ClearFlow: An Enhanced Image Dehazing Workflow Using Model Ensembling and CNN Postprocessing

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Hemanth Harikrishnan
School of Computer Science
University of Galway

Supervisor(s)
Dr. Frank Glavin

In partial fulfillment of the requirements for the degree of
MSc in Computer Science (Artificial Intelligence)
August 2024

DECLARATION I, Hemanth Harikrishnan, hereby declare that this thesis, titled "ClearFlow: An Enhanced Image Dehazing Workflow Using Model Ensembling and CNN Postprocessing", and the work presented in it is entirely my own except where explicitly stated otherwise in the text, and that this work has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signature: H.Hemanth

Abstract

Haze exists in the environment due to pollution or humidity. Images captured in such an environment have degradation in the quality of the image. This negatively impacts various high-end computer vision tasks such as object detection which can be one of the most important components in the domain of autonomous vehicles. Haze reduces the contrast of the image, brightens the haze affected regions, and blurs the edge details. Various single-image dehazing systems that rely on different methodologies are present. This thesis aims to investigate the usage of two end to end Convolutional Neural Network (CNN) models in an ensemble setting and use a fixed discrete wavelet transform in one of the models to add more information to the input image to procure plausible output without increasing the complexity of the model with additional traditional denoising methods to procure a haze free image.

Keywords: Image Dehazing, Computer Vision, Encoder Decoder Model, Convolutional Neural Networks

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Haze is a natural phenomenon that occurs mostly due to the existence of particles in the atmosphere. When images are captured in a hazy environment, they result in brightened, low contrast with no clear texture. It is a problem that can easily affect any high-end visual systems to perform tasks as the image clarity is lost. The work done in this thesis is focused on procuring a haze-free image from a hazy image. This is achieved by using a dehazing workflow which consists of deep learning and traditional denoising approaches. The deep learning models learn the relations between the visual features between the hazy and clear images. In this thesis, this is achieved using Convolutional Neural Networks (CNN) with a supervised learning approach. The convolutional models can help reduce the scope of overfitting as it encourages sparse connectivity, translation invariance, and uses a lower number of parameters due to parameter sharing. The dataset contains pairs of hazy and clear images, and two encoder-decoder models with residual connections are used to provide the dehazed image for a given hazy image. One of the models in the ensemble uses Discrete Wavelet Transform (DWT) to enrich the intermediate features. Additionally, other traditional denoising algorithms are used to denoise the output image from the models to enhance the

quality of the output image.

1.1 Motivation

Encoder-decoder models have been significantly used in research works in data-based approaches to dehaze an image. The motivation to apply this method is that the prior-based methods rely on the statistics of the image and are not robust to all kinds of haze. However, these methods are designed manually and do not need any form of training to perform the dehazing process. Convolutional Neural Networks (CNN) help exploit the relationships between various regions of the image and they can help effectively model the hazy image into a clear image. Additionally, the frequency domain information of the hazy image is used to add more information to the input when passed into the model.

Deep learning models in general require large amounts of data for better generalisation. One of the major issues in the past decade is the lack of quality and quantity of hazy and clear image pairs, and these have been mitigated in recent times by collecting data in a controlled environment where the haze is generated using machines and has been made openly available.

Considering the abundance of data present, a significant amount of work has been done in the area of encoder-decoder models to improve the quality of images affected by haze. The published models [5] such as AOD-Net [6] can reduce the number of artefacts and colour shifts present in the output of the model (dehazed image).

1.2 Purpose

This thesis aims to present the reader with a novel dehazing workflow (ClearFlow) using encoder-decoder models and other postprocessing techniques where the models use the combined input of spatial and frequency domain information of the image. The model is trained on homogeneous and non-homogeneous image pairs. The results of the methodology adopted will be compared with other relevant research published earlier. The following research directions are being addressed in this thesis.

1. Creating a model with lower training and processing time to perform the dehazing process with a lower response time such that it can be used as a preliminary component in complex computer vision tasks such as image segmentation.

2. Creating a robust model workflow that can handle varying densities and nature of haze on the image.

1.3 Structure

This thesis is divided into 6 chapters. Chapter 1 provides introductory information on the domain of image dehazing. Chapter 2 provides detailed information on the background knowledge required to address the problem statement. Chapter 3 reviews the relevant research literature for image dehazing such as prior-based methods, and data based methods. Chapter 4 discusses the datasets present for the task in detail, then discusses the methodology used for this research work and provides details of the experiments performed and the model architectures used. Chapter 5 reports the results achieved, and the comparative analysis for the approaches experimented. Chapter 6 provides the contributions of this thesis

and discusses the scope for future improvements. Finally, Appendix A provides the details on the Python scripts used for the thesis.

Chapter 2

Background

In this chapter, the focus is on defining and mathematically formulating haze, and the components used in this research for the removal of haze from the image. The chapter specifically discusses the Atmospheric Scattering Model (ASM) which helps mathematically define haze, and Discrete Wavelet Transform (DWT) which helps procure the frequency domain components of the image. Additionally, the description of the Convolutional Neural Networks (CNN) and the associated building blocks of the model along with the metrics used to evaluate a dehazing model will be discussed. Most of the existing literature evaluates the model performance using the Peak Signal Noise Ratio (PSNR), and Structural Similarity Index Metric (SSIM).

2.1 Atmospheric Scattering Model

Various models have been proposed to mathematically model the haze present in the image. The Atmospheric Scattering Model [7, 8] is a widely used model to mathematically represent the haze present in the image. Some of the paired dehazing datasets have been generated based on this approach [9]. A large amount

of past research approaches that perform the dehazing process are heavily based on ASM where the image affected by haze (I) can be expressed mathematically as:

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

Where J is the haze-free image, A is the atmospheric light, t is the transition map associated, and x is the pixel location. The transmission map can be mathematically represented as:

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

Here β represents the attenuation coefficient [1] and d(x) represents the depth of the scene at pixel x. It is believed that the distance of the camera from the scene has an impact and larger distances face more attenuation. Additionally, a higher attenuation coefficient degrades the image. All the dehazing methods aim to estimate the parameters A and t for the given depth to procure the dehazed image as output.

2.2 Discrete Wavelet Transform

Analysing images in the frequency domain helps understand the rate of change of values of pixels in the spatial domain. The image is transformed to a frequency distribution and is then provided for further processing. The high frequency components help derive the information on the edges and low frequency refers to the other smooth regions. Discrete Wavelet Transform (DWT) [10] can provide both frequency and temporal details. In DWT the signal is passed into two

[1] Describes the rate at which the intensity of the light decreases when it propagates through the atmospheric medium. The magnitude defines how easily the medium can scatter light.

filters which are a high pass and a low pass respectively which decomposes the image to high frequency details and low frequency components. There are four outputs derived after downsampling and passing to two other high-pass and low-pass filters. In this thesis, Daubechies 4 tap wavelet (DB4) is used to procure the frequency domain information of the image. This wavelet is used due to the following properties:

- DB4 wavelet is relatively smooth, orthogonal and computationally less intense without compromising on the output feature quality.
- It helps balance the time and frequency localisation.
- Higher tap wavelet versions and other higher quality wavelets (Symlet, biorthogonal) are more computationally intensive.

This process, illustrated in Figure 2.1 below, can help capture features in four directions of the image.

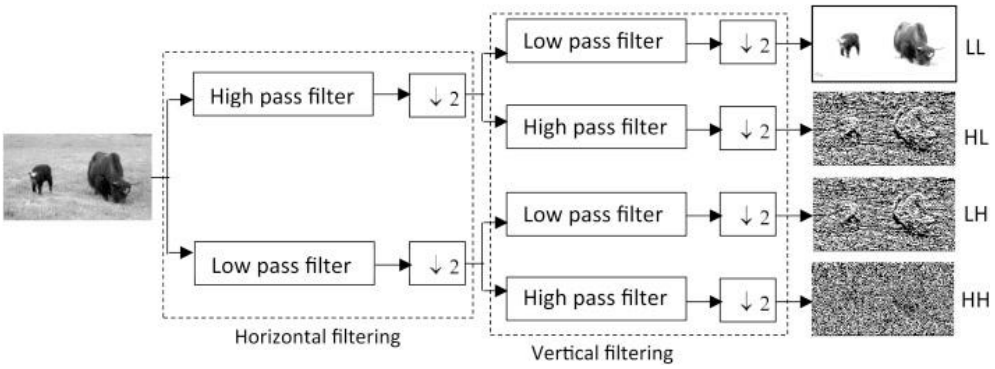

Figure 2.1: Discrete Wavelet Transform of an image. Image from [1].

2.3 Convolutional Neural Network

Convolutional Neural Networks (CNN) [11] are used majorly in the tasks associated with computer vision. This class of neural network architectures can interpret visual data and can complete of specific tasks (classification, segmentation, etc.) based on the image data provided as input. Here, mathematical operations are applied to the image to extract useful features by also including the neighbouring pixels in the operations. The output from this can be passed into a Feed-forward Neural Network (FNN). CNN is specifically used as they are good in detecting patterns in image data, and translation invariant, and there is no need to manually design the feature extraction kernels. The main building blocks of the model are the convolution layers, instance normalisation layers, and residual connections. We will now look at these concepts in more detail.

2.3.1 Convolution Layers

The primary property of a CNN is that the weights are shared. The convolution layer takes the image or the feature extracted as the output from the previous convolution layer as the input. Multiple kernels are used at each layer where each kernel has a matrix of weights, a bias term, and an activation function associated with it. Each kernel has the property to extract a specific feature from the image. A kernel can be one dimensional matrix $(1 \times 1 \times n)$, or can also be a 2D matrix of any size $(a \times b \times n)$ where n is the number of kernels. The output from the layer is an n-dimensional feature map where there are n kernels used. This can help extract more than one specific feature at one single step. All the values in the kernel are learned during the backpropagation step and the values are adjusted accordingly. Figure 2.2 illustrates the convolution process which is performed in the convolution layers.

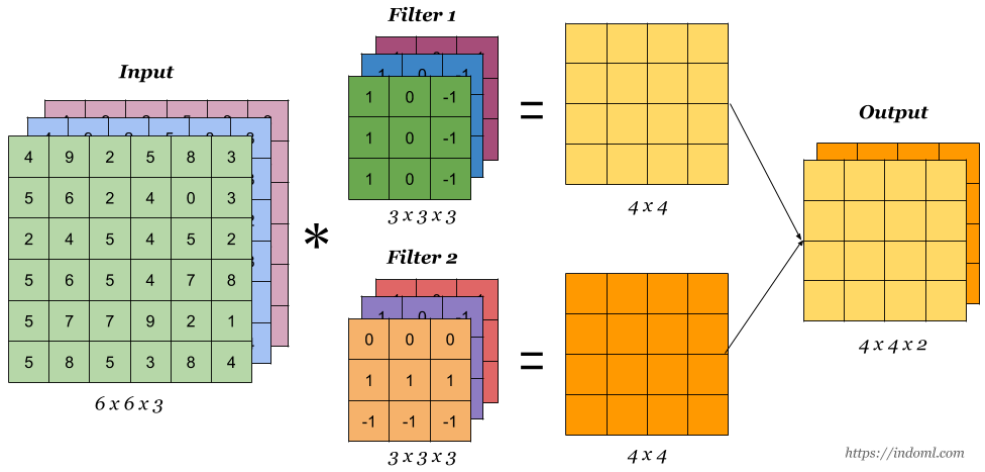

Figure 2.2: Convolution Layers. Image from [2].

