DEEP LEARNING-BASED IMAGE DEHAZING ACROSS DIFFERENT COLOR SPACES

THESIS SUBMITTED IN
PARTIAL FULFILLMENT OF THE
REQUIREMENTSFOR THE AWARD OF
THE DEGREE OF

Master of
Technology
in
Intelligent Automation and Robotics
(Electronics and Telecommunication Engineering Department)
Jadavpur University

by

SINGHAN GANGULY
Class Roll No: 002110704017
Examination Roll No: M6IAR24015
Registration No: 160233 of 2021-2022

Under the guidance of

Prof. Sheli Sinha Chaudhuri
Professor
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 Technology in Intelligent Automation and Robotics under 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 : Singhan Ganguly
Exam Roll No : M6IAR24015
Thesis Title : DEEP LEARNING-BASED IMAGE DEHAZING ACROSS DIFFERENT COLOR SPACES

Signature of Candidate

FACULTY OF ENGINEERING & TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled — **"DEEP LEARNING-BASED IMAGE DEHAZING ACROSS DIFFERENT COLOR SPACES"** has been carried out by **Singhan Ganguly** bearing **Class Roll No: 002110704017, Examination Roll No**: **M6IAR24015** and **Registration No: 160233** of **2021-2022**, under my guidance and supervision and be accepted in partial fulfillment of the requirement for the degree of **Master of Technology** in **Intelligent Automation and Robotics** in the **Department of Electronics and Telecommunication Engineering , Jadavpur University**

Prof. Sheli Sinha Chaudhuri

Supervisor
Department of Electronics
and Telecommunication
Engineering
Jadavpur University

Prof. Manotosh Biswas

Head of the Department
Department of Electronics
and Telecommunication
Engineering
Jadavpur University

Prof. Dipak Laha

Dean
Faculty Council of
Engineering
Jadavpur University
Kolkata- 700032

FACULTY OF ENGINEERING & TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL

The forgoing thesis titled **"DEEP LEARNING-BASED IMAGE DEHAZING ACROSS DIFFERENT COLOR SPACES"** 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

ACKNOWLEDGEMENT

This thesis, titled **"DEEP LEARNING-BASED IMAGE DEHAZING ACROSS DIFFERENT COLOR SPACES"**, is the culmination of my efforts, supported by many individuals, including my advisor and lab seniors. I am grateful for the opportunity to acknowledge their contributions.

First and foremost, I extend my heartfelt thanks to my thesis supervisor, **Prof. Sheli Sinha Chaudhuri**, Professor in the Department of Electronics and Telecommunication Engineering at Jadavpur University. Her invaluable guidance, insightful suggestions, and unwavering support were instrumental in both conducting this research and writing this thesis. I consider myself fortunate to have had such a positive and inspiring mentor. Her confidence and innovative ideas were key to the successful completion of my work, and she has been a true guardian throughout this journey. I am also deeply thankful to our Head of Department, **Prof. Manotosh Biswas**, for his constant assistance and for providing access to all necessary resources for my research.

I would like to extend my heartfelt gratitude to **Mr. Asfak Ali**, a research scholar in our department and a figure akin to an elder brother, for his invaluable assistance with ideation, implementation, and analysis.

Finally, I am profoundly grateful to my father, **Mr Subrata Kumar Ganguly** and my mother **Mrs Sutapa Ganguly**. Their sacrifices, love, encouragement, and support have been indispensable, shaping who I am today.

Singhan Ganguly

Intelligent Automation & Robotics
Department of Electronics and Telecommunication Engineering
Jadavpur University
Kolkata- 700032

Abstract

A difficult problem in computer vision is image dehazing. Since haze alters the color of objects in an image and weakens the intensity of certain light wavelengths, the color becomes a crucial factor to consider when we are trying to remove haze. Primarily, significant performance is achieved using deep learning-based computer vision in RGB, HSV and LAB color models. This study examines the effects of deep convolutional neural networks based on GAN architectures on the elimination of haze in various color spaces. The GAN architecture has generator and discriminator com- ponents 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.

After training and testing in twelve different color spaces, the six deep-learning models show that, for most cases, XYZ, RGB, RGBCIE, and YDbDr consistently beat other color models. Due to differences in pixel intensity information across different color models, the color space translation causes change in the image's color characteristics under foggy situations. In this study, the previously described idea is used to remove haze from foggy images using deep learning networks. The outcomes show a range of responses, with some networks outperforming RGB, HSV, and LAB in terms of SSIM and PSNR. The dehazing performance in RGB, HED, HSV, LAB, RGBCIE, XYZ, YUV, YCbCr, YDbDr, YIQ, YPbPr, and YUV color spaces is discussed in this paper. The goal of the study is to demonstrate how deep-learning models perform in eliminating haze in various color spaces.

Contents

Abstract **vi**

Contents **vii**

List of Tables **x**

List of Figures **xi**

Chapter 1: Introduction **1**

1.1 Haze Imaging Model 2

1.2 Importance of Dehazing 4

1.3 Potential of GAN Based Dehazing 6

Chapter 2: Motivation and Contribution **7**

2.1 Motivation 7

2.2 Contribution 7

Chapter 3: Literature Review **9**

3.1 Existing Dehazing Techniques 9

3.1.1 Prior Based Dehazing Techniques 9

3.1.2 Learning Based Dehazing Techniques 10

3.1.3 GAN Based Dehazing Techniques 12

3.2 Strengths and Limitations of the Existing Techniques 13

3.3 Dehazing in Different Color Spaces 15

Chapter 4: Color Spaces and Transformations **19**

4.1 RGB 19

4.2 HSV 21

4.3 HED 23

4.4 LAB 24

4.5 LUV 26

4.6 RGBCIE 26

4.7 XYZ 27

4.8 YCbCr 29

4.9 YDbDr 30

4.10 YIQ 31

4.11 YPbPr 32

4.12 YUV 33

Chapter 5: Methodology **35**

5.1 Dataset 35

5.2 Color Space Conversion 35

5.3 Base Models 36

5.4 Encoder Models 36

5.4.1 MobileOne 36

5.4.2 ResNet 38

5.4.3 EfficientNet 38

5.4.4 Mix Vision Transformer 41

5.4.5 ResNet 41

5.4.6 GerNet 44

5.5 U-Net Decoder 46

5.6 Discriminator 47

5.7 Loss Function 47

5.8 Training Details 48

Chapter 6: Results and Analysis 49

6.1 Performance Evaluation of Various Deep Learning Models for Image Dehazing Across Different Color Spaces 51

6.1.1 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Mobileone Encoder Model 51

6.1.2 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with ResNet Encoder Model 52

6.1.3 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Efficientnet Encoder Model 52

6.1.4 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Mix Vision Transformer (MIT) Encoder Model 53

6.1.5 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Resnext Encoder Model 54

6.1.6 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with GetNet Encoder Model 55

Chapter 7: Conclusion **56**

Chapter 8: List of Publications **57**

Bibliography **58**

List of Tables

Table 1: PSNR and SSIM Comparison of the Proposed Dehazing Techniques in Different Color Spaces Using Various Encoders on RESIDE6K Indoor, RESIDE6K Outdoor and RESIDE6K Datasets. 50

Table 2: The Results of Comparing Proposed Dehazing Techniques on RESIDE6K with Various State-of-the-Art. 50

List of Figures

Figure 1.1: Image formation under hazy weather conditions and atmospheric scattering model. 2

Figure 1.2: Different components in Atmospheric Scattering Mode 4

Figure 4.1: RGB in Graphical View 20

Figure 4.2: RGB Color Space 20

Figure 4.3: HSV in Graphical View 22

Figure 4.4: HSV Color Space 22

Figure 4.5: LAB in Graphical View 24

Figure 4.6: LAB Color Space 25

Figure 4.7: RGBCIE Color Space 27

Figure 4.8: YCbCr Color Space 29

Figure 4.9: YUV Color Space 33

Figure 4.10: Comparison of Input Hazy Images in Different Color Spaces 34

Figure 5.1: Illustration of MobileOne Block 37

Figure 5.2: Illustration of ResNet Block 39

Figure 5.3: Illustration of EfficientNet Block 40

Figure 5.4: Illustration of Mix Vision Transformer 42

Figure 5.5: Illustration of ResNext Block 43

Figure 5.5: Illustration of GerNet Block 45

Figure 5.7: Illustration of Discriminator 47

Figure 6.1: Comparative Performance of the MobileOne Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces. 51

Figure 6.1: Comparative Performance of the MobileOne Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces. 51

Figure 6.2: Comparative Performance of the ResNet Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces. 52

Figure 6.3: Comparative Performance of the Efficientnet Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces. 53

Figure 6.4: Comparative Performance of the Mix Vision Transformer (MIT) Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces. 53

Figure 6.5: Comparative Performance of the Resnext Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces. 54

Figure 6.6: Comparative Performance of the GerNet Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces. 55

Chapter 1

Introduction

Haze refers to a natural atmospheric condition characterized by the presence of minute particles like dust, smoke, and pollutants. These particles scatter and absorb light, resulting in reduced clarity, contrast, and color saturation of images. This deterioration in image quality, manifesting as decreased visibility, diminished contrast, and altered color accuracy, presents considerable challenges in both computer vision and image processing. The inclusion of haze leads to undesirable artifacts and diminishes the visual appeal of images, representing a widespread problem that significantly degrades their quality and complicates various applications in computer vision.

Haze manifests its effects through several key observations. Initially, it diminishes light intensity, leading to decreased contrast and a faded appearance of distant objects compared to clear conditions. Additionally, haze causes a color cast, typically appearing as a bluish or grayish tint, which alters the accurate representation of colors in scenes. Lastly, haze diminishes the visibility of intricate details and textures, complicating the accurate perception of crucial image features.

The presence of haze in computer vision and remote sensing applications can significantly affect algorithm performance and data interpretation. In tasks like object detection, recognition, and tracking, haze diminishes the clarity of objects, making it challenging to discern fine details and accurately identify objects. This can result in errors in object localization and recognition, undermining the reliability of computer vision systems. Similarly, in remote sensing applications such as aerial and satellite imagery, haze obstructs the observation of crucial features on the Earth's surface. It limits visibility of geographical landmarks, vegetation, water bodies, and other environmental elements, making it difficult to extract meaningful information and derive accurate insights. This limitation can impede tasks like land cover classification, environmental monitoring, and disaster assessment, hindering our ability to understand and manage our environment.

Developing robust dehazing techniques can significantly enhance the visual quality of hazy images, improve the performance of image analysis algorithms, and enable better understanding of scenes captured under such conditions. These techniques help mitigate the adverse effects of haze, such as reduced contrast and color distortion, thereby enhancing the clarity and detail of images. By improving visibility and restoring true colors, dehazing methods can aid in more accurate object detection, recognition, and tracking in computer vision applications.

Additionally, they can enhance the interpretability of remote sensing data, allowing for better analysis of geographical features and environmental conditions. Overall, addressing the challenges posed by haze through advanced dehazing techniques can lead to more effective image analysis and interpretation in various fields.

This thesis investigates the effectiveness of using Generative Adversarial Networks (GANs), known for creating realistic images, to enhance the removal of haze from images. Rather than relying solely on conventional color representations like RGB, the study explores a range of color spaces including HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV. These color spaces offer unique characteristics that can facilitate more effective haze removal. For instance, LAB and YCbCr separate brightness from color information, aiding in estimating and correcting haze without distorting colors. HSV, on the other hand, separates color into hue, saturation, and value, which is particularly useful for addressing how haze affects color saturation. The study employs a cutting-edge approach that leverages these diverse color spaces in conjunction with GANs to ensure that dehazed images not only appear clear but also retain their true colors and details. Ultimately, the research aims to advance haze removal techniques by harnessing the capabilities of GANs within various color spaces. Through extensive testing, the study seeks to develop innovative methods to enhance the vibrancy and realism of hazy images.

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.

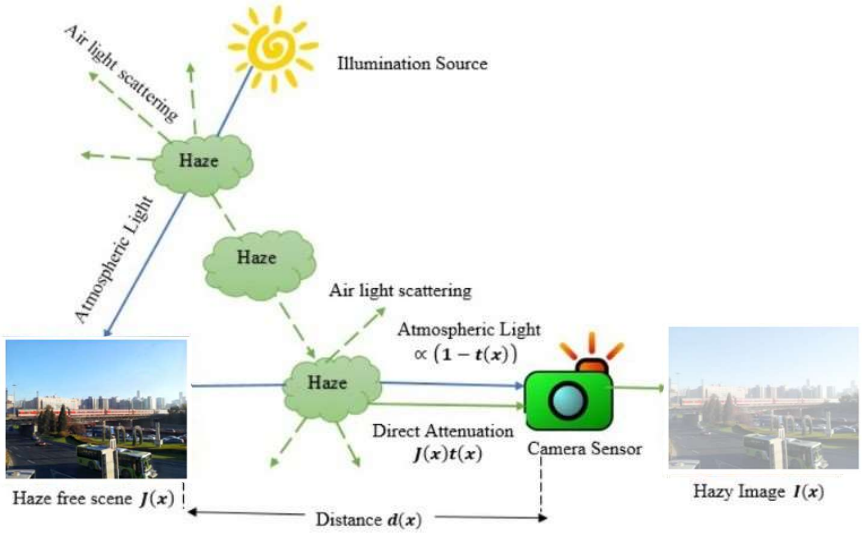


Figure 1.1: *Image formation under hazy weather conditions and atmospheric scattering model.*

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.

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 removal. 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 removal in different dehazing methods.

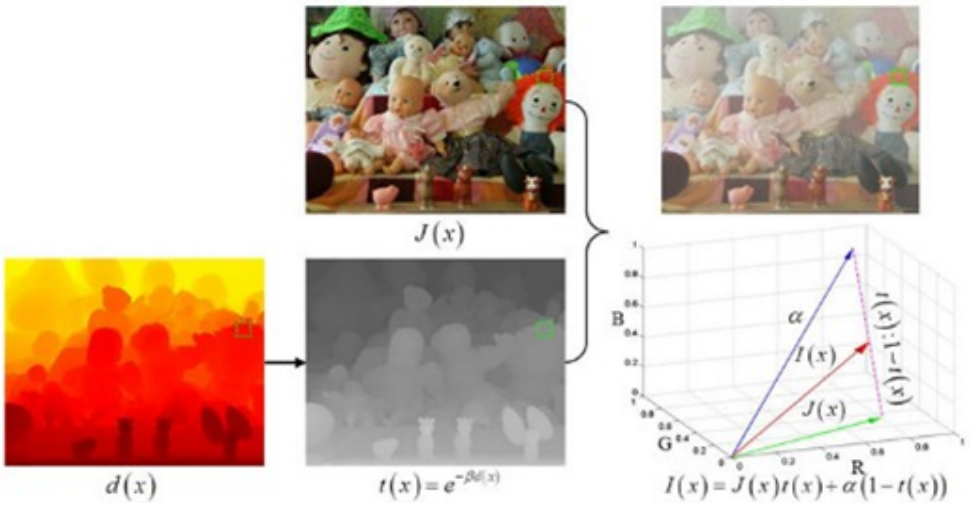


Figure 1.2: *Different components in Atmospheric Scattering Model.*

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 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 brings along 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 is one of the simplest dehazing techniques and involves increasing the hazy image's contrast and brightness. Histogram equalization [1], gamma correction [2], and color balance adjustments [3] are some of the methods that may be used to accomplish this. However, for 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 [4] 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 [5] 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 [6,7,8] 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). Utilizing 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 often rely on predefined assumptions about the haze degradation process, which can lead to loss of image details or require manual feature extraction, limiting their universality. Methods like the Dark Channel Prior (DCP) and Color Attenuation Prior (CAP) utilize statistical properties and priors related to haze and scene characteristics, but struggle with complex scenes and varying haze conditions. As a result, there is a growing need for more advanced dehazing techniques that can adapt to different scenarios.

