SINGLE IMAGE DEHAZING FOR REAL COLORED-HAZE IMAGES

Single image dehazing for real colored-haze images

By

Zijun Wu,

A Thesis submitted to

the Department of Electrical and Computer Engineering

and the School of Graduate Studies

of McMaster University

In partial fulfilment of the requirements

for the Degree of Master of Applied Science

September, 2021

McMaster

Master of Applied Science (2021)

Hamilton, Ontario (Electrical Engineering)

TITLE: Single image dehazing for real colored-haze images

AUTHOR: Zijun Wu (McMaster)

SUPERVISOR: Dr. Jun Chen

NUMBER OF PAGES: x, 49

Abstract

Since the problem of image dehazing is of great significance and challenging, it has attracted many researchers for a long time. Many excellent dehazing algorithms have emerged one after another. However, few works focus on dehazing tasks with color deviations. Existing methods do not perform as well as expected on real image dehazing, especially on images with color deviations. The proposed method mainly focuses on this task and achieves competitive performance. In this paper, a two step network is proposed, which consists of the Colour Balance Module and the Dehazing Module. Hazy images are first fed into the Colour Balance Module to generate color corrected hazy images, which are then sent to the pretrained dehazing net to get the real color corrected haze-free results. A colored haze image database is also constructed. Experiments and comparisons are implemented to show the effectiveness of the proposed method, not only on real colored-haze images, but also on other dehazing tasks.

To my dear parents, for their love and support

Acknowledgements

Firstly, I would like to express my sincerest gratitude to my supervisor Dr. Jun Chen, not only for his continuous support and help but also for his his patience and guidance during my whole master program.

Furthermore, I would like to thank Dr. Shirani, Shahram and Dr. Xiaolin Wu not only for being members of my defense committee but also for reviewing my thesis, and providing technical comments.

In addition, I am also grateful for Mr. Xiaohong Liu, who gave me a lot of advice and help during many projects. Last but not least, I am also grateful to my parents and boyfriend for their love and encouragement, and I wish to thank my friend Yanning Li for her company.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Problem Statement

1.1 Introduction and Problem Statement

In the current situation where hazy weather often occurs, people need to extract as much information from hazy images as possible. Then the image dehazing technology is derived based on image processing. This technology is is a result of interdisciplinary research involving multiple knowledge categories such as atmospheric imaging, physics and mathematics.

Image is an objective mapping of the natural landscape through a certain system, and it is the visual basis for humans to perceive the world.The visual quality of the image will have an important impact on the analysis, understanding and use of the target for human. Images captured by sensors are often not directly used for processing and applications such as target recognition. The sensor imaging process will be affected by many uncertain factors, such as image defocusing, camera movement, atmospheric turbulence, etc., which will cause different degrees of blur in the acquired image. Severe weather, including haze weather, is the main interference to outdoor imaging. Some

preprocessing operations are required to improve the quality and effects. This process is image processing a technology to perform various operations on digital images through computers. In the past 30 years, with the rapid development of electronic technology, computer science and other disciplines, digital image processing technology has made remarkable progress, which has greatly changed people's lives and brought convenience to a wide range of fields such as medicine, industry, and military.[16].

As mentioned above, existing image acquisition equipment is very sensitive to the interference of the external environment. When there is heavy haze, particles suspended in the atmosphere will reflect and scatter the radiation of the scene. This phenomenon will cause the contrast of the image to decrease and cause the loss of edge and texture information. The visibility of the image is also significantly reduced. Thus, images taken under such weather conditions have low contrast, small dynamic range, and unclear image details. At the same time, the color will be partially distorted. Such an image will affect the outdoor vision system. The functions and effects of target recognition and tracking, monitoring and so on are limited. Fig.1.1 gives examples of the effect of haze for images. Taking the highway monitoring system as an example, the visibility of the road is greatly reduced, causing visual obstacles to drivers, and also causing great difficulties to the outdoor traffic monitoring system. Therefore, to improve the availability of data, there is an urgent practical need to study how to effectively reconstruct the original clear image from the image obtained in the haze environment, and to improve the robustness of the visual system.

In the hazy weather, when the reflected light of the scene object passes through the atmosphere, the absorption and scattering of the scene radiation will happen. Part of it is absorbed along the way, resulting in attenuation of the final light intensity. The light

Hazy Image Dehazed Image

FIGURE 1.1: Example of Clear and Hazy Images

in the atmospheric environment will be scattered, which causes a negative confounding effect on the original scene radiation. The attenuation of the scene radiation and the scattering of atmospheric light will gradually accumulate as the propagation distance increases. This kind of phenomenon can be described by the Atmosphere Scattering Model(ASM)[26][6]:

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

where $I(x)$ and $J(x)$ denote the hazy image and clear image respectively. x indicates the coordinate position of the image pixel while A is the ambient light and $t(x)$ represents the transmittance.

Besides, $t(x)$ is closely related to depth information:

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

where, β is the attenuation parameter and $d(x)$ is the scene depth.

Restoring the original image from the haze map is a very challenging task, as it requires recovering unknown atmospheric ambient light, scene depth, and haze-free images based only on a given single input haze map. Although the research started relatively late about two decades ago, it has become one of the important research hotspots in related fields[43]. At the same time, due to the randomness and complexity of weather changes faced by dehazing technology, traditional image enhancement algorithms, such as contrast stretching and histogram equalization, often have limited processing effects and are not sufficient to deal with the image quality degradation caused by physical reasons. Almost all the effective methods currently known have certain limitations and require continuous improvement[45].

Except[8], few researchers focus on dehazing tasks with color deviations, although the existing dehazing methods performed badly in this task (as is shown in chapter4). This work mainly deals with hazy images with color deviations due to light and other reasons. Combined with application scenarios, a new dehazing scheme is developed and verified through a large number of experiments, which greatly improved the visual effect of dehazing pictures.

1.2 Thesis Structure

In this thesis, Chapter 2 will make a quick review of existing dehazing models and related works, while Chapter 3 will illustrate the details of the proposed color correction based dehazing model. In addition, Chapter 4 will analyze of the proposed network and make comparisons of this work and state-of-the-art methods and Chapter 5 is a conclusion.

Chapter 2

Related Works

Since the problem of image dehazing is of great significance and challenging, it has attracted many researchers for a long time. Many excellent dehazing algorithms have emerged one after another. According to the principle of the dehazing algorithms, this chapter will divide them into two categories: physical models based dehazing algorithms and dehazing algorithms based on neural network. They are currently the mainstream solution in this field.