2.3.2 Instance Normalisation

Normalisation is done to handle the covariant shift in the neural network which affects the existing learned representations of the model negatively. Normalisation methods help maintain the uniformity of the distribution being passed into the layers. In the case of instance normalisation, the mean and the standard deviation are computed across each channel for every single example hence, there is no need for stalling the normalisation calculation until a batch is completed. It normalises the intensity in each channel in the feature map of the image.

2.3.3 Residual Connections

Residual connections help avoid overfitting of a model, and high computation costs making the flow of gradients very smooth. As the convolutions take place, especially in an encoder to decoder architecture, there is a loss of information and fine grained details of the image when the downsampling operation is being performed. To avoid this the layers from the encoder are skip-connected to the decoder to help fine grain features not to be lost.

2.4 Metrics

Metrics are important to help compare and evaluate the performance of the model created.

2.4.1 Peak Signal to Noise Ratio (PSNR)

Peak Signal to Noise Ratio (PSNR) is a referential metric to measure the quality of the procured image. The higher the value, the better the quality of the reconstructed image. The Mean Squared Error (MSE) is the sum of the squared error between the output image produced by the model and the input image, while Peak Signal to Noise Ratio (PSNR) measures the peak error present in the reconstructed image. MSE is desired to be lower for better quality. The result is expressed in decibels (dB).

$$\text{PSNR} \;=\; 10 \;\log_{10}\left(\; \frac{R^2}{\text{MSE}} \;\right)$$

Here R is the maximum pixel value present in the input image, and MSE is the mean squared error calculated between the two images.

2.4.2 Structural Similarity Index Metric (SSIM)

Structural Similarity Index Metric (SSIM) [12] can capture the similarity in structures and patterns between two images. It considers an image patch between two images to compute the structural similarity by comparing the luminance, contrast, and structure. The scores from the three aspects are combined to be represented in a range between -1 and 1. If the Structural Similarity Index Metric (SSIM) is closer to 1 then there is a high structural similarity between the images. The metric is also less sensitive to random noise present in the image.

$$L(x,y) = \frac{(2\mu_x\mu_y + C_1)}{\left(2\mu_x^2 + \mu_y^2 + C_1\right)}$$
$$C(x,y) = \frac{(2\sigma_x\sigma_y + C_2)}{\left(2\sigma_x^2 + \sigma_y^2 + C_2\right)}$$
$$S(x,y) = \frac{(\sigma_{xy} + C_3)}{(\sigma_x\sigma_y + C_3)}$$

Here µ refers to the mean over the window, σ refers to the standard deviation, and σ_{xy} refers to the covariance between the images. C refers to constants.

$$\text{SSIM}(x,y) = [L(x,y)]^{\alpha} \cdot [C(x,y)]^{\beta} \cdot [S(x,y)]^{\gamma}$$

Chapter 3

Related Work

Single image dehazing tasks have gained a lot of attention recently due to their importance in providing accurate high-end computer vision task results. Researchers have proposed a wide range of methods specifically to tackle single-image dehazing. The methods can rely on the Atmospheric Scattering Model (ASM) [7, 8] or can be completely dependent on the relationships learned by the neural networks. Different methods use different datasets, which can be synthetically generated or captured from a scene in the real world. The methods can be divided into two categories broadly: Prior based approaches, and Learning based approaches.

3.1 Prior based Approaches

Prior based approaches extensively use the statistical knowledge obtained from the image. Fattal *et al.* [13] proposed an approach that does not employ any deep learning based approach to dehazing the image. The method determines the optical transmission in a hazy image and then the scattered light is eliminated to retrieve the dehazed image. The model accounts for surface shading along with the transmission function. The haze degraded image is modelled as the

components of airlight and surface radiance. These components are added with the help of a transmission coefficient to determine the value of each pixel. The goal is to ensure the shading and the transmission functions are statistically uncorrelated. A graphical model is used to utilise the scene transmission and surface shading along with the transmission function to convert the solution to pixels. All the components are estimated using the uncorrelation principle and this model is derived analytically and makes sure the signal to noise ratio drops below a threshold. The method is efficient in retrieving the colour of the haze and can also be used in tasks such as image focusing, and novel view synthesis without needing multiple images. The scene depth of the image is estimated and need not be provided as an additional input. The model is not robust to thick haze and blurring effect caused by haze. Kaiming *et al.* [14] used Dark Channel Prior (DCP) where the dehazed image is retrieved using the dark channel prior of the image. The prior computed is used to estimate the thickness of haze in the image. This contains the information of atmospheric light which can be used to compute the dehazed image with the help of the atmospheric scattering model (ASM). It generates a few halo artefacts and fails to retrieve images with high quality of the object having the same colour as the atmospheric light with no shadows. It also cannot accommodate complex haze relationships present in the image.

Qingsong *et al.* [15] used colour attenuation prior based on haze lines. The method estimates the transmission medium which is useful to remove the haze from the images. The study found a strong variance of pixels in the image correlated with the strength of the haze. The method was evaluated using the O-Haze [16], I-Haze [17], FRIDA [18], and some additional real-world images. The method is similar to Dark Chanel Prior (DCP), where the depth image computation takes place. The depth of the scenes is estimated using a linear model. The top 0.1

percent bright pixels are considered and the pixels in them with the highest L2 normalisation are considered as atmospheric light. The scene radiance and transmission map can be estimated with the atmospheric light and depth data. Since the depth image is blurred due to the filter, a guided filter is then used to produce the final dehazed image. The method cannot work well if background noise or colour distortions are present in the image.

Li *et al.* [19] uses a novel Atmospheric Scattering model where an additional scattering compensation term is introduced to the heurisitc model [7, 8]. The scattering compensation term can suppress the colour cast produced by the atmospheric light term. A Global Information Search (GIS) is used to retrieve the transmission map using the minimum value channel. The method lacks a better approximation of parameters to procure the haze free image. Other methods are closely inspired by Colour Attenuation Prior (CAP) and Dark Channel Prior (DCP) which use modified version of the algorithms proposed along with some additional processing steps [20, 21].

3.2 Learning based Approaches

3.2.1 Supervised Learning Approaches

Supervised approaches rely on paired data which can guide the learning process in the form of supervisory signals. The models can learn based on the generated haze-free images, and transmission maps. Models can be based on the ASM or can completely rely on the mapping given by the neural network to map hazy images to dehazed images.

3.2.1.1 End to End CNN Based Approaches

End-to-End methods have used CNNs to learn the pixel relation between the hazy and clear image. They focus more on the aspect of feature representation and an encoder decoder architecture is used. Frants *et al.* [22] proposed an encoder-decoder model that uses a quaternion[1] neural network to perform the dehazing task. The model uses a special quaternion pixel-wise loss function and a quaternion instance normalisation layer. The architecture models the relationship between the channels and the image data which is generally ignored in other approaches. It also has a lesser number of parameters and hence is very fast. The model is also subject to limited overfitting and can capture more colour relationships in the data. Dilated convolutions are used to retrieve the dehazed output. Wu *et al.* [3] introduced a knowledge transfer-based approach to perform the process of dehazing where there is a teacher network and a dehazing network. The teacher network learns on clear images and provides a robust image prior as output. The teacher and the dehazing networks have an identical architecture. The intermediate features are supervised in the teacher network and the role of

[1] A mathematical entity that extends the dimension of complex number system space. They introduce three complex entities (i, j, k) bounded by specific mathematical rules. They are effective to represent rotations in 3D space [23].

the dehazing network is to imitate the teacher network. An attention mechanism is used to combine the channel and pixel attention which can help capture information effectively. An enhancing module is used to fuse the global context into the output and improve the quality. The prior knowledge from the teacher network is passed into the dehazing network through the intermediate feature maps. The model uses feature-level loss to evaluate the knowledge transfer. The encoder portion uses a pre-trained Res2net architecture [24] to procure feature-level information and also has skip connections to preserve the information. The output from the encoder of the teacher network is passed into the dehazing network. The enhancement module is used in the last convolution layer. The overall performance is highly varying and is dataset dependent.

Chen *et al.* [25] introduced a model based on gated context aggregation to restore the hazy image. The method uses smooth dilation in the context aggregation step to remove the grid artefacts which are present in the output while dilated convolution is performed. A gated subnetwork is used to fuse the features without compromising the spatial information. The gated subnetwork determines the feature importance and weightage is assigned and the features are fused accordingly. The model is in the form of an encoder-decoder architecture where the input image is encoded into feature maps, and post-encoding the features are fused on various levels without downsampling. The enhanced feature maps are decoded back to the original image and the haze residue is procured. This haze residue is added back to the input image to procure the final image. To enhance the learning between the encoder and decoder multiple residual blocks are placed. To facilitate the learning, additional edge information is also fed into the model. Guo *et al.* [26] propose a model based on semi-circular attention to handle non-homogeneous haze where the focus is on the pixels in the haze-affected regions. The model is based on two components which are the attention generation net-

work and the scene recreation network. The attention map is generated based on the luminance difference between the hazy and dehazed images. The attention network learns complex features between non-homogeneous haze and the underlying scene. It has a self-paced semi-circular attention map to handle the varying luminance in the image and also helps in model convergence by reducing learning ambiguities that come with multi-objective learning. In this case, the attention map and dehazed image need to be learned. The attention generation unit uses channel attention, and multi-scale pixel attention to generate feature maps, and the scene reconstruction unit uses encoder-decoder architecture. The method supervises the learning of attention maps along with learning the dehazed image as inherently learning the representation can lead to irregular weight assignments on the map. In semi-circular learning, the ground truth attention map is more relied on for the first 25 percent of the epochs, and then for the rest, the generated feature map is more relied on.

Sun *et al.* [27] introduced a method to fuse multiple features which are shallow and deep. The non-local and local information are merged using an enhancement attention module to learn the channel attention. This has proven to have better recovery of dehazed images with encoded quality. The other methods heavily rely on the Unet based architectures and have a loss in edge information, and scene information due to ignoring encoding layer information. The squeeze to excitation based channel attention affects the weight prediction of channels in a negative manner which is the reason for using multi-level feature interaction and non-local information enhanced channels. The model also contains a Generative Adversarial Network (GAN) which utilises multiple features to drive the model to provide high-quality dehazed images. Li *et al.* [6] proposed a model built on CNNs, and dependent on reformulated ASM. It uses a lightweight CNN to model the hazy image to the dehazed image, and can also be cascaded before

any high-level computer vision model. There are no intermediate steps while performing the dehazing step. The model is trained on synthetic and natural images. The model has the K estimation module and clean image generation module. The K estimation module estimates the depth and relative haze levels which is comprised of 5 convolution layers. The clean image module performs element-wise multiplication and addition layers. There is a continuous integration of information from the previous layers and multiple intermediate connections to compensate for the loss from convolutions. The model had good generalisation with both outdoor and indoor images. It was also able to remove the halation[1] effect from the images inherently.

Zhang *et al.* [28] proposed a model which directly maps the hazy images to non-hazy images. There are three components present in the architecture where the point-wise convolution extracts local statistical regularities, the feature combination learns the spatial relation in the image, and the reconstruction module helps recover the haze-free image. The model is trained on synthetic and real-world images. The model is a fully convolutional dehazing network where the global information is exploited to perform haze removal. A non-local loss is introduced and can drastically reduce the colour variations in the model output. The point-wise subnetwork with non-local loss is used to generate the training data as the existing methods misinterpret the depth regions and lose the edge content of the image as inaccuracies in the depth pixels can affect the neighbours. Pooling is not used in the model as it is found to affect the learning process negatively. Cai *et al.* [29] provided a model that produces the dehazed output based on ASM where the model predicts the transmission map and retrieves the dehazed output image. A novel activation layer called Bilateral ReLU (BReLU) is used to recover dehazed images. The haze transmission map is heavily dependent on the

[1] A glow encircling the bright areas of the image. The effect of light is present beyond the natural boundaries. This introduces a foggy effect around the edges of an object in the image.

depth map present for the image. The architecture is closely aligned with prior-based methods that are used for image dehazing. The model majorly consists of cascaded convolution and pooling layers with non-linear activation functions. The model can inherently remove the halation effect from the images however, it cannot perform well on images with thick haze and has issues with sky regions in the image as they contain the same qualities as haze.