Despite the positive outcomes of deep learning-based dehazing techniques, many still evaluate ambient light and transmittance values using physical scattering models through CNNs. These methods often use mean square loss on a per-pixel basis, resulting in image smoothing and loss of detail. To address these limitations, more effective learning algorithms and loss functions are being explored, including the use of Generative Adversarial Networks (GANs) and perceptual loss functions.

GANs have shown promise in dehazing by capturing complex image patterns and generating visually realistic outputs. They leverage 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 visually pleasing results. By training on large datasets of paired hazy and clear images, GANs can effectively remove haze and enhance visibility.

Comparative analysis plays a crucial role in evaluating the effectiveness of GAN-based models in dehazing. It helps in understanding the strengths and limitations of different GAN models and their components, guiding future research and application development. By evaluating different architectural choices, encoder blocks, loss functions, and datasets, researchers can identify the most effective configurations and techniques for GAN-based dehazing. This analysis also helps in assessing the performance of GAN-based dehazing models on different datasets, providing insights into their generalization capabilities and robustness under varying hazy conditions.

Chapter 2

Motivation and Contribution

2.1 Motivation

Haze removal has been a focal point for numerous researchers, primarily concentrating on the RGB, HSV, and LAB color spaces. However, there exist twelve distinct color spaces, including RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV. The presence of haze alters the color of objects in images due to its attenuation of different light wavelengths, emphasizing the critical role of color in haze removal. Haze tends to scatter shorter wavelengths of light, such as blue and violet, more than longer wavelengths like red and orange, leading to reduced color saturation and contrast. Algorithms designed for haze removal must account for these color shifts to accurately restore the original colors of objects in the scene. By considering color information, haze removal techniques can significantly enhance the visual quality of the final image and more faithfully reconstruct the original appearance of objects.

This underscores the pressing need to investigate whether there is an inherent relationship between the characteristics of deep learning models and their performance in specific color spaces. It also raises the question of whether alternative color spaces can outperform the RGB color space in haze removal tasks. Therefore, a comprehensive analysis is essential to evaluate the impact of color spaces on the performance of deep-learning models for haze removal. Such an analysis could provide valuable insights into the optimal choice of color space for haze removal tasks, potentially leading to improvements in the effectiveness and efficiency of haze removal algorithms.

2.2 Contribution

This thesis makes several key contributions in the field of image dehazing:

- Exploration of Dehazing in Twelve Color Spaces: The paper thoroughly investigates the effectiveness of dehazing across twelve different color spaces, including RGB, YCbCr, YDbDr, YIQ, YPbPr, LAB, LUV, RGBCIE, XYZ, HED, and HSV. This analysis provides valuable insights into how different color representations impact the ability to remove haze from images.

- Utilization of Six Encoder Models: To enhance the dehazing process, the paper employs six different encoder models: MobileOne, ResNet, EfficientNet, Mix Vision Transformer (MIT), ResNeXt, and GerNet. By leveraging these diverse models, the study aims to achieve superior dehazing performance and compare the efficacy of different architectures.

- Comparison with State-of-the-Art Models: Finally, the paper compares the performance of the GAN-based dehazing model across the twelve color spaces with state-of-the-art techniques. This comparison is conducted using the RESIDE6K dataset, a benchmark dataset in the field of image dehazing. By benchmarking against established methods, the paper provides a comprehensive evaluation of the proposed approach's effectiveness.

- Overall, this paper contributes significantly to the field of image dehazing by exploring a wide range of color spaces, leveraging diverse encoder models, and providing a thorough comparison with state-of-the-art techniques. These contributions enhance our understanding of effective dehazing methods and provide valuable insights for future research in this area.

Chapter 3

Literature Review

3.1 Existing Dehazing Techniques

Haze removal from hazy images has become a dynamic and thriving research area in the realms of computer vision and image processing. A variety of techniques have been put forth, ranging from prior-based methods [9, 4, 5] to learning-based methods [6,7,8] and the innovative use of generative adversarial networks (GANs) [10, 11, 12, 13, 14]. In this overview, we delve into these approaches, shedding light on their respective contributions and advancements in 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 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. Multiscale fusion methods extend the prior-based approach by considering information from multiple scales to improve transmission map estimation. These methods perform haze 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) [5], introduced by He et. al, leverages the dark channel prior to estimate the transmission map. DCP-based methods have demonstrated strong performance in haze removal, providing a foundation for subsequent research in the field. FFA-Net [6] (Zhang et. al, 2019) utilized feature fusion modules and attention blocks to capture contextual information and emphasize relevant features. AECR- Net [7] developed by Zhang, incorporated adaptive enhancement and context reasoning modules to improve dehazing performance. PFF-Net (proposed by Cai et. al [8], uses a pyramid feature fusion module to extract multi-level features from the ground truth image for estimating the transmission pattern and then generating the dehazed image. Liu et. al [15] 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 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.

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 modelling of intermediate variables. The network learns the transformations required to remove haze 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 presented AOD-Net [16], 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 introduced a multi-scale fully convolutional network (MS-CNN) [17] 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's [16] method on benchmark datasets. DehazeNet [18] employed a CNN architecture that employs a transmission-guided layer-wise propagation strategy to eliminate the haze. Gated Fusion Network (GFN) [19], 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 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 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 Dehaze- GAN [13], introduced by Hu, addressed haze removal by utilizing a generator- discriminator framework. Dehaze GAN employed a conditional GAN to learn the map- ping 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 [10] 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. [11] 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. Dehaze Former [12] 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 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 the Existing Techniques

The limitations of prior-based dehazing methods stem from their reliance on specific assumptions about haze and scene properties. These assumptions may not always hold true, leading to inaccurate transmission map estimation and compromised dehazing results, especially in complex scenes where assumptions about haze

distribution or color distortion may not be valid. Despite these limitations, prior-based techniques have shown promising outcomes in various practical applications and offer valuable insights into the dehazing process.

On the other hand, learning-based methods have demonstrated superior performance compared to traditional approaches in terms of accuracy and visual quality. They can capture complex relationships between hazy and clear images, adapt to various types of haze, and handle complex scenes. However, learning-based methods require substantial training data and computational resources. The availability of suitable datasets and computational complexity are important considerations when applying learning-based dehazing techniques. GAN-based dehazing techniques present several key advantages over prior-based and learning-based methods, making them an exciting and promising approach for enhancing image quality and visibility in dehazing scenarios. Firstly, GAN-based methods can 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 achieved significant success in removing haze from images. CNN-based methods have shown favourable outcomes in single-image dehazing, while GAN-based methods have achieved state-of-the-art results by generating high- quality haze-free images. Researchers have proposed several variations of these techniques to address their limitations and achieve improved performance. GAN-based feature extraction methods, in particular, 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 extract 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 previous works, this paper introduces a novel single-image dehazing algorithm that utilizes a generative adversarial network based on a Unet network. Our proposed method improves the feature utilization within the generator, resulting in enhanced image dehazing quality and efficiency.

3.3 Dehazing in Different Color Spaces

Traditional methods often operate in grayscale or other color spaces, but this research has explored the effectiveness of image dehazing directly in the twelve-color spaces. This literature review dives into key methodologies that con-tribute to the understanding and advancement of image dehazing in all such color spaces.

In 2015 [20] introduced a novel image dehazing method in the HSV color space, demonstrating that the H channel is haze-invariant. By treating dehazing as a contrast enhancement problem, the proposed algorithm seamlessly combined dehazing and contrast enhancement, yielding superior performance over existing methods. Experimental results validated the efficacy and potential of this innovative approach. Othman et. al [21] addressed image degradation caused by different weather conditions, proposing a novel method to estimate airlight in hazy images. Utilizing the HSV color space, the approach identified the haziest area and then calculated the airlight accordingly. Experimentation on the RESIDE SOTS dataset demonstrated superior performance in peak signal-to-noise ratio and structural similarity index compared to existing techniques. Wang et. al [22] presented a non-local saturation algorithm for single image dehazing, addressing block and halo effects associated with the dark channel prior. The RGB to HSV color space conversion incorporated a non-local method, with saturation weight based on window size.

The proposed method effectively restored image color, enhanced contrast, and outperformed existing approaches, as validated through quantitative and qualitative evaluations. Experimental results demonstrated superior visual effects and high efficiency. Mahdei et. al [23] came up with a novel de-hazing method involving pre-processing with an average filter, dark channel prior estimation, and a unique approach to airlight value estimation through RGB and HSV standard deviations. The method employed rules derived from experiments, integrated local contrast enhancement using HSV's V channel, and exhibited promising results in visual and metric assessments, outperforming other algorithms. Ranota et. al [24] stated the challenge of dehazing degraded images in bad weather. Introducing a method using a modified dark channel prior and an adaptive Gaussian filter, it operated in the LAB color space, applied Adaptive Histogram Equalization for contrast improvement, refined transmission maps effectively, and employed an Adaptive Gaussian filter for further enhancements. The results demonstrated efficient haze removal and clear reconstruction of fine details in the original scene.

Neumann et. al [25] advocated the application of imaging polarimetry in general photography. It has introduced high-resolution methods for determining linearly polarized components (Q,U) using the CIE Luv color space, enabling visualization and representation of polarization attributes. The study has proposed a polarization-based de-reflection method for image enhancement, capable of removing haze and amplifying specular effects, offering versatile applications in photography for optimal image quality and artistic effects. The work of [26] has addressed the significance of image dehazing for societal applications. It emphasizes the need to identify color channels that are most affected by haze in various color spaces for accurate dehazing. The research conducted a comprehensive analysis of histograms and intensities in RGB, HSV, and LAB color spaces for both daytime and night time hazy images, aiming to distinguish the most influenced channels for effective restoration of image visibility. Employing particle swarm optimization for parameter optimization in [27], the method has utilized Retinex-based processing in the HSI color space, overcoming texture preservation issues. Incorporating multi-scale detail enhancement and bilateral filtering has enhanced overall visual quality, demonstrated through favourable results in natural and synthetic images, emphasizing de-fogging effectiveness, particularly in underwater and aerial scenarios. The study in [28] has addressed halo artifacts in the widely used dark channel prior (DCP) image dehazing method by introducing an adaptive mask size. This proposed approach has employed the inverse ratio of the gradient, determining mask size based on gradient magnitude. Smoothing gradients, correcting the y function with a Gaussian filter, and optimizing parameters using ant colony optimization have minimized halo artifacts, as validated by the proposed dehazing performance index (DPI) in experimental results. Bui et. al [29] puts forward a novel single-image dehazing method using color ellipsoids in RGB space. The proposed method has calculated transmission values by statistically fitting ellipsoids to haze pixel clusters and maximizing contrast without over-saturation. It has embedded a fuzzy segmentation process to simultaneously execute transmission calculation and refinement, avoiding complex post-processing.

Experimental results have demonstrated its effectiveness across various haze and noise levels without visible artifacts, making it suitable for real-time applications due to its relatively low complexity. Li et. al [30] presented a fusion-based technique for underwater image enhancement, addressing color distortion and haze. It has categorized underwater images based on hue channel statistics, applied category-specific color correction algorithms, employed a haze-line prior for dehazing, and utilized channel-wise fusion in the CIE Lab color space. Experimental results have demonstrated the robustness and state-of-the-art performance of the proposed method on various underwater image datasets. Afifi et. al [31] addressed the challenge of "un-processing" nonlinear images by introducing CIE XYZ Net, a deep learning framework. It leverages the camera's canonical CIE XYZ color space, not available as an output, to un-process images and subsequently apply low-level computer vision tasks. The proposed approach proves effective across various vision tasks, mitigating the sensor-specific limitations of existing methods. Tufail et. al [32] introduced an improved dark channel prior (DCP) method for defogging images affected by environmental factors like fog and haze. The method refines transmission maps in both RGB and YCbCr color spaces, preserving edge information. Evaluation against existing methods have demonstrated superior results in terms of fog reduction, structural similarity, and visual appeal, making it effective for applications like intelligent vehicles and surveillance. Das et. al [33] addressed the challenge of poor visibility caused by various atmospheric particles in adverse weather conditions. It has introduced a modified dark channel prior-based image dehazing system that enhances the transmission map using RGB and YCbCr color systems. The method incorporates the Relativity of Gaussian and Guided filter, aiming to improve transmission map accuracy without over-enhancing images, resulting in distortion-free, color artifact-free, and visually appealing reconstructed images. In [34], they have introduced a novel method for dehazing aerial images by enhancing the light value (Y) in the YCbCr color space using Sigmoid function mapping and Adaptive Histogram Equalization. The chromatic components (CbCr) are improved through Dark Channel Prior (DCP). Evaluation using entropy and average gradient measures have demonstrated the proposed method's effectiveness compared to existing techniques. Hashim et. al [35] presented a novel haze removal algorithm using the YIQ color space and Dark Channel Prior (DCP). The method combines contrast and lightness enhancement with DCP applied to RGB images. Con-verting the result to YIQ and combining components, yielded high-quality, haze-free images, outperforming other methods based on various quality measures in outdoor image applications. Khalaf et. al [36] addressed the contrast issues in aerial images caused by dust and pollutants. Employing modified dark channel prior (DCP) and adaptive histogram equalization (AHE) algorithms in the YIQ color space, it achieved superior image enhancement. The proposed method outperformed alternatives and demonstrated higher quality measures, including entropy (7.771), average gradient (14.207), and naturalness image quality evaluator (3.700). Zhang et. al [37] addressed cataractous retinal image dehazing by combining image denoising and sigmoid functions. Introducing a double-pass fundus reflection model in the YPbPr color space, the MUTE strategy employed multilevel stimulated denoising. The cataract layer's transmission matrix integrated denoised raw images via pixel-wise sigmoid functions, refined through adaptive momentum estimation.

The proposed method outperformed existing techniques in visual and objective assessments, with potential applications in blood vessel segmentation, retinal image registration, and enhanced diagnostic imaging. Fang et. al [38] addressed single-image dehazing, prioritizing luminance over chrominance based on the haze formation model. Utilizing the YUV color space, a variational model is proposed for the Y channel, preserving chrominance information during dehazing. The numerical procedure, employing ADMM, yielded optimal solutions. Experimental results on real and synthetic hazy images showcased the method's effectiveness in unveiling details, recovering vibrant color, and serving as a competitive dehazing and image enhancement approach. Liu et. al [39] addressed issues in single-image haze removal algorithms, presenting a weight-based quad-tree hierarchical search for more accurate selection of atmospheric light and region-growing seed points. The method has optimized transmission in segmented sky or fog regions, employed an edge-preserving filter, and reduced time complexity through down sampling and YUV color space. Experimental results have confirmed the algorithm's effectiveness in accurately evaluating atmospheric light, segmenting images, and producing clear, natural haze-free images. Wen et. al [40] presented a novel approach for removing thin clouds (haze) in satellite imagery. In the proposed method, a YUV-GAN, has utilized a Wasserstein generative adversarial network in the YUV color space, enabling efficient removal of clouds by learning luminance and chroma components independently. A transfer-learning based training strategy has optimized the model's applicability to real cloudy images, demonstrating state-of-the-art performance on RICE1 and Sentinel-2A datasets.

Chapter 4

Color Spaces and Transformations

A specific method of representing and structuring colors in a digital image or video is known as a color space. This refers to a mathematical model which are mapped to numerical values. Different color spaces are used in image processing, including RGB (Red, Green, Blue), HSV (Hue, Saturation, Value), and LAB (Lightness, a (green to red), b (blue to yellow). Each color space has its own unique properties and can be used for different purposes. This section provides an overview of the twelve color spaces such as RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YPbPr, YIQ, YUV.

4.1 RGB

RGB color space is the most commonly used color space in image classification. It represents color as a combination of red, green, and blue values. In this color space, each pixel in an image is represented by three numbers ranging from 0 to 255, which corresponds to the intensity of the red, green, and blue colors.

