progressively learns to capture and represent different levels of abstraction. The skip connections enable the preservation of low-level details and gradients, which can be important for dehazing tasks.By using ResNet as an encoder, the generator benefits from its ability to extract and encode meaningful features from the hazy image, capturing both low-level and high-level information. The depth and complexity of the ResNet model influence the level of abstraction and complexity of the extracted features. Deeper ResNet variants can capture more intricate haze patterns and scene structures, potentially leading to improved dehazing performance.The choice of a specific ResNet variant depends on factors such as the desired trade-off between model complexity and computational resources. Deeper ResNet models may offer better representation capacity but require more memory and computational power. It is important to consider the available resources and the specific requirements of the dehazing task when selecting the ResNet variant.

ResNet serves as a powerful encoder in the GAN generator for dehazing, leveraging its residual blocks and skip connections to extract and encode features from the hazy image. The encoder's ability to capture hierarchical information aids in enhancing the quality and visibility of the dehazed images.

### 4.2.4 Mixed Vision Transformer

Mixed Vision Transformer (MixViT) is a state-of-the-art architecture that combines the strengths of Vision Transformers (ViTs) and convolutional neural networks (CNNs). In

the context of the GAN generator for dehazing, MixViT is utilized as an encoder to extract and encode features from the input hazy image.

The MixViT architecture incorporates both self-attention mechanisms, characteristic of ViTs, and convolutional layers, commonly used in CNNs. This hybrid design allows MixViT to effectively capture both global and local features, making it well-suited for handling image-based tasks like dehazing.At a high level, MixViT consists of multiple blocks, where each block consists of a self-attention layer followed by a convolutional layer. The self-attention layer captures the global context of the input image, allowing the model to understand the relationships between different image regions. On the other hand, the convolutional layer focuses on capturing local details and patterns within the image.The self-attention mechanism in MixViT enables the model to attend to relevant features across the entire image, promoting long-range dependencies and capturing global context. It allows the model to identify and understand the hazy regions and their relationships with other parts of the image. The convolutional layers, on the other hand, capture local features and fine-grained details, which can be crucial for effectively removing haze and enhancing image visibility.

By leveraging MixViT as an encoder, the generator benefits from its ability to capture both global and local features. This holistic representation allows the model to effectively learn the complex relationships between haze, scene structures, and important image details. The combination of self-attention and convolutional layers enables MixViT to exploit both global and local context, leading to enhanced dehazing performance.The specific architecture and hyperparameters of MixViT, such as the number of blocks, the size of self-attention heads, and the number of channels in convolutional layers, can be adjusted based on the requirements of the dehazing task and the available computational resources. Fine-tuning these parameters allows for customization and optimization of the MixViT model for improved performance.

MixViT serves as a powerful encoder in the GAN generator for dehazing, incorporating both self-attention mechanisms and convolutional layers. This hybrid architecture enables the model to capture both global and local features, promoting a comprehensive understanding of the hazy image. By leveraging MixViT, the generator is equipped with the capability to effectively remove haze effect and enhance image quality and visibility.

## 4.3 Discriminator

The discriminator is an essential component of the proposed GAN model for dehazing. Its purpose is to discriminate between real clear images and the generated dehazed images produced by the generator. The discriminator consists of a series of convolutional layers that extract features from the input images and make predictions regarding their authenticity.

The discriminator architecture typically comprises several convolutional layers followed by activation functions such as ReLU (Rectified Linear Unit) or LeakyReLU (Leaky Rectified Linear Unit). These layers are responsible for learning hierarchical representations of the input images, capturing both low-level and high-level features.Each convolutional layer applies a set of filters to convolve with the input feature maps, extracting specific patterns and information. The number of filters and their sizes may vary depending on the specific discriminator architecture. As the input passes through these convolutional layers, the feature maps become more abstract, representing increasingly complex features related to the authenticity of the input images.To reduce the spatial dimensions of the feature maps while increasing the receptive field, pooling layers such as max pooling or average pooling may be employed. These layers downsample the feature maps, consolidating the learned information and enabling the discriminator to capture more global patterns.Following the convolutional layers and downsampling operations, the discriminator may employ fully connected layers or global pooling to further aggregate the extracted features. These layers aim to capture high-level representations and summarize the learned information from the convolutional layers.The final layer of the discriminator is a sigmoid activation function, which maps the aggregated features to a probability score. This score represents the discriminator's confidence in the input image being real or generated. A value close to 1 indicates a high probability of the input being a real clear image, while a value close to 0 indicates a high probability of it being a generated dehazed image.

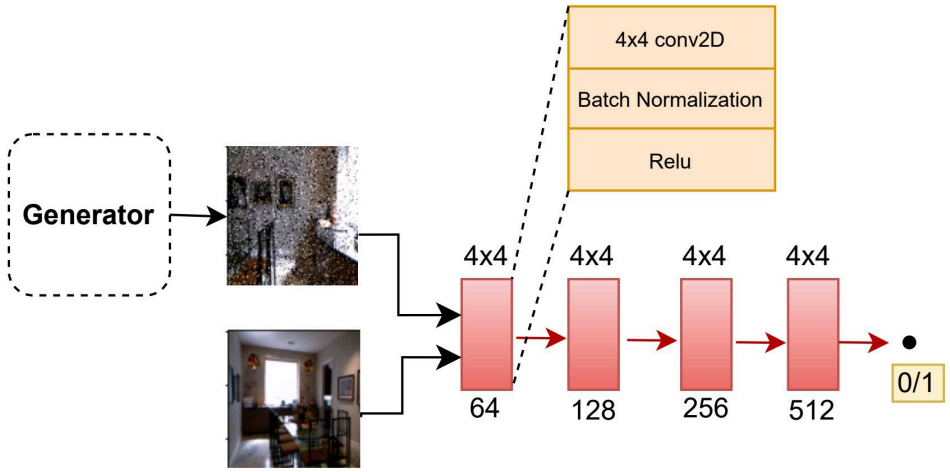


FIGURE 4.3: Discriminator Architecture.

During training, the discriminator is simultaneously trained alongside the generator in an adversarial manner. The generator aims to generate dehazed images that can fool the discriminator into classifying them as real, while the discriminator aims to accurately distinguish between real clear images and generated dehazed images.By optimizing the

discriminator's parameters through backpropagation and gradient descent, it becomes increasingly adept at discriminating between real and generated images. This process encourages the generator to improve its dehazing capabilities and produce outputs that are more realistic and indistinguishable from real clear images.

The discriminator plays a critical role in the GAN model, providing feedback to the generator and facilitating the adversarial learning process. Its ability to discern real clear images from generated dehazed images contributes to the overall training and improvement of the GAN model for dehazing.

## 4.4 Loss Function

The proposed method uses multiple loss functions to optimize the generator and discriminator networks during the training process. The generator's total loss is formulated as follows:

$$L_G = \frac{1}{3}(L_{L1} + L_{MS-SSIM} + L_{BCE})$$

The L1 loss, sometimes referred to as the mean absolute error (MAE) loss, calculates the pixel wise difference between the dehazed image and the original image. It is defined as:

$$L_{L1} = \frac{1}{N}\sum_{i=1}^{N} \|G(I_i) - J_i\|$$

where $G(I_i)$ is the dehazed ,output of the generator network $G$ for the input hazy image $I_i$, $J_i$ is the corresponding ground truth, and $N$ is the training size.

The ground truth image and the dehazed image are compared at various scales using a method called multi-scale structural similarity (MS-SSIM) loss. It is defined as:

$$L_{MS-SSIM} = 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M} MS\_SSIM\left(G(I_i), J_i\right)$$

where $MS\_SSIM\left(G(I_i), J_i\right)$ is the MS-SSIM score between the generated dehazed image and the orginal image, calculated at $M$ different scales.

The binary cross-entropy with logits (BCEWithLogits) loss quantifies the dissimilarity between the predicted probability generated by the discriminator network and the

corresponding ground truth label. It is defined as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(D(J_i)) + (1 - y_i)\log(1 - D(G(I_i)))]$$

where $D$ is the discriminator network, $y_i$ is the ground truth label (1 for clear images and 0 for hazy images), and $N$ is the training size.

The discriminator loss is computed solely using the BCEWithLogits loss function. Its purpose is to enable the discriminator to differentiate between the dehazed images generated by the generator and the orginal ground truth images. The loss is defined as:

$$L_D = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(D(J_i)) + (1 - y_i)\log(1 - D(G(I_i)))]$$

where $D$ is the discriminator network, $y_i$ is the ground truth label (1 for clear images and 0 for hazy images), and $N$ is the training size.

# Chapter 5

# Real-Time Implementation

In this chapter, a detailed account of the real-time implementation of the proposed dehazing architecture has been discussed, specifically focusing on the selection of the U-Net architecture with the MobileNet encoder and the steps taken to optimize the model for deployment on the Raspberry Pi.

## 5.1 Architecture Selection

After evaluating various architectures and encoders, the U-Net architecture with the MobileNet encoder was chosen for its combination of performance and computational efficiency. The U-Net architecture has demonstrated promising results in restoring image quality and preserving fine details in haze effect removal tasks. The MobileNet encoder, on the other hand, is renowned for its lightweight design, making it highly suitable for deployment on low-power devices such as the Raspberry Pi.

The selection of the U-Net architecture with the MobileNet encoder aimed to achieve a compromise between dehazing performance and computational resource utilization, thus enabling real-time processing on the Raspberry Pi.

## 5.2 Conversion to ONNX Format and Model Compression

To enable easy deployment on the Raspberry Pi, the generator component of the U-Net architecture was converted to the ONNX format. The ONNX format offers compatibility across various frameworks and platforms, ensuring smooth integration with the Raspberry Pi.

Alongside the ONNX conversion, model compression techniques were applied to reduce the computational demands of the model. Specifically, the model was compressed to 8-bit floating-point precision (float8) to optimize its size and complexity while minimizing the impact on dehazing performance. This compression technique aids in reducing the memory footprint and computational overhead, thereby enabling faster processing on the resource-constrained Raspberry Pi.