Liu *et al.* [30] proposed an end-to-end CNN that can be used to procure the dehazed image. It consists of three modules which are preprocessing, backbone, and post-processing. The model does not rely on the ASM to generate the dehazed images. It has three modules (preprocessing, backbone, and post-processing). The trained preprocessing module can easily generate more high quality features than handcrafted methods. The backbone is an attention based multi scale estimation which is done on a grid network. This is used to mitigate the bottleneck issue which is seen mostly in Unet models. The post-processing module reduces the number of artefacts in the final processed image. The proposed model performs well in both synthetic and real-world domain images. Due to the data scarcity, the preprocessing module augments various versions of the image making the relevant features present in the image more exposed to the model to learn better. The channel attention network is used to mitigate the bottleneck issue. The backbone is an enhanced Gridnet model where the grid elements in the same row have the same dimensions and the flow is connected to other rows with upsampling and downsampling blocks. The dehazed image from the backbone module contains artefacts and these are removed by the post-processing module.

Ren *et al.* [31] introduced a residual encoder-decoder model to procure the dehazed image and did not estimate the transmission map and atmospheric light. Since it is not dependent on the ASM the chance of error propagation to affect

the final image is drastically reduced. The encoder captures the image context while the decoder captures the pixel contribution to the final output. The white balance, contrast-enhancing, and gamma corrected images are procured as intermediate outputs when a hazy image is passed into the model to blend the regions and preserve regions with visibility and colour cast. The multi-scale approach helps mitigate halo effects which are found during image dehazing. The model uses early feature fusion to fuse the intermediate outputs procured from the input image. The dehazed image output is procured by gating the important features from the three inputs. A multi-scale approach artefact is used to avoid halo effects. To retrieve the final output the confidence maps procured from the decoder are multiplied with the inputs to procure the final output. The model cannot handle images with thick haze.

Kaur *et al.* [32] introduced an end-to-end network that can remove haze while keeping the spatial and spectral characteristics of the image intact. Residual connections and Mean Squared Error (MSE) with consistency loss are used to maintain the consistency and content of the image. The model uses two layers which are the haze removal layer, and the spectral consistency layer. The model works well except for extreme hazy conditions and also lacks adaptability to other hazy conditions. The training process can be challenging as it is resource intensive. Rani *et al.* [33] aimed to perform haze removal and detecting traffic signs simultaneously. An Unet (HRU-net) is used to perform the haze removal task. The output from the dehazed network is passed into an object detection network. It is aimed at making sure the autonomous system can process the data in real time to ensure smooth and safe functioning. An Adaboost detector with a convolution network is used to capture the high-level features, and a parallel Deformable convolution module is used to increase the effectiveness of the feature extraction. The HRU-net is cascaded with an image processing module as the

direct output is distorted in nature. The haze is removed with the white balance module and ASM. The method was able to detect at 99 per cent accuracy and was aimed to be deployable in autonomous vehicles.

3.2.1.2 Frequency Domain Based Approaches

Image dehazing leveraging the frequency domain characteristics has been successful and has been incorporated into various data-driven methods. Yang *et al.* [34] emphasised that the edges and colours of the image are the most important factors to be restored in the process of dehazing. The model is a two stage end to end network where a wavelet Unet replaces the procedures of the discrete wavelet transform, and inverse discrete wavelet transforms using a 2D Haar wavelet to perform the procedures of upsampling and downsampling. The process of DWT and IDWT is capable of capturing low and high-frequency features which help capture the edge-level features of the image in the frequency domain. The decomposed wavelets help divide the overall image spectrum to help capture the images better. The image gets split into four bands and a scaling function is applied. The residual blocks are used in the Unet to strengthen the accuracy. A chromatic adaptation transform is then used to enhance the images.

Wang *et al.* [35] introduced using a Multi-Wavelet Residual Dense CNN (MWRDCNN) for the task of image denoising. The model uses residual blocks in each layer with a multi-wavelet CNN (MWCNN) [36] model as the backbone. This helps balance the tradeoffs between the denoising quality and the computation power. It uses a short-term residual strategy to increase the efficiency of learning. The residual dense block will be useful as it can extract better representations due to the hierarchical structure of the network. The model has DWT and IDWT in the CNN architecture and is implemented by using convolutions.

Dong *et al.* [37] highlighted that existing methods do not focus on the aspect

of signal degradation but only on the feature representation. A multi-scale architecture is proposed to extract the contaminated features from varying spatial scales to restore the dehazed image in a coarse to fine manner. A Gabor Wavelet module is used as just multi-scale architecture cannot remove all the artefacts. It can capture contaminated features in four orientations which helps in easier learning. A refining unit is then used to remove the contamination in features extracted in all orientations. An IDWT step is performed in the end to combine all the features. The model also uses Channel attention in the encoder to extract features across various scales. The decoder upsamples and a Gabor wavelet module is used to process the features. It decomposes into eight parts and the IDWT step is trainable to reconstruct the features with efficient texture analysis. There are real and imaginary branches to represent the real and the imaginary parts which are represented using convolution kernels. The CNN in the refining module is responsible for finding the optimal frequency to remove degradation. There is not much difference between using a Haar wavelet transform and a Gabor wavelet transform, and there was no colour distortion observed. Liu *et al.* [38] introduced a model that focused on non-homogeneous haze and identifying dense haze regions. The model learns image to image mapping and consists of three subnetworks which are encoder-decoder, detail refinement, and haze density map subnetwork. The amalgam of these networks helps distinguish thick haze, and thin haze regions to avoid applying the wrong amount of dehazing. A frequency domain loss (FFT loss) is used to make sure all the frequency bands of the image are uniform. Along with it, other loss functions are used while training to keep the spatial domain intact. The encoder-decoder model is based on a pre-trained model ResNet [39] which is trained on the ImageNet dataset. The decoder has skip connections to incorporate features from the encoder and also uses instance normalisation. The haze density map from the image is learned in the haze den-

sity subnetwork which is an Unet based architecture. An inverse pixel shuffle is done to convert the depth data from the haze density map to spatial, and a residual scaling is done to avoid training instability. The detail refinement subnetwork is used to enhance the quality of the image and it does nonlinear mapping from pixel to pixel.

Wang *et al.* [40] introduced a model to retain the image texture details using wavelet transform. The ill effect of downsampling the image was discussed and was the primary reason for using wavelet transform as it can preserve both the low and high frequency information. The model consists of two major parts which are a three scale residual CNN, and an ensemble attention CNN. The former uses wavelet transforms to obtain a downsampled image. In each scaling branch res2net modules [24] are connected in series to procure the hierarchical information from the hazy images. The model extracts multiple features using DWT and the features are concatenated and passed into the convolution network to procure the final image. A residual network with wavelet transform can perform well as sparsity is introduced leading to better convergence. The channel attention is used in the ensemble attention CNN to procure the final output and also fuses the hazy image from the previous stage. The model predicts the residue between the hazy image and the dense image, and the ensemble module considers multiple features at a time and fuses them based on the weights of the channels from the channel attention module.

3.2.1.3 Transformer Based Approaches

Transformers have recently become the state of the art approach not only in the domain of language but also in the domain of images. Song *et al.* [41] introduced Dehazeformer which used customised vision transformers for the application of single-image dehazing. The paper conveys that using the original

vision transformers has not achieved good results for the dehazing task as the Swin transformer [42] does not have a suitable architecture for it. Dehazeformer uses a modified normalisation layer, activation function, and spatial information aggregation step. The model has minimal computation cost when compared to other dehazing frameworks. The model uses rescale normalisation instead of layer normalisation as the latter loses the spatial relationship between the image patches. ReLU activation was found to work better in the domain of dehazing when used in the decoder layers. The cyclic shift used in the Swin Transformer [42] loses a lot of edge features, hence a shifted window partitioning scheme is used along with reflected padding of input to mitigate this and maintain a constant patch size. The reconstruction module uses multiple Unet transformers to reconstruct the image from feature representations. While computing the global representations, SoftReLU is used instead of ReLU, LeakyReLU, or GeLU as the non-linearity of the model costs the quality of output. The model was found to have weak inductive bias and quadratic computation cost.

Li *et al.* [43] introduced the Guided Transmission Map Network (GTMNet) which is aimed at dehazing remote sensing images that have dense haze. The results from other models are prone to over enhancement, colour distortion, and the presence of artefacts. The model uses a convolution neural network (CNN), and Vision Transformers (ViT) along with the dark channel prior approach to model the haze. The spatial feature transform layer is used to procure a guided transmission map which is used as the input to the neural network. Using the transmission map can help the network to capture the haze thickness details. The restored image from the process is passed into the strengthen-operate-subtract (SOS) boosted module. The model is an encoder decoder-based model where the encoder architecture is used to procure the feature maps.

3.2.2 Unsupervised Learning Approaches

Unlike supervised learning methods which are heavily dependent on the existence of paired data that are difficult to obtain, unsupervised methods do not rely on paired data to learn the mapping between the hazy environment and clear environment. A limited amount of high quality training data causes overfitting and instability of learning to remove haze.

Fu *et al.* [44] introduced a discrete wavelet transform GAN which aims to remove non-homogeneous haze from the image. It is a two-branch GAN where the first branch does wavelet downsampling which helps reduce the number of parameters and scope of overfitting of the model. The second branch Res2net [24] is used to extract multi-level features. An attention mechanism using pixel-wise attention, and channel-wise attention is used. A discriminator is then used to introduce an adversarial loss in training and pushes the model to generate high quality samples. The model uses 2D discrete wavelet transform (DWT) to capture the low and high-frequency knowledge in feature maps. It also helps find the colour mapping of the image from hazed to dehazed. The DWT branch is constructed using Unet architecture. Residual connections are present to enhance the information capturing. The frequency domain operations are combined with CNN to make sure both the spatial and frequency features are learned by the model. To handle the lack of information, Res2net [24] is used in the knowledge adoption branch. A patch-based discriminator reduces the presence of artefacts in the output image. A higher emphasis is put on restoring the edges as haze erodes the high-quality edge information.

Qu *et al.* [45] introduced an enhanced pix2pix dehazing network where the output is generated independent of the physical scattering model. The model is embedded with a GAN and an enhancer module sequentially. It contains three parts which are generator, enhancer, and discriminator. The generator along

with the enhancer module produces realistic images for a given input, and the discriminator guides the network to output high quality images. Before passing the output to the enhancer block, a pyramid pooling block is used to extract multi-scale features and pass these features as input to the enhancer block. The enhancer module performs dehazing on both colour and details. It is based on the receptive field model. A shortcut is present to preserve the colour information of the images. The generator and the discriminator are multi-resolution where there are two modules of generator and discriminator embedded where one is focused on capturing local features while the other captures global features. The overall model performs conversion from a hazy image to a dehazed image by modelling pixel to pixel along with the method of style transfer provided by the GAN architecture. The training is aimed at optimising both the GAN model and the generator enhancer module. It was observed that the output without the enhancer block was over coloured, and lacked a lot of details. This is due to the inability of the discriminator to drive the output to contain high structural information. The method is not effective for heavy haze scenes and the edges cannot be recovered.