RGB (Red, Green, Blue) is a color model used to represent and display colors in digital devices, such as computer monitors, televisions, and cameras. It is based on the additive color theory, which suggests that when red, green, and blue light are combined at full intensity, they create pure white light.

In the RGB color space, colors are represented by three values that indicate the intensity of each primary color. Each value ranges from 0 to 255, where 0 represents no intensity and 255 represents full intensity. For example, pure red would be represented as (255, 0, 0) in RGB, indicating that the red component has full intensity and the green and blue components have no intensity.

The RGB color space can produce a wide range of colors using just three primary colors, making it a popular choice for digital devices. However, it is not ideal for printing, as the ink used in printers is based on the subtractive color model, which is different from the additive color model used in RGB. Therefore, RGB colors may appear differently when printed.

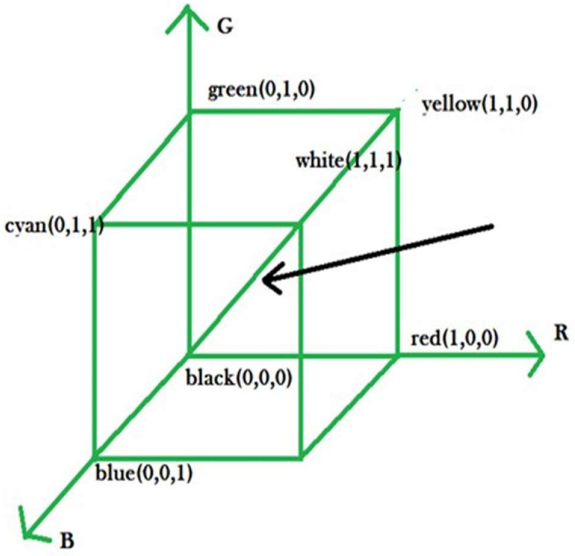


Figure 4.1: *RGB in Graphical View*

There are several variations of the RGB color space, including sRGB, Adobe RGB, and ProPhoto RGB. sRGB is the most widely used RGB color space and is commonly used for web graphics and digital images. Adobe RGB has a larger color gamut than sRGB and is used in professional photography and graphic design. ProPhoto RGB has an even larger color gamut and is used for high-end printing and professional photography.

In addition to the RGB color space, there are other color models used in digital imaging, such as CMYK (Cyan, Magenta, Yellow, Black) for printing and HSL/HSV (Hue, Saturation, Lightness/Value) for color selection and adjustment. Each color model has its strengths and weaknesses and is suited for different applications.

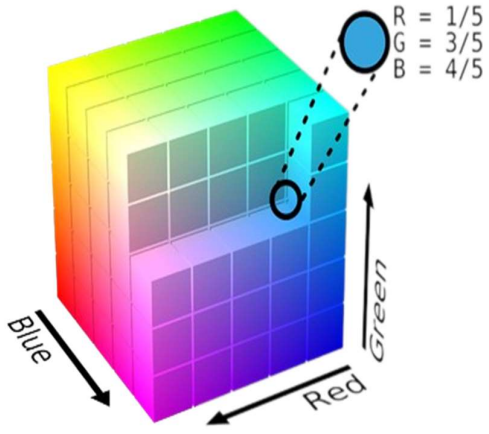


Figure 4.2: *RGB Color Space*

4.2 HSV

HSV (Hue, Saturation, and Value) is a color space that is commonly used in computer graphics, image processing, and color analysis. The HSV color space represents colors in terms of their hue, saturation, and value, providing a more intuitive way of describing color than the RGB color space.

Hue refers to the color of the light, and is measured in degrees around a color wheel. In the HSV color space, hues range from 0 to 360 degrees, with red at 0 degrees, green at 120 degrees, and blue at 240 degrees.

Saturation refers to the purity of the color, or how much gray is mixed with the hue. Saturation ranges from 0 (gray) to 100 (fully saturated). Value refers to the brightness of the color, or how much light is reflected. Value ranges from 0 (black) to 100 (white). RGB to HSV calculated as:

$$\begin{aligned}
\mathrm{R}' &= \mathrm{R}/255 \\
\mathrm{G}' &= \mathrm{G}/255 \\
\mathrm{B}' &= \mathrm{B}/255 \\
\mathrm{Cmax}' &= \max\left(\mathrm{R}', \mathrm{G}', \mathrm{B}'\right) \\
\mathrm{Cmin} &= \min\left(R', G', B'\right)
\end{aligned}$$

Δ = Cmax - Cmin Hue calculation:

$$H := \begin{cases} 0, & \text{if } C = 0 \\ 60^\circ \cdot \left(\frac{G-B}{C} \bmod 6\right), & \text{if } V = R \\ 60^\circ \cdot \left(\frac{B-R}{C} + 2\right), & \text{if } V = G \\ 60^\circ \cdot \left(\frac{R-G}{C} + 4\right), & \text{if } V = B \end{cases}$$

Saturation calculation:

$$S_V := \begin{cases} 0, & \text{if } V = 0 \\ \frac{C}{V}, & \text{otherwise} \end{cases}$$

$$S_L := \begin{cases} 0, & \text{if } L = 0 \text{ or } L = 1 \\ \frac{C}{1-|2V-C-1|} = \frac{2(V-L)}{1-|2L-1|} = \frac{V-L}{\min(L,1-L)}, & \text{otherwise} \end{cases}$$

Value calculation: $\mathrm{V} = \mathrm{Cmax}$

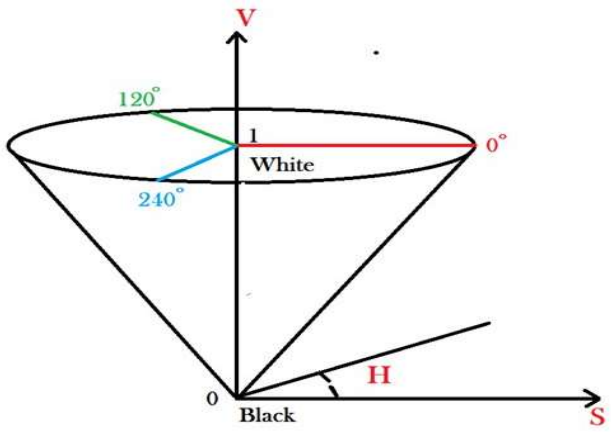


Figure 4.3: *HSV in Graphical View*

In the HSV color space, colors are represented as a cylindrical coordinate system, with the hue angle around the center axis and the saturation and value axes extending outward from the center. This allows for easy manipulation of colors by changing their hue, saturation, or value independently. The conversion from RGB to HSV is a non-linear transformation that involves normalizing the RGB values to a range of 0 to 1, and then calculating the hue, saturation, and value based on the normalized values. The formula for converting RGB to HSV can vary depending on the software or application used, but typically involves some form of normalization and calculation of the hue, saturation, and value components. The HSV color space is useful for a variety of color-related tasks, such as color selection, color manipulation, and color segmentation. For example, in computer graphics, the HSV color space can be used to create color palettes that are more aesthetically pleasing or to adjust the brightness and saturation of images. In image processing and computer vision, the HSV color space can be used for color-based segmentation and object detection, as it provides a more intuitive way of describing color than the RGB color space.

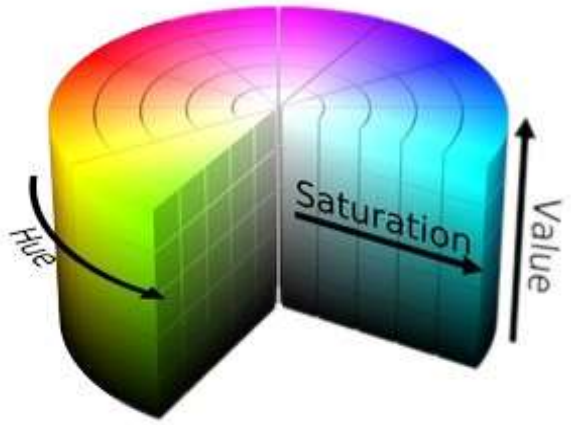


Figure 4.4: *HSV Color Space*

HSV color space, on the other hand, represents color as a combination of hue, saturation, and value. Hue represents the color itself, saturation represents the purity of the color, and value represents the brightness of the color.

LAB color space is another color space that is widely used in image processing. It is based on the opponent process theory of color vision, which states that there are three pairs of complementary colors: red-green, blue-yellow, and black-white. In this color space, the values of L, a, and b represent the lightness, red-green axis, and blue-yellow axis, respectively.

4.3 HED

The analysis of biomedical images commonly employs the HED (Hematoxylin-Eosin-DAB) color space. This particular color space effectively segregates an image into three channels, aligning with the staining components of hematoxylin (blue-purple), eosin (pink), and DAB (brown). To create the HED color space, the initial step involves transforming an RGB image into an optical density space. In this space, the values of each channel reflect the logarithm of the ratio of light transmitted through the stained tissue to that transmitted through an unstained reference. Facilitating this transformation is a color deconvolution algorithm, responsible for effectively separating the image into three channels, each corresponding to the staining components. RGB to HED is calculated as:

$$\begin{bmatrix} H \\ E \\ D \end{bmatrix} = \begin{bmatrix} 0.65 & 0.70 & 0.29 \\ 0.07 & 0.99 & 0.11 \\ 0.27 & 0.57 & 0.78 \end{bmatrix} \cdot \begin{bmatrix} R^{\phi} \\ G^{\phi} \\ B^{\phi} \end{bmatrix}$$

[28] proposed a fast, artifact-free method for color dehazing based on visual perception models, applicable potentially in real-time. Operating in the HSV color space, the method has enhanced images by reducing the value component with globally constrained histogram flattening, adjusting the saturation component accordingly, and locally enhancing the value component. This approach avoids color artifacts and the need to invert physical equations. This method has competed well with the state-of-the-art for standard hazy images and has surpassed it for challenging cases. Notably, it can dehaze a Full HD image on a GPU in just 90 ms, highlighting its efficiency and effectiveness.

4.4 LAB

LAB (or sometimes written as CIELAB) is a color space that was developed by the International Commission on Illumination (CIE) as a device-independent way to describe colors. It is based on the CIE 1931 XYZ color space, which defines a set of color values that correspond to the three types of cone cells in the human eye that are responsible for color vision.

The LAB color space consists of three channels: L, a, and b. The L channel represents the lightness or brightness of a color, and ranges from 0 (black) to 100 (white). The channel represents the color on a red-green axis, with positive values representing red and negative values representing green. The b channel represents the color on a blue-yellow axis, with positive values representing yellow and negative values representing blue.

The LAB color space is device-independent, which means that it is not affected by the characteristics of the device used to create or display the color. This makes it a useful color space for color management, where it is used to accurately match colors between different devices, such as monitors and printers.

One of the key advantages of the LAB color space is that it is designed to be perceptually uniform, meaning that equal differences in color values are perceived as equal differences in color by the human eye. This makes it easier to make precise color adjustments and corrections, as changes in the LAB values correspond more closely to changes in human perception of color.

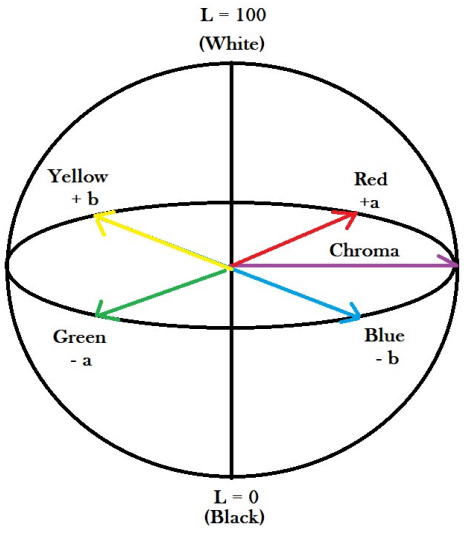


Figure 4.5: *LAB in Graphical View*

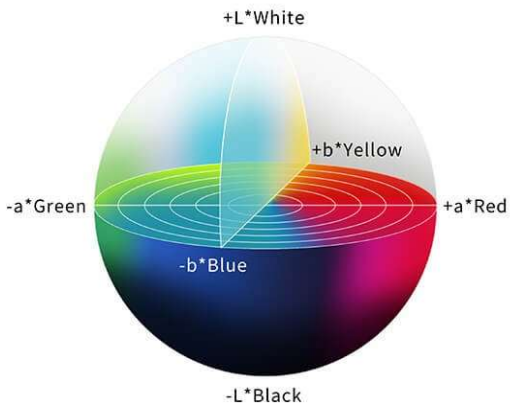


***Figure 4.6**: LAB Color Space*

The LAB color space is used in a variety of applications, including digital imaging, graphic design, and color management. It is particularly useful in color correction and color matching applications, where it allows for precise adjustments to be made to the colors in an image or design.

What sets Lab apart from other color models is its wide color spectrum, allowing it to differentiate between very similar shades that are perceptible to the human eye. The model uses three parameters or axes: L for lightness, ranging from 0 to 100; A for red to green, with negative values representing green and positive values representing red; and B for yellow to blue, with negative values indicating blue and positive values indicating yellow.

$$
\begin{aligned}
L^\star &= 116 f\left(\frac{Y}{Y_\mathrm{s}}\right) - 16 \\
a^\star &= 500\left(f\left(\frac{X}{X_\mathrm{s}}\right) - f\left(\frac{Y}{Y_\mathrm{s}}\right)\right) \\
b^\star &= 200\left(f\left(\frac{Y}{Y_\mathrm{s}}\right) - f\left(\frac{Z}{Z_\mathrm{s}}\right)\right) \\
f(t) &= \begin{cases} \sqrt[3]{t} & \text{if } t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{4}{29} & \text{otherwise} \end{cases} \\
\delta &= \frac{6}{29}
\end{aligned}
$$

Here X,Y,Z are the standard color stimulus and Xs, Ys, Zs is a specific white, achromatic reference illuminant.

4.5 LUV

In 1976, the International Commission on Illumination (CIE) introduced the CIELUV color space as a simpler transformation of the CIE XYZ color space, aiming for perceptual uniformity.

Unlike CIELAB, which uses a von Kries trans-form, CIELUV employs Judd-type white point adaptation. While CIELUV can be effective for working with a single light source, it may predict colors that don't exist in reality when used for chromatic adaptation.

Image dehazing is a process aimed at recovering image details lost due to fog, haze, or smoke during image capture. These atmospheric factors reduce contrast and color information, making enhancement necessary for various imaging applications and consumer photography. While the perceptual quality of color information has been evaluated, there is a lack of research focusing on this aspect. [41] demonstrates the impact of dehazing models on color information using both simulated and real images. A convergence model, simulating haze, is employed to evaluate color loss based on hue angle in the IPT color space, saturation in the CIE LUV color space, and perceived color difference in the CIE LAB color space. The findings show significant changes in saturation and hue, particularly for achromatic and blue/yellow colors, where conventional image processing spaces fail to maintain consistent hue lines. The study proposes a correction model based on color transparency perception to supplement dehazing algorithms and recover color information effectively.

4.6 RGBCIE

RGBCIE is a color space that combines the RGB (Red, Green, Blue) and CIE (Com- mission Internationale de l'Eclairage) color spaces. It is a popular color space used in digital imaging and color management systems.

The RGB color space is based on the additive color model, where colors are created by combining red, green, and blue light. The CIE color space, on the other hand, is a mathematical model that describes how humans perceive color. It is based on the XYZ color space, which defines a set of color values that correspond to the three types of cone cells in the human eye that are responsible for color vision.

The RGBCIE color space combines the RGB color space and the CIE color space to create a more accurate and comprehensive color representation system. It adds two additional color channels to the standard RGB model, which represent the chromaticity and luminance of a color.