## 5.3 Real-Time Dehazing on the Raspberry Pi

During the implementation and testing phase, challenges arose due to the limited computational power of the Raspberry Pi. The processing capabilities of the Raspberry Pi are considerably lower compared to high-end computers or dedicated GPUs, resulting in slower processing speeds for the dehazing model.

However, despite these limitations, real-time dehazing performance was achieved on the Raspberry Pi, albeit at a reduced frame rate. The implementation processed hazy videos at an approximate rate of 1 frame per second (fps), which falls below the desired real-time performance threshold of 25-30 fps.

## 5.4 Addressing Computational Constraints

To improve the real-time processing speed of the dehazing architecture on the Raspberry Pi, several strategies can be considered:

1. **Quantization and Pruning**: Further optimization techniques, such as quantization and pruning, can be applied to reduce the model's size and computational requirements. These techniques enable faster execution by decreasing the number of parameters and operations needed for inference.

2. **Hardware Acceleration**: Leveraging hardware accelerators, such as Neural Processing Units (NPUs) or Graphics Processing Units (GPUs), can significantly enhance the processing speed of the dehazing model on the Raspberry Pi. These specialized processors are designed to handle complex computations efficiently, enabling real-time performance even on low-power devices.

3. **Model Simplification**: Exploring ways to simplify the architecture or reduce its complexity can improve processing speed. This may involve modifying the network's depth, width, or incorporating lightweight modules specifically tailored for low-power devices.

4. **Edge Inference Frameworks**: Utilizing specialized inference frameworks designed for edge devices can further optimize the model's execution on the Raspberry Pi. These

frameworks often incorporate hardware-specific optimizations and provide efficient runtime environments for deploying deep learning models on resource-constrained platforms.

## 5.5 Future Considerations

Achieving real-time dehazing performance on the Raspberry Pi requires a careful balance between computational efficiency and dehazing accuracy. Further research and optimization efforts are necessary to enhance the model's speed without compromising its effectiveness.

Additionally, exploring hardware-specific optimizations, such as leveraging the capabilities of the Raspberry Pi's GPU or incorporating custom hardware accelerators, can lead to significant improvements in processing speed.

It is important to note that the computational constraints of the Raspberry Pi may impose limitations on the real-time performance of the dehazing architecture. Therefore, it is crucial to consider the specific requirements of the target application and strike a balance between computational efficiency and the desired level of dehazing quality.

By continuously refining and optimizing the dehazing architecture for real-time implementation on devices like the Raspberry Pi, we can unlock its potential for various practical applications, including video surveillance, autonomous vehicles, and real-time image enhancement.

# Chapter 6

# Experimental Setup

## 6.1 Datasets

To address the challenges of acquiring a large-scale dataset of real atmospheric hazy images, the RESIDE6K dataset [25] was utilized, providing a valuable resource for training and evaluating the dehazing model. The RESIDE dataset [11] is a significant contribution in the field of image dehazing, offering a large-scale collection of hazy images for training and evaluation purposes. The dataset comprises five subsets, each serving a specific purpose in the evaluation process.

The RESIDE dataset contains several subsets that were utilized for training and evaluation. The Indoor Training Set (ITS) and Outdoor Training Set (OTS) subsets are synthetic datasets that include hazy images captured in indoor and outdoor environments, respectively. These subsets provide a diverse range of hazy conditions encountered in different settings, making them valuable for training the dehazing model. The Synthetic Objective Testing Set (SOTS) is another synthetic subset within RESIDE that is specifically designed for objective evaluation. It includes ground truth clear images, allowing for the quantitative assessment of the dehazing model's performance. By comparing the generated clear images with the ground truth images in the SOTS subset, we can measure the effectiveness of the proposed approach in removing haze and enhancing visibility in synthetic hazy images. To evaluate the model's performance in real-world scenarios, the Real-World task-driven Testing Set (RTTS) is included in RESIDE. This subset consists of real-world hazy images captured under different conditions, providing a more realistic and challenging test environment. By testing the model on the RTTS subset, we can assess its ability to handle real-world haze and validate its performance under practical conditions.Lastly, the Hybrid Subjective Testing Set (HSTS) combines both synthetic and real-world hazy images. This subset is designed for subjective evaluation, involving human observers who assess the quality of dehazed images. By including

subjective evaluations, we can gain insights into the visual perception of dehazing results and validate the overall visual quality of the proposed method.

The hazy images in the RESIDE6K dataset were generated synthetically by applying atmospheric haze effects to the corresponding clear images. This synthetic generation process ensured that the hazy images exhibited realistic and diverse characteristics, including varying intensities of haze, distinct color tones, and spatial distributions. Training and evaluating the model on this dataset allowed us to effectively address different hazy conditions encountered in real-world scenarios. The synthetic nature of the dataset provided a controlled environment for training and testing the dehazing model, enabling us to study the model's performance under a wide range of hazy conditions and assess its generalization capabilities.

In this research, the Indoor Training Set (ITS) and Synthetic Objective Testing Set (SOTS) from the RESIDE6K dataset were utilized for training and evaluation, respectively. These subsets provided a comprehensive representation of hazy images in indoor and outdoor settings, enabling the examination of various hazy conditions encountered in practical scenarios. Leveraging the RESIDE dataset ensured that the proposed dehazing model underwent training and evaluation on a diverse and realistic collection of hazy images, leading to robust and reliable results. The training set of the RESIDE6K dataset included a total of 6,000 hazy images, evenly split between indoor and outdoor scenes, resulting in 3,000 images for each category. Paired with each hazy image was its corresponding ground truth clear image, which served as a reference during the training process. This pairing enabled the establishment of a supervised learning framework, allowing the model to learn the mapping between hazy images and their respective clear counterparts. For comprehensive evaluation of the proposed method, a separate test set was created using the RESIDE6K dataset. This test set encompassed 1,000 hazy images, evenly divided into 500 indoor hazy images and 500 outdoor hazy images. By maintaining the distinction between indoor and outdoor hazy images, the evaluation process covered a broad range of environmental conditions and factors contributing to haze formation, ensuring a thorough assessment of the model's performance.

In addition to the RESIDE dataset, other benchmark datasets were utilized for testing and comparing the performance of the proposed dehazing method. These datasets include I-Haze [26], DENSE-Haze [27], and NH-Haze [28], which are widely used in the field of image dehazing research.

**I-Haze** dataset [26] contains a significant number of hazy images captured from various real-world scenes. It consists of approximately 50 indoor and outdoor scenes, resulting in a dataset size of around 500 hazy-clear image pairs.

**DENSE-Haze** dataset [27] is specifically designed to address the challenge of dense and thick haze. It contains a large collection of hazy images with extremely low visibility,

making it a suitable benchmark for assessing the performance of dehazing algorithms under severe haze conditions. The dataset consists of over 1000 hazy-clear image pairs.

**NH-Haze** dataset [28] focuses on natural haze conditions, including fog, mist, and other atmospheric phenomena. It contains hazy images captured in different weather conditions and geographical locations. The dataset consists of several hundred hazy-clear image pairs, providing a diverse range of hazy scenes for evaluation.

By utilizing these benchmark datasets, including I-Haze, DENSE-Haze, and NH-Haze, the proposed dehazing method was tested and its performance was evaluated in various hazy conditions. This evaluation allowed for comparisons with existing state-of-the-art techniques in the field. The diverse range of hazy images provided by these datasets enabled the validation of the generalization and robustness of the proposed approach across different real-world scenarios. The large dataset sizes also facilitated rigorous evaluations and reliable conclusions regarding the effectiveness and superiority of the proposed dehazing method. The availability of a substantial number of hazy images from diverse environments, along with the corresponding ground truth clear images, played a crucial role in effectively training the dehazing model. Furthermore, the utilization of multiple datasets allowed for the assessment of the model's generalization capabilities and the validation of its performance across various hazy conditions.

## 6.2 Preprocessing

In the preprocessing stage, a normalization step was applied to the input hazy images before feeding them into the network. Specifically, the transforms.Normalize(0.5, 0.5) function was utilized to normalize the pixel values of the hazy images. This normalization process ensures that the pixel values are scaled to a standardized range, which aids in the training and optimization of the network.The transforms.Normalize(0.5, 0.5) function subtracts the mean value of 0.5 from each pixel and then divides it by a standard deviation of 0.5. This operation effectively shifts the pixel values towards zero-centered normalization, resulting in a distribution that is more suitable for the subsequent network operations.By normalizing the hazy images in this manner, it helps in reducing the impact of varying illumination conditions and intensity ranges across different images in the dataset. Additionally, it can facilitate faster convergence during the training process and enhance the stability of the network's optimization.

After the normalization step, the preprocessed hazy images are ready to be passed through the generator network for dehazing. The application of this preprocessing step ensures that the input data is appropriately prepared for the subsequent network operations, thereby aiding in the generation of accurate and visually appealing dehazed outputs.

## 6.3 Training Details

In this section, an overview of the experimental setup and training details for the proposed dehazing method is provided. The model was initialized by incorporating pretrained ImageNet weights into the encoder block. This initialization step allowed the model to leverage the learned visual patterns from the large-scale ImageNet dataset, enhancing its ability to capture relevant features from the hazy input images effectively.