3.3 Summary

In this section two prominent methodologies to dehaze the image were discussed. The prior-based approach relies only on the statistical data of the image to dehaze the image. It does not need to be explicitly trained to dehaze an image however, it is not robust to non-homogeneous and dense haze environments. Learning based approaches use encoder-decoder architecture built using CNNs to learn the relationships between a hazy and clear image. Supervised learning approaches require pairs of hazy and clear images to effectively dehaze. Methods also use frequency domain information along with the input image to compensate for the loss in details due to convolutions. Unsupervised learning based approaches are based on GANs which do not require explicit pairs of hazy and clear images to learn the relationships. However, the issues present in the models are that the models demand higher computing power and memory, and they lack generalisability to the density and nature of haze.

Chapter 4

Experiments and Methodology

This chapter discusses the datasets present for this task in detail along with the methodology and various experiments performed for the dehazing task. The chapter focuses on the data, methodology to approach the task, experimentation settings and the model architecture.

4.1 Data

This section discusses the datasets used for the model's training and evaluation. Multiple datasets have been used which consist of image pairs from the real world.

The model is trained using the supervised learning strategy where pairs of haze and haze-free images are considered. The datasets considered are REalistic Single Image DEhazing (RESIDE) (Indoor Training Set (ITS) and Synthetic Objective Training Set (SOTS)) [46], I-HAZE (NTIRE 2018) [17], O-HAZE (NTIRE 2018) [16], DENSE-HAZE (NTIRE 2019) [47], and NH-HAZE (NTIRE 2020) [48]. All the datasets except for RESIDE contain hazy images that are not synthetically generated but are generated using a haze generating machine. The RESIDE dataset's hazy images are generated based on the depth maps and images in

the NYU Depth-v2 [49] and Middlebury [50] datasets. RESIDE-ITS contains 1399 unique scenes with 13990 homogeneous hazy images. RESIDE-SOTS Indoor training set is considered for testing the model and has 50 unique scenes with 500 hazy images. I-HAZE contains 35 pairs of images that focus on indoor scenes, O-HAZE contains 45 pairs of images that focus majorly on outdoor scenes, DENSE-HAZE contains 33 pairs of dense and homogeneous hazy images, and NH-HAZE contains 55 non-homogeneous hazy and haze-free image pairs.
Real world image pairs are predominantly preferred to be used rather than synthetic [9, 18, 51] attributing to the underperformance of models in the real world which were trained on the synthetic datasets. However, RESIDE provides a larger number of pairs which helps to easily train the models and prevents overfitting. Table 4.1 details the list of other datasets which are for the single image dehazing task.

Dataset Name	Year	Nature	Number of Image Pairs	Haze generation Method	Description	Has Depthmap?
D-Hazy [52]	2016	Augmented	1449	Koschmieder's Model	The data is based out of real world images instead of generated ones. The depth map of the images is used to construct the hazy images. Since there is a sufficient amount of information, it is suitable to train. It is built on the Middlebury [50] and NYU Depth-v2 [49] datasets.	Yes
FRIDA [18]	2010	Virtual	90	Koschmieder's Model	Simulated Images for evaluating contrast restoration and visibility. It focuses on urban road scenes. Has images of ground truth, uniform fog, heterogeneous fog, cloudy fog, and cloudy heterogeneous fog.	Yes
FRIDA2 [51]	2012	Virtual	330	Koschmieder's Model	Simulated Images for evaluating contrast restoration and visibility. It focuses on urban road scenes. Has images of ground truth, uniform fog, heterogeneous fog, cloudy fog, and cloudy heterogeneous fog.	Yes
FROSI [9]	2014	Virtual	1620	ASM	It contains image pair of clear and hazy. The images are generated using a 3D modelling software. The road signs are placed at various distances and the hazy data is generated using the Atmospheric Scattering model to augment images with uniform haze. The visibility is modelled from 50 meters to 400 meters.	No
HazeRD [53]	2017	Augmented	14	Mie Scattering model	The dataset contains the images, and corresponding depth maps. The dataset author uses a Matlab function based on the Mie scattering model to augment hazy images with different intensities. It provides a realistic effect of haze to the existing images.	Yes
RESIDE-RTTS [46]	2019	Augmented	76457	ASM	It consists of both real-world and hazy images which are generated from the NYU Depth-v2 [49] and Middlebury [50] datasets. The metrics referenced here are the Structural Similarity Index Metric (SSIM) and Peak Signal to Noise Ratio (PSNR). It utilises the depth map, and the dehazed image to procure the hazed image. The dataset has 5 subsets, and each subset has its task associated.	Yes
REVIDE [54]	2021	Real World	1698	Haze generator	It is a video dataset aimed at performing dehazing tasks. The hazy scenes are generated by haze generators and a robotic arm is used to navigate the camera such that the illumination is maintained the same across the video. Here the overall goal is to leverage multiple frames of data to perform the dehazing.	No
Waterloo IVC [55]	2015	Augmented	25	Koschmieder's Model	The dataset contains 22 outdoor scenes and 3 indoor scenes. The haze is artificially added to only three of the indoor images and the remaining are from the real world.	No
BeDDE [4]	2019	Real World	832	NA	It is a real-world dataset where the haze free images along with the corresponding hazy images are provided. It is collected from 23 provincial cities in China. The data has images along with their projective transformation which are used to match the scene. The images are classified based on the haze densities.	No

Table 4.1: Additional datasets available for Single Image Dehazing.

4.2 Dataset Preparation

The dataset settings are consistent with the data settings used by the Trident Dehazing Network (TDN) [38]. The models used for the task of dehazing are complex and they need a large number of image pairs to have any aspect of overfitting. The dataset settings are mentioned below in Table 4.2:

Dataset	Type	Training Set	Validation Set	Testing Set
RESIDE-ITS [46]	Indoor	13990	-	-
RESIDE-SOTS [46]	Indoor	-	-	500
I-HAZE [17]	Indoor	20	5	5
O-HAZE [16]	Outdoor	35	5	5
DENSE-HAZE [47]	22 (Indoor), 33 (Outdoor)	45	5	5
NH-HAZE [48]	Outdoor	45	5	5

Table 4.2: Dataset Information used in this Thesis.

4.3 Methodology

With the chosen datasets which contain the hazy and clear image pairs, an end to end convolutional neural network which will be able to comprehend low level, and high level features to produce a haze free output can be used. This research aims to train two encoder-decoder models that can produce a dehazed image as output for a given input hazy image along with specific postprocessing steps. The overall methodology employed involves:

1. Using Discrete Wavelet Transformation (DWT) information in addition to the image data to help the model capture both the spatial and frequency characteristics. This is useful as most of the high frequency and details of

the image are lost due to the repeated convolutions.

2. Creating a model workflow which ensembles outputs from multiple models resulting in small and robust performance to be easily deployed as a module in complex computer vision tasks. This can aid autonomous vehicles to function safely even in low visibility conditions.

3. Training the model on homogeneous and non-homogeneous images for better generalisability. This is achieved by using channel and pixel attention in the model architecture.

4.4 Loss Functions

This section discusses the details of the loss functions used in the training process. The methods are aimed at convergence at both pixel and feature levels.

4.4.1 Mean Squared Error (MSE)

Mean Squared Error (MSE) in the context of image dehazing, is used to find the sum of the averaged squared difference between the clear and the dehazed output from the model. This helps quantify the discrepancy between the images on a pixel level. The model aims to minimise the difference between the pixels produced by the model and the clear image. The mathematical representation of the Mean Squared Error is :

$$\mathbf{MSE} = \frac{1}{\mathbf{mn}} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - \hat{I}(i,j)]^2$$

Here $I(i,j)$ is the clear image, and $\hat{I}(i,j)$ is the output procured from the model.

4.4.2 Perceptual Loss (VGG-16)

Perceptual loss helps evaluate the output from visual and perceptual quality. Rather than pixel based comparison, perceptual loss compares the result utilising the high level features extracted from pretrained models. In this thesis, the perceptual loss uses the VGG-16 pretrained model to compute the loss. The high level features from the model can be meaningful as the pretrained models are trained on large datasets.

The perceptual loss computes the difference (Euclidean distance [1]) between the feature maps of the dehazed image and the clear image. Reducing this loss helps not only reduce the pixel difference between the images but also encourages the retention of the structural and style aspects of the image.

4.4.3 Fast Fourier Transform Loss (FFT)

Fast Fourier Transform (FFT) loss is a loss function based on the Fourier transform. It takes both clear and dehazed images as the input. FFT transforms the image in the spatial domain to the fourier domain (frequency domain) where the difference is computed in the fourier domain focusing on both the amplitude and phase characteristics of the images in the form of L1 loss [2]. It helps the model perform the manipulations on the frequency domain of the image which helps enhance the image quality and reduce the noise. The frequency domain can help extract useful features which can help improve the process of dehazing.

[1] Euclidean distance is the square root of the sum of the squared differences between the data points. $d(\mathbf{F_1}, \mathbf{F_2}) = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (F_{1(}i,j)^2 - F_{2(}i,j)^2}$

[2] It provides the mean of absolute differences in magnitude of pixels between the hazy and clear images. $d(\mathbf{F_1}, \mathbf{F_2}) = \sum_{i=1}^{n} \sum_{j=1}^{n} |(F_{1(}i,j) - F_{2(}i,j)|$

4.5 Experiment Settings

During the entire training process, all the images were resized to (256×256). The model was trained on RGB images with no additional augmentations. The model was optimised using Adam optimiser [56] where the running average of moments are ($\beta 1 = 0.9$, $\beta 2 = 0.999$), with numerical stability term ($\epsilon = 10^{-8}$). This setting can efficiently minimise the loss functions. The training batch size is 32, and the validation and testing batch sizes are 4 and 1 respectively. The model is trained on 15 epochs with learning rate initialised to 10^{-4}. Additionally, to prevent overfitting the training process, the learning rate is reduced upon plateauing of the validation loss, and early stopping is used with a patience of 5. The entire training process is logged using the Weights and Biases framework [1]. The training process uses two Tesla T4 GPUs which are available on Kaggle [2].

4.6 Model Architecture

This section discusses the components present in the model and the postprocessing modules used in the entire model workflow (ClearFlow). Each model is an end to end CNN model with an image enhancement technique where the dehazed image is provided as the input to the model. The workflow adopted two models which are DehazerNet and Discrete Wavelet transform (DWT) with DehazerNet. AOD-Net inspires the model architectures [6] as it is the fastest image dehazing model.

[1] `https://api.wandb.ai/links/hemanthh17/k66xssih`
[2] `https://www.kaggle.com/`

4.6.1 Attention Module

The attention module is adapted from the Knowledge Transfer Dehazing Network (KTDN) [3] and Feature Fusion Attention Network (FFA-Net) [57]. The module was used in the paper to enhance the capturing of relevant features on channel and pixel levels in the intermediate feature maps. It helps improve the performance in cases where the non-homogeneous haze is present and the distribution of haze density is uneven across the pixels in the image. The attention module contains channel and pixel attention blocks with residual connections between them. The skip connection allows for preserving details and passing this information to the deeper layers. The feature map passes initially to the Channel Attention (CA) and then to the Pixel Attention (PA) module.

4.6.1.1 Channel Attention

Channel Attention (CA) is a module which helps focus details present in various channels in the feature map. It is useful to capture finer details in feature maps with many channels. Channel Attention (CA) plays the role of emphasising channels which contain relevant information and suppressing channels which have minimal impact on the outcome of the task. Channel Attention (CA) can help focus on helping the model learn the most relevant information. The overall module has a lesser number of parameters and has less impact on the overall performance of the network. This helps it to be easily integrated with other high end computer vision models. The channel attention module has an average pooling layer which helps the network to aggregate information from each channel, then has two convolution layers with ReLU and Sigmoid activation. The Sigmoid activation allows the network to assign weight to the channels. Figure 4.1 illustrates the workflow of the channel attention module.