The first additional channel, the chromaticity channel, represents the hue and saturation of a color. It is based on the CIE xy chromaticity diagram, which maps all visible colors onto a two-dimensional plane. The chromaticity channel in the RGBCIE color space uses these coordinates to represent the color's hue and saturation. The

second additional channel, the luminance channel, represents the brightness or intensity of a color. It is based on the CIE Y luminance function, which defines the perceived brightness of a color. The luminance channel in the RGBCIE color space uses this function to represent the color's brightness or intensity. By combining the RGB color space and the CIE color space, the RGBCIE color space provides a more comprehensive representation of color than the standard RGB model. It allows for more accurate color selection, manipulation, and reproduction in digital imaging and color management systems. The RGBCIE color space is used in a variety of applications, including digital photography, video, and color management software. It is also used in color calibration systems, where it is used to accurately match colors between different devices, such as monitors and printers.

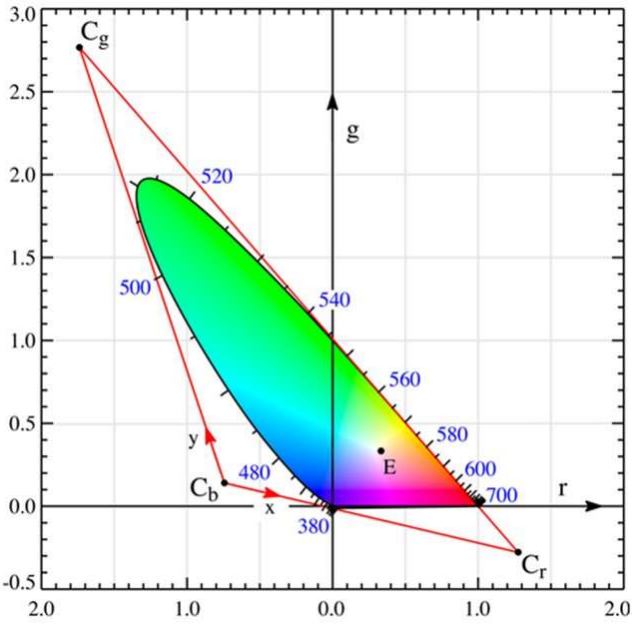


Figure 4.7: *RGBCIE Color Space*

4.7 XYZ

The XYZ color space, introduced by the International Commission on Illumination (CIE) in 1931, is a widely used and influential color model in the field of color science. It provides a mathematical representation of human perception of colors and serves as a foundation for various color-related applications.

In the XYZ color space, colors are defined by three components: X, Y, and Z. These components represent the tristimulus values of the color, which quantify the amount of three primary colors required to reproduce the color. The X, Y, and Z values are determined based on the spectral power distribution of the light source and the sensitivity of the human visual system to different wavelengths.

The Y component in the XYZ color space represents the brightness or luminance of the color. It quantifies the perceived intensity of the color and plays a crucial role in determining the overall visual appearance of an image or scene. The X and Z components represent the chromaticity or color information of the color. They capture the relative balance of the different primary colors and specify the color's hue and saturation.

One of the key advantages of the XYZ color space is its device-independence. It is designed to be independent of any specific color reproduction system or display device. This allows for consistent color representation and communication across different devices, ensuring that colors appear the same regardless of the medium or technology used.

The XYZ color space encompasses the entire range of human vision and includes both real and imaginary colors. The real colors represent actual visible colors that can be reproduced by light sources, while the imaginary colors are beyond the range of human perception and do not correspond to any physical emission spectra. Although imaginary colors are not perceptible, they are mathematically defined within the XYZ color space to ensure a comprehensive and consistent color model. RBG to XYZ is calculated as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \cdot \begin{bmatrix} R^{\phi} \\ G^{\phi} \\ B^{\phi} \end{bmatrix}$$

The XYZ color space has been widely adopted as a fundamental reference in color science and related fields. It serves as the basis for defining other color spaces, such as the CIE RGB color space and the CIE Lab color space. These color spaces are derived from the XYZ color space using specific transformations that optimize color representation for different purposes, such as additive color mixing or perceptual uniformity.

In practical applications, the XYZ color space is often utilized alongside other color spaces for various color-related tasks. Color space conversions are commonly performed to translate colors between XYZ and other color models, such as RGB, CMYK, or Lab. These conversions allow for seamless integration with existing color workflows and enable accurate color measurement, analysis, manipulation, and reproduction.

4.8 YCbCr

YCbCr (also written as YC´BCR) is a color space that is commonly used in digital video and image processing. It is similar to YUV, but with a slightly different color conversion formula. The YCbCr color space consists of three channels: Y, Cb, and Cr. The Y channel represents the luma or brightness component, and is calculated in the same way as in the YUV color space. The Cb and Cr channels represent the chroma or color information, and are calculated as the difference between the luma component and the blue and red components, respectively.

The conversion from RGB to YCbCr is typically performed using a matrix multiplication, with different matrices used for different video standards. For example, the ITU-R BT.601 standard uses a conversion matrix that is optimized for standard definition television, while the ITU-R BT.709 standard uses a conversion matrix that is optimized for high-definition television. The YCbCr color space is widely used in digital video compression and transmission, as it allows for efficient coding of color information. Similar to YUV, the chroma channels (Cb and Cr) can be subsampled to reduce the amount of data needed to represent the color information, while the luma channel (Y) is typically left at full resolution to preserve image quality.

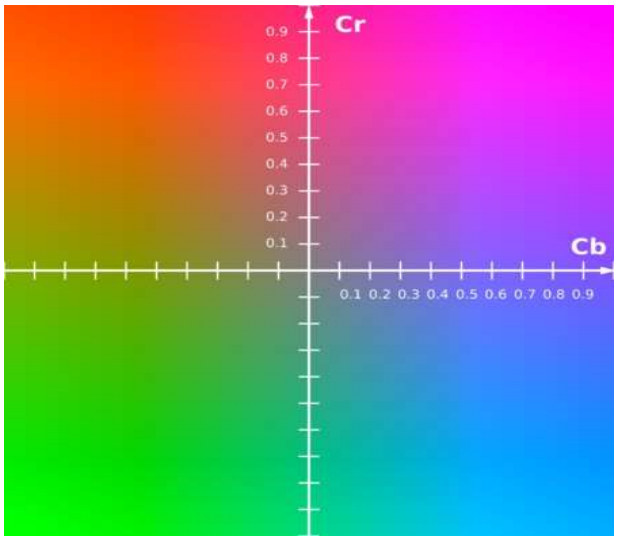


Figure 4.8: *YCbCr Color Space*

The YCbCr color space is also used in digital image processing, where it can be used for tasks such as color correction and manipulation. In this context, the Y channel can be used to adjust the brightness and contrast of an image, while the Cb and Cr channels can be used to adjust the color balance and saturation. RGB to YCbCr is calculated as:

$$
\begin{aligned}
\mathrm{Y} &= 0.299 * \mathrm{R}^{\phi} + 0.587 * \mathrm{G}^{\phi} + 0.114 * \mathrm{B}^{\phi} \\
\mathrm{Cb} &= 128 + \left(-0.169 * \mathrm{R}^{\phi} - 0.331 * \mathrm{G}^{\phi} + 0.500 * \mathrm{B}^{\phi}\right) \\
\mathrm{Cr} &= 128 + \left(0.500 * \mathrm{R}^{\phi} - 0.419 * \mathrm{G}^{\phi} - 0.081 * \mathrm{B}^{\phi}\right)
\end{aligned}
$$

4.9 YDbDr

YDbDr (also known as Y'Db'Dr) is a color space that is commonly used in analog video and digital video compression. It is a three-channel color space that separates the image into the luminance (Y) and two color-difference channels (Db and Dr).

The Y channel represents the luminance, or brightness, of the image, while the Db and Dr Channels represent the color difference between the blue and the luminance, and the red and the luminance, respectively.

The YDbDr color space is derived from the YUV color space, which separates the image into the luminance (Y) and two color-difference channels (U and V). The YUV color space was originally developed for analog television broadcasting and is still widely used in video compression standards such as MPEG and H.264.

The conversion from YUV to YDbDr involves a simple matrix transformation that preserves the luminance information and changes the color difference information to be more compatible with digital video compression. The conversion from RGB to YDbDr can also be performed using a matrix transformation. RGB to YDbDr is calculated as:

$$
\begin{bmatrix} Y \\ D_B \\ D_R \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.450 & -0.883 & 1.333 \\ -1.333 & 1.116 & 0.217 \end{bmatrix} \begin{bmatrix} R^{\phi} \\ G^{\phi} \\ B^{\phi} \end{bmatrix}
$$

4.10 YIQ

YIQ color space is a color representation system used in the analog NTSC television broadcasting system. It was developed by the National Television System Committee (NTSC) to enable the transmission of color TV signals over existing black-and-white TV channels. The YIQ color space represents colors as a combination of three components: Y, I, and Q. The Y component represent the luminance or brightness of the color, whereas the I and Q components represent the chrominance or color information. The I component represents the difference between the red and the luminance, and the Q component represents the difference between the blue and the luminance.

In the YIQ Y represents the black-and-white portion of a signal, I and Q represents the information of a color. It means that color TV signals can be transmitted over the same channel as black-and-white signals, which was important for the transition from black-and-white to color TV. Changing the range in the orange–blue(I) to the purple-green range(Q), in this condition human eye is more sensitive. Broadcast NTSC limits I to 1.3 MHz and Q to 0.4 MHz.

Because of the high costing in implementation some television can perform the true I and Q decoding. Another advantage of the YIQ color space is that it is perceptually uniform. This means that equal changes in the I and Q components result in equal perceptual changes in color. This makes it easier to manipulate and process color images and videos in a way that preserves their visual appearance. RGB to YIQ is calculated as:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} \approx \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.5959 & -0.2746 & -0.3213 \\ 0.2115 & -0.5227 & 0.3112 \end{bmatrix} \begin{bmatrix} R^{\phi} \\ G^{\phi} \\ B^{\phi} \end{bmatrix}$$

In the YIQ color space, the Y component represents the black-and-white portion of the signal, while the I and Q components represent the color information. The I and Q components are usually normalized to the range of -0.5 to +0.5, so they can be transmitted as analog signals over the same channel as the Y component.

One of the advantages of the YIQ color space is that it is perceptually uniform. This means that equal changes in the I and Q components result in equal perceptual changes in color. This makes it easier to manipulate and process color images and videos in a way that preserves their visual appearance.

4.11 YPbPr

YPbPr is a color space that is commonly used in analog and digital video applications. It is similar to YCbCr and YUV, but with a different color conversion formula.

The YPbPr color space consists of three channels: Y, Pb, and Pr. The Y channel represents the luma or brightness component, and is calculated in the same way as in the YCbCr and YUV color spaces. The PB and PR channels represent the chroma or color information, and are calculated as the difference between the luma component and the blue and red components, respectively.

YPbPr is typically used in analog video applications, where the color information is carried on separate cables from the brightness information. The Y channel is carried on one cable, while the Pb and Pr channels are carried on two separate cables. In digital video applications, YPbPr is often used as an intermediate color space for processing video signals before they are converted to another color space, such as RGB or YCbCr.

The conversion from RGB to YPbPr is typically performed using a matrix multiplication, with different matrices used for different video standards. For example, the SMPTE 170M standard uses a conversion matrix that is optimized for standard definition television, while the SMPTE 240M standard uses a conversion matrix that is optimized for high-definition television. RGB to YPbPr is calculated as:

$$
\begin{aligned}
\mathrm{Y} &= 0.213 * \mathrm{R}^{\phi} + 0.715 * \mathrm{G}^{\phi} + 0.072 * \ \mathrm{B}^{\phi} \\
\mathrm{Pb} &= -0.115 * \mathrm{R}^{\phi} - 0.385 * \mathrm{G}^{\phi} + 0.500 * \ \mathrm{B}^{\phi} \\
\mathrm{Pr} &= 0.500 * \mathrm{R}^{\phi} - 0.454 * \mathrm{G}^{\phi} - 0.046 * \ \mathrm{B}^{\phi}
\end{aligned}
$$

The YPbPr color space is also used in digital image processing, where it can be used for tasks such as color correction and manipulation. In this context, the Y channel can be used to adjust the brightness and contrast of an image, while the Pb and Pr channels can be used to adjust the color balance and saturation.

4.12 YUV

A color image pipeline frequently uses the YUV color model. As opposed to a ”direct” RGB representation, it allows for a smaller bandwidth for the chrominance components when encoding a color image or video. The designations YUV and YUV have historically been used to refer to a particular analogue encoding of color information in broadcast systems. Currently, color spaces that are encoded using YCbCr are referred to as YUV in the computer industry.

One luminance component (Y), which represents actual linear-space brightness, and two chrominance components, designated as U (blue projection) and V (red projection), respectively, are defined by the YUV model. Different color spaces and the RGB model can both be converted using this method.

YUV is a color space that is widely used in video and digital image processing. It is based on the concept of separating color information (chrominance) from brightness information (luminance) in an image or video signal.

The YUV color space consists of three channels: Y, U, and V. The Y channel represents the luminance or brightness of the image, and is often referred to as the luma component. The U and V channels represent the chrominance or color information, and are often referred to as the chroma components.

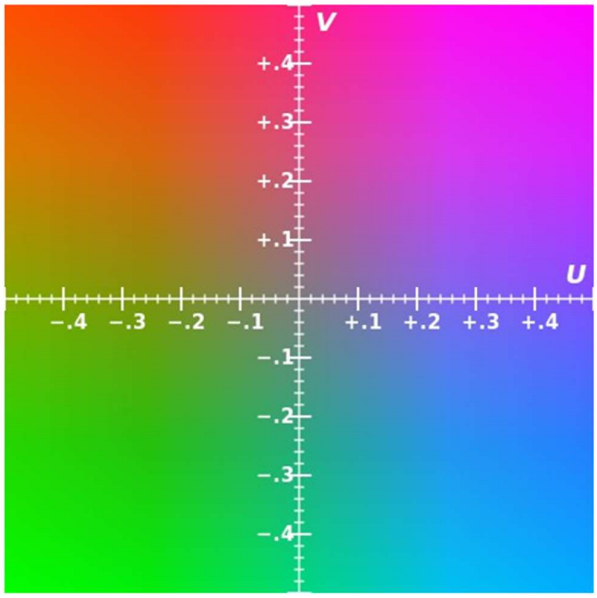


Figure 4.9: *YUV Color Space*

RBG to YUV is calculated as:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & -0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R^{\phi} \\ G^{\phi} \\ B^{\phi} \end{bmatrix}$$

The Y channel is calculated as a weighted sum of the red, green, and blue values of a pixel, with higher weights given to the green component, as the human eye is more sensitive to green light. The U and V channels represent the difference between the blue and luma channel, and the red and luma channel, respectively. This allows for a compact representation of color information, as the U and V channels can be represented with less data than the full RGB color space. The YUV color space is used in video compression and transmission, as it allows for more efficient coding of color information. By separating the luminance and chrominance components, it is possible to apply different levels of compression to each component, with less compression applied to the luminance channel to preserve the image quality.

The YUV color space is also used in digital image processing, where it can be used for tasks such as color correction and manipulation. In this context, the Y channel can be used to adjust the brightness and contrast of an image, while the U and V channels can be used to adjust the color balance and saturation.