During the training process, a combination of loss functions was employed to guide the learning of the generator and discriminator components of the GAN. The generator loss function consisted of Mean Structural Similarity (MSS) loss, L1 loss, and Binary Cross-Entropy with Logits loss. This combination of loss functions aimed to promote perceptual similarity, pixel-level reconstruction, and adversarial realism in generating high-quality clear images. The MSS loss encouraged the preservation of structural information and texture details in the dehazed outputs, while the L1 loss enforced pixel-wise similarity between the generated clear images and the ground truth clear images. The Binary Cross-Entropy with Logits loss guided the generator in producing clear images that can deceive the discriminator.On the other hand, the discriminator loss solely focused on Binary Cross-Entropy with Logits (BCEWithLogits) loss. This loss function aided the discriminator in distinguishing between real clear images and the generated clear images produced by the generator. By optimizing the discriminator to accurately classify real and generated images, the generator was encouraged to produce more realistic and visually convincing dehazed outputs.To optimize the learning process, an initial learning rate of 0.001 was set, and the Adam optimizer with default parameters was utilized to update the network weights. The Adam optimizer adaptively adjusted the learning rate based on the gradients of the model parameters, facilitating efficient convergence during training. A batch size of 16 was employed for mini-batch training, allowing for the simultaneous processing of multiple samples and effective updates of the model weights based on computed losses and gradients.For training and evaluation, the RESIDE6K dataset was utilized, comprising 6,000 hazy images for training and 1,000 hazy images for testing. This dataset provided a diverse range of hazy scenes, encompassing both indoor and outdoor environments, ensuring that the model learned to handle various hazy conditions encountered in real-world scenarios. The testing dataset enabled the assessment of the generalization and performance of the proposed method on unseen hazy images.

By implementing the aforementioned training setup and leveraging the diverse loss functions, pretrained weights, and optimization techniques, the aim was to train the GAN-based dehazing model to effectively learn the mapping between hazy input images and high-quality clear images, ultimately enhancing the visibility and quality of hazy images.

## 6.4 Performance Metrics

To evaluate the performance of the proposed dehazing method, several commonly used metrics that provide objective measurements of image quality, have been employed. The following performance metrics were used:

1. Peak Signal-to-Noise Ratio (PSNR): PSNR is a widely used metric to assess the reconstruction quality of images. It measures the ratio between the maximum possible signal power (based on the pixel intensity range) and the distortion caused by the dehazing process. Higher PSNR values indicate better image quality, as they indicate lower levels of distortion.

2. Structural Similarity Index (SSIM): SSIM is a perceptual metric that measures the structural similarity between two images. It takes into account the luminance, contrast, and structure of the images. SSIM values range from 0 to 1, with 1 indicating perfect similarity. Higher SSIM values imply better image quality, as they suggest a closer resemblance to the ground truth clear images.

By evaluating dehazing results using PSNR and SSIM metrics, the aim was to quantitatively measure the effectiveness of the proposed method in enhancing image quality and visibility. These metrics provide objective measures of image fidelity and perceptual similarity, allowing for the comparison of different dehazing techniques and analysis of the impact of various architectural choices and training strategies on the quality of the dehazed images.It is worth noting that while PSNR and SSIM are commonly used metrics for image quality assessment, they have their limitations. PSNR focuses on pixel-level differences and may not fully capture the perceptual quality of an image. SSIM, although more perceptually motivated, may not always align with human perception. Therefore, in addition to these metrics, visual inspections and qualitative assessments were also conducted to gain a holistic understanding of the performance of the dehazing method.

# Chapter 7

# Evaluation

In this Chapter, a detailed evaluation of the proposed method for haze effect removal and image enhancement has been presented. The evaluation aims to assess the performance, effectiveness, and robustness of the approach on multiple datasets and compare it with existing state-of-the-art methods. The metrics used for evaluation include structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR), which are widely used to measure image quality and similarity to ground truth.

## 7.1 Quantitative Evaluation

The evaluation results, as presented in Table 7.1, offer comprehensive insights into the performance of the proposed dehazing techniques compared to various state-of-the-art methods on the RESIDE6K dataset. Objective evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) were employed to assess the quality of the dehazed images.

The most efficient architecture of this research, the Unet model, exhibited remarkable performance on the outdoor dataset of Reside6k. It achieved an average PSNR of 29.37 and an average SSIM of 0.965, indicating significant improvement in image quality and clarity. On the indoor dataset of Reside6k, the Unet model attained an average PSNR of 25.82 and an average SSIM of 0.926. These results clearly demonstrate the superiority of the proposed method over other existing approaches, such as AOD-Net, DehazeNet, DCP, and DCPDN, in terms of both PSNR and SSIM metrics.The higher PSNR values achieved by the Unet model indicate reduced noise and improved image fidelity, while the higher SSIM values signify better structural similarity and preservation of important image details. This signifies the effectiveness of the proposed dehazing technique in enhancing image quality and visibility, particularly in hazy outdoor scenarios. These evaluation results validate the efficacy of the proposed approach in tackling the challenges

posed by haze and highlight its potential for real-world applications. By outperforming state-of-the-art methods on the Reside6k dataset, the Unet model demonstrates its capability to accurately restore hazy images and produce visually pleasing results.

| Method | | DCP[17] | DehazeNet[21] | AOD-NET[19] | DCPDN[12] | GFN[22] | Proposed Unet |
|---|---|---|---|---|---|---|---|
| Indoor | PSNR | 16.62 | 21.14 | 19.06 | 15.85 | 22.3 | **25.82** |
| | SSIM | 0.8179 | 0.8472 | 0.8504 | 0.8175 | 0.88 | **0.926** |
| Outdoor | PSNR | 19.13 | 22.46 | 20.29 | 19.93 | 21.55 | **29.37** |
| | SSIM | 0.8148 | 0.8514 | 0.8765 | 0.8449 | 0.8444 | **0.965** |

TABLE 7.1: The outcomes of comparing Unet architecture on Reside6k with various state of the art techniques.

| Model | Encoder | RESIDE6K[25] | | RESIDE6K Indoor | | RESIDE6K Outdoor | |
|---|---|---|---|---|---|---|---|
| | | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR |
| Unet | Mix Vision Transformer | 0.903 | 26.37 | 0.879 | 24.74 | 0.928 | 27.99 |
| Unet | Resnet152 | 0.935 | 25.94 | 0.919 | 24.85 | 0.952 | 27.03 |
| Unet | Vgg19 | 0.930 | 25.85 | 0.912 | 25.09 | 0.947 | 26.60 |
| Unet | Modified-Mobilenet | 0.949 | **27.86** | 0.926 | 25.82 | **0.965** | **29.37** |
| Unet | Efficientnet | **0.950** | 27.81 | **0.935** | **26.51** | 0.965 | 29.10 |

TABLE 7.2: Evaluation Results of Different Encoders with the Unet Model on the Reside Dataset.

The evaluation results presented in Table 7.2 provide insights into the performance of the Unet model with various encoder blocks, including MobileNet, EfficientNet, VGG19, Mix Vision Transformer, and ResNet, on the Reside dataset. These results offer valuable information regarding the impact of different encoders on dehazing performance.

On the outdoor dataset of Reside6k, the modified MobileNet encoder demonstrated superior performance, achieving the highest average PSNR of 29.37 and an average SSIM of 0.965. These results indicate significant improvements in image quality and structural similarity, showcasing the effectiveness of the MobileNet encoder in capturing fine details and preserving image characteristics.Conversely, on the indoor dataset of Reside6k, the EfficientNet encoder emerged as the top performer, achieving the highest average PSNR of 25.82 and an average SSIM of 0.926. These findings highlight the superiority of the EfficientNet encoder in handling haze effect removal tasks in indoor environments, enabling the restoration of image clarity and enhanced visibility.The evaluation of the resnet and Vgg19 encoders revealed satisfactory performance, with competitive SSIM and PSNR values across all datasets. These encoders demonstrate their capability to handle dehazing tasks and improve image quality.It is important to note that the mit_b5 encoder exhibited relatively lower SSIM and PSNR values compared to the other encoders, indicating its comparatively lower performance in terms of haze effect removal and image quality enhancement.

The evaluation results in Table 7.2 demonstrate the varying performance of different encoders when used with the Unet model on the Reside dataset. The MobileNet and EfficientNet encoders showcase superior performance, while the resnet and Vgg19 encoders deliver satisfactory results. The mit_b5 encoder exhibits relatively lower performance in comparison. These findings contribute to my understanding of the effectiveness of different encoders in the context of dehazing tasks.

| Model | Encoder | I-Haze[26] | | DENSE-Haze[27] | | NH-Haze[28] | |
|---|---|---|---|---|---|---|---|
| | | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR |
| **Unet** | **Mix Vision Transformer** | 0.716 | 15.98 | 0.408 | **11.97** | 0.497 | 12.72 |
| **Unet** | **Resnet152** | 0.684 | 14.96 | 0.353 | 11.19 | 0.490 | **12.92** |
| **Unet** | **Vgg19** | 0.711 | 15.67 | 0.366 | 10.90 | 0.464 | 12.27 |
| **Unet** | **Modified-MobileNet** | 0.756 | 16.59 | 0.384 | 11.21 | 0.482 | 12.68 |
| **Unet** | **Efficientnet** | **0.775** | **16.84** | **0.411** | 11.79 | **0.512** | 12.72 |

TABLE 7.3: Outcomes of comparing Unet architecture with different encoders on various datasets.

The evaluation results presented in Table 7.3 showcase the performance of the Unet model with various encoders on three different haze datasets: I-Haze, Dense Haze, and NH Haze. It is important to note that the Unet model was not specifically trained on these datasets but rather on the Reside dataset. Therefore, the model's performance may vary on these datasets, particularly on Dense Haze and NH Haze, which contain highly dense and challenging hazy images.