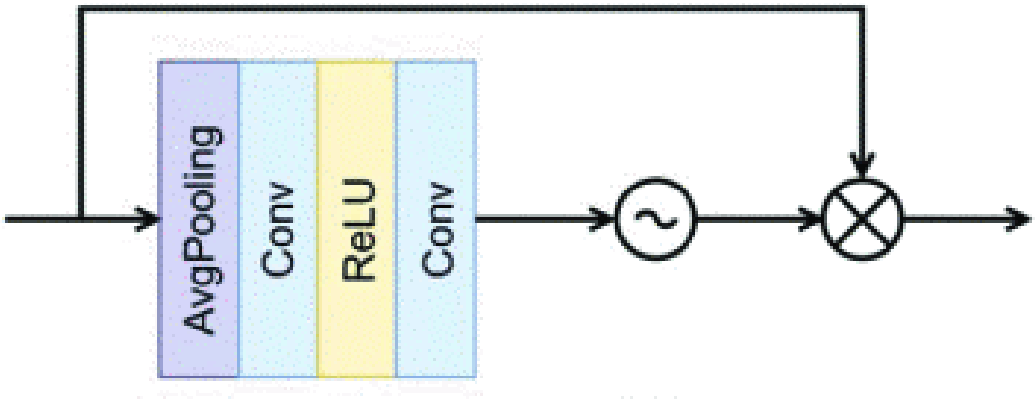

Figure 4.1: Channel Attention Module. Image from [3]

4.6.1.2 Pixel Attention

Pixel Attention (PA) is used to enhance the performance of the model by focusing on the important spatial aspect of the feature map. Pixel Attention (PA) focuses on the specific pixels and allows the model to concentrate on specific regions in the feature map by assigning weightage to pixels. In the context of an image affected by non-homogeneous haze, some pixel regions can be affected more than others and require more emphasis to dehaze. Similar to Channel Attention (CA), Pixel Attention (PA) can help capture the spatial dependencies and can help significantly improve performance in high end image tasks. The module has a lesser number of parameters making it easy to be integrated into other computer vision models. It has two convolution layers with a ReLU and Sigmoid activation. Figure 4.2 illustrates the workflow of the pixel attention module.

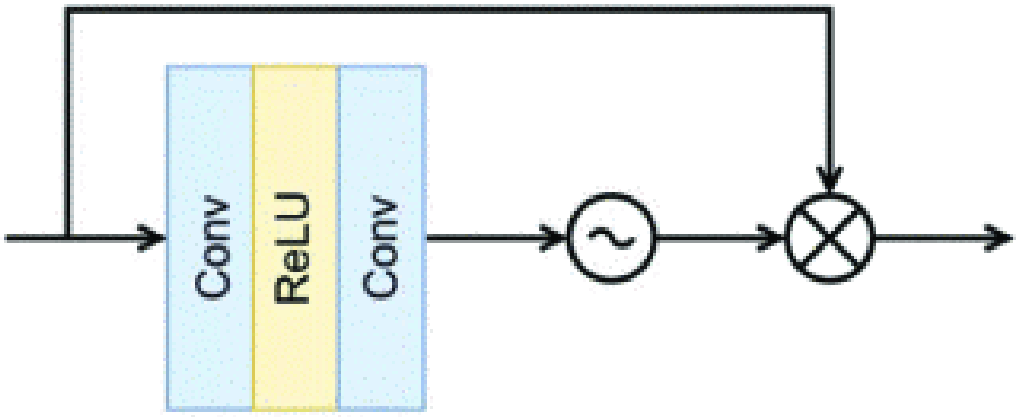

Figure 4.2: Pixel Attention Module. Image from [3]

4.6.1.3 Final Attention Module

The final attention module is inspired by KTDN [3]. The attention module integrates both Channel and Pixel Attention modules sequentially with a skip connection. The feature weightage is learnt from the attention module and the dehazing network pays more attention to the important feature regions which include colour, haze, and texture. The versatility of the overall module along with an emphasis on specific features helps deal with non-homogeneous haze conditions. Figure 4.3 illustrates the workflow of the final attention module.

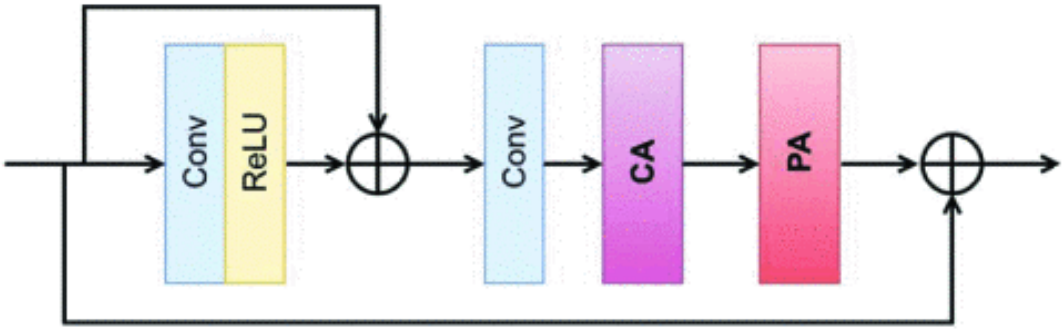

Figure 4.3: Attention Module. Image from [3]

4.6.2 DehazerNet Model

DehazerNet model is an end to end CNN which takes the input hazy image and aims to provide a dehazed image. The model is trained on Mean Squared Error and Perceptual loss. Compared to other dehazing models, it has a smaller architecture helping it provide the output in a shorter time. It uses an attention module which comprises Channel and Pixel attention. Channel attention can effectively identify the important channels in the feature maps obtained in the intermediate CNN layers. Pixel attention emphasises the weightage for specific pixel regions in the feature maps which have higher importance in the final dehazed image.

The model takes in a fixed size input image of size (256×256). There are two attention modules present in the network along with convolutional layers of different sizes. The initial convolutional layers are aimed at extracting the basic features of the dehazed image. The attention modules present are aimed to enhance and refine the features. The final convolutional layer converts the feature map to a dehazed output. The model as a whole aims to predict the dehazed image but is not performed on a pixel to pixel basis but by predicting the transmission map which is used to procure the dehazed image by mathematically processing it with the haze affected image. This step was inspired by the AOD-Net [6] paper.

4.6.3 DWT with DehazerNet Model

The model combines wavelet transformation along with Pixel and Channel Attention to enrich the feature extraction process by helping the model capture both local and global information. It takes a fixed input image size of (256×256). Discrete Wavelet Transform (DWT) along with the existing DehazerNet model is used where the frequency domain characteristics are captured using the

Daubachies wavelet (DB4) and this information is added to intermediate layers to help the model retain the overall structural characteristics of the image. The use of DWT is aimed at mitigating the disadvantage of loss of details in end to end CNN models.

The model contains various convolutional layers of different sizes. Two attention modules are present in the network in the encoder and decoder stages of the network. The result from the wavelet transform layer is used to enrich the feature extraction process from the hazy image. The initial convolution layers extract the basic features from the image, the attention blocks help enhance the features and identify the important pixel areas in the image. The intermediate convolution layers help combine and refine features. The Discrete Wavelet Transform (DWT) step helps decompose the image and procure the frequency domain information. This information is added to the network in intermediate layers to enhance the image details. The final convolution layer converts the feature map to the dehazed image.

4.6.4 Image Postprocessing

The postprocessing modules use traditional denoising techniques to enhance the given input image. The techniques can effectively improve the Peak Signal Noise Ratio (PSNR) metric of the image.

4.6.4.1 Primary Image Postprocessing

The primary image postprocessing is performed on the outputs procured from the two dehazing models. The steps involve applying unsharp masking and Contrast Limited Adaptive Histogram equalisation (CLAHE). Unsharp masking is a technique which is used to improve the sharpness of the image. This is done by performing the following steps.

1. Blurring the image using a Gaussian filter.
2. Subtracting the blurred image from the original image to create a mask.
3. Adding the mask to the original image

CLAHE is aimed at improving the contrast of the image. It uses the histogram equalisation technique to improve the contrast. The overall process is done by the following steps:

1. Divide the image into smaller tiles
2. Apply histogram equalisation [1] on each tile.
3. Perform contrast limiting [2] to reduce the amplification of noise.
4. Perform interpolation of results from each tile to reduce the presence of artefacts.

Figure 4.4 illustrates the workflow of the primary postprocessing step.

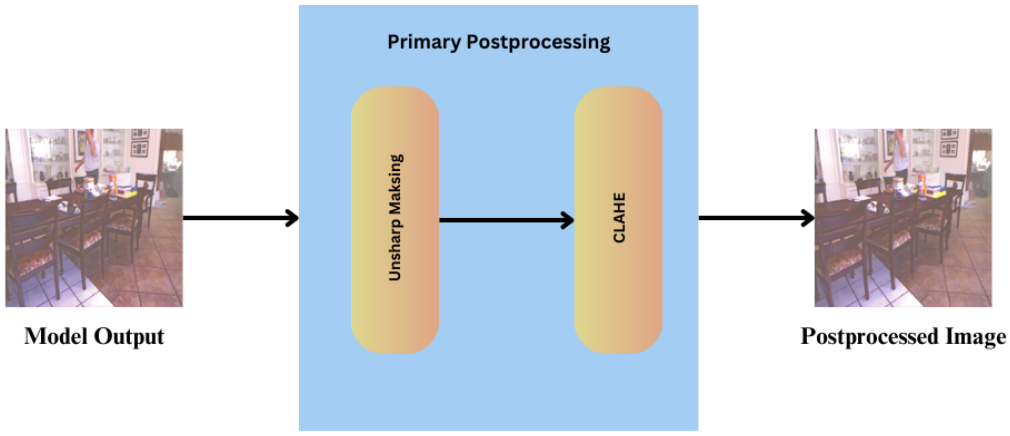

Figure 4.4: Primary Image Preprocessing workflow.

[1] method used to adjust the pixel values distribution to make the image uniform by making the distribution uniform.

[2] limiting the contrast enhancement to reduce the impact of noise. The height of the histogram bins is regulated.

4.6.4.2 Secondary Image Postprocessing

The secondary image postprocessing is performed on the alpha blended image. It consists of the same workflow as the primary preprocessing except that a fast non-local means denoising algorithm is applied before the primary preprocessing steps are involved. The non-local means algorithm is effective in reducing noise by averaging the similar patches in the image. To maintain the colour integrity, the algorithm uses a specific coloured denoising algorithm which considers the chrominance and luminance channels in the image. Figure 4.5 illustrates the workflow of the secondary postprocessing step.