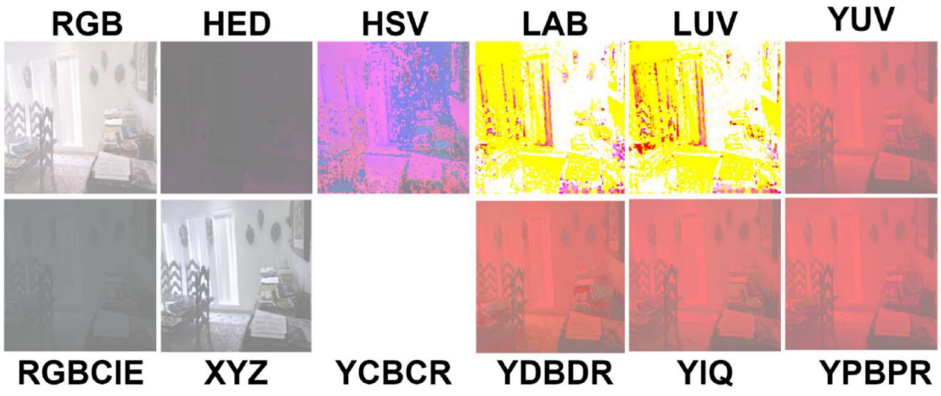


Figure 4.10: *Comparison of Input Hazy Images in Different Color Spaces*

Chapter 5

Methodology

This section explains the analytic method used in this work. The analysis has been divided into five steps: data preparation, image conversion into various color spaces such as RGB, HED, HSV, LAB, RGB-CIE, XYZ, YCbCr, YDbDr, YPbPr, YIQ, and YUV; selection of encoder models from six options, MobileOne[42], ResNet[43], MIT[44], EfficientNet[45], ResNeXt[46] and GerNet[47]; connecting the chosen encoder to the UNet decoder; training of the models on different color spaces; and finally, performance analysis of the respective models.

5.1 Dataset

The RESIDE6K dataset is a recent benchmark created for image dehazing tasks, consisting of 6,000 pairs of hazy images and their clear counterparts. This dataset offers a wider range of haze conditions than previous benchmarks, making it valuable for evaluating image dehazing algorithms. Its key feature is its scale, providing a large number of image pairs for training and evaluation. Notably, the hazy images in RESIDE6K are created using a physically-based model, adding a level of realism and difficulty not found in other datasets. Each hazy image is paired with a clear, haze-free image, allowing for precise evaluation of dehazing algorithms.

5.2 Color Space Conversion

The RESIDE6K[48] dataset comprises indoor and outdoor hazy images with respective ground truth. Before model training, the input hazy images are transformed into twelve different color spaces, including RGB, HED, HSV, LAB, RGB-CIE, XYZ, YCbCr, YDbDr, YPbPr, YIQ, YUV and LUV. As mentioned earlier, this transformation has been accomplished by using the matching color space transformation matrix.

5.3 Base Models

The dehazing model consists of two parts, a generator and a discriminator. The generator is also divided into two subparts, encoder and decoder as we have adopted pix-to-pix GAN [49] for haze removal. For the encoder part, we have used six different architectures such as MoblieOne[42], ResNet[43], EfficientNet[45], Mix Vision Transformer (MIT)[50], ResNeXt[46] and GerNet[47] and combined it with the UNet[51] decoder block, which is described in the following subsection. For the decoder, we have used a patch discriminator as mentioned in the original pix-to-pix architecture [49].

5.4 Encoder Models

5.4.1 MobileOne

MobileOne represents a cutting-edge neural network architecture tailored for image classification tasks within the realms of deep learning and computer vision. This recently developed model is meticulously optimized to operate seamlessly on mobile devices, achieving remarkable efficiency with a runtime of under 1 millisecond. This swift-ness distinguishes MobileOne from many existing models, ensuring superior speed without compromising state-of-the-art accuracy.

Breaking down the key features of MobileOne, it boasts high accuracy by achieving top-1 accuracy on ImageNet, a widely recognized benchmark dataset for image classification. This places it on par with other efficient models. Furthermore, its low latency is a standout feature, enabling real-time applications on mobile devices. The model's efficiency is attributed to its innovative architecture, incorporating techniques such as squeeze-and-excitation.

These methods strike a balance between achieving top-notch accuracy and minimizing computational costs. In this paper, we have used the MobileOne-S0 model as the encoder part of the generator for dehazing. The generator architecture using MobileOne is shown in Figure 5.1.

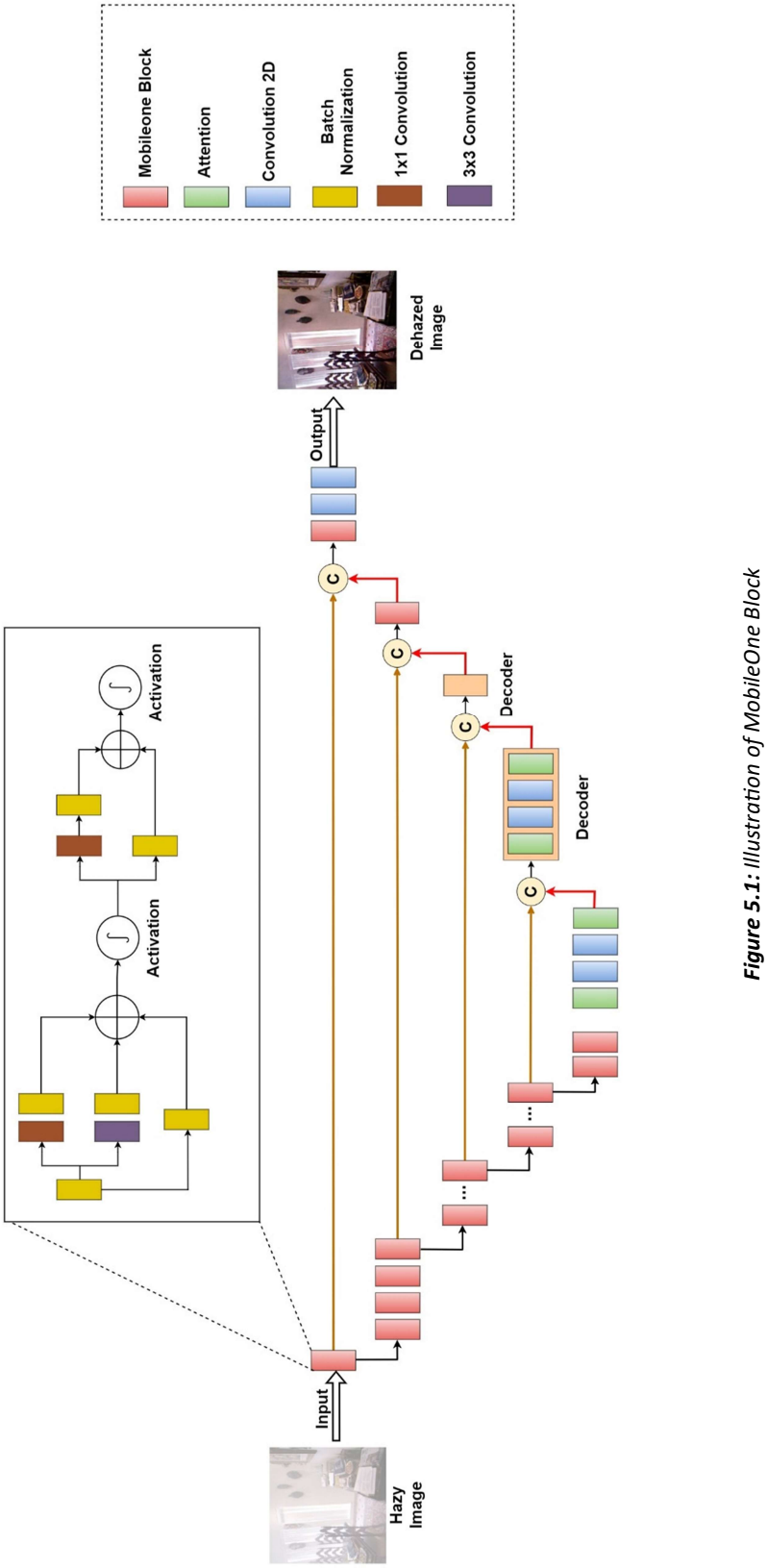


Figure 5.1: *Illustration of MobileOne Block*

5.4.2 ResNet

ResNet, short for Residual Network with different numbers of hidden layers, is designed to overcome challenges in training very deep neural networks by introducing the concept of residual learning. The ResNet architecture comprises several key components, the input layer accepts images of a specified size, typically 224x224 pixels. Initial features are extracted through a 7x7 convolutional layer with 64 filters, followed by 3x3 max pooling with a stride of 2 to down sample the feature maps. The network consists of four stages, each containing multiple residual blocks. These blocks include basic blocks, utilizing two 3x3 convolutional layers with batch normalization and a shortcut connection to address the vanishing gradient problem. Additionally, bottleneck blocks introduce 1x1 convolutions before and after 3x3 convolutions to enhance performance and reduce the computational cost, and lastly, global average pooling replaces fully connected layers, reducing parameters and improving generalization. In this paper, we have used ResNet-18 as the encoder of the generator, which consists of 18 hidden layers and stands out as a versatile and efficient convolutional neural network architecture, leveraging its moderate depth, residual connections, and bottleneck blocks to achieve notable success in deep learning and computer vision. The generator using ResNet-18 is shown in Figure 5.2.

5.4.3 EfficientNet

EfficientNet, is a convolutional neural network (CNN) architecture crafted for image classification tasks in deep learning and computer vision. It is renowned for achieving heightened accuracy while maintaining a significantly smaller and faster profile compared to preceding CNNs, EfficientNet stands out in the following aspects: firstly, it adeptly navigates the traditional trade-off between accuracy and efficiency observed in many CNNs. Unlike the typical scenario where increasing layers and parameters enhances accuracy but inflates computational costs, EfficientNet utilizes a unique scaling method. Employing compound scaling, it adjusts various architectural aspects such as depth, width, and resolution in a specific manner, ensuring the preservation of accuracy while reducing overall model size. Secondly, the architecture incorporates innovative features, including Squeeze-and-Excitation (SE) blocks dynamically adjusting channel weights based on importance, Inverted Residual blocks combining depth-wise and pointwise convolutions for cost reduction, and Mobile Inverted Bottleneck blocks optimizing efficiency and accuracy. Its efficiency makes it suitable for resource-constrained environments like mobile devices and embedded systems. In this paper, we have used the EfficientNet-B0 model as the encoder of the generator which is shown in Figure 5.3.

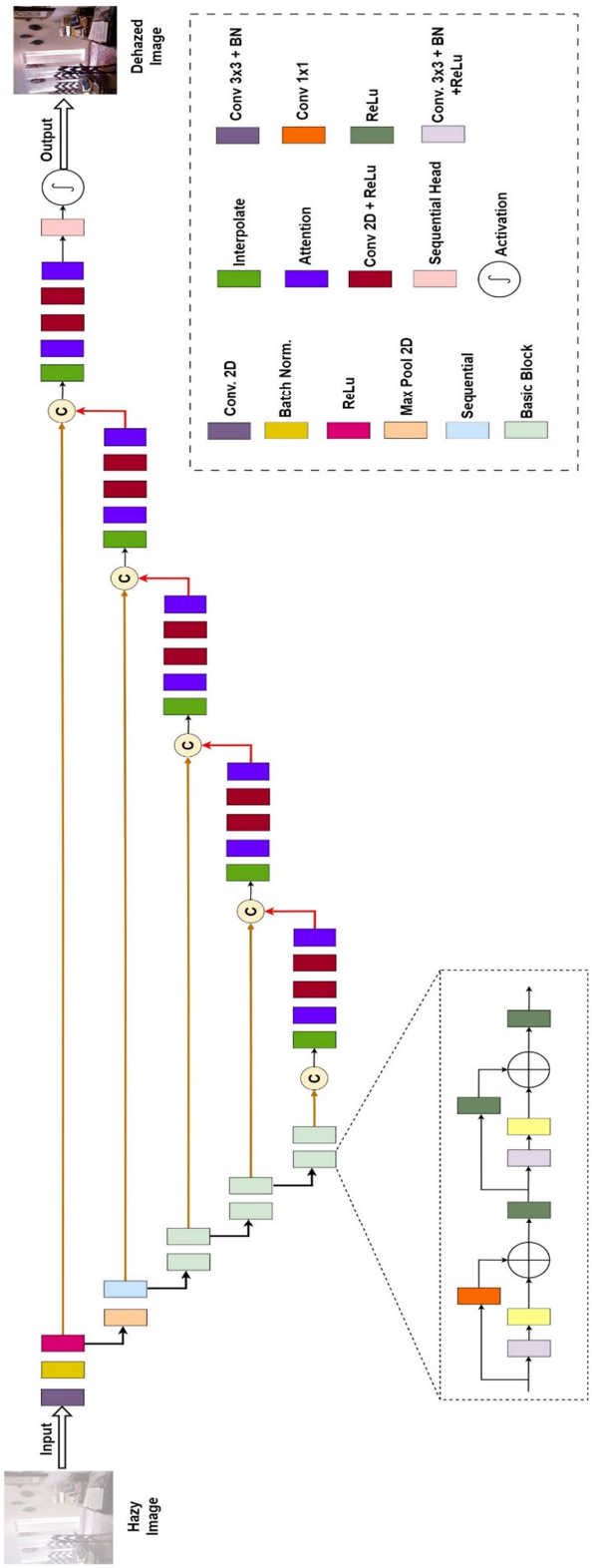


Figure 5.2: *Illustration of ResNet Block*

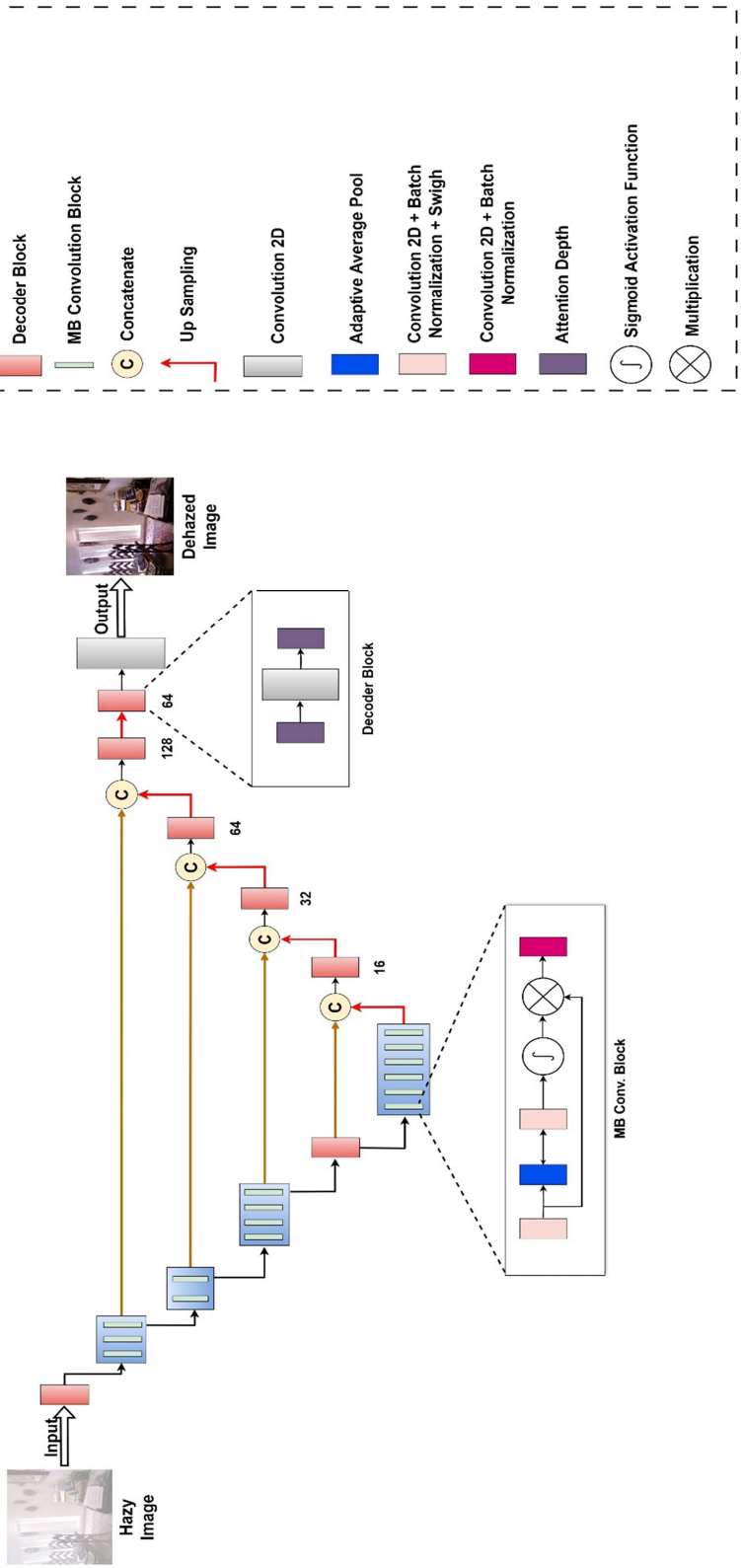


Figure 5.3: *Illustration of EfficientNet Block*

5.4.4 Mix Vision Transformer

Vision transformers have gained a prominent position in the field of computer vision with their unique approach. Unlike traditional methods that rely on convolutional neural networks, ViTs utilize self-attention, allowing them to understand relationships between distant parts of an image. This makes them particularly adept at tasks like object detection and image generation, often surpassing CNNs in performance.