Despite the model not being trained on these specific datasets, it still demonstrates notable performance on the I-Haze dataset. The Unet model with the efficientnet encoder achieves the highest SSIM of 0.775 and PSNR of 16.84, indicating its ability to effectively restore image details and enhance visibility in hazy conditions. Additionally, the Unet model with the mix vision transformer encoder shows promising performance with an SSIM of 0.716 and PSNR of 15.98 on the I-Haze dataset.However, it is observed that the Unet model with the efficientnet encoder achieves relatively lower SSIM and PSNR values on the Dense Haze and NH Haze datasets compared to the I-Haze dataset. This can be attributed to the dense and challenging nature of these hazy images, which may require specialized training or adaptations to achieve optimal results.On the Dense Haze dataset, the Unet model with the efficientnet encoder exhibits the highest SSIM of 0.411 and PSNR of 11.79 among the evaluated combinations. While the performance may not be as strong as desired, it still indicates some level of improvement in image quality in dense haze scenarios.Similarly, on the NH Haze dataset, the Unet model with the efficientnet encoder achieves the highest SSIM of 0.512 and PSNR of 12.72. Although the results may not be as favorable as on the Reside dataset, they demonstrate the model's ability to enhance image quality and reduce haze to some extent.

| Encoder | Unet* | | MAnet | | FPN | | PSPnet | |
|---|---|---|---|---|---|---|---|---|
| | **SSIM** | **PSNR** | **SSIM** | **PSNR** | **SSIM** | **PSNR** | **SSIM** | **PSNR** |
| **mit_b0** | 0.928 | 27.99 | 0.875 | 26.65 | 0.724 | 22.87 | 0.617 | 21.26 |
| **mit_b1** | 0.882 | 25.40 | 0.875 | 25.92 | **0.756** | **24.03** | **0.624** | **21.71** |
| **mit_b2** | 0.909 | 27.85 | 0.892 | 26.30 | 0.726 | 22.84 | 0.619 | 21.51 |
| **mit_b3** | 0.884 | 27.32 | **0.927** | **27.58** | 0.723 | 22.78 | 0.616 | 21.60 |
| **mit_b5** | **0.931** | **28.22** | 0.888 | 27.16 | 0.749 | 23.93 | 0.561 | 19.76 |

TABLE 7.4: Outcomes of comparing different dehazing architectures with various Mix Vision Transformers encoder on the Reside outdoor dataset.

The evaluation results in Table 7.4 highlight the performance of different dehazing architectures in combination with various Mix Vision Transformers on the Reside outdoor dataset. The metrics used

Among the architectures, U-Net consistently demonstrates superior performance across all encoder variations. When paired with the mit_b5 encoder, U-Net achieves the highest SSIM of 0.931 and PSNR of 28.22, indicating its ability to produce high-quality dehazed images.MAnet also shows competitive performance, particularly when combined with the mit_b3 encoder, achieving an SSIM of 0.927 and PSNR of 27.58. This highlights the effectiveness of MAnet in capturing image details and enhancing visibility in hazy conditions.On the other hand, FPN and PSPnet exhibit lower performance compared to U-Net and MAnet across all encoder variations.The choice of encoder has a noticeable impact on the dehazing performance. While encoders such as mit_b0 and mit_b2 yield relatively lower SSIM and PSNR values, encoders like mit_b5 and mit_b3 demonstrate better performance, indicating their effectiveness in improving image quality and reducing haze.

| Encoder | Unet | | MAnet | | FPN | | PSPnet | |
|---|---|---|---|---|---|---|---|---|
| | **SSIM** | **PSNR** | **SSIM** | **PSNR** | **SSIM** | **PSNR** | **SSIM** | **PSNR** |
| **mit_b0** | 0.785 | 23.54 | 0.823 | 23.31 | 0.672 | 20.21 | 0.543 | 18.73 |
| **mit_b1** | 0.852 | 23.84 | 0.817 | 23.44 | 0.707 | 20.96 | 0.551 | 18.73 |
| **mit_b2** | 0.867 | 24.68 | 0.863 | **24.89** | 0.668 | 20.30 | 0.550 | 18.25 |
| **mit_b3** | 0.804 | 23.73 | **0.893** | 24.51 | 0.674 | 20.47 | **0.560** | **19.01** |
| **mit_b5** | **0.892** | **24.72** | 0.829 | 24.52 | **0.712** | **21.00** | 0.538 | 18.62 |

TABLE 7.5: Outcomes of comparing different dehazing architectures with various Mix Vision Transformers encoder on the Reside indoor dataset.

The evaluation results presented in Table 7.5 compare the performance of different dehazing architectures U-Net, FPN, MAnet, PSPnet with various Mix Vision Transformers on the Reside indoor dataset.The results also highlight the superior performance of U-Net, particularly with the mit_b5 encoder, in terms of SSIM and PSNR metrics on the Reside outdoor dataset.MAnet, with the mit_b3 encoder, shows competitive performance as well, indicating its effectiveness in addressing haze-related challenges and

enhancing visibility in outdoor environments.On the other hand, FPN and PSPnet generally exhibit lower performance compared to U-Net and MAnet across all encoders, albeit with varying SSIM and PSNR values.Therefore, it can be concluded that U-Net and MAnet consistently outperform FPN and PSPnet in terms of dehazing performance on the Reside outdoor dataset, with U-Net showing the highest performance, especially when paired with the mit_b5 encoder.

| Encoder | Reside Indoor | | Reside Outdoor | | Ihaze | | Dense Haze | | NH Haze | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR |
| **mit_b0** | 0.7855 | 23.54 | 0.928 | 27.99 | 0.608 | 15.64 | 0.345 | **12.30** | 0.402 | 12.59 |
| **mit_b1** | 0.8523 | 23.84 | 0.882 | 25.40 | 0.697 | 16.41 | 0.384 | 11.98 | 0.395 | 12.56 |
| **mit_b2** | 0.8670 | 24.68 | **0.909** | 27.85 | 0.721 | **16.85** | 0.394 | 12.08 | 0.457 | 12.67 |
| **mit_b3** | 0.8042 | **23.73** | 0.884 | 27.32 | 0.666 | **16.03** | **0.345** | 12.21 | 0.448 | **12.73** |
| **mit_b5** | **0.8920** | **24.72** | **0.931** | **28.22** | **0.734** | 16.18 | **0.411** | 12.17 | **0.477** | 12.68 |

TABLE 7.6: Comparison of U-Net Architecture with Mix Vision Transformer Encoder on Various Datasets

The evaluation results presented in Table 7.6 compare the performance of the U-Net architecture with various Mix Vision Transformer encoders such as mit_b0, mit_b1, mit_b2, mit_b3,and mit_b5 on different datasets.

For the Reside Indoor dataset, the mit_b5 encoder demonstrates the highest performance with an SSIM of 0.8920 and a PSNR of 24.72, indicating its ability to effectively remove haze effect and improve image quality in indoor environments. The other encoders also show respectable performance, with SSIM and PSNR values ranging from 0.7855 to 0.8670.On the Reside Outdoor dataset, the mit_b5 encoder once again achieves the highest performance, achieving an SSIM of 0.931 and a PSNR of 28.22. This demonstrates its effectiveness in enhancing image details and capturing structural similarities in outdoor hazy scenes. The other encoders also perform well, with SSIM and PSNR values ranging from 0.882 to 0.909.Moving on to the I Haze dataset, the mit_b5 encoder maintains its superior performance, attaining an SSIM of 0.734 and a PSNR of 16.18. This indicates its ability to handle haze effect removal tasks in the I Haze dataset effectively. The other encoders also exhibit satisfactory performance, with SSIM and PSNR values ranging from 0.608 to 0.721.In the Dense Haze dataset, the mit_b5 encoder achieves the highest SSIM of 0.411 and a PSNR of 12.17. The other encoders also demonstrate reasonable performance, with SSIM and PSNR values ranging from 0.345 to 0.394.Finally, on the NH Haze dataset, the mit_b5 encoder once again outperforms other encoders with an SSIM of 0.477 and a PSNR of 12.68. The other encoders also exhibit decent performance, with SSIM and PSNR values ranging from 0.402 to 0.457.

Overall, the evaluation results in Table 7.6 highlight the varying performance of the U-Net architecture with different Mix Vision Transformer encoders across different datasets. The mit_b5 encoder consistently demonstrates superior performance, while

other encoders also show satisfactory results in handling haze effect removal tasks in different scenarios.

| Encoder | Reside Indoor | | Reside Outdoor | | Ihaze | | Dense Haze | | NH Haze | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR |
| **mit_b0** | 0.672 | 20.21 | 0.724 | 22.87 | 0.611 | 15.76 | **0.378** | **12.46** | 0.325 | **12.57** |
| **mit_b1** | 0.707 | 20.96 | **0.756** | **24.03** | **0.632** | **16.17** | 0.378 | 12.01 | 0.329 | 12.37 |
| **mit_b2** | 0.668 | 20.30 | **0.726** | 22.84 | 0.591 | **15.10** | 0.370 | 12.00 | 0.311 | 12.22 |
| **mit_b3** | 0.674 | **20.47** | 0.723 | 22.78 | 0.605 | **15.65** | **0.365** | 12.23 | 0.312 | **12.51** |
| **mit_b5** | **0.712** | **21.00** | **0.749** | **23.93** | **0.622** | 15.78 | **0.371** | 11.35 | **0.326** | 12.32 |

TABLE 7.7: Comparison of FPN Architecture with Mix Vision Transformer Encoder on Various Datasets

The table 7.7 presents the outcomes of comparing the FPN architecture with various Mix Vision Transformer encoders (mit_b0, mit_b1, mit_b2, mit_b3, mit_b5) on different datasets including Reside Indoor, Reside Outdoor, I Haze, Dense Haze, and NH Haze.