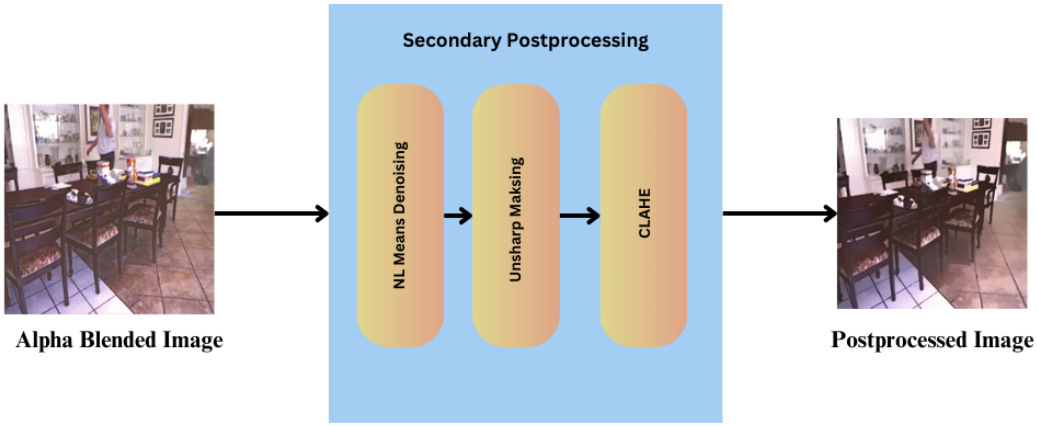

Figure 4.5: Secondary Image Preprocessing workflow.

4.6.5 Enhancer CNN Model

The enhancer CNN is used at the final step of the entire process to enhance the image. The module consists of three convolution layers. The presence of this module helped improve the Peak Signal Noise Ratio (PSNR) and the Structural Similarity Index Metric (SSIM) by a significant margin. The model is trained on the GPU for 20 epochs with a training batch size of 16. The learning rate is initialised to 10^{-4}. The Mean Squared Error (MSE) is the loss function to train the model. Figure 4.6 illustrates the workflow of the Enhancer CNN module.

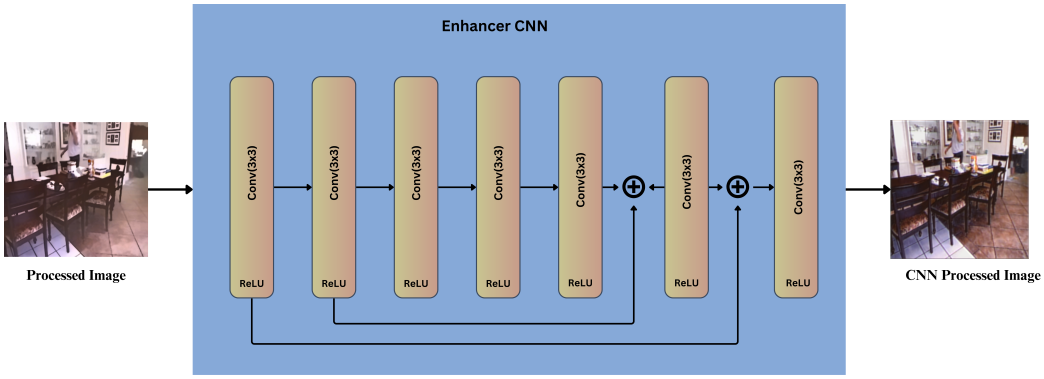

Figure 4.6: Enhancer CNN process.

4.7 ClearFlow: The Dehazing Workflow

The dehazing process involves taking a (256×256) hazy image and it is parallel processed by both the DehazerNet and DWT with DehazerNet models. The results obtained from the models are enhanced using statistical methods. The processed results are alpha blended with a coefficient of 0.6 where 60 percent of the output from DehazerNet contributes to the blended output. A secondary postprocessing step is applied to the blended image. The final step involves passing the input into an enhancer CNN aiming to enhance the output procured from the postprocessing step. Figure 4.7 illustrates the overall process involved.

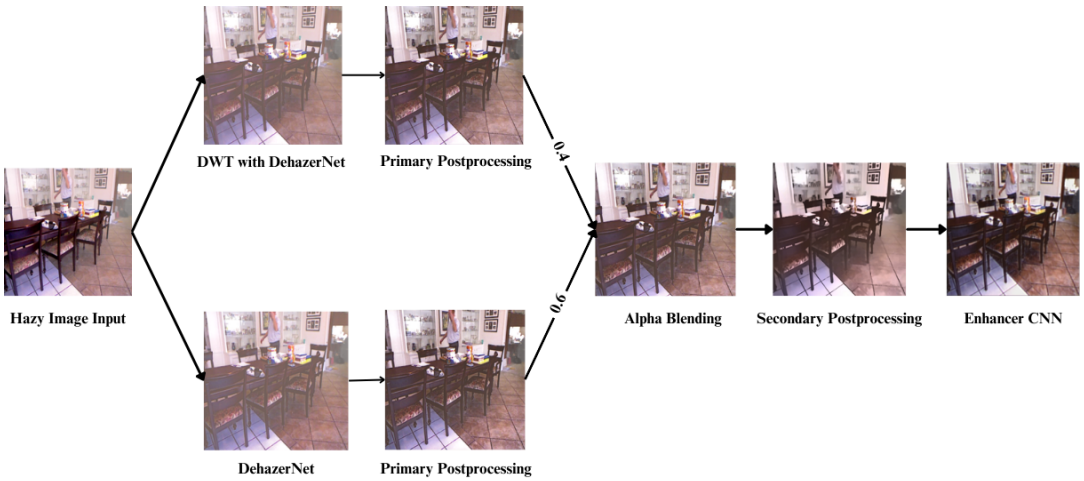

Figure 4.7: Workflow adopted for the dehazing process.

4.8 Summary

This chapter highlighted the existing datasets present which can help with the task of training a model to perform image dehazing. Some of the datasets were used to train the model from the existing datasets, and some of the image pairs were recorded without using any computerised methods to generate the corresponding hazy image for a clear image. However, the RESIDE dataset was used due to the larger amount of image pairs present for training the models. From the existing methodologies, it could be observed that models which were only trained on synthetic image pairs could not perform well in real world scenarios. The chapter also highlighted the data splitting strategy used to create the training, validation and testing datasets which are closely aligned to the Trident Dehazing Network (TDN) paper [38]. The loss functions used for performing all the experiments, and the experiment settings which comprise the details on the learning rate, batch sizes, number of epochs, optimiser, and schedulers were discussed. The various components needed for the entire model workflow (ClearFlow) were mentioned along with details on how they work with necessary architecture details. Finally, the entire workflow diagram and the sequence of these components were highlighted. The details on the trends can be found in the Weights and Biases framework [1].

[1] `https://api.wandb.ai/links/hemanthh17/k66xssih`

Chapter 5

Results

This chapter discusses the qualitative and quantitative results obtained from the model. The models were evaluated at each step of the proposed workflow. At each step, all the output from the model on the test dataset was stored. It was observed that at each step there is an improvement in either of the chosen evaluation metrics.

5.1 Evaluation Metrics

To evaluate the model qualitatively, the performance is measured using the Structural Similarity Index (SSIM) and Peak Signal Noise Ratio (PSNR). These metrics have been predominantly used across the majority of the literature work as it helps in effectively evaluating the performance of a model for low level vision tasks. These metrics are discussed in detail in Section 2.4. While Peak Signal to Noise Ratio (PSNR) majorly focuses on the amount of noise present in the image data, while Structural Similarity Index Metric (SSIM) focuses on the details and other structural aspects of the image.

5.2 Model Evaluation

The models have been evaluated at each stage have been shown in comparison to the state of the art models chosen for this thesis. The state of the art models chosen in this thesis are AOD-Net [6], TDN [38], and Knowledge Transfer Dehazing Network (KTDN) [3]. The quantitative comparison has been done on the testing dataset which was created.

The experiments involved training the models on RGB [1] image pairs and by converting the images to YCbCr [2] image pairs. The YCbCr format was chosen as the image channels represent the chrominance and brightness components of an image which are heavily affected due to haze. Hence, each model was involved in two experiments where one was purely using RGB image pairs, and the other using YCbCr image pairs. The performance of the models in the RGB domain was better than the models trained in the YCbCr domain.

The preliminary experiments involved using ComplexNet, a novel architecture which uses a U-Net based architecture with Discrete Wavelet Transform (DWT) whose output is passed to the model in the intermediate layers. The experiments involved models being trained on only some or all of the loss functions discussed earlier. In the case of ComplexNet, the model was overfitting for the chosen dataset. During the second experiment, the model architecture of AOD-Net [6] was used and was retrained on the dataset chosen for this thesis. The model was not facing the issue of overfitting however, the model could not perform well in both the evaluation metrics compared to the benchmark models considered in this thesis.

The next stage of experiments used the DehazerNet model which is a novel

[1](Red, Green, Blue) where the image is represented based on the colour intensities of red green and blue channels. The intensity is represented in the range of 0 and 255.

[2](Luminance, Blue Chroma difference, Red Chroma difference) where the image is represented using chrominance and luminance components Each data is represented in the range of 0 and 255.

architecture inspired by AOD-Net [6] and contained the attention mechanism essential for handling non-homogeneous haze. The performance of the model significantly improved. In the next experiment to further enhance the quality of output, the frequency domain information was induced as additional information to the model by appending the frequency domain features to the intermediate layer features. This was done through the Discrete Wavelet Transform (DWT). The next experiment was done on DehazerNet with Position Normalisation [1] and Moment Shortcut (MS) [2] was used however, there was no significant improvement for the increase in the complexity of the architecture in the results when compared to the DehazerNet model. From all these experiments it was observed that using Mean Squared Error (MSE) and Perceptual Loss to train the model provided better performance in both Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM) metrics.

The primary postprocessing improved the Peak Signal to Noise Ratio (PSNR) metric by a small margin however it reduced the Structural Similarity Index Metric (SSIM) metric. This is attributed to the loss of fine details of the image due to the usage of statistical denoising algorithms. Considering that DehazerNet and DWT with DehazerNet in the RGB domain were able to perform well after the primary postprocessing stage, The outputs from these models were considered for the alpha blending stage for experimentation. The alpha blended image from the postprocessed images was able to improve both the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Metric (SSIM) metric. The blended image was passed to the secondary image postprocessing step and this improved the Peak Signal to Noise Ratio (PSNR) metric further. After the secondary postprocessing stage, these outputs were provided to the enhancer CNN which

[1]Position Normalisation normalises the spatial dimension for the position of the region in the image.

[2]Moment Shortcut passes the statistical information (mean and variance) of the feature map to the next layers to help generalise the learning process.

provides the final result with improved Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Metric (SSIM) metrics. The existing method has a difference of 2.46 decibels when compared to the best state of the art performing model (TDN [38]) with respect to the Peak Signal to Noise Ratio (PSNR) metric. However, the proposed approach outperforms the benchmark by an improvement of 0.05 (8.3 percent increase) in the Structural Similarity Index Metric (SSIM) metric. This demonstrates the enhanced visual quality and structural integrity in the output dehazed image compared to the existing benchmark. The information on the experiments and the existing benchmark are provided below. Table 5.1 provides the details of the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Metric (SSIM) metric for each model along with the details of the loss function and the domain in which the model was trained. The table clearly shows the better performance of the models in the RGB image domain with respect to the Structural Similarity Index Metric (SSIM) metric.