ViTs work by splitting an image into patches, analyzing them individually, and then using self-attention to understand how they connect. This global awareness gives them an edge over CNNs, which process images in a localized manner. In this paper, we have used MIT-B0 as the encoder of the generator, which is shown in Figure 5.4.

5.4.5 ResNext

ResNeXt is a convolutional neural network architecture also known as Residual Next, designed to enhance learning capacity through the use of parallel pathways or "cardinality" within each residual block. The cardinality, representing the number of paths or groups in a block, is a key feature explored in ResNeXt to evaluate its impact on network performance.

ResNeXt is based on the ResNet architecture but incorporates cardinality to boost representational power. The cardinality in ResNeXt signifies the number of parallel pathways or groups within a residual block, whereas a higher cardinality, for instance, 32, implies 32 different paths or groups of convolutional layers within each block.

The network employs a bottleneck design in its residual blocks, similar to ResNet, utilizing 1x1, 3x3, and 1x1 convolutional layers. The 1x1 convolutions serve to reduce and then increase dimensionality, enhancing computational efficiency. ResNeXt incorporates skip connections, or residual connections, addressing the vanishing gradient problem and aiding in training very deep networks by allowing the model to learn residual mappings. In this paper, we have incorporated the ResNeXt-50 as the encoder of the generator for dehazing, which is shown in Figure 5.5.

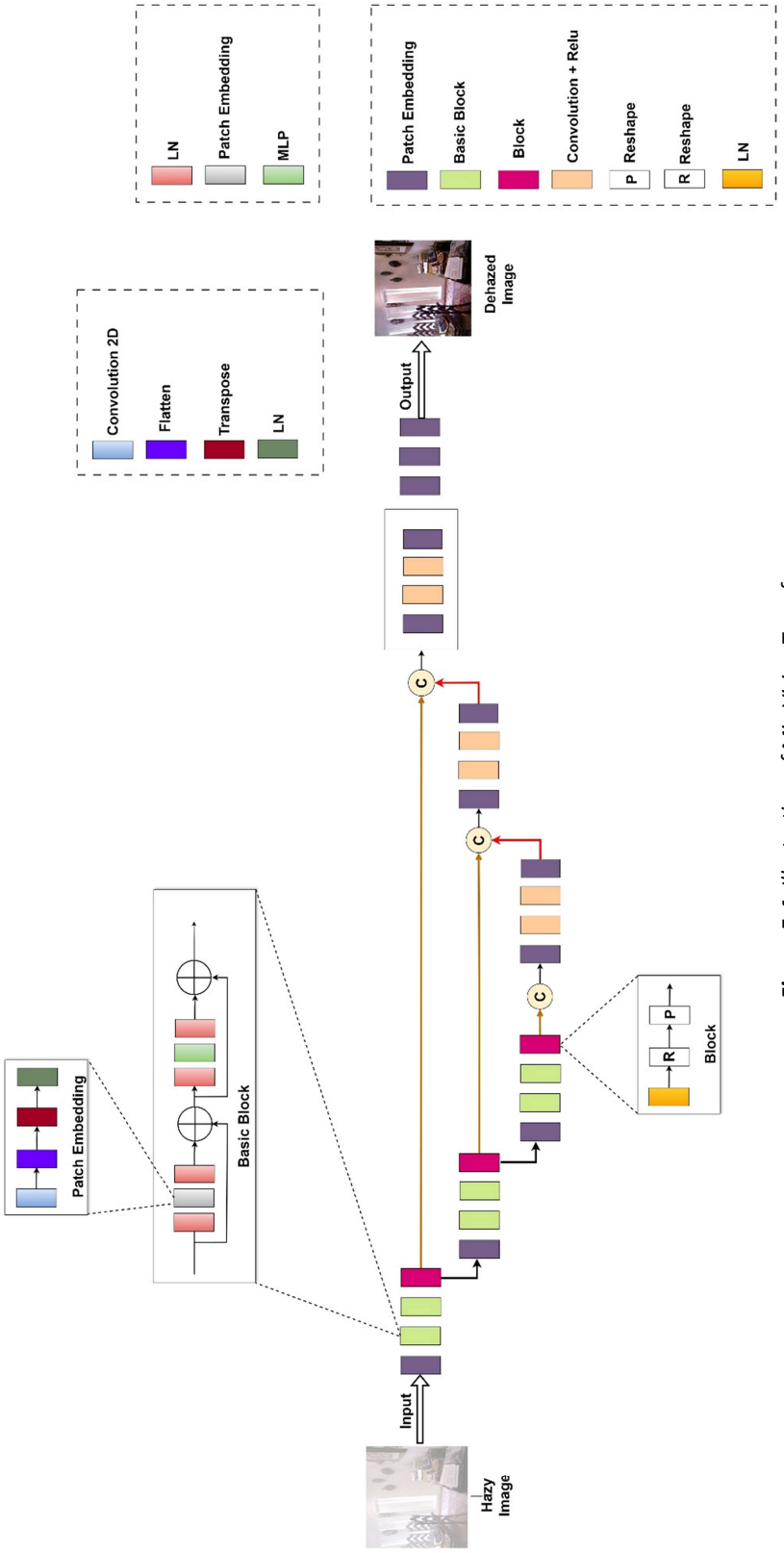


Figure 5.4: *Illustration of Mix Vision Transformer*

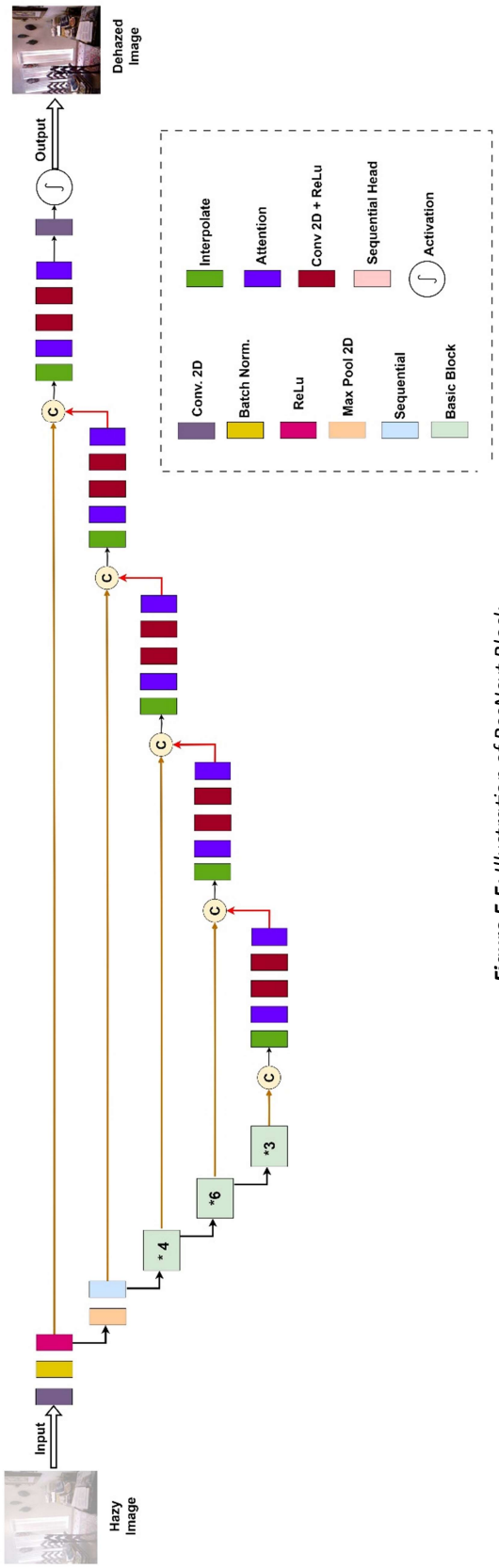


Figure 5.5: *Illustration of ResNext Block*

5.4.6 GerNet

GERNet, an acronym for Global Embedding Representation Network, is a specialized deep learning architecture tailored for computer vision tasks such as image classification and retrieval. The distinctive features of GERNet include global context encoding, where it prioritizes capturing the overall context of an image through a global pooling layer that consolidates information from all convolutional layer-generated feature maps. This global approach enhances the encoding of the image's semantic meaning, resulting in more resilient representations. Additionally, GERNet incorporates a channel attention mechanism to assign weights to each channel based on its significance in capturing semantic information, enabling the network to focus on informative channels and suppress irrelevant ones, thus improving the discriminative nature of the global representation. The architecture further employs a discriminative loss function to ensure that the learned global representations are well-suited for image retrieval tasks, encouraging the network to embed semantically similar images closer in the embedding space while pushing dissimilar images further apart. GERNet's benefits encompass improved accuracy in image classification and retrieval tasks compared to traditional CNN architectures, scalability due to its efficient handling of large-scale image datasets, and robustness achieved through its emphasis on the global context, making it less susceptible to noise and local variations in images. In this paper, we have used GERNet as an encoder part of the generator for haze removal, which is shown in Figure 5.6.

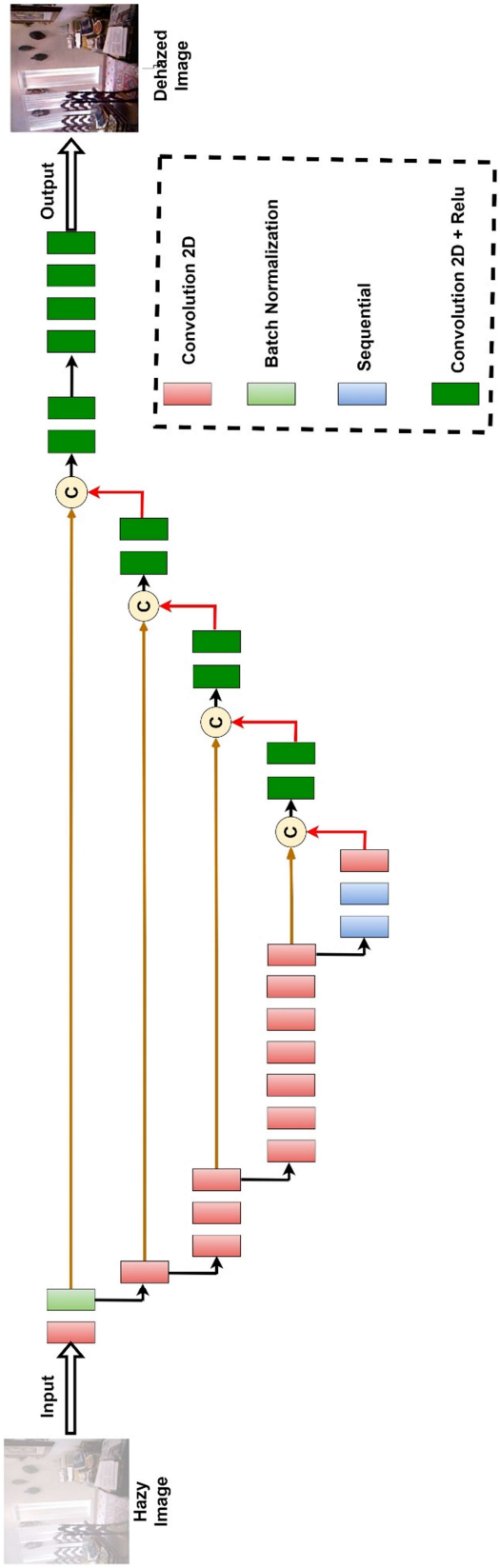


Figure 5.5: *Illustration of GerNet Block*

5.5 U-Net Decoder

The U-Net architecture is a specialized convolutional neural network (CNN) designed for semantic segmentation within the realm of computer vision, with a particular emphasis on its widespread use in medical image segmentation tasks. The U-Net derives its name from its distinctive U-shaped structure. This architecture encompasses a contracting path, responsible for context capture and spatial resolution reduction, followed by an expansive path, which aids in spatial information recovery and segmentation mask generation. The symmetry of the architecture is integral to its efficacy.

The U-Net comprises key components, starting with the contracting path called the encoder. In this phase, the input image undergoes convolutional layers with max pooling operations, reducing spatial dimensions. Each convolutional layer is typically accompanied by a rectified linear unit (ReLU) activation function for introducing non-linearity. The bottleneck layer acts as a crucial bridge connecting the contracting and expansive paths, capturing essential features. The expansive path involves up-sampling layers, often utilizing transposed convolutions or up-sampling followed by regular convolutions called the decoder. Each up-sampling step is concatenated with corresponding feature maps from the contracting path, enhancing detailed spatial information. To counter information loss during down-sampling, skip connections directly link layers from the contracting path to their counterparts in the expansive path, preserving fine-grained details. The architecture culminates in the output layer, utilizing a convolutional layer with a softmax activation function to yield the segmentation mask.

The U-Net has demonstrated effectiveness in not only segmentation tasks but also in image-to-image generation tasks, delivering precise and detailed results. Its capability to capture both local and global context, coupled with the strategic implementation of skip connections, renders it particularly adept for tasks necessitating accurate object localization, such as in the domain of medical imaging, image restoration, and image-to-image generation.

5.6 Discriminator

The discriminator has a very significant role to play in differentiating between real (clear) and fake (generated dehazed) images. Functioning as an adversarial network, it provides feedback to the generator, thereby improving the overall quality and fidelity of the output. The discriminator architecture consists of four convolutional layers, followed by fully connected layers. The patch discriminator, as depicted in Figure 5.7, receives the generator network's output and the corresponding original image as inputs, predicting the input image's probability of authenticity. During training, the discriminator is optimized to minimize the binary cross-entropy with logits (BCEWithLogits) loss. This loss function aids the discriminator in accurately categorizing real and synthetic images. Through an adversarial framework involving the generator and discriminator, the proposed method effectively learns to produce high-quality dehazed images that closely resemble realistic clear images.