In terms of the Reside Indoor dataset, the FPN architecture with the mit_b1 encoder achieves the highest SSIM of 0.707 and PSNR of 20.96. The FPN architecture with the mit_b5 encoder shows the highest SSIM of 0.749 and PSNR of 23.93 on the Reside Outdoor dataset.For the I Haze dataset, the FPN architecture with the mit_b1 encoder achieves the highest SSIM of 0.756 and the FPN architecture with the mit_b5 encoder achieves the highest PSNR of 15.78.On the Dense Haze dataset, the FPN architecture with the mit_b1 encoder demonstrates the highest SSIM of 0.378, and the FPN architecture with the mit_b3 encoder achieves the highest PSNR of 12.23.Lastly, on the NH Haze dataset, the FPN architecture with the mit_b5 encoder shows the highest SSIM of 0.326, and the FPN architecture with the mit_b1 encoder achieves the highest PSNR of 12.32.These results indicate the performance of the FPN architecture with different Mix Vision Transformer encoders on each dataset in terms of SSIM and PSNR metrics.

| Encoder | Reside Indoor | | Reside Outdoor | | Ihaze | | Dense Haze | | NH Haze | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR |
| **mit_b0** | 0.823 | 23.31 | 0.874 | 26.65 | 0.680 | 16.04 | **0.362** | **11.33** | 0.442 | **12.48** |
| **mit_b1** | 0.816 | 23.43 | **0.874** | **25.91** | **0.700** | **16.55** | **0.404** | 11.56 | **0.473** | 12.31 |
| **mit_b2** | 0.862 | 24.89 | **0.891** | 26.29 | 0.691 | **16.28** | 0.403 | **12.00** | 0.447 | 12.54 |
| **mit_b3** | 0.892 | **24.51** | **0.926** | **27.58** | **0.720** | **16.10** | **0.395** | 11.62 | 0.458 | **12.68** |
| **mit_b5** | **0.829** | **24.51** | **0.888** | **27.16** | **0.7051** | 16.23 | **0.392** | 11.50 | **0.442** | 12.30 |

TABLE 7.8: Comparison of MANet Architecture with Mix Vision Transformer Encoder on Various Datasets

Table 7.8 presents the outcomes of comparing the MAnet architecture with various Mix Vision Transformer encoders on multiple datasets: Reside Indoor, Reside Outdoor, Ihaze, Dense Haze, and NH Haze.

Across the datasets, MAnet demonstrates varying levels of performance depending on the encoder used. In the Reside Indoor dataset, MAnet achieves relatively high SSIM values ranging from 0.817 to 0.893, and PSNR values ranging from 23.31 to 24.89. Among the encoders, mit_b3 shows the highest performance in terms of both SSIM and PSNR.Similarly, in the Reside Outdoor dataset, MAnet performs well, with SSIM values ranging from 0.875 to 0.927 and PSNR values ranging from 25.92 to 27.58. Again, mit_b3 stands out as the encoder with the highest performance.When evaluated on the Ihaze dataset, MAnet's performance decreases, with SSIM values ranging from 0.681 to 0.720 and PSNR values ranging from 16.05 to 16.56. The mit_b3 encoder achieves the highest SSIM and PSNR values in this case.In the Dense Haze and NH Haze datasets, MAnet exhibits relatively lower performance compared to the other datasets. The SSIM values range from 0.363 to 0.404 in Dense Haze and from 0.442 to 0.473 in NH Haze, while the PSNR values range from 11.33 to 12.01 in Dense Haze and from 12.31 to 12.69 in NH Haze. Among the encoders, mit_b1 and mit_b3 show comparatively better performance in these challenging haze conditions.

| Encoder | Reside Indoor | | Reside Outdoor | | Ihaze | | Dense Haze | | NH Haze | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR |
| **mit_b0** | 0.543 | 18.73 | 0.617 | 21.26 | 0.532 | 15.16 | **0.349** | **11.85** | **0.261** | **12.18** |
| **mit_b1** | 0.551 | 18.73 | **0.624** | **21.71** | **0.503** | **14.44** | **0.343** | 11.44 | **0.259** | 11.85 |
| **mit_b2** | 0.550 | 18.25 | **0.619** | 21.51 | 0.519 | **14.93** | 0.336 | **11.40** | 0.253 | 12.04 |
| **mit_b3** | **0.560** | **19.01** | **0.616** | **21.60** | **0.538** | **15.35** | **0.337** | 11.01 | 0.262 | **11.97** |
| **mit_b5** | **0.538** | **18.62** | **0.561** | **19.76** | **0.500** | 14.72 | **0.322** | 11.62 | **0.260** | **12.32** |

TABLE 7.9: Comparison of PSPnet Architecture with Mix Vision Transformer Encoder on Various Datasets

Table 7.9 presents the outcomes of comparing the PSPnet architecture with various Mix Vision Transformer encoders mit_b0, mit_b1, mit_b2, mit_b3, mit_b5 on different datasets, including Reside Indoor, Reside Outdoor, Ihaze, Dense Haze, and NH Haze.

The evaluation results are measured using the SSIM (Structural Similarity Index) and PSNR (Peak Signal-to-Noise Ratio) metrics, which indicate the quality and fidelity of the dehazed images.Across all datasets, the PSPnet architecture combined with different Mix Vision Transformer encoders shows varying performance. The choice of encoder has a notable impact on the dehazing results.When considering the Reside Indoor dataset, the PSPnet with mit_b1 encoder achieves an SSIM of 0.551 and PSNR of 18.73, while on the Reside Outdoor dataset, the PSPnet with mit_b1 encoder achieves an SSIM of 0.624 and PSNR of 21.71. For the Ihaze dataset, the PSPnet with mit_b2 encoder achieves an SSIM of 0.519 and PSNR of 14.93.However, overall, the PSPnet architecture combined with the Mix Vision Transformer encoders demonstrates relatively lower performance compared to other dehazing architectures (U-Net, FPN, and MAnet) on these datasets.

## 7.2 Qualitative Evaluation

In this subsection, a qualitative evaluation of the dehazing architecture is presented using a set of images. The output images generated by the U-Net architecture, along with those from different state-of-the-art methods, are compared on the RESIDE indoor and outdoor datasets. Furthermore, the output images of different models with Mix Vision Transformer encoders are compared across multiple datasets, including RESIDE indoor, RESIDE outdoor, I Haze, Dense Haze, and NH Haze.

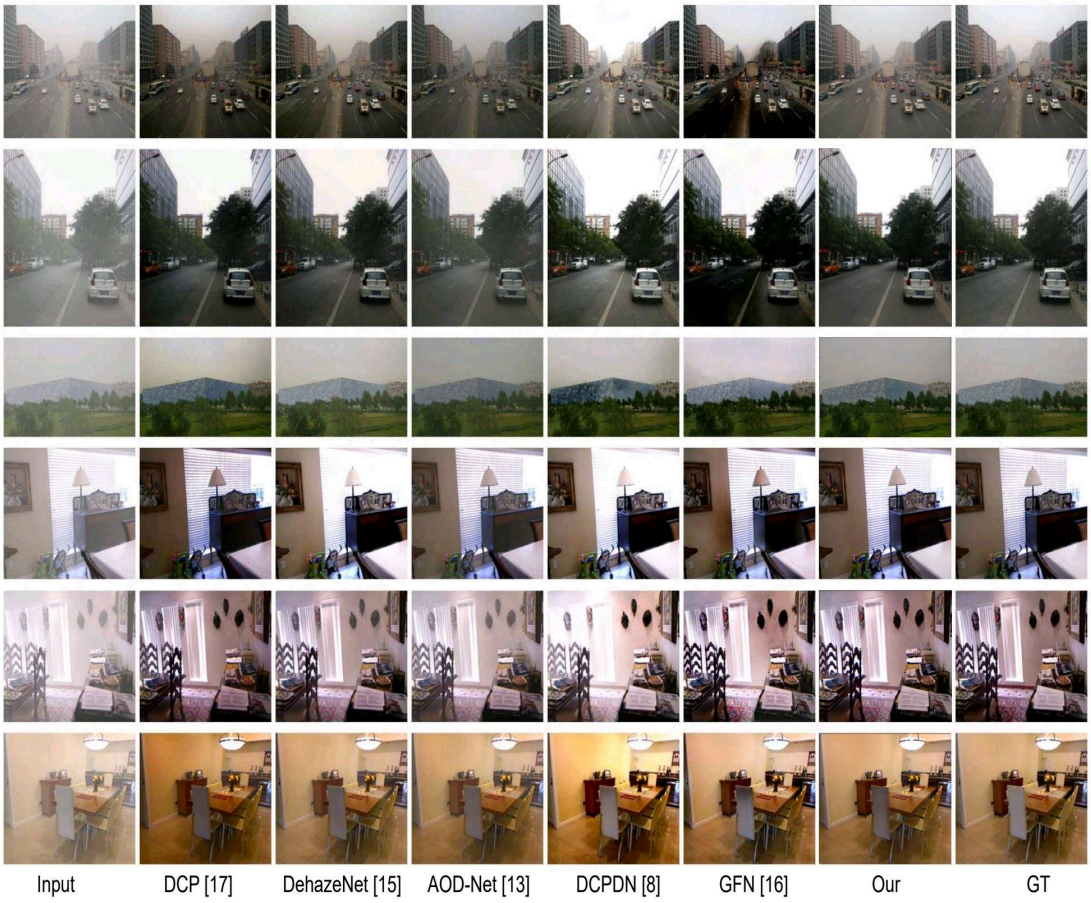


FIGURE 7.1: Comparison with State-of-the-Art Methods on Reside Indoor and Outdoor Datasets

Figure 7.1 showcases the output images generated by the U-Net architecture and several state-of-the-art dehazing methods on the Reside indoor and outdoor datasets. The comparison provides insights into the effectiveness of the proposed approach compared to existing methods in restoring image quality and enhancing visibility in hazy conditions.