Model	Loss function	Image Format	PSNR	SSIM
AOD-Net (Pretrained) [6]	MSE	RGB	11.2857	0.5677
TDN (Pretrained) [38]	L1, and FFT Loss	RGB	**21.41**	0.71
KTDN (Pretrained) [3]	L1, Laplace, and Knowledge Transfer Loss	RGB	20.85	0.69
ClearFlow**	-	RGB	-	-
DehazerNet*	MSE and Perceptual	RGB	13.3097	0.6668
DWT with DehazerNet*	MSE and Perceptual	RGB	13.336	0.6609
ComplexNet	MSE, Perceptual, FFT	RGB	10.6614	0.6308
DWT with DehazerNet	MSE, Perceptual, FFT	RGB	12.4295	0.658
DehazerNet	MSE, Perceptual, FFT	RGB	12.5937	0.6547
DehazerNet with PONO and MS	MSE, Perceptual, FFT	RGB	12.5728	0.6387
DehazerNet	MSE and Perceptual	YCbCr	13.0838	0.6156
DWT with DehazerNet	MSE, Perceptual, FFT	YCbCr	12.1813	0.6147
AOD-Net	MSE and Perceptual	YCbCr	13.1084	0.6135
DWT with DehazerNet	MSE and Perceptual	YCbCr	13.1157	0.5967
AOD-Net	MSE	YCbCr	13.3223	0.586
AOD-Net	MSE, Perceptual, FFT	YCbCr	7.5735	0.3065
ComplexNet	MSE, Perceptual, FFT	YCbCr	9.6974	0.5973

Table 5.1: List of Experiments with comparison to state of the art models. (Raw output from the models)

Table 5.2 provides the details of the metrics measured on the output images procured after passing each model output to the primary postprocessing stage. It can be seen the DehazerNet and DWT with DehazerNet models have performed well in this stage and are further considered for the remaining workflow of ClearFlow.

Model	Loss function	Image Format	PSNR (Processed)	SSIM (Processed)
AOD-Net (Pretrained) [6]	MSE	RGB	-	-
TDN (Pretrained) [38]	L1, and FFT Loss	RGB	-	-
KTDN (Pretrained) [3]	L1, Laplace, and Knowledge Transfer Loss	RGB	-	-
ClearFlow** [1]	-	RGB	18.9468	**0.7689**
DehazerNet*	MSE and Perceptual	RGB	13.6845	0.6527
DWT with DehazerNet*	MSE and Perceptual	RGB	13.7185	0.6499
ComplexNet	MSE, Perceptual, FFT	RGB	12.051	0.6172
DWT with DehazerNet	MSE, Perceptual, FFT	RGB	12.8374	0.6075
DehazerNet	MSE, Perceptual, FFT	RGB	12.9368	0.6025
DehazerNet with PONO and MS	MSE, Perceptual, FFT	RGB	12.8899	0.5916
DehazerNet	MSE and Perceptual	YCbCr	13.5392	0.5937
DWT with DehazerNet	MSE, Perceptual, FFT	YCbCr	12.7067	0.5772
AOD-Net	MSE and Perceptual	YCbCr	13.552	0.5939
DWT with DehazerNet	MSE and Perceptual	YCbCr	13.4942	0.5612
AOD-Net	MSE	YCbCr	13.3715	0.5496
AOD-Net	MSE, Perceptual, FFT	YCbCr	7.4146	0.3065
ComplexNet	MSE, Perceptual, FFT	YCbCr	11.2211	0.5791

Table 5.2: List of Experiments with comparison to state of the art models. (Post-processing the Image)

[1] This metric is after the completion of the entire workflow.

Table 5.3 provides the details on the average training time and processing time associated with each model. The data clearly shows that the proposed approach (ClearFlow) involves limited training costs by taking 7 hours and 45 minutes to train the models sequentially unlike the chosen benchmark model which takes around a day to train. However, the model takes 3.74×10^{-2} seconds more than the benchmark to provide the dehazed output.

Model	**Loss function**	**Image Format**	**Average Model Training Time (hours)**	**Average Model Processing Time (seconds)**
AOD-Net (Pretrained) [6]	MSE	RGB	-	0.33009
TDN (Pretrained) [38]	L1, and FFT Loss	RGB	About 1 day	0.64000
KTDN (Pretrained) [3]	L1, Laplace, and Knowledge Transfer Loss	RGB	-	0.30000
ClearFlow[1]**	-	RGB	7:45 [2]	0.67738
DehazerNet*	MSE and Perceptual	RGB	3:51	0.02706
DWT with DehazerNet*	MSE and Perceptual	RGB	3:50	0.04209
ComplexNet	MSE, Perceptual, FFT	RGB	3:14	1.89428
DWT with DehazerNet	MSE, Perceptual, FFT	RGB	3:51	0.04742
DehazerNet	MSE, Perceptual, FFT	RGB	3:42	0.03002
DehazerNet with PONO and MS	MSE, Perceptual, FFT	RGB	3:44	0.05245
DehazerNet	MSE and Perceptual	YCbCr	1:35	0.02581
DWT with DehazerNet	MSE, Perceptual, FFT	YCbCr	4:19	0.04838
AOD-Net	MSE and Perceptual	YCbCr	1:02	**0.01532**
DWT with DehazerNet	MSE and Perceptual	YCbCr	1:42	0.04639
AOD-Net	MSE	YCbCr	0:54	0.01781
AOD-Net	MSE, Perceptual, FFT	YCbCr	1:01	0.01675
ComplexNet	MSE, Perceptual, FFT	YCbCr	2:35	0.78818

Table 5.3: List of Experiments with comparison to state of the art models. (Time)

[1]The time was computed after the completion of the entire workflow. This was done using the inference script.

[2]calculated as sum of all the model training times of the three models used. (3:51+3:50+0:04)

The existing method was applied to the Benchmark Dataset for Dehazing Evaluation (BeDDE) [4]. This dataset was chosen as it contains hazy image pairs of various densities for the same scene from the real world without any synthetic ways to generate haze such as haze generating machines and synthesising haze using the depth maps. The proposed model workflow (ClearFlow) was able to produce a good Structural Similarity Index Metric (SSIM) score till the alpha blending stage. The coefficient of alpha blending at which the model gave the best result is not the same as the existing workflow. The metrics at each stage show that the workflow till the alpha blending stage can provide structurally intact images however, the Peak Signal to Noise Ratio (PSNR) metric shows that the image is affected by noise. The metrics at each stage of the proposed workflow are mentioned below.

Model Stage	PSNR (decibels)	SSIM
DehazerNet	13.1094	0.6109
Postprocessing DehazerNet	13.5047	0.6310
DWT with DehazerNet	13.48490	0.6229
Postprocessing DWT with DehazerNet	13.7045	0.6498
Alpha Blending	**13.8350**	**0.6511**
Postprocessing Alpha Blending	13.6737	0.5990
Enhancer CNN	9.3596	0.4849

Table 5.4: Experiment performed on the BeDDE Dataset [4] . (Raw output from the models)

5.3 Model Predictions

A sample of the test dataset which contains images from each dataset used for the thesis was used to observe the qualitative results. From the results, it can observed that the final image is darker than the clear image. Concerning the performance of the model on various types of haze, it can be observed that the model can perform dehazing efficiently in the case of homogeneous haze irrespective of the location being outdoor or indoor. The model can only remove some portions of the haze from the non homogeneous haze affected images. However, a very

limited amount of haze was removed in dense haze affected images. This can be attributed to the limited number of cases present for the case of non-homogeneus haze used in the training stage. The following model predictions can be found in Figure 5.1 below clearly showcases the output from the various stages of the model workflow (ClearFlow).

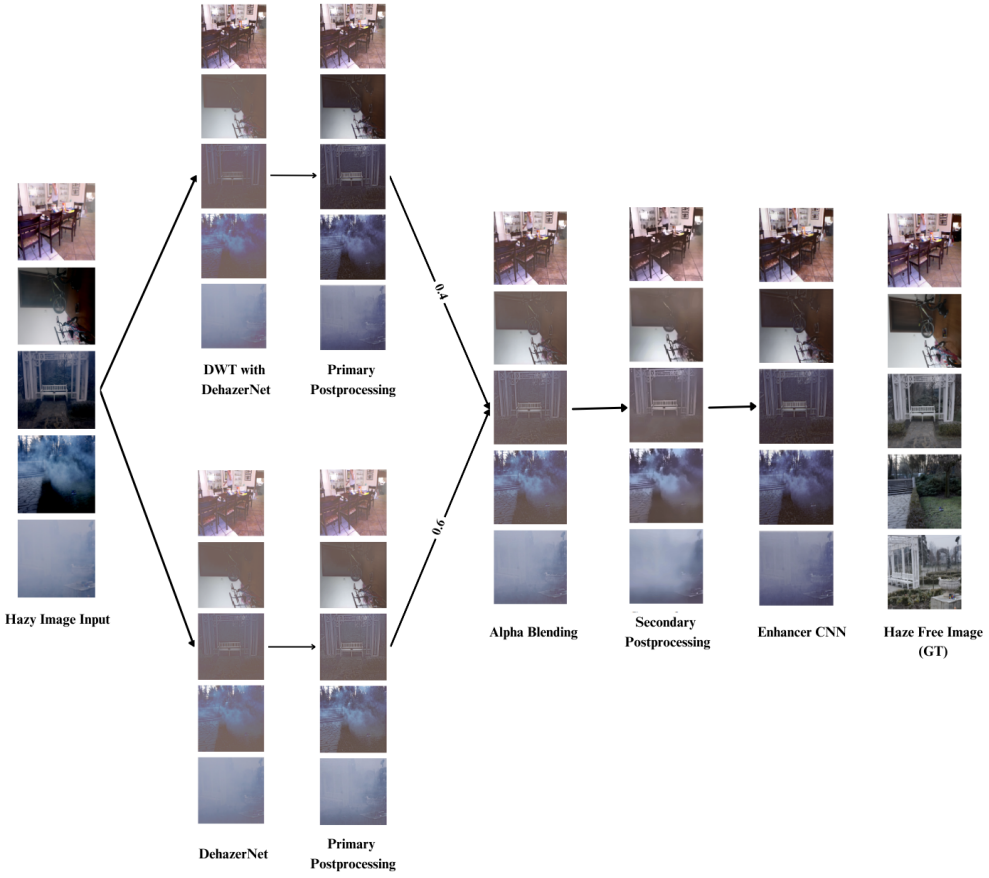

Figure 5.1: Results (row wise) obtained on homogeneous indoor image (RESIDE-SOTS Indoor), homogeneous indoor image (I-HAZE), homogeneous outdoor image (O-HAZE), non-homogeneous outdoor image (NH-HAZE), Dense haze affected outdoor image (DENSE-HAZE).

Chapter 6

Conclusion and Future Work

This thesis focused on developing a novel dehazing workflow **ClearFlow**, to dehaze and enhance the quality of a hazy image. The procedure involves integrating deep learning and traditional denoising methods to achieve the goal.

The primary objective of this research was to design and implement a robust dehazing model that can effectively remove haze from the images regardless of the type and density of haze within a limited time frame to be used as a primary module in a complex computer vision workflow. To achieve this, ClearFlow uses two end to end CNN dehazing models along with postprocessing techniques and an enhancer CNN module.

6.1 Contributions

The following are the overall contributions made:

1. **Development of an end to end dehazing workflow (ClearFlow)**
   A novel dehazing workflow ClearFlow was introduced to dehaze a hazy image. ClearFlow is an amalgam of deep learning and traditional denoising methods to dehaze and enhance the image. The workflow begins with passing the image into two end to end CNN models where the output from the models is passed into the primary postprocessing phase. The outputs from this stage are blended using the alpha blending strategy with 60 percent of the output from the Dehazernet contributed to form the final blended image. The alpha blended result is then passed into the secondary postprocessing stage and then is passed into the enhancer CNN module. The proposed novel workflow demonstrates the ability to handle both homogeneous and non-homogeneous haze conditions effectively.

2. **Performance on Homogeneous and Non-Homogeneous Haze**
   The ClearFlow workflow can effectively handle homogeneous haze conditions and remove a significant amount of haze from images affected by non-homogeneous haze. This shows that ClearFlow can improve the quality of haze affected images irrespective of the type. However, the lesser amount of images belonging to the non-homogeneous category might have impacted the performance.