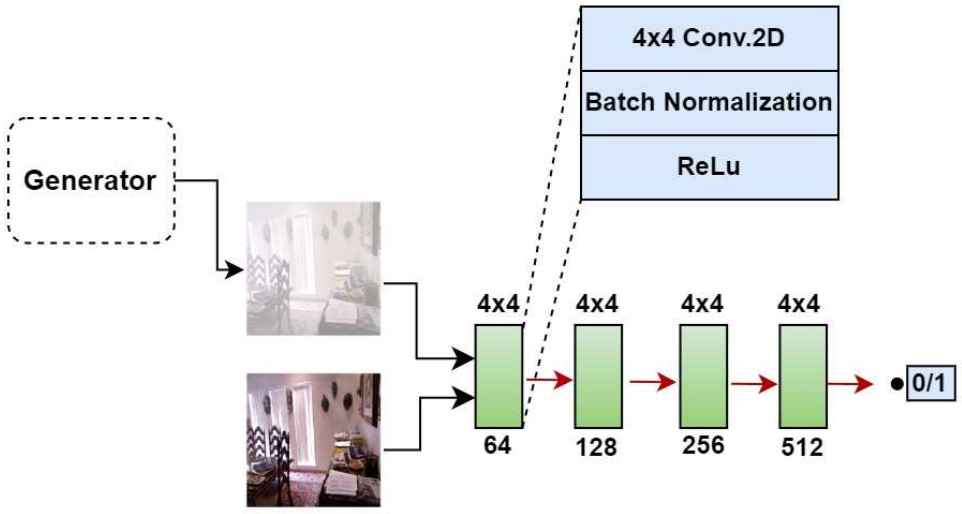


Figure 5.7: *Illustration of Discriminator* Architecture

5.7 Loss Function

The binary cross-entropy loss is used to train the generator, which measures how different the predicted probability, generated by the discriminator network, is from the actual ground truth label. It is defined as:

$$L_G = -\frac{1}{N}\sum^{N}[\log(D(I_i)) + \log(1 - D(G(J_i)))]$$

where D is the discriminator network, G is the generator network, J_i is the ground truth image, I_i is the input hazy image, and N is the training size.

For training the discriminator we use, discriminator loss which is given as:

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

5.8 Training Details

The models were trained within the Kaggle Notebook environment, leveraging the computational power of a 16GB Nvidia P100 GPU and 16GB of RAM. (The Nvidia P100 GPU is a high-performance computing accelerator designed for deep learning and other computationally intensive tasks. It features 16GB of high-bandwidth memory (HBM2) and is based on the Pascal architecture, offering significant performance improvements over previous generations.)

The development of the models was carried out using the PyTorch Python library. (PyTorch is an open-source machine learning library for Python, developed by Facebook's AI Research lab (FAIR). It provides a flexible and dynamic computational graph system, making it suitable for research and production deployment in various domains, especially in deep learning.)

During the training process, the ADAM optimizer was employed, coupled combination of GAN loss and SSIM loss as the chosen loss function. (The Adaptive Moment Estimation or ADAM in short is a popular optimization algorithm for training deep learning models. It combines the advantages of both AdaGrad, which adapts the learning rate for each parameter, and RMSProp, which uses a moving average of squared gradients to normalize the learning rate. ADAM is known for its efficiency and effectiveness in training a wide range of neural network architectures.)

The learning-rate used throughout the training was set to 0.001. (Learning- Rate is a hyperparameter that controls the step size during the optimization process of training a neural network. It determines how much the model weights are updated in response to the loss gradient.)

The training process spanned 100 epochs. (Epoch is a single pass through the entire training dataset during the training of a machine learning model. Training for multiple epochs allows the model to see the entire dataset multiple times, which is often necessary for the model to converge to a good solution.)

Chapter 6

Results and Analysis

This thesis uses RESIDE6K dataset to analyze the model performance in twelve color spaces: RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, YUV of six different encoder models such as MobileOne, ResNet, EfficientNet, MIT, ResNeXt, GerNet. We have compared both indoor and outdoor images qualitatively and quantitatively. For quantitative comparison, Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) metric are used, which are described as follows:

- **Structural Similarity Index Measure (SSIM):** Structural Similarity Index Measure is mainly used for finding the similarity between two images. It can reflect more complex features like textures and boundaries. The SSIM is written as:

$$SSIM = \frac{(2\mu_I\mu_O + C_1)}{(\mu_I^2 + \mu_O^2 + C_1)} \cdot \frac{(2\sigma_{IO} + C_2)}{(\sigma_I^2 + \sigma_O^2 + C_2)} = l(x) \cdot cs(x)$$

  where μ_I *is* the average of I; μ_O *is* the average of O; σ_I^2 is the variance of I; σ_O^2 is the variance of O; σ_{IO} *is* the covariance of I and O; C_1, C_2 are the constant for stabilizing the division.

- **Peak Signal-to-Noise Ratio (PSNR):** To introduce sharpness to the generated image PSNR loss has been used. It also reduces noise in reconstructed images and improves the signal-to-noise ratio. PSNR is given as,

$$PSNR = 10\log_{10}\left(\frac{I_{\max}^2}{MSE}\right)$$

  In our case I_{max} is 1, as we re-scaled the ground truth in the range [0.0, 1.0].

Table 1 provides us with the SSIM and PSNR values of different models in all twelve color spaces for both indoor and outdoor conditions. RESIDE6K provides the average value of the SSIM and PSNR for each respective color space in the case of a particular model.

In Table 2, the performance of the best color space for a particular model along with the state-of-the-art models are described along with its runtime.

Encoder	Type	Performance	RGB	HED	HSV	LAB	LUV	RGBCIE	XYZ	YCBCR	YDBDR	YIQ	YPBPR	YUV
MobileOne	door	SSIM	0.830	0	0.363	0	0	0.237	0.675	0	0.518	0.214	0.346	0.289
		PSNR	21.866	5.076	9.292	0	0	1.770	12.937	0	5.499	1.335	8.267	0
	Outdoor	SSIM	0.929	0	0.533	0	0	0.325	0.727	0	0.598	0.221	0.350	0.372
		PSNR	28.172	4.018	9.774	0	0	1.777	13.771	0	6.389	1.339	8.272	0
	RESIDE6K	SSIM	0.880	0	0.448	0	0	0.281	0.701	0	0.558	0.218	0.348	0.331
		PSNR	25.019	4.546	9.292	0	0	1.774	13.354	0	5.944	1.337	8.270	0
ResNet	Indoor	SSIM	0.604	9.774	0.688	0.029	0.020	0.826	0.875	0.021	0.835	0.810	0.832	0.8 .85
		PSNR	17.590	22.809	20.114	0	1.041	22.772	23.489	0	23.097	19.492	22.857	22.516
	Outdoor	SSIM	0.820	0.901	0.861	0.042	0.027	0.914	0.937	0.030	0.907	0.914	0.888	0.919
		PSNR	22.611	26.045	25.466	1.391	0.528	26.376	27.531	0	26.375	26.471	25.770	26.961
	RESIDE6K	SSIM	0.712	0.863	0.775	0.036	0.024	0.870	0.906	0.026	0.871	0.862	0.860	0.889
		PSNR	20.101	24.427	22.790	0.696	0.785	24.574	25.510	0	24.736	22.982	24.314	24.739
EfficientNet	Indoor	SSIM	0.870	0.786	0.625	0.055	0.034	0.847	0.843	0.004	0.812	0.866	0.863	0.851
		PSNR	23.635	22.450	19.771	0.160	0	23.860	23.061	3.687	22.706	23.488	24.132	23.891
	Outdoor	SSIM	0.938	0.892	0.817	0.086	0.064	0.927	0.927	0.010	0.901	0.917	0.924	0.920
		SNR	28.363	25.763	24.333	0.382	1.183	28.431	27.937	0	26.116	27.211	27.561	27.891
	RESIDE6K	SSIM	0.904	0.839	0.721	0.071	0.049	0.887	0.885	0.007	0.857	0.892	0.894	0.886
		PSNR	25.999	24.1065	19.771	0.271	0.5915	26.1455	25.499	1.8435	24.411	25.3495	25.8465	25.891
MIT	Indoor	SSIM	0.801	0.760	0.621	0.018	0.013	0.734	0.771	0.006	0.795	0.779	0.770	0.791
		PSNR	22.486	20.236	20.341	0	0	21.778	16.016	0	23.489	19.228	22.678	22.639
	Outdoor	SSIM	0.872	0.859	0.778	0.040	0.035	0.841	0.909	0.015	0.879	0.879	0.867	0.873
		PSNR	26.168	25.412	24.464	1.233	0	25.284	26.824	0	26.846	27.099	26.495	26.631
	RESIDE6K	SSIM	0.837	0.810	0.700	0.029	0.024	0.788	0.840	0.011	0.837	0.829	0.819	0.832
		PSNR	24.327	22.824	22.403	0.617	0	23.531	21.420	0	25.168	23.164	24.587	24.635
ResNext	Indoor	SSIM	0.865	0.770	0.693	0.032	0.040	0.855	0.754	0.011	0.833	0.873	0.831	0.836
		PSNR	19.642	15.824	21.926	1.591	0	23.598	16.497	0	23.904	23.107	23.556	22.396
	Outdoor	SSIM	0.941	0.893	0.873	0.051	0.053	0.918	0.902	0.025	0.913	0.917	0.898	0.906
		PSNR	28.070	24.737	26.065	1.700	1.135	27.019	25.336	0	27.034	26.189	25.738	25.385
	RESIDE6K	SSIM	0.903	0.832	0.783	0.042	0.047	0.887	0.828	0.018	0.873	0.895	0.865	0.871
		PSNR	23.856	20.281	23.996	1.646	0.568	25.309	20.917	0	25.469	24.648	24.647	23.891
GerNet	Indoor	SSIM	0.803	0.764	0.590	0.021	0.021	0.786	0.824		0.811	0.839	0.817	0.814
		PSNR	22.470	19.139	19.8032,	0.348	0	21.365	22.974	0	22.599	22.062	22.618	22.475
	Outdoor	SSIM	0.897	0.882	0.795	0.055	0.022	0.886	0.927	0.014	0.897	0.904	0.880	0.896
		PSNR	26.425	25.237	23.342	2.398	0.733	25.990	27.465	0	26.682	26.547	25.912	25.706
		SSIM	0.850	0.823	0.693	0.038	0.022	0.836	0.876	0.007	0.854	0.872	0.849	0.855
	RESIDE6K	PSNR	24.448	22.188	23.342	1.373	0.367	23.678	25.220	0.000	24.641	24.305	24.265	24.091

Table 1: *PSNR and SSIM Comparison of the Proposed Dehazing Techniques in Different Color Spaces Using Various Encoders on RESIDE6K Indoor, RESIDE6K Outdoor and RESIDE6K Datasets*

Encoder	Best	Indoor		Outdoor		RESIDE6K		Runtime
		SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	
Mobileone S0	RGB	0.83	21.866	0.929	28.172	0.88	25.019	0.141
Resnet 18	XYZ	0.875	23.489	0.937	27.531	0.906	25.51	0.083
EfficientNet B0	RGBCIE	0.863	24.132	0.927	28.431	0.887	26.146	0.121
Mit B0	YDBDR	0.795	23.489	0.879	27.099	0.837	25.168	0.088
ResNext50_32x4d	YDBDR	0.833	23.904	0.941	28.07	0.873	25.469	0.092
GerNet (timm-gernet_s)	XYZ	0.817	22.618	0.927	27.465	0.876	25.22	0.088
DCP	-	0.818	16.62	0.815	19.13	0.817	17.875	3.002
DehazeNet	-	0.847	21.14	0.851	22.46	0.849	21.8	0.831
AODNet	-	0.85	19.06	0.877	20.29	0.864	19.675	0.106
DCPDN	-	0.818	15.85	0.845	19.93	0.832	17.89	6.053
GFN	-	0.88	22.3	0.844	21.55	0.862	21.925	-

Table 2: *The Results of Comparing Proposed Dehazing Techniques on RESIDE6K with Various State-of-the-Art*

6.1 Performance Evaluation of Various Deep Learning Models for Image Dehazing Across Different Color Spaces

This section presents a rigorous performance evaluation of color spaces like RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV over various models. The assessment is conducted using SSIM and PSNR values for both indoor and outdoor conditions on the RESIDE6K dataset.

6.1.1 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Mobileone Encoder Model

Upon examining the efficacy of various color spaces including RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV on the Mobileone model based on SSIM and PSNR values, it is found that RGB demonstrates superior performance under both indoor-outdoor environments as shown in Table 1. Conversely, YCBCR exhibits notably poor performance across both scenarios shown in Table 1. Upon computing the average SSIM and PSNR values for both indoor and outdoor conditions in both Tables 2 and 1, RGB outperforms better than all other color spaces, securing the top position for the Mobileone model.

On doing visual analysis by observing the output dehazed images in different color spaces for both indoor and outdoor conditions as shown in Figure 6.1, it is observed that only RGB has shown promising results while the other color spaces have performed poorly.

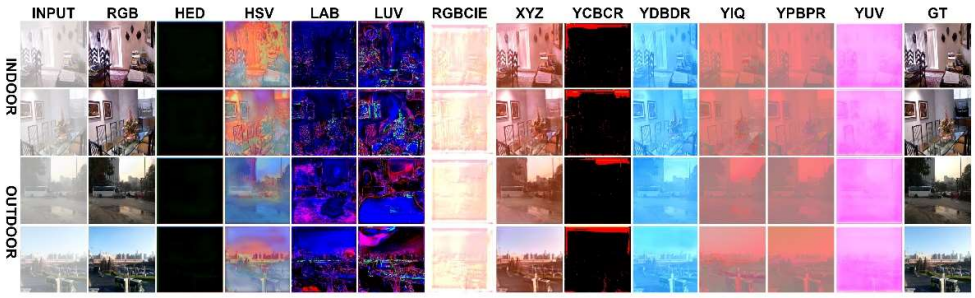


Figure 6.1: *Comparative Performance of the MobileOne Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces.*

6.1.2 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with ResNet Encoder Model

On evaluating the performance of the RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV color spaces on the Resnet18 model based on SSIM and PSNR values, it is found that XYZ has performed most effectively under both indoor and outdoor conditions as shown Table 1. Thus, upon computing the average SSIM and PSNR values for both indoor and outdoor conditions as shown in both Tables 1 and 2, XYZ has stood out to be the best. Conversely, YCbCr has yielded the least satisfactory results for both scenarios.

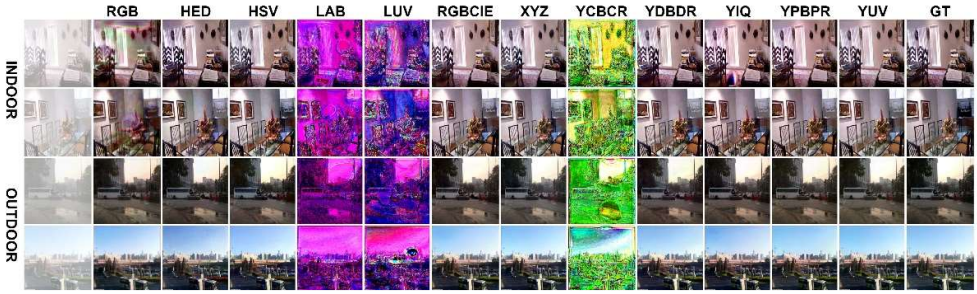


Figure 6.2: *Comparative Performance of the ResNet Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces.*

On the other hand, in visual analysis by observing the output dehazed images in different color spaces for both indoor and outdoor conditions as depicted in Figure 6.2, it has been observed that HED, HSV, RGBCIE, XYZ, YDbDr, YIQ, YPbPr, YUV have shown promising results. In the case of RGB in indoor conditions, we have observed color distortion while LAB, LUV, and YCbCr have shown a lot of artifacts in the output.

6.1.3 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Efficientnet Encoder Model

After assessing the effectiveness of various color spaces, including RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV with the Efficientnet encoder model based on SSIM and PSNR values, it has been seen that YPbPr shows superior performance for indoor, while RGBCIE performs exceptionally well for outdoor as shown in Table 1. Conversely, LUV under indoor conditions and YCBCR under outdoor conditions have demonstrated notably poor performance shown.

On averaging SSIM and PSNR values for both indoor and outdoor conditions as shown in both Tables 2 and 1, RGBCIE has outperformed YPbPr, securing the top position not only for the Efficientnet model but also for all other models.

On doing qualitative analysis, by observing the output dehazed images in different color spaces for both indoor and outdoor conditions as shown in Figure 6.3, it has been observed that RGB, HED, HSV, RGBCIE, XYZ, YDbDr, YIQ, YPbPr, YUV have shown satisfactory results. While LAB, LUV, and YCbCr have produced results with artifacts.