2.1 Physical Model-based Dehazing Algorithm

The degradation process of the haze map ais captured by the atmospheric model as shown in Eq.1.1 shows, the atmospheric scattering model. Based on this model, many excellent dehazing algorithms have been proposed. Since the model is an underdetermined equation, researchers need to add some reasonable priori constraints, to get the unique solution for the model. So algorithms based on physical models are often referred to as priori algorithm. The most critical step in the dehazing algorithm based on the physical model is the atmospheric ambient light and transmittance evaluation, which is directly

related to the final dehazing effect, and the evaluation of these two physical quantities is an open question.

As it is difficult to calculate an accurate transmission map and atmospheric ambient light with a single input hazy map, some researchers attempt to improve the visual quality of the image through color and contrast enhancement to avoid this model. The histogram equalization algorithm is used to improve the contrast of hazy images. Then adaptive histogram equalization[37] and partially overlapping sub-blocks[18] are proposed respectively. To improve the visual quality, a hazy image enhancement algorithm based on multi-image fusion[1][12] is proposed by Ancuti et al. Retinex is a color vision model that simulates human vision under different lighting conditions. Based on this model, Adrian et al[1] proposed an effective haze map enhancement algorithm. However, These methods do not consider the essential principle of haze map degradation. Because of this, the enhancement effect is limited and the robustness is often poor.

In recent years, researchers have proposed many excellent methods for evaluating atmospheric light and transmittance. Tan et al [38] observed that the contrast of a haze-free image was higher than that of a degraded image in a hazy day. The air light, whose depth of field is positively correlated, is usually globally smooth. Based on this, a cost function based on Markov random field is proposed[38]. The haze-free image can be obtained by optimizing this cost function. However, the image restored by this method will show Halo effects in discontinuous places.

He et al. [13] proposed the well-known dark channel algorithm in 2009, which is based on a statistical prior. That is, in a haze-free image, the minimum channel value of the local regions of the image is close to 0. This channel is called the dark channel.

While due to the shadow of air light, the dark channel of a hazy image does not approach 0. According to this clue, the degree of concentration of haze can be judged by the dark channel value of the haze map so as to obtain the atmospheric light and transmission map. For a haze-free image J, its dark channel mathematical expression is as follows[13]:

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in (r,g,b)} J^c(y)) \tag{2.1}$$

where J^c represents one colour channel of J, Ω is a local area centered on x. Then we get the dark channel image. The dark channel prior can be expressed as:

$$J^{dark}(x) \to 0. \tag{2.2}$$

Then clear images can be restored by the following equation:

$$J(x) = \frac{I(x) - A}{t(x)} + A \tag{2.3}$$

where the brightest pixels of a hazy image can be considered as A and t can be expressed as:

$$t = 1 - \min_{\Omega}(\min_{c} \frac{I^c}{A^c}). \tag{2.4}$$

In view of the halo problem that He's dark channel algorithm encounters in the transmission map evaluation, based on local block evaluation of transmission map, Berman et al. [3][4] proposed a method for globaly evaluating the transmission map, which is based on a new non-local prior (Non-local Prior). That is, in a haze-free image, all the colors can be quantified by hundreds of colors without affecting the visual perception, and these colors can be gathered in clusters in the RGB space. Pixels of the same color

at different positions (that is, different depths of field) will also cluster into the same color group. This prior is global and does not rely on local tiles. And in the hazy image, these bright color groups will form a line in the RGB space due to the difference in the emission rate related to the depth, that is called the haze line. The two ends of the haze line are the original scene light and the ambient light. However, when the atmospheric light is extreme, this method will fail due to the unsuccessful detection of the haze line. Fattal[10] found that the local pixels in natural images without haze are generally distributed in one dimension in the RGB space. That is called the color line. The color line of the haze map will have an offset, so the scene transmission map can be solved by this a prior.

2.2 CNN Based Dehazing Algorithm

Convolutional neural network algorithm has a strong learning ability, it can effectively capture potential mapping relationship between input signal and output signal. Therefore, researchers have successively jumped out of traditional dehazing methods based on hypothetical priors. In recent years, there have been many dehazing methods based on deep learning. According to whether the dehazing framework relies on the classic atmospheric scattering model, the algorithm can be roughly divided into two categories, two-stage methods and single-stage dehazing algorithms(also known as end-to-end methods).

For two-stage neural network dehazing, although the algorithm introduces a neural network structure, it still strictly follows the atmospheric scattering model, and the entire dehazing framework is focused on the estimation of the scene transmission map t and the atmospheric ambient light A by neural networks and the prior knowledge. Then obtain the haze-free image according to the atmospheric scattering model[6]. In this

kind of algorithm framework, the neural network is only used as an intermediate tool for obtaining transmission maps or atmospheric light.

The single-stage dehazing algorithm is completely separated from the physical model and directly simulated by the neural network. By combining the data to obtain the mapping function between hazy images and haze-free images, and finally realize the haze removal through the mapping function.

The early applications of neural network algorithms in the field of dehazing are all using the two-stage framework mentioned above. In 2016, Cai et al. [5] proposed a network called DehazeNet to estimate the transmission map of the scene. A novel nonlinear activation function BReLU (Bilateral Rectified Linear Unit) is proposed. The author uses Convolution Neural Network (CNN) to get transmission image information by non-linear regression, and the global atmospheric light is obtained by assuming that the atmospheric light is the brightest point of the image. Finally, a haze-free image is obtained. DehazeNet significantly improves the effect of the dehazing algorithm, compared with traditional prior-based algorithms.

In the same year, Ren et al. [30] proposed the MSCNN algorithm at the at the 2018 European Computer Vision Conference (ECCV). Similar to DehazeNet, it is also a convolution neural network based algorithm. The algorithm is also to find the mapping between the input hazy image and the corresponding transmission map, but the difference is that MSCNN passes multi-scale network structure, constructs coarse sub-networks and fine sub-networks, and realizes evaluation and refinement of rough transmission map at the same time. The transmission map obtained by DehazeNet needs to be additionally smoothed by guided filtering. Therefore, the final transmission map obtained by

MSCNN effectively suppresses halo artifacts on the dehazing result.