3. **Benchmark Performance in the SSIM Metric**
   The proposed workflow (ClearFlow) outperformed the benchmark models considered in this thesis with respect to the Structural Similarity Index Metric (SSIM) metric. This was computed with respect to the test data considered to evaluate the performance.

The ClearFlow approach was able to perform relatively well when evaluated on the BeDDE [4] dataset as it contained real-world hazy and haze free image pairs. However, the workflow was able to perform well till the alpha blending stage. The proposed workflow establishes the novelty of the work in this thesis.

6.2 Future Work

The potential areas for future work is as follows:

1. **Collecting Non-Homogeneous Image Data**

   There is a scarcity of the number of non-homogeneous image data pairs to help the model learn to dehaze non-homogeneous image data. The existing limitation might have reduced the capability of the model to dehaze images affected by non-homogeneous haze.

2. **Improving ClearFlow Processing Speed**

   Another future direction is to improve the processing speed of the ClearFlow workflow. The future goal is oriented towards providing robust and accurate dehazed images without compromising on the speed which can help the existing workflow to be easily used for real-time applications.

3. **Task Specific Model Training**

   The aspect of training the dehazing model based on the task and recent works have shown improvement in the performance of task while using dehazing models which were trained for the specific task such as object detection [33].

4. **Trainable Postprocessing Modules**

   The existing approach uses traditional denoising approaches in the postprocessing stage where the components parameters are manually tuned. This

leads to model performance being biased to a particular dataset. The future work would be aimed at making all of the postprocessing modules trainable making the workflow robust for any type of data.

References

[1] P. Parida and N. Bhoi, "Wavelet based transition region extraction for image segmentation," *Future Computing and Informatics Journal*, vol. 2, no. 2, pp. 65–78, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2314728817300508 vi, 7

[2] "Student notes: Convolutional neural networks (cnn) introduction," Mar 2018. [Online]. Available: https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/ vi, 9

[3] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma, "Knowledge transfer dehazing network for nonhomogeneous dehazing," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, pp. 1975–1983, 6 2020. vi, 15, 35, 36, 37, 46, 49, 51, 52

[4] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao, "Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines," *IEEE Transactions on Image Processing*, vol. 29, pp. 6947–6962, 2020. vii, 30, 53, 54, 58, 70

[5] J. Gui, X. Cong, Y. Cao, W. Ren, J. Zhang, J. Zhang, J. Cao, and D. Tao, "A comprehensive survey and taxonomy on single image dehazing based on deep learning," *ACM Computing Surveys*, vol. 55, 7 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3576918 2

[6] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, pp. 4780–4788, 12 2017. 2, 17, 34, 38, 46, 47, 49, 51, 52

[7] S. Nayar and S. Narasimhan, "Vision in bad weather," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2. IEEE, 1999, pp. 820–827 vol.2. 5, 12, 14

[8] S. Narasimhan and S. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003. 5, 12, 14

[9] R. Belaroussi and D. Gruyer, "Impact of reduced visibility from fog on traffic sign detection," *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 1302–1306, 2014. 5, 29, 30

[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc., 2006. 6

[11] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson, 2020. [Online]. Available: http://aima.cs.berkeley.edu/ 8

[12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. 10

[13] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, p. 1–9, aug 2008. [Online]. Available: https://doi.org/10.1145/1360612.1360671 12

[14] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Conference on Computer Vision and Pattern Recognition, 6 2009, pp. 1956–1963. 13

[15] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, pp. 3522–3533, 11 2015. 13

[16] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: A dehazing benchmark with real hazy and haze-free outdoor images," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 867–8678. 13, 28, 31

[17] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-haze: A dehazing benchmark with real hazy and haze-free indoor images," in *Advanced Concepts for Intelligent Vision Systems*, J. Blanc-Talon, D. Helbert, W. Philips, D. Popescu, and P. Scheunders, Eds. Cham: Springer International Publishing, 2018, pp. 620–631. 13, 28, 31

[18] J.-P. Tarel, N. Hautière, A. Cord, D. Gruyer, and H. Halmaoui, "Improved visibility of road scene images under heterogeneous fog," in *2010 IEEE Intelligent Vehicles Symposium*, 2010, pp. 478–485. 13, 29, 30

[19] Z. Li, X. Xiao, and N. Zhang, "Idacc: Image dehazing avoiding color cast using a novel atmospheric scattering model," *IEEE Access*, vol. 12, pp. 70 160–70 169, 2024. 14

[20] M. I. Anwar and A. Khosla, "Vision enhancement through single image fog removal," *Engineering Science and Technology, an International*

Journal, vol. 20, no. 3, pp. 1075–1083, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2215098616305067 14

[21] L. Liu, G. Cheng, and J. Zhu, "Improved single haze removal algorithm based on color attenuation prior," in *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, vol. 2, 2021, pp. 1166–1170. 14

[22] V. Frants, S. Agaian, and K. Panetta, "Qcnn-h: Single-image dehazing using quaternion neural networks," *IEEE Transactions on Cybernetics*, vol. 53, pp. 5448–5458, 9 2023. 15

[23] J. Voight, *Quaternion Algebras*. Springer International Publishing, 2021. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-56694-4 15

[24] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, p. 652–662, Feb 2021. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2019.2938758 16, 23, 25

[25] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Los Alamitos, CA, USA: IEEE Computer Society, jan 2019, pp. 1375–1383. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/WACV.2019.00151 16

[26] Y. Guo, Y. Gao, R. W. Liu, Y. Lu, J. Qu, S. He, and W. Ren, "Scanet: Self-paced semi-curricular attention network for non-homogeneous image de-

hazing," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 6 2023, pp. 1885–1894. 16

[27] H. Sun, B. Li, Z. Dan, W. Hu, B. Du, W. Yang, and J. Wan, "Multi-level feature interaction and efficient non-local information enhanced channel attention for image dehazing," *Neural networks : the official journal of the International Neural Network Society*, vol. 163, pp. 10–27, 6 2023. 17

[28] S. Zhang, F. He, and W. Ren, "Nldn: Non-local dehazing network for dense haze removal," *Neurocomputing*, vol. 410, 2020. 18

[29] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, pp. 5187–5198, 11 2016. 18

[30] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 10 2019, pp. 7313–7322. 19

[31] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 6 2018, pp. 3253–3261. 19

[32] M. Kaur, D. Singh, V. Kumar, U. Rawat, and M. Amoon, "Dsscnet: Deep custom spatial and spectral consistency layer-based dehazing network," *IEEE Access*, vol. 12, 2024. 20

[33] A. R. Rani, Y. Anusha, S. K. Cherishama, and S. V. Laxmi, "Traffic sign detection and recognition using deep learning-based approach with haze removal for autonomous vehicle navigation," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 7, 2024. 20, 58

[34] H.-H. Yang and Y. Fu, "Wavelet u-net and the chromatic adaptation transform for single image dehazing," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 9 2019, pp. 2736–2740. 21

[35] S.-F. Wang, W.-K. Yu, and Y.-X. Li, "Multi-wavelet residual dense convolutional neural network for image denoising," *IEEE Access*, vol. 8, pp. 214 413–214 424, 2020. 21

[36] P. Liu, H. Zhang, W. Lian, and W. Zuo, "Multi-level wavelet convolutional neural networks," *IEEE Access*, vol. 7, pp. 74 973–74 985, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:195225324 21

[37] H. Dong, X. Zhang, Y. Guo, and F. Wang, "Deep multi-scale gabor wavelet network for image restoration," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5 2020, pp. 2028–2032. 21

[38] J. Liu, H. Wu, Y. Xie, Y. Qu, and L. Ma, "Trident dehazing network," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 6 2020, pp. 1732–1741. 22, 31, 44, 46, 48, 49, 51, 52

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 22

[40] J. Wang, C. Li, and S. Xu, "An ensemble multi-scale residual attention network (emra-net) for image dehazing," *Multimedia Tools and Applications*, vol. 80, pp. 29 299–29 319, 8 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11042-021-11081-x 23

[41] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Transactions on Image Processing*, vol. 32, pp. 1927–1941, 2023. 23

[42] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, oct 2021, pp. 9992–10 002. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/ICCV48922.2021.00986 24

[43] H. Li, Y. Zhang, J. Liu, and Y. Ma, "Gtmnet: a vision transformer with guided transmission map for single remote sensing image dehazing," *Scientific Reports*, vol. 13, p. 9222, 6 2023. 24

[44] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang, "Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 6 2021, pp. 203–212. 25

[45] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6 2019, pp. 8152–8160. 25

[46] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, pp. 492–505, 1 2019. 28, 30, 31

[47] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images," in *2019*

IEEE International Conference on Image Processing (ICIP). IEEE, 9 2019, pp. 1014–1018. 28, 31

[48] C. O. Ancuti, C. Ancuti, and R. Timofte, "Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 6 2020, pp. 1798–1805. 28, 31

[49] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7576 LNCS, pp. 746–760, 2012. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-33715-4_54 29, 30

[50] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, *High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth*. Springer International Publishing, 2014, p. 31–42. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11752-2_3 29, 30

[51] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, vol. 4, pp. 6–20, 4 2012. 29, 30

[52] C. Ancuti, C. O. Ancuti, and C. D. Vleeschouwer, "D-hazy: A dataset to evaluate quantitatively dehazing algorithms," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 9 2016, pp. 2226–2230. 30

[53] Y. Zhang, L. Ding, and G. Sharma, "Hazerd: An outdoor scene dataset and benchmark for single image dehazing," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 9 2017, pp. 3205–3209. 30

[54] X. Zhang, H. Dong, J. Pan, C. Zhu, Y. Tai, C. Wang, J. Li, F. Huang, and F. Wang, "Learning to restore hazy video: A new real-world dataset and a new method," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6 2021, pp. 9235–9244. 30

[55] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 9 2015, pp. 3600–3604. 30

[56] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014. [Online]. Available: https://arxiv.org/abs/1412.6980v9 34

[57] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11 908–11 915, Apr. 2020. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/6865 35

Appendix A

Project Repository

The project repository for the thesis can be found on GitHub [1]. The scripts and their purpose are listed below:

- The **data-creation-for-dehazing.ipynb** notebook is used to create the Comma Separated Values (CSV) files for training, validation, and testing dataset.
- The **gpu-information.ipynb** provides detailed information on the GPU used for the training.
- The **DWT with DehazerNet Training Script.ipynb** and **DehazerNet Training Script.ipynb** notebooks are used to train the DehazerNet and DWT with DehazerNet models respectively.
- The **weights** folder consists of the weights of the two models used in the overall process. The weights of other models can be found in Kaggle [2]
- The **data-prep-enhancer-cnn.ipynb** is used to prepare the training data for training the enhancer CNN module. The training data and the output

[1] `https://github.com/hemanthh17/CT5129-Thesis-Image-Dehazing`
[2] `https://www.kaggle.com/datasets/hemanthhari/dehazing-models-ct5129`

post the secondary postprocessing are stored along with the corresponding image pairs.

- The **enhancer-cnn-training and final inference-script.ipynb** is used to train the enhancer CNN model and evaluate the final output of the workflow on the test dataset. This model is used in the final step of the entire process.

- The **model-testing-script.ipynb** is used to test multiple models and report the quantitative performance of each model.

- The **inference-script.ipynb** is used for inference purposes where for a given hazy image path, the output image from the entire process will be saved in a folder.

- The **bedde-inference-script.ipynb** is used for evaluating the workflow's performance on the Benchmark Dataset for Dehazing Evaluation (BeDDE) [4] dataset where for a given hazy image path from the dataset, the output image from the entire process will be saved in a folder and the evaluations will be performed.