Figure 7.2 provides a visual representation of the output images obtained from different models with Mix Vision Transformer encoders on the Reside indoor dataset.Upon careful examination of the images, it becomes evident that the Unet and MAnet models consistently produce better results compared to the FPN and PSPnet models. The

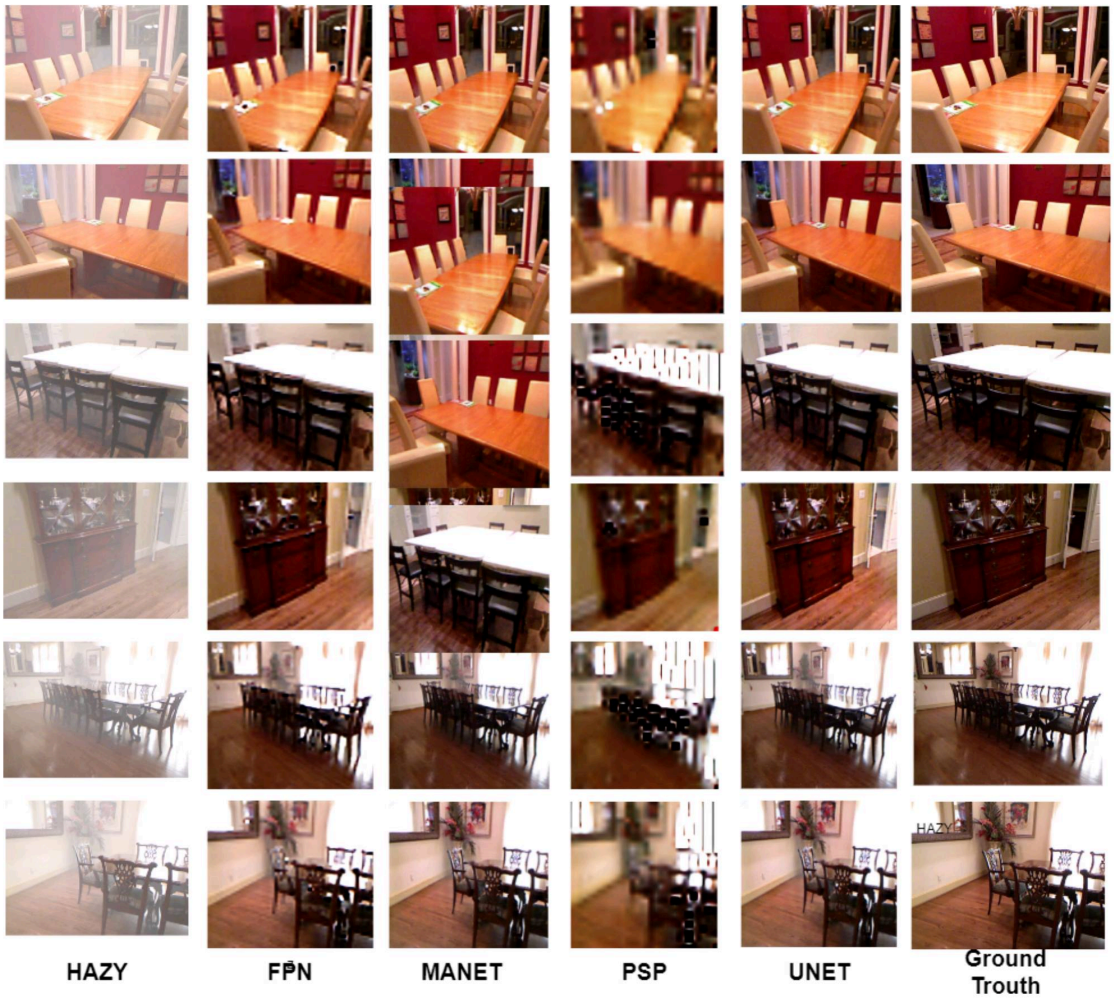


FIGURE 7.2: Comparison of Models with Mix Vision Transformer Encoders on Reside Indoor Dataset

output images generated by Unet and MAnet exhibit clearer details, reduced haze, and improved visibility, thereby indicating their superior performance in handling hazy outdoor scenarios.On the other hand, the FPN and PSPnet models demonstrate relatively lower performance in terms of haze effect removal and image enhancement. The output images generated by these models exhibit comparatively more residual haze and less visible details, indicating their limitations in effectively handling hazy outdoor scenarios. The clear and visually appealing results obtained from these models underscore their effectiveness in capturing fine details and enhancing visibility in outdoor hazy scenarios, making them promising choices for practical applications in computer vision and image processing domains.

Figure 7.3 showcases the output images generated by different models with Mix Vision Transformer encoders on the Reside outdoor dataset. This visual comparison allows us to assess the performance of each model-encoder combination in handling hazy outdoor conditions and enhancing image clarity. Upon examination, it is evident that both the

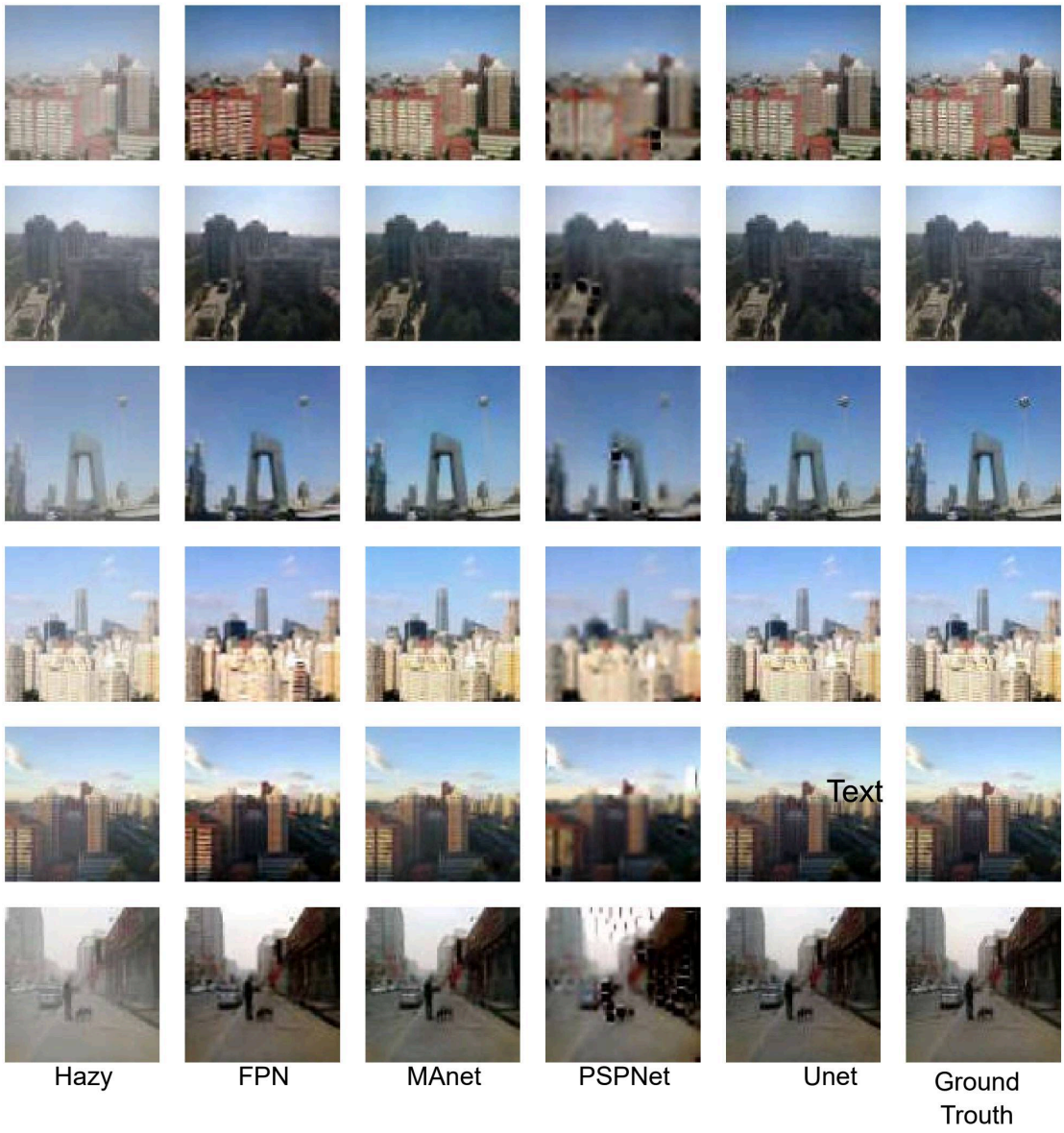


FIGURE 7.3: Comparison of Models with Mix Vision Transformer Encoders on Reside Outdoor Dataset

Unet and MAnet models outperform the FPN and PSPnet models, displaying superior results in terms of haze effect removal and image enhancement. The Unet and MAnet models demonstrate their effectiveness in addressing the challenges posed by outdoor haze and significantly improving visibility and image quality.

Figure 7.4 showcases the output images obtained from different models with Mix Vision Transformer encoders on the I Haze dataset. Despite being trained on a different dataset (Reside), the Unet model consistently outperforms other models in terms of haze effect removal and image enhancement on the I Haze dataset. This demonstrates the Unet model's robustness and ability to handle diverse hazy conditions, producing high-quality dehazed images while preserving important details. The visual comparison reinforces