At the 2017 International Computer Vision Conference (ICCV), Li et al. [20] proposed an AOD-Net, which was the first to achieve end-to-end dehazing. It leads the dehazing algorithm framework from two stages to a single stage. The author believes that evaluating the transmission map and atmospheric light separately, then substituting the two into the atmospheric scattering model for dehazing directly will cause errors. Because the entire process will be affected by the errors of these three steps at the same time, the author integrates these three steps in the neural network structure. It should be noted that the structure of AOD-Net is is different from the single stage dehazing. The design does not completely abandon the physical scattering model. As shown in Eq.2.5, The atmospheric scattering model is deformed by the following formula in this paper:

$$J(x) = K(x)I(x) - K(x) + b \tag{2.5}$$

with,

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1} \tag{2.6}$$

where $I(x)$ and $J(x)$ denote the hazy image and clear image respectively. x indicates the coordinate position of the image pixel while A is the ambient light, $t(x)$ represents the transmittance and b is the bias.

Thus, the transmittance and atmospheric light are combined into an intermediate variable K. The author tests the algorithm in vehicle inspection task in a hazy environment. The results show that AOD-Net as a pre-processing module can significantly improve the performance of subsequent detection algorithms.

Since AOD-Net, single-stage dehazing algorithms have sprung up. Zhang et al. in DCPDN[42](Densely Connected Pyramid Dehazing Network) constructed a densely connected feature pyramid decoder sub-network to estimate the transmittance, and at the same time used a U-shaped sub-network to evaluate the atmospheric light. Finally they use a joint discriminator to conduct adversarial learning for the dehazing network to yield the dehazing results. Ren et al. [31] combined image fusion strategy with deep network for the first time. In the dehazing task, they first perform white balance on the input image, then synthesize the input image and apply gamma correction. By put them into a decoder respectively, the confidence map can be calculated. By multiplying the input images and their confidence maps, final haze-free images are obtained.

The GCANet, proposed by Chen et al. [7], uses the latest smooth dilated convolution to eliminate the halo effect generated by the convolution process, and leverages gated subnets to integrate features from different scales. It achieves good performance in dehazing and rain removal tasks at the same time.

As Generative Adversarial Networks (GANs) perform good on various computer vision tasks, many GAN-based dehazing algorithms have been proposed [22][28]. Besides, the GridDehazeNet[23] showed excellent performance on synthetic image dehazing, while DANet[33] focused on improving the quality of real haze image dehazing. Details will be introduced as follows.

2.2.1 Enhanced Pix2pix Dehazing Network

EPDN[29] includes a multi-scale generator, a multi-scale discriminator and an enhancer. The discriminator supervises the multi-scale generator on the coarse scale to generate

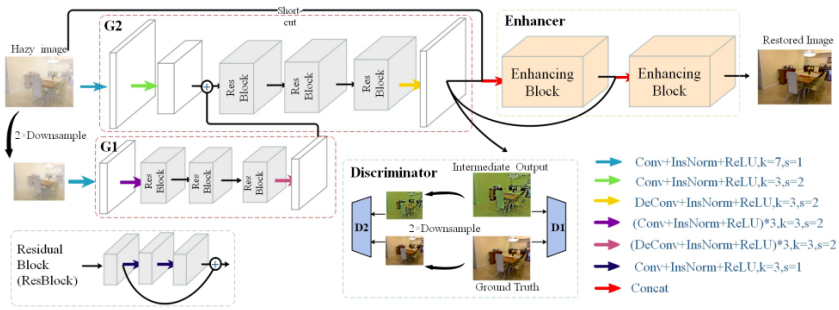

FIGURE 2.1: Enhanced Pix2pix Dehazing Network[29]

haze-free images, followed by the enhancer on the fine scale to do the refinement. The network structure is shown in Fig.2.1.

The multi-resolution generator includes a global generator G_1 and a local generator G_2. G_1 generates images on small scales while G_2 generates image on the original size. Combine G_2 and G_1, restored haze-free can be obtained and will be fed into the next enhancer step.

Similarly, the multi-scale discriminator contains two discriminators on the aforementioned two different scales D_1 and D_2. D_2 is used to supervise generator to generate a real image, while D_1 will help to restore details.

The enhancer module introduces the pyramid pooling module in DCPDN[42]. This module can integrate feature details of different scales into the final fine result.

2.2.2 GridDehazeNet

The author proposed an end-to-end multi-scale network with a preprocessing model and a post image refinement block[23]. This network can efficiently exchange information

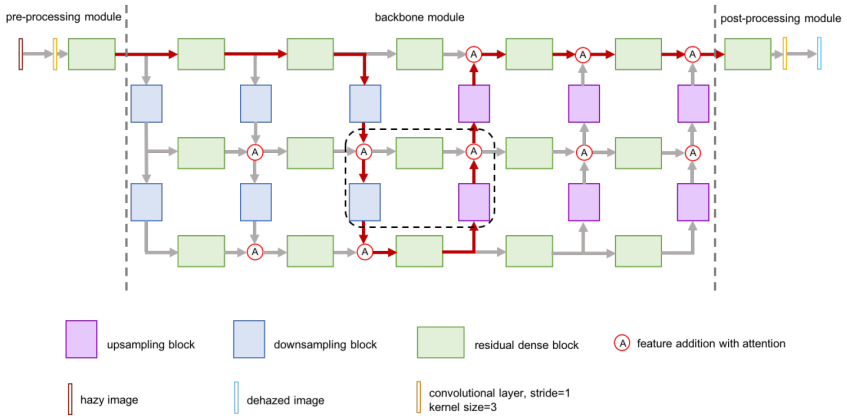

FIGURE 2.2: GridDehazeNet[23]

of different scales, thereby effectively alleviating the bottleneck problem of multi-scale estimation.

The author stated that grid network has obvious advantages over encoder-decoder network. Because encoder-decoder network is susceptible to bottleneck effect. The method also uses the channel-level attention mechanism, which makes the exchange and aggregation of information more flexible. Experiments show that the attention mechanism can better deal with the preprocessing module.

The overall network architecture of this method is shown in Fig.2.2 which is a three-row and six-column structure and is based on the GridNet[11] network architecture (GridNet was originally used for semantic segmentation). Each layer is composed of five residual dense blocks, and each column in the figure can be seen as a bridge for multi-scale operations (implemented through up-sampling and down-sampling).