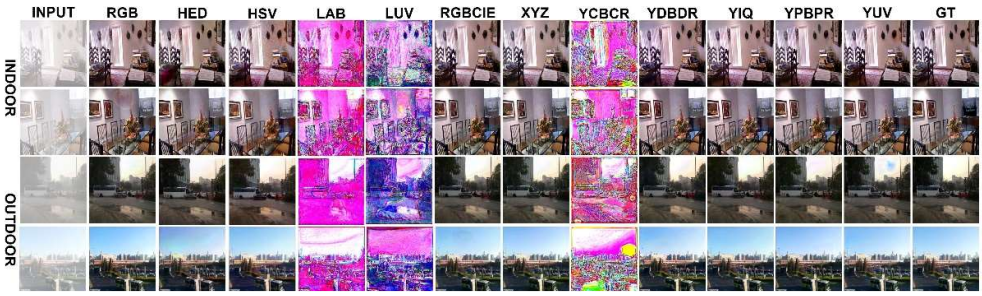


***Figure 6.3**: Comparative Performance of the Efficientnet Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces.*

6.1.4 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Mix Vision Transformer (MIT) Encoder Model

Upon evaluating various color spaces, including RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV on the Mit S0 model based on SSIM and PSNR values, it has been found that YDbDr shows superior performance indoors, while YIQ performs exceptionally well in outdoor settings as shown in Table 1. In contrast, YCbCr has performed poorly in both scenarios.

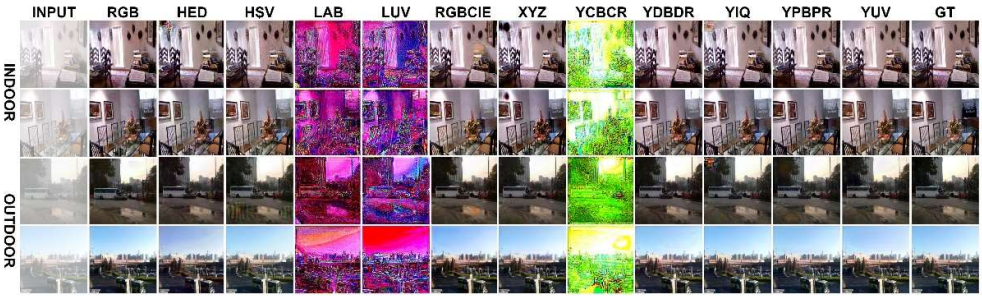


***Figure 6.4:** Comparative Performance of the Mix Vision Transformer (MIT) Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces.*

On averaging the SSIM and PSNR values for indoor and outdoor conditions as shown in Tables 2 and 1, YDbDr has outperformed YIQ, securing the top position for the Mit model.

On doing visual analysis by observing the output dehazed images in different color spaces for both indoor and outdoor conditions in Figure 6.4, it has been observed that RGB, HED, HSV, RGBCIE, XYZ, YDbDr, YIQ, YPbPr, YUV show acceptable results. While LAB, LUV, and YCbCr show poor results.

6.1.5 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with Resnext Encoder Model

After the assessment of various color spaces, encompassing RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV within the Resnext 50 model framework based on SSIM and PSNR values, it has been observed that YDbDr exhibits superior efficacy in indoor environments, whereas RGB demonstrates exceptional performance in outdoor settings as shown in Table 1. Conversely, YCbCr has exhibited suboptimal performance in both scenarios. Upon averaging SSIM and PSNR values across indoor and outdoor conditions, as shown in Tables 2 and 1, YDbDr has surpassed RGB, securing the foremost position for the Resnext model.

On doing visual analysis by observing the output dehazed images in different color spaces for both indoor and outdoor conditions in Figure 6.5, it has been observed that RGB, HED, HSV, RGBCIE, XYZ, YDbDr, YIQ, YPbPr, YUV show acceptable results. But in the first outdoor image of HSV and the second indoor image of XYZ, we can see a patch of distorted colors. In LAB, LUV, and YCbCr we find poor results.

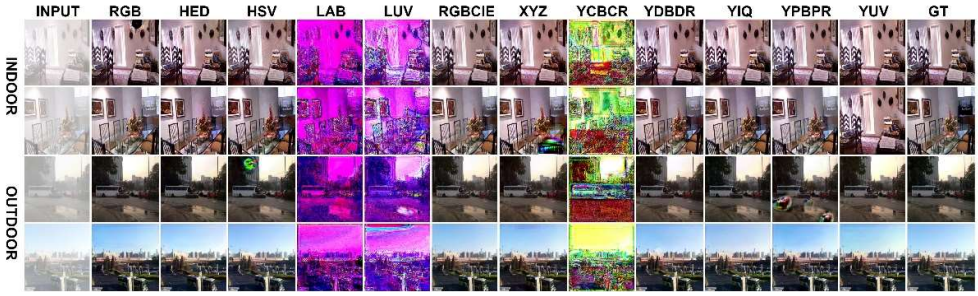


Figure 6.5: *Comparative Performance of the Resnext Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces.*

6.1.6 Performance Assessment of RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr and YUV color space on UNet with GerNet Encoder Model

On doing a thorough examination of the efficacy of various color spaces like RGB, HED, HSV, LAB, LUV, RGBCIE, XYZ, YCbCr, YDbDr, YIQ, YPbPr, and YUV, within the framework of the GerNet model, based on SSIM and PSNR values, it has been determined that YPbPr demonstrates superior performance in indoor settings, while XYZ exhibits excellence in outdoor environments as shown in Table 1. Conversely, YCbCr has displayed notably inadequate performance across both scenarios. Upon computation of the average of Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) values for both indoor and outdoor conditions, in Tables 2 and 1, it has been found evident that XYZ has surpassed YPbPr, thereby securing the foremost position within the GerNet model.

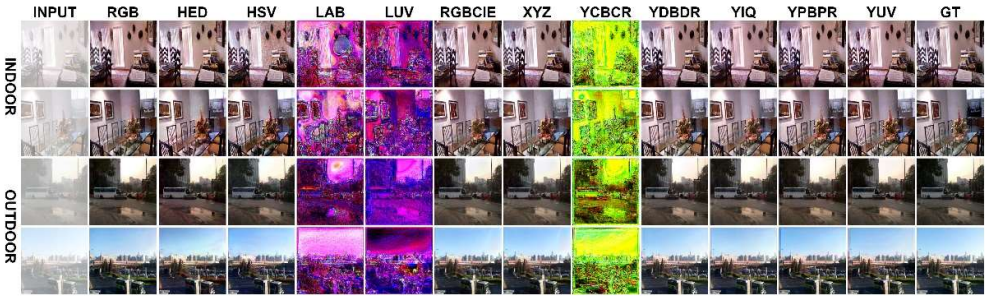


***Figure 6.6**: Comparative Performance of the GerNet Block in Different Color Spaces for Both Indoor and Outdoor Conditions. The First Column Represents the Input Hazy Images, While the Last Column Represents the Haze-free Ground Truth Images. The Other Columns Show the Outputs of the Dehazing Process in Different Color Spaces.*

On doing visual analysis by observing the out-put dehazed images in different color spaces for both indoor and outdoor conditions in Figure 6.6, it has been observed that RGB, HED, HSV, RGBCIE, XYZ, YDbDr, YIQ, YPbPr, YUV show acceptable results while LAB, LUV, and YCbCr show poor results.

Chapter 7

Conclusion

This thesis conducts an in-depth analysis of deep-learning models based on GAN architecture for haze removal. Specifically, evaluating the performance of the UNet model with different encoder such as MoblieOne, ResNet, EfficientNet, Mix Vision Transformer (MIT), ResNeXt and GerNet models across twelve different color spaces like RGB, HED, HSV, LAB, RGB-CIE, XYZ, YCbCr, YDbDr, YPbPr, YIQ, YUV and LUV. The key findings for future researchers are as follows:

- While RGB, HSV, and LAB color spaces are commonly utilized in research, XYZ, RGBCIE and YDbDr color spaces have also proven to be the most effective choices along with RGB.
- The Efficientnet model in RGBCIE color space has shown outstanding results, outperforming all other color spaces not only for the Efficientnet model but all of the other five models.
- In YCbCr and LUV color space most of the model fails to generalize the haze removal task and fails to generate realistic haze-free images.
- In comparison with state-of-the-art the ResNet in XYZ color space achieves better results than other models in terms of PSNR and SSIM, which reflects the superiority of other color spaces like XYZ in image restoration.

It’s important to note that this paper relies on the widely-used computer vision backbones and it hasn't introduced any new model. The primary focus of this research lies in analyzing the performance of existing models across various color spaces rather than achieving state-of-the-art results on the RESIDE6K dataset. Additionally, exclusive exploration of various color spaces is taken into consideration, with no investigation into other feature combinations to enhance the haze removal performance. Future research should incorporate color feature fusion to further improve of the dehazing performance.

Chapter 8

List of Publications

- “Deep Learning-Based Image Dehazing Across Color Spaces” (**S. Ganguly**, A. Ali, B. Das, S. Sk and S.S. Chaudhuri). Expert Systems.

- Patent: “A Real-time Portable Low-Power Skin Cancer Detection Device” (A. Ali, A. Paul, **S. Ganguly**, C. Ghorai, and S.S. Chaudhuri, Application no: 202431005005, Date:24/01/2024).

- Copyright: HazyVidSegNet (A. Ali, H. Shaw, **S. Ganguly**, B. Das, and S. S.Chaudhuri, Diary No: 353/2024-CO/L, Date: 04/01/2024)

- “Arabic Sign Language Recognition and Voice Translation Device Using Embedded Machine Learning” (A. Ali, B. Das, S. Sk, **S. Ganguly**, and S.S. Chaudhuri). 2nd Doctoral Symposium on Human Centered Computing (HUMAN 2024).

Bibliography

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] Yi Wan and Qiqiang Chen. Joint image dehazing and contrast enhancement using the hsv color space. In 2015 Visual Communications and Image Processing (VCIP), pages 1–4. IEEE, 2015.

[21] Mohammed Khalid Othman and Alan Anwer Abdulla. Enhanced single image dehazing technique based on hsv color space. UHD Journal of Science and Technology, 6(2):135–146, 2022.

[22] Linlin Wang, Qian Zhang, Deyun Yang, Yingkun Hou, and Xiaoting He. Image dehazing based on non-local saturation. In Ninth International Conference on Graphic and Image Processing (ICGIP 2017), volume 10615, pages 710–718. SPIE, 2018.

[23] Hussein Mahdi, NIDHAL EL ABBADI, and HIND RUSTUM. Single image de-hazing through improved dark channel prior and atmospheric light estimation. Journal of Theoretical & Applied Information Technology, 95(15), 2017.

[24] Harmandeep Kaur Ranota and Prabhpreet Kaur. A new single image dehazing approach using modified dark channel prior. In Advances in Intelligent Informatics, pages 77–85. Springer, 2015.

[25] L´aszl´o Neumann, Ram´on Heged¨us, G´abor Horv´ath, and Rafael Garc´ıa. Applications of high precision imaging polarimetry. In CAe, pages 89–97, 2008.

[26] Sriparna Banerjee, Pritam Kumar Ghosh, Pranay Kumar Singha, Swati Chowdhuri, and Sheli Sinha Chaudhuri. Relevance of color spaces and color channels in per-forming image dehazing. In 2021 Innova-tions in Energy Management and Renewable Resources (52042), pages 1–6. IEEE, 2021.

[27] Li-Ping Yao and Zhong-liang Pan. The retinex-based image dehazing using a particle swarm optimization method. Multimedia Tools and Applications, 80:3425–3442, 2021.

[28] Wen-Chang Cheng, Hung-Chou Hsiao, Wei-Lin Huang, and Cheng-Hsiung Hsieh. Image haze removal using dark channel prior technology with adaptive mask size. Sensors & Materials, 32, 2020.

[29] Trung Minh Bui and Wonha Kim. Single image dehazing using color ellipsoid prior. IEEE Transactions on Image Processing, 27(2):999–1009, 2017.

[30] Yiming Li, Chunli Zhu, Junxin Peng, and Liheng Bian. Fusion-based underwater image enhancement with category-specific color cor-rection and dehazing. Optics Express, 30(19):33826–33841, 2022.

[31] Mahmoud Afifi, Abdelrahman Abdelhamed, Abdullah Abuolaim, Abhijith Punnappurath, and Michael S Brown. Cie xyz net: Unprocessing images for low-level computer vision tasks. IEEE Transactions on Pattern Anal-ysis and Machine Intelligence, 44(9):4688–4700, 2021.

[32] Zahid Tufail, Khawar Khurshid, Ahmad Salman, Imran Fareed Nizami, Khurram Khurshid, and Byeungwoo Jeon. Improved dark channel prior for image defogging using rgb and ycbcr color space. IEEE Access, 6:32576–32587, 2018.

[33] M Kokila Das, P Dinulal, G Koshy, and Philomina Simon. Image dehazing using improved dark channel prior and relativity of gaussian. Procedia Computer Science, 165:442–448, 2019.

[34] Rana T Saihood et al. Aerial image enhancement based on ycbcr color space. International Journal of Intelligent Engineering & Systems, 14(6), 2021.

[35] Ahmed Rafid Hashim, Hazim G Daway, and Hana H Kareem. Single image dehazing by dark channel prior and luminance adjustment. The Imaging Science Journal, 68(5-8):278–287, 2020.

[36] Nadia A Khalaf, Hazim G Daway, and Baida M Ahmed. Hazy image enhancement using dcp and ahe algorithms with yiq colour space. International Journal of Intelligent Engineering & Systems, 16(2), 2023.

[37] Shuhe Zhang, Ashwin Mohan, Carroll AB Webers, and Tos TJM Berendschot. Mute: A multilevel-stimulated denoising strategy for single cataractous retinal image dehazing. Medical Image Analysis, 88:102848, 2023.

[38] Faming Fang, Tingting Wang, Yang Wang, Tieyong Zeng, and Guixu Zhang. Variational single image dehazing for enhanced visualization. IEEE Transactions on Multimedia, 22(10):2537–2550, 2019.

[39] Wei Liu, Ping Ye, and Hanxu Sun. Image dehazing based on region growing. In 2017 4th International Conference on Systems and Informatics (ICSAI), pages 192–197. IEEE, 2017.

[40] Xue Wen, Zongxu Pan, Yuxin Hu, and Jiayin Liu. Generative adversarial learning in yuv color space for thin cloud removal on satellite imagery. Remote Sensing, 13(6):1079, 2021.

[41] Jessica El Khoury, Jean-Baptiste Thomas, and Alamin Mansouri. Does dehazing model preserve color information. In 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Sys-tems, pages 606–613. IEEE, 2014.

[42] Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel, and Anurag Ranjan. Mobileone: An improved one mil-lisecond mobile backbone. In Proceedings of the IEEE/CVF Conference on Com-puter Vision and Pattern Recognition, pages 7907–7917, 2023.

[43] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[44] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

[45] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning, pages 6105–6114. PMLR, 2019.

[46] Saining Xie, Ross Girshick, Piotr Doll´ar, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural net-works. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1492–1500, 2017.

[47] Cristian A Gallo, Jessica A Carballido, and Ignacio Ponzoni. Gernet: A framework for inference, visualization and manipulation of gene regulatory networks based on association rules. In VI Congreso Argentino de Bioinform´atica y Biolog´ıa Computacional, 29-31 de octubre de 2013, Rosario, Argentina, 2013.

[48] 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.

[49] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translтion with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1125–1134, 2017.

[50] Jie-Neng Chen, Shuyang Sun, Ju He, Philip HS Torr, Alan Yuille, and Song Bai. Transmix: Attend to mix for vision trans-formers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12135–12144, 2022.

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th Interna-tional Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pages 234–241. Springer, 2015.