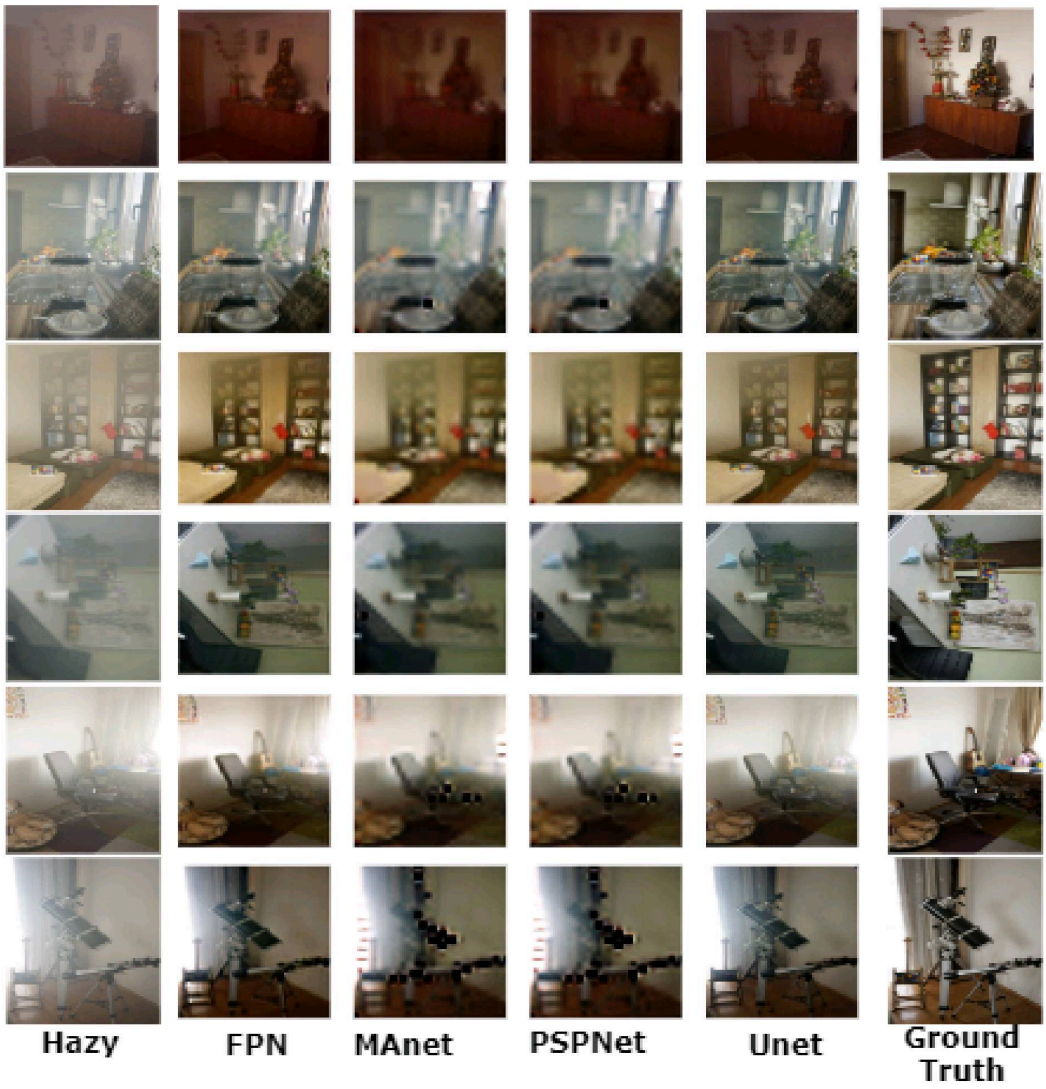


FIGURE 7.4: Comparison of Models with Mix Vision Transformer Encoders on I Haze Dataset

the evaluation results, highlighting the effectiveness of the Unet model with Mix Vision Transformer encoders for haze effect removal tasks, including the specific hazy conditions of the I Haze dataset..

Figure 7.5 presents the output images generated by different models with Mix Vision Transformer encoders on the Dense Haze dataset. This comparison provides insights into the performance of each model in addressing dense haze conditions and its influence on the restoration of image quality. The Dense Haze dataset is characterized by denser haze compared to other datasets, presenting a more challenging scenario for dehazing algorithms. As the models were trained on a different dataset (Reside), none of the models achieved significantly better results in terms of haze effect removal and image restoration on the Dense Haze dataset. This suggests that the models may require

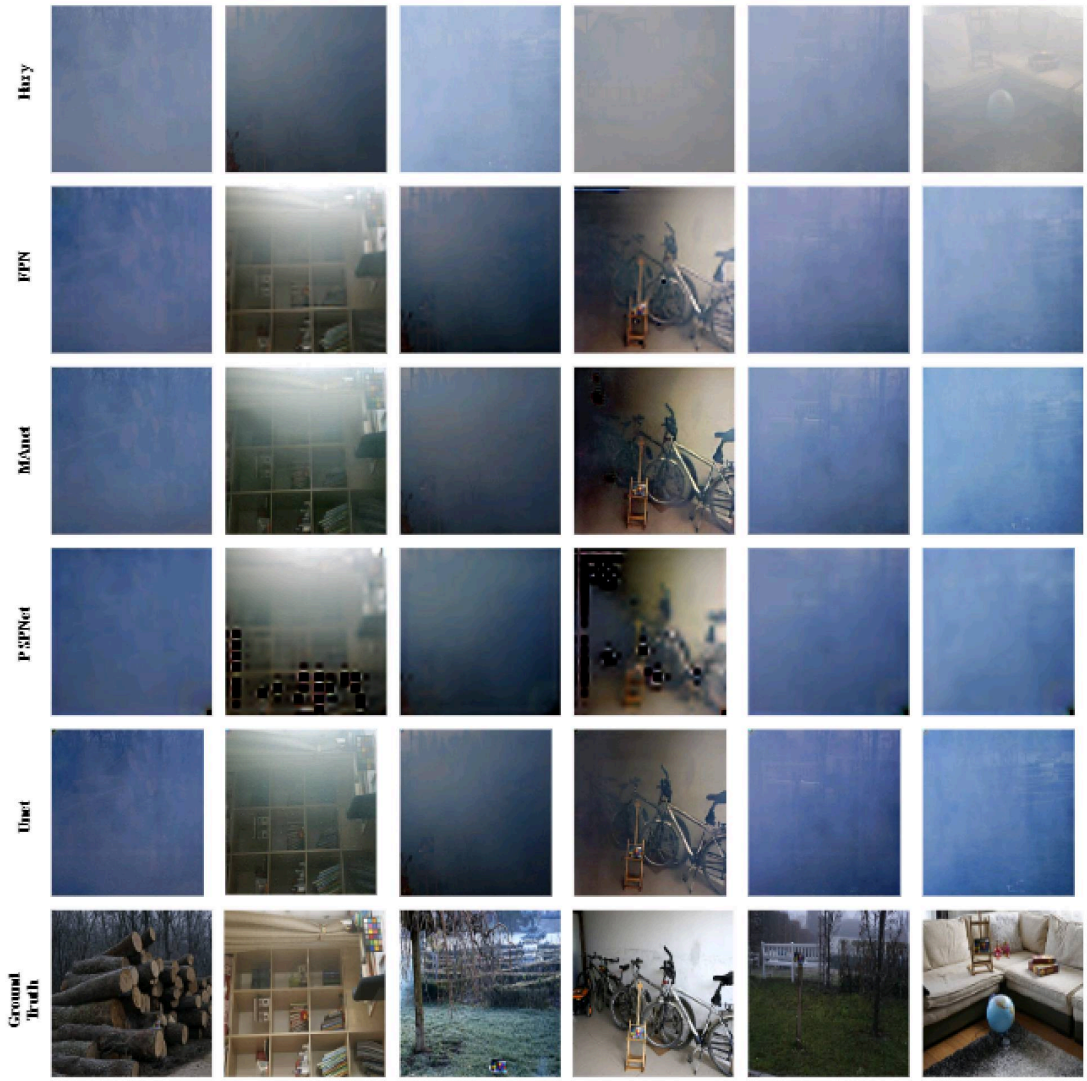


FIGURE 7.5: Comparison of Models with Mix Vision Transformer Encoders on Dense Haze Dataset

specialized training or adaptations to effectively handle the unique challenges posed by dense haze conditions.

Figure 7.6 showcases the output images obtained from different models with Mix Vision Transformer encoders on the NH Haze dataset. This comparison allows us to evaluate the performance of each model in handling the haze conditions specific to this dataset and gain insights into their applicability in real-world scenarios. Since the models were trained on the Reside dataset, none of the models achieved superior results on the NH Haze dataset. This indicates the need for further fine-tuning or adaptation of the models to effectively address the challenges presented by the haze conditions in this particular dataset.


FIGURE 7.6: Comparison of Models with Mix Vision Transformer Encoders on NH Haze Dataset

## 7.3 Evaluation Results

The higher PSNR values achieved by Unet model indicate reduced noise and improved image fidelity, while the higher SSIM values signify better structural similarity and preservation of important image details. This signifies the effectiveness of the proposed dehazing technique in enhancing image quality and visibility, particularly in hazy outdoor scenarios. The superior performance of the Unet model is attributed to its ability to capture spatial information and preserve important details during the dehazing process. By leveraging the power of deep learning and the U-Net architecture, the model effectively removes haze and enhances the visual clarity of outdoor images. These evaluation results validate the potential of the proposed dehazing technique for various applications, such as surveillance, autonomous driving, and outdoor photography, where visibility is critical. The achieved improvements in PSNR and SSIM metrics demonstrate the capability of the Unet model to generate high-quality dehazed images with

enhanced visibility and reduced noise. This contributes to a better user experience and facilitates more accurate analysis and decision-making in real-world scenarios.

Overall, the evaluation results support the effectiveness of the proposed Unet model in addressing the challenges associated with hazy outdoor scenarios and highlight its potential for practical applications in computer vision and image processing domains. Among the architectures, U-Net consistently demonstrates superior performance across all encoder variations. When paired with the mit_b5 encoder, U-Net achieves the highest SSIM and PSNR values, indicating its ability to produce high-quality dehazed images. MAnet also shows competitive performance, particularly when combined with the mit_b3 encoder, highlighting its effectiveness in capturing image details and enhancing visibility in hazy conditions. On the other hand, FPN and PSPnet exhibit lower performance compared to U-Net and MAnet across all encoder variations.The choice of encoder has a noticeable impact on the dehazing performance. Encoders like mit_b5 and mit_b3 demonstrate better performance, indicating their effectiveness in improving image quality and reducing haze. However, the PSPnet architecture combined with different Mix Vision Transformer encoders generally demonstrates relatively lower performance compared to other dehazing architectures (U-Net, FPN, and MAnet) on the evaluated datasets.

These evaluation results highlight the potential and effectiveness of this proposed Unet model in addressing hazy outdoor scenarios and demonstrate its superiority compared to other architectures in terms of dehazing performance.