The image output directly from backbone may have artifacts. Therefore, a post-processing module is proposed to process the image.

2.2.3 DANet

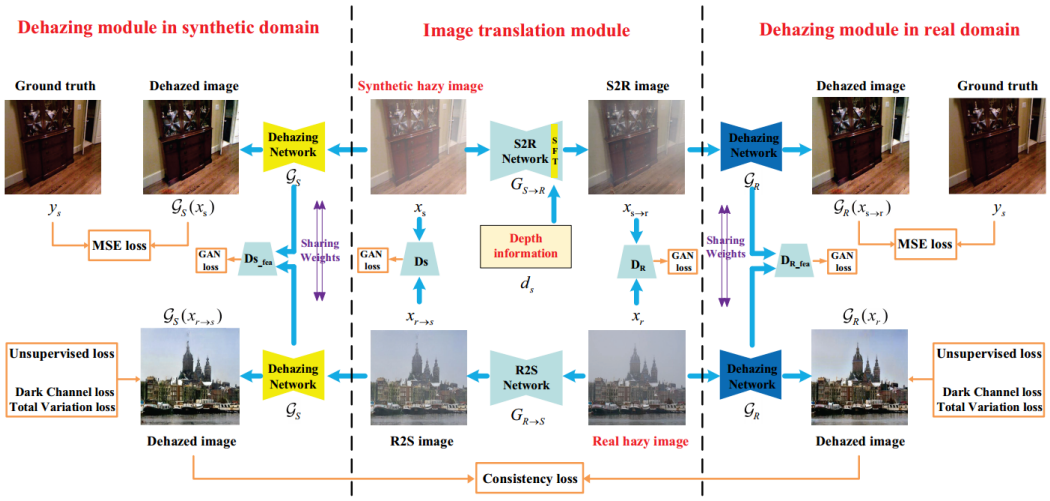

FIGURE 2.3: DANet[33]

The author uses the advantages of CycleGAN[44] structure to adapt the real hazy image to the dehazing model trained on synthetic data. The framework[33] consists of two parts, an image translation module and two domain-related dehazing modules (one for the synthetic domain and the other for the real domain). In order to reduce domain gaps, this method first uses a two-way image translation network to convert images from one domain to another. Since the hazy image inconsistently depends on the depth of the scene, the author incorporates the depth information into the conversion network to guide the conversion into a real hazy image.

Taking synthetic to real subnet as an example, the dehazing network related to synthetic domain takes the image from synthetic domain (including the original synthetic image and the translated fake synthetic image) as input to perform image dehazing. In addition, consistency loss is used to ensure the two-way dehazing networks produce consistent results. At this training stage, in order to further improve the generalization

ability of the network in the real domain, real hazy images are incorporate into the training. Dark channel priors and image gradient smoothing are used as the regularizers to ensure the dehazing results of real hazy images to have some characteristics of clear images. By training the image conversion network and the dehazing network at the same time, they can improve on each other.

The feature-level adaptation method is employed to adjust the feature distribution between the source domain and the target domain by minimizing the maximum average difference or adopting an adversarial learning strategy in the feature space. Another research area focuses on pixel-level adaptation. This method solve the problem of domain shift by applying inter-image conversion learning or style conversion methods to increase the database in the target domain.

Chapter 3

Proposed Method

The proposed method consists of two parts: the colour correction part and image dehazing part.

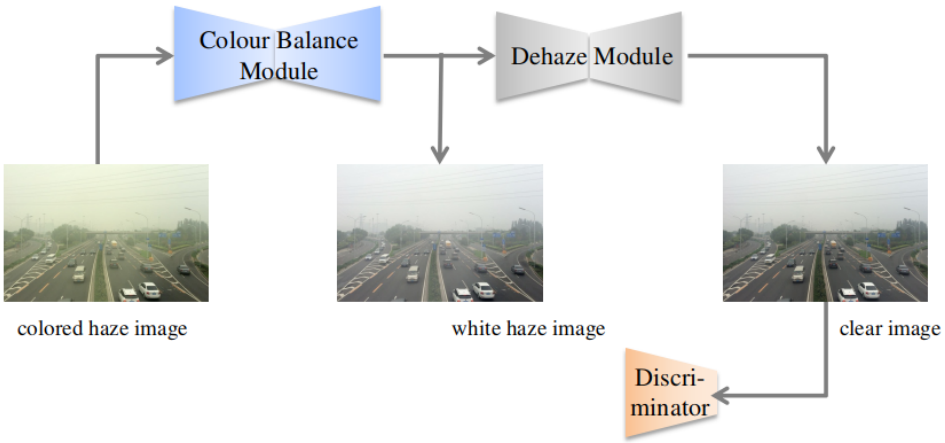

FIGURE 3.1: Net Structure of Proposed Method

As illustrated in 3.1, we train the colour balance module and the dehazing module respectively, then we use the two pretrained network to generate the final colour corrected dehazed image.

3.1 Colour Correction Module

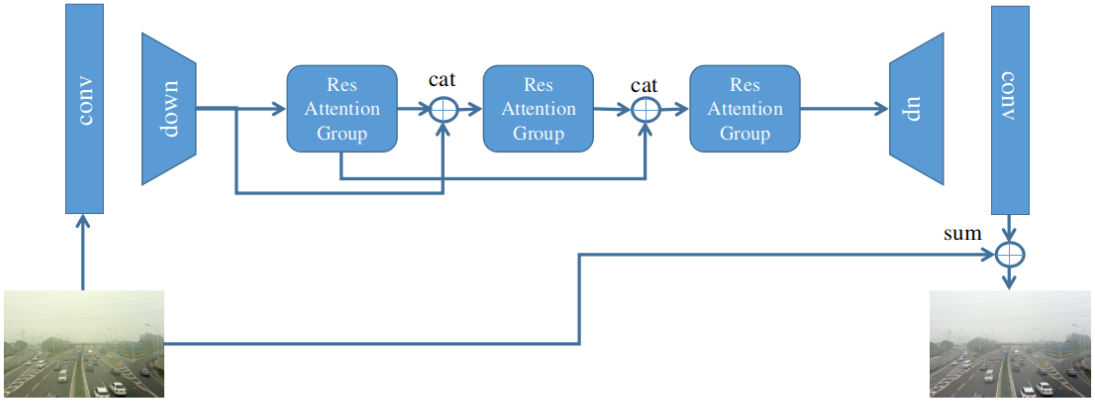

FIGURE 3.2: Network Structure of Colour Correction Module

As illustrated in (Fig.3.2), a designed end to end network is used to convert colored haze image into white haze image. Table.3.1 gives some examples of real colored haze images and the corresponding results of color balanced network.

Colored	Color Corrected	Colored	Color Corrected

TABLE 3.1: Results of Color Balance Module

3.1.1 Network Structure

As is shown in Fig.3.2, a colored hazy image first enters a 5×5 convolution layer followed by a ReLU activation function to extract 16-channel features. Then the features are sent to

two residual downsample blocks to get higher dimensional feature maps. These features will be refined by 3 residual attention groups. Skip-connections are applied to avoid the loss of essential detailed information and to keep the training progress more stable. After upsampling, residual features will be refined by a 3×3 convolution layer. After element-wise sum with the input image, color corrected images will be generated.

Downsample Block

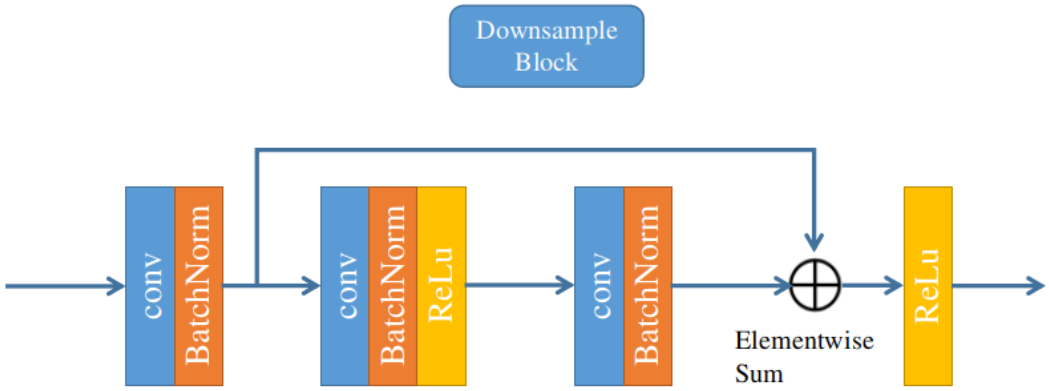

FIGURE 3.3: Downsample Block

The downsample block employs the resnet[14] structure, consisting of 3 convolution layers, all followed by a batch norm layer. The setting for kernel size, stride and padding are [3,2,1], [3,1,1] and [1,2,0] respectively.

Residual Attention Block

In order to flexibly deal with uneven color distribution we use a residual pixel[27] attention block to extract features. The output of the previous layer is fed into two convolution layers with ReLU and sigmoid activation function. Then we use the original input to the element-wise product with the output of attention block to get the enhanced features.

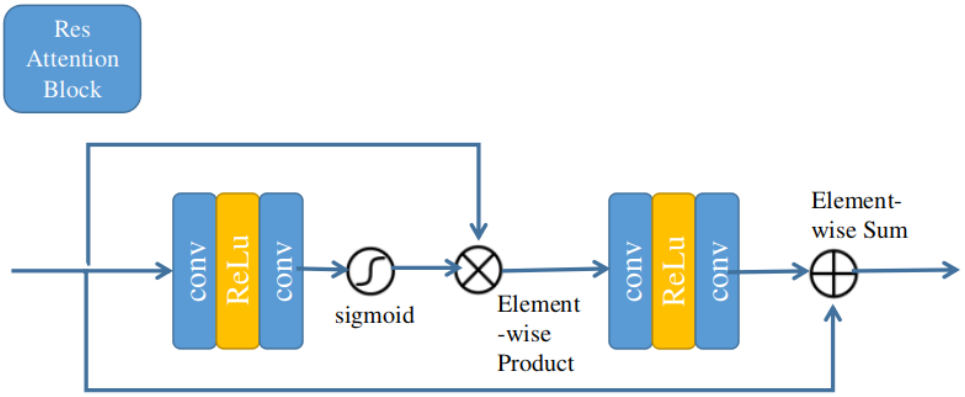

FIGURE 3.4: Res Attention Block

3.1.2 Loss Functions

Smooth L1 Loss

Let I_i^w denotes the value of white haze in the ith color channel, I_i^c represents the value of colored haze in the ith color channel, x is the pixel location. The Smooth L1 Loss can be expressed as:

$$L_{l1} = \frac{1}{N}\sum_{i=1}^{3}\sum_{x=1}^{N}\sigma(I_i^w(x) - I_i^c(x)) \tag{3.1}$$

where,

$$\sigma(e) = \begin{cases} 0.5e^2, & if|e| < 1 \\ |e| - 0.5, & otherwise \end{cases} . \tag{3.2}$$

As is shown in 3.5, this function performs better than $L1$ and $L2$ loss functions. It is smoother than $L1$ loss and is more stable than $L2$ loss. As the derivative of $smoothl1$ and $l2$ with respect to x are:

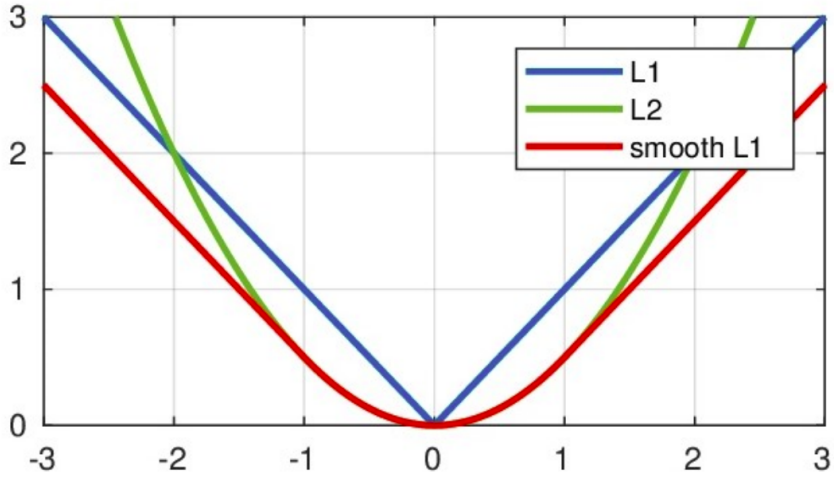

FIGURE 3.5: Comparison of $l1$ loss, $l2$ loss and smooth$l1$ loss

$$\frac{\partial L_2(x)}{\partial x} = 2x,$$
$$\frac{\partial smooth_{l1}(x)}{\partial x} = \begin{cases} x, & if|x| < 1 \\ \pm 1, & otherwise \end{cases}.$$

It is clear that in the early stage of training , when there is a large difference between the predicted value and the ground truth, the gradient of the $L2$ loss function is too large and too sensitive to make the training progress stable.

Perceptual Loss

In order to optimize the visual effect, we need to quantify the difference between features extracted from model results and ground truths. Then the perceptual loss[17] is applied. As illustrated in Eq.3.3, C_l, H_l and W_l are the label, height and width of the lth channels of the image respectively. I_w is the white haze image and I_c is the result of the proposed

network.

$$L_p = \sum_{l=1}^{3} ||\frac{1}{C_l H_l W_l}(\phi_l(I_w) - \phi_l(I_c))||_2^2. \tag{3.3}$$

A pretrained Vgg16[36] is used to extract features of I_w and I_c.

Overall Loss Function for Colour Balance Net

The total loss for the colour balance module is combined by 2 parts, the $smoothl1$ loss L_{l1} and perceptual lossL_p:

$$L_{color} = L_{l1} + \gamma L_p \tag{3.4}$$

where, γ is the weight coefficient, set as 0.04.

3.1.3 Synthetic Colored Haze Image Dataset

Inspired by [8], although there have been many synthetic datasets with real clear image and the corresponding synthetic hazy images, there is no datasets with paired clean images and colored hazy images. Then we followed the atmospheric scattering model[26], based on the RESIDE[21] Outdoor dataset, generating the Synthetic Colored Haze Images. As is shown in the (ASM)[26] formula:

$$I(x) = J(x) \cdot t(x) + [1 - t(x)] \cdot A \tag{3.5}$$

where $I(x)$ and $J(x)$ denote the hazy image and clear image respectively. x indicates the coordinate position of the image pixel while A is the ambient light and $t(x)$ represents the transmittance.

Besides, $t(x)$ is closely related to depth information:

$$t(x) = e^{-\beta d(x)} \tag{3.6}$$

where, β is the attenuation parameter and $d(x)$ is the scene depth.

The scene depth map $d(x)$ and clear image $J(x)$ have already been given by the RESIDE dataset. In order to generate hazy image dataset, all we need are the attenuation parameter β and the ambient light A. The conventional method in the past is to assume the values of these two parameters. β and A are always assumed as the same value for 3 color channels(RGB). After experiments, we find that different values of A for different channels can change the color of a hazy image.

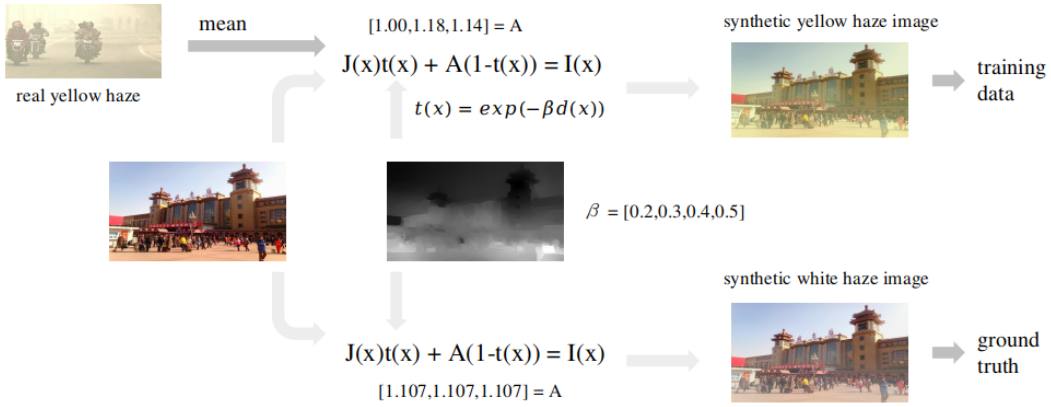

FIGURE 3.6: Method of Generating Colored Haze Image

As is shown in 3.6, we use the mean of every channel from a real colored haze image as the value of A. For the diversity of the data, we choose β from $[0.2, 0.3, 0.4, 0.5]$ to generate hazy images with different density.

To generate the ground truth white haze data, we set $\hat{A}$ as the mean of A and keep all other elements unchanged. Then according to Eq.3.5, with 50 randomly selected real colored haze images from the RESIDE URHI[21], training data for the proposed colour balance net can be generated. Fig.3.7 and Fig.3.8 give examples of the generated training data and the comparison results of their colour histograms.

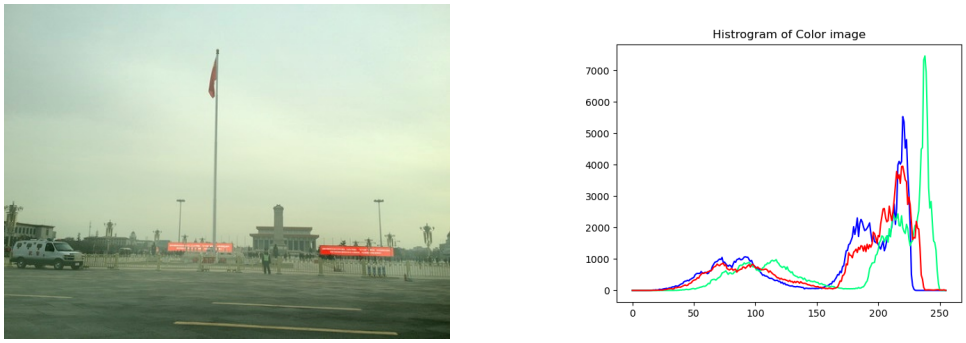

FIGURE 3.7: Histogram of Generated Color Haze Image

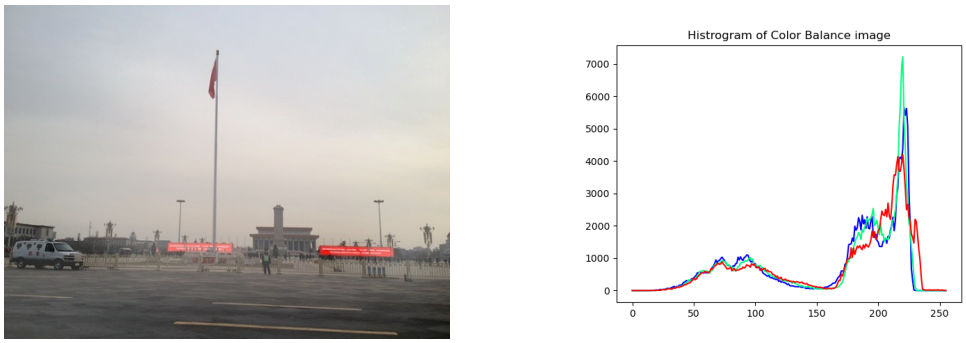

FIGURE 3.8: Histogram of Generated White Haze Image

3.2 Dehazing Module

3.2.1 Network Structure

As is shown in Fig.3.9, a hazy image first come into a 5×5 convolution layer with a ReLU activation function to extract 16 channels of features. Then the features are sent

to 4 residual downsample blocks to get higher dimensional information. Deformable Feature extractor and 5 resdiual groups are used to fully extract features. Attention based skip-connections are built between downsample blocks and upsample blocks to avoid the loss of detailed information in low dimensional features. Finally, features will be refined by a 3×3 convolution layer and restored haze-free images will be generated.

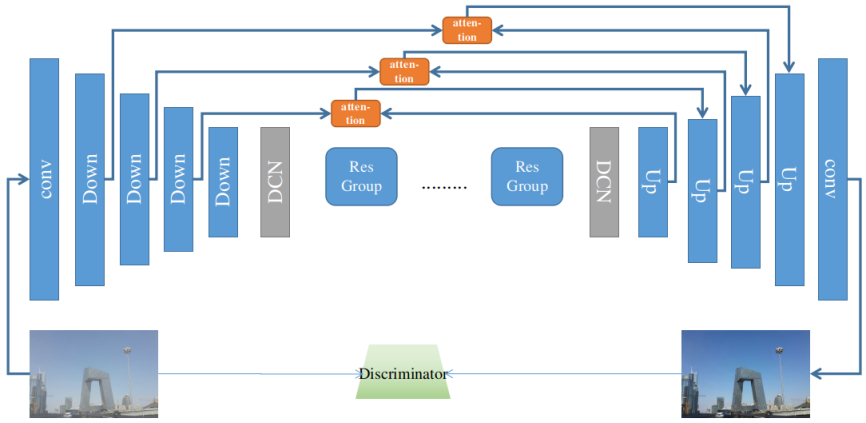

FIGURE 3.9: Dehaze Net Sturcture

Downsample and Upsample Block

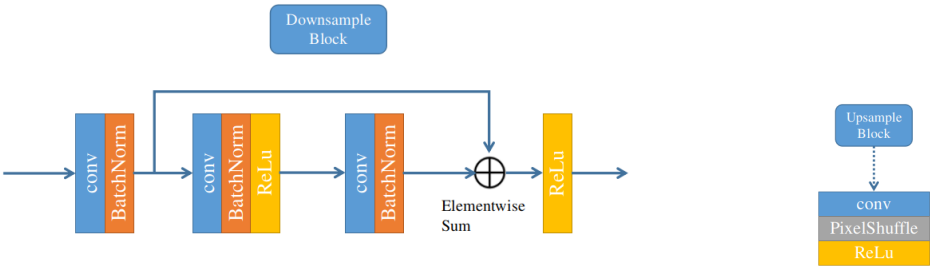

FIGURE 3.10: Details for the Downsample and Upsample Block

The downsample block employs the resnet[14] structure, consisting of 3 convolution layers, all followed by a batch norm layer.The setting for kernel size, stride and padding

are [3,2,1], [3,1,1] and [1,2,0] respectively.

Then we use pixelshuffle[34] method for upsampling. The pixel shuffle method is to

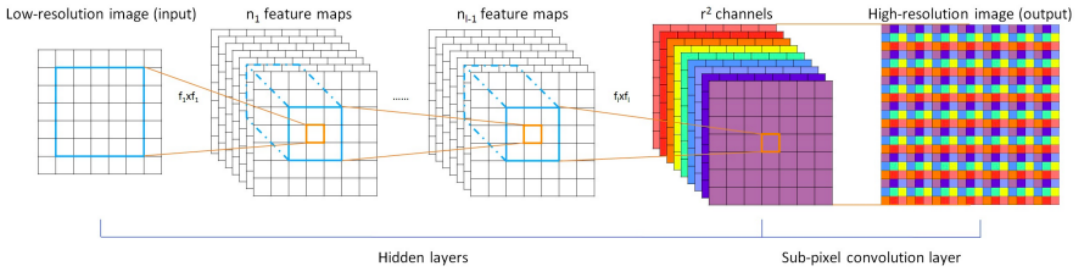

FIGURE 3.11: PixelShuffle

convert the feature map from low-resolution space to high-resolution space. In Fig.3.11, The first part on the left side is used to extract features. Then generate r^2 channels of feature maps. Final up sampling results will obtained by converting the r^2 feature maps to size $w*r$, $h*r$ features. That is, to transform image from shape $N*(C*r*r)*W*H$ to shape $N*C*(H*r)*(W*r)$. This operation is implemented by inserting low-resolution features into high-resolution images periodically according to specific positions, which is called the periodic shuffling. Weights of the r^2 features can be adjusted to optimize the result.

Attention Based Skip-connection

The input of this block is the output d_i concatenated with the output of u_i, where d_i and u_i denote the ith downsample and upsample block respectively.

In order to handle uneven haze distribution, channel and spatial attention methods[40] are applied.

Denote the input as F with size $H \times W \times C$, we first perform a spatial global average pooling and maximum pooling to obtain two $1 \times 1 \times C$ features. Then feed them to a shared two-layer convolution block. The number of output channels in the first layer is C/r, while the second is C and the activation function is ReLU. r is the reduction ratio. Next, perform element-wise sum of the two obtained features and pass the result through a sigmoid activation function to obtain the weight coefficient. Finally, multiply the weight coefficient and the original feature F to get the enhanced feature map F^* with size $H \times W \times C$.

The obtained F^* is the input of the following spatial attention operation. Similar to channel attention, given a feature F^* with size $H \times W \times C$, we first perform average pooling and maximum pooling on channel dimension to obtain two $1 \times 1 \times C$ features. Then concatenate them and send the result into a 7×7 convolution layer with the sigmoid activation function to obtain the weight coefficient. Finally, multiply the weight coefficient and the feature F^*, the final enhanced feature is generated.

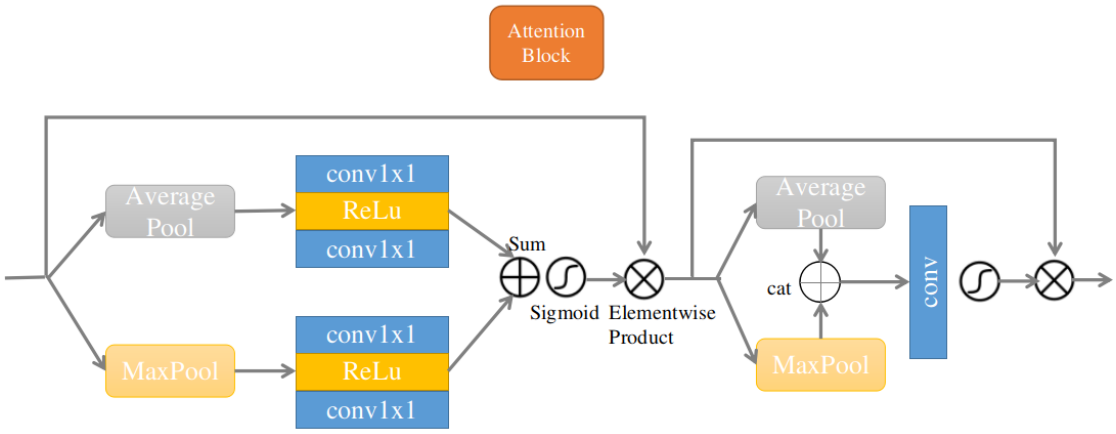

FIGURE 3.12: Attention Block

Deformable Feature extractor

Instead of the fixed kernel shape in traditional convolution layers, deformable convolution can apply more flexible deformable kernel shape to do the computation[41]. That is, as the Fig3.13 shows, the extracted features will be more accurately.

FIGURE 3.13: Deformable Kernel

3.2.2 Loss Function

Adversarial Loss

Here we adopted Least Squares GAN (LSGAN) [25] to calculate the adversarial loss as it is more stable during the learning process. Eq.3.7 shows the function of the adversarial loss for the discriminator:

$$L_D = E_{x_{gt} \sim P_{gt}}(D(x_{gt}) - 1)^2 + E_{x_{dh} \sim P_{dh}}(D(x_{dh}))^2 \tag{3.7}$$

where D denotes the discriminator, x_{gt} and x_{dh} are sampled from clear image distribution P_{gt} and dehazed image distribution P_{dh}.

Smooth L1 Loss

Let I_i^{gt} denote the value of ground truth clear image in the ith color channel, I_i^{dh} represents the value of dehazed image in the ith color channel, x is the pixel location. The Smooth L1 Loss can be expressed as:

$$L_{l1} = \frac{1}{N}\sum_{i=1}^{3}\sum_{x=1}^{N}\sigma(I_i^{gt}(x) - I_i^{dh}(x)) \tag{3.8}$$

$$\sigma(e) = \begin{cases} 0.5e^2, & if|e| < 1 \\ |e| - 0.5, & otherwise \end{cases} . \tag{3.9}$$

Perceptual Loss

In order to optimize the visual effect, the perceptual loss[17] is applied. As illustrated in Eq.3.10, C_l, H_l and W_l are the label, height and width of the lth channels of the corresponding image. I_{gt} is the ground truth clear image and I_{dh} is the dehazed image.

$$L_p = \sum_{l=1}^{3} ||\frac{1}{C_l H_l W_l}(\phi_l(I_{gt}) - \phi_l(I_{dh}))||_2^2. \tag{3.10}$$

A pretrained Vgg16[36] is used as the loss network and uses the first, second and third layers to calculate the loss value.

Overall Loss Function for Dehaze Net

The total loss for the Dehaze Net combines 3 parts, the $smoothl1$ loss L_{l1}, perceptual loss L_p and adversarial loss L_D:

$$L_{color} = L_{l1} + \lambda_1 L_p + \lambda_2 L_D \tag{3.11}$$

where λ_1 and λ_2 are trade-off weights, set as 0.04 and 1 respectively.

Chapter 4

Experiments

4.1 Training and Testing Datasets

4.1.1 RESIDE Dataset

The RESIDE Dataset[21] is a new benchmark consisting of synthetic and real-world hazy images. Besides, the synthetic data contains three parts, the Indoor Training Set(ITS), the Synthetic Objective Testing Set(SOTS), and the Outdoor Training Set(OTS). The ITS training set contains 13,990 hazy indoor images while the OTS training set contains 72135 hazy outdoor images. These images are generated by using the atmospheric scattering model[6], with the depth maps are from the Middlebury stereo data set[32] and the NYU Depth V2 data set[35].

4.1.2 Synthetic to Real Haze Image Dataset

This dataset[24] is generated from randomly selected 3000 ITS hazy images and 3000 OTS hazy images. By applying the CycleGAN[44] model, setting synthetic hazy images

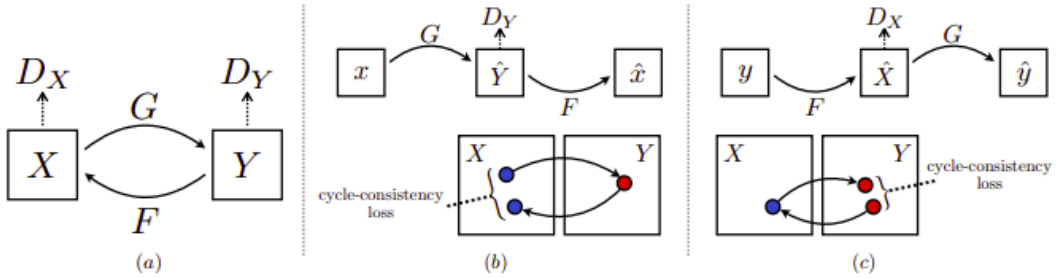

FIGURE 4.1: The CycleGAN Network Structure[44]

and real hazy images as X and Y in Fig.4.1 respectively, to make the synthetic haze distribution and real haze distribution closer. The 1000 real images are randomly selected from the URHI[21] dataset.