# Chapter 8

# Conclusion

A GAN-based dehazing model architecture has been proposed, which includes U-Net, PSPNet, MANet, and FPN, along with different encoder blocks such as MobileNet, EfficientNet, ResNet, and Vision Transformer. Extensive evaluations were conducted on multiple datasets, namely Reside indoor, Reside outdoor, I Haze, Dense Haze, and NH Haze, to assess the performance of the architecture and compare it with state-of-the-art methods.The quantitative evaluation results consistently demonstrated the favorable performance of the U-Net architecture with the MobileNet encoder. This architecture consistently outperformed other configurations in terms of the SSIM and PSNR metrics across various datasets, showcasing its effectiveness in restoring image details and improving visibility in hazy conditions.In addition to the quantitative evaluation, a qualitative assessment was carried out by comparing the output images of the proposed architecture with different state-of-the-art methods and model-encoder combinations. These comparisons provided visual evidence of the superior performance of the proposed approach in removing haze and enhancing image quality across different datasets and hazy conditions.Furthermore, the real-time implementation of the dehazing architecture on resource-constrained devices, such as the Raspberry Pi, was addressed. Although the generator output was successfully converted to the ONNX format and the model was optimized for deployment, the limited computational power of the Raspberry Pi posed challenges in terms of processing speed. The implementation achieved a processing speed of 1 FPS, which is not suitable for real-time applications.

Research on dehazing can be further improved by addressing several constraints and challenges encountered during this research. Firstly, expanding the dataset diversity is essential. Although multiple datasets were evaluated, obtaining a larger and more diverse dataset with real-world hazy images would enhance the authenticity and generalizability of the research. However, acquiring such datasets can be challenging due to the scarcity of readily available real hazy images.Additionally, the constraint of limited computational power for training the model should be addressed. Exploring techniques to optimize the model architecture, such as reducing complexity and employing model

compression methods, can help mitigate this constraint. Leveraging parallel computing frameworks or hardware accelerators may also be beneficial in speeding up the training process.Another important aspect to consider is the limited computational power for real-time implementation on devices like the Raspberry Pi. Optimizing the model for deployment on resource-constrained devices can be achieved through techniques like model quantization, pruning, and distillation. Moreover, exploring edge computing solutions or utilizing hardware accelerators can help improve the processing speed on these devices.

## 8.1 Future Scope

The future scope of this research on dehazing using GAN-based models is multi-faceted and encompasses several key areas for improvement and exploration.

Firstly, expanding the dataset diversity is essential. While the Reside dataset was used in this study, obtaining a larger and more diverse dataset with real-world hazy images would enhance the authenticity and generalizability of the research. Collecting such datasets can be challenging due to the scarcity of readily available real hazy images. Collaborations with institutions, organizations, or the research community can facilitate data sharing and acquisition, enabling access to a broader range of hazy scenes and conditions for training and evaluation purposes.

Furthermore, the exploration of different architectural variations holds promise for improving dehazing performance. While architectures like U-Net, PSPNet, MANet, and FPN were investigated in this study, future research can delve into attention-based models, generative adversarial architectures, and novel network structures to uncover more effective and efficient models.The investigation of different encoder blocks is another avenue for improvement. While MobileNet, EfficientNet, ResNet, and Vision Transformer were evaluated, other encoder architectures should be considered to assess their impact on dehazing performance. This exploration can lead to identifying the most suitable encoder block for achieving high-quality dehazed outputs.Optimizing the choice of loss functions is crucial for model training. While this study employed MS-SSIM, L1 loss, and BCEWithLogits loss, future research can explore alternative or domain-specific loss functions. Perceptual loss metrics, adversarial losses, or content-specific constraints can be investigated to further enhance training and improve the visual quality of dehazed images.

Additionally, addressing the constraint of limited computational power for training the model is crucial. Exploring techniques to optimize the model architecture, such as reducing complexity, employing model compression methods (e.g., pruning, quantization), or using knowledge distillation, can help mitigate this constraint. Leveraging parallel computing frameworks or hardware accelerators may also be beneficial in speeding up

the training process, enabling the training of larger models or more extensive experimentation.

Furthermore, optimizing the model for real-time implementation on resource-constrained devices like the Raspberry Pi is an important future direction. Techniques such as model quantization (reducing the precision of weights and activations), network compression (reducing the size of the model), or efficient inference algorithms can be explored to improve the processing speed and resource utilization on such devices. Leveraging hardware accelerators or specialized inference frameworks tailored for low-power devices can also significantly enhance the real-time performance of the dehazing model.

Apart from technical considerations, establishing standardized evaluation metrics and benchmark datasets is crucial for fair comparisons and progress in the field. Developing comprehensive evaluation protocols that consider perceptual quality, haze effect removal effectiveness, and robustness to different hazy conditions will facilitate objective comparisons between different dehazing methods. Similarly, creating benchmark datasets that cover a wide range of hazy scenes, including various atmospheric conditions and different types of hazy distortions, will enable researchers to evaluate and compare their models comprehensively.

Lastly, exploring application-specific adaptations of dehazing models is an important future direction. Tailoring the dehazing algorithms to address specific challenges and requirements in domains such as autonomous driving, surveillance systems, or underwater imaging can lead to significant advancements. Adapting the models to handle motion blur, depth estimation, or scene-specific characteristics can enhance their effectiveness and applicability in real-world scenarios.

By addressing these future scope areas, researchers can make significant strides in improving the robustness, practicality, and effectiveness of dehazing methods. It is important to recognize that research is an iterative process, and each improvement builds upon the knowledge gained from previous work, even within the constraints of limited resources. Continued exploration and improvement in these areas will ultimately lead to better visibility enhancement and improved image quality in hazy conditions across various domains and applications.

# Chapter 9

# List Of Publication

- **M. S. Akhtar**, A. Ali, S. S. Chaudhuri (2023). "Mobile-UNetGAN: A Single Image Dehazing Model." Submitted to Signal, Image and Video Processing.

# Bibliography


[1] Hira Khan, Bin Xiao, Weisheng Li, and Nazeer Muhammad. Recent advancement in haze removal approaches. *Multimedia Systems*, 28, 06 2022.

[2] Anh Van, Tuan V Vu, Hien Hien, Le-Ha Vo, Nhung Le, Phan Nguyen, Prapat Pongkiatkul, and Bich-Thuy Ly. A review of characteristics, causes, and formation mechanisms of haze in southeast asia. *Current Pollution Reports*, 8, 06 2022.

[3] Yangyang Xiang, Rajiv R Sahay, and Mohan S Kankanhalli. Hazy image enhancement based on the full-saturation assumption. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–4. IEEE, 2013.

[4] Le Thanh, Dang Thanh, Nguyen Hue, and Surya Prasath. Single image dehazing based on adaptive histogram equalization and linearization of gamma correction. 11 2019.

[5] Mingye Ju, Can Ding, Y Jay Guo, and Dengyin Zhang. Idgcp: Image dehazing based on gamma correction prior. *IEEE Transactions on Image Processing*, 29:3104–3118, 2019.

[6] Rasmita Lenka, Asimananda Khandual, Koustav Dutta, and Soumya Nayak. *Image Enhancement: Application of Dehazing and Color Correction for Enhancement of Nighttime Low Illumination Image*, pages 211–223. 09 2019.

[7] Mingye Ju, Can Ding, Wenqi Ren, Yi Yang, Dengyin Zhang, and Y Jay Guo. Ide: Image dehazing and exposure using an enhanced atmospheric scattering model. *IEEE Transactions on Image Processing*, 30:2180–2192, 2021.

[8] Sungmin Lee, Seokmin Yun, Ju-Hun Nam, Chee Sun Won, and Seung-Won Jung. A review on dark channel prior based image dehazing algorithms. *EURASIP Journal on Image and Video Processing*, 2016:1–23, 2016.

[9] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:11908–11915, 04 2020.

[10] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021.

[11] Kangfu Mei, Aiwen Jiang, Juncheng Li, and Mingwen Wang. Progressive feature fusion network for realistic image dehazing. In *Asian Conference on Computer Vision (ACCV)*, 2018.

[12] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018.

[13] Kanghui Zhao, Tao Lu, Yu Wang, Yuanzhi Wang, and Xin Nie. Single image dehazing based on enhanced generative adversarial network. In *2020 5th International Conference on Control, Robotics and Cybernetics (CRC)*, pages 129–133. IEEE, 2020.

[14] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023.

[15] N Bharath Raj and N Venketeswaran. Single image haze removal using a generative adversarial network. In *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pages 37–42. IEEE, 2020.

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

[17] Sungmin Lee, Seokmin Yun, Ju-Hun Nam, Chee Sun Won, and Seung-Won Jung. A review on dark channel prior based image dehazing algorithms. *EURASIP Journal on Image and Video Processing*, 2016:1–23, 2016.

[18] Yang Zhao and Yigang Wang. Single image dehazing based on contrastive learning and transformer. In *Journal of Physics: Conference Series*, volume 2450, page 012085. IOP Publishing, 2023.

[19] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4780–4788, 2017.

[20] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 154–169. Springer, 2016.

[21] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[22] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3253–3261, 2018.

[23] Wenyi GE, Yi Lin, Zhitao WANG, Guigui WANG, and Shihan TAN. An improved u-net architecture for image dehazing. *IEICE Transactions on Information and Systems*, E104.D:2218–2225, 12 2021.

[24] Amina Khatun, Mohammad Reduanul Haque, Rabeya Basri, and Mohammad Shorif Uddin. Single image dehazing: An analysis on generative adversarial network. *Journal of Computational Chemistry*, 8:127–137, 2020.

[25] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019.

[26] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe 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.

[27] Javier Hidalgo-Carrió, Daniel Gehrig, and Davide Scaramuzza. Learning monocular dense depth from events. In *2020 International Conference on 3D Vision (3DV)*, pages 534–542. IEEE, 2020.

[28] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 444–445, 2020.

[29] Anil Singh Parihar, Yash Kumar Gupta, Yash Singodia, Vibhu Singh, and Kavinder Singh. A comparative study of image dehazing algorithms. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 766–771. IEEE, 2020.

[30] Zhaoqing Pan, Weijie Yu, Xiaokai Yi, Asifullah Khan, Feng Yuan, and Yuhui Zheng. Recent progress on generative adversarial networks (gans): A survey. *IEEE access*, 7:36322–36333, 2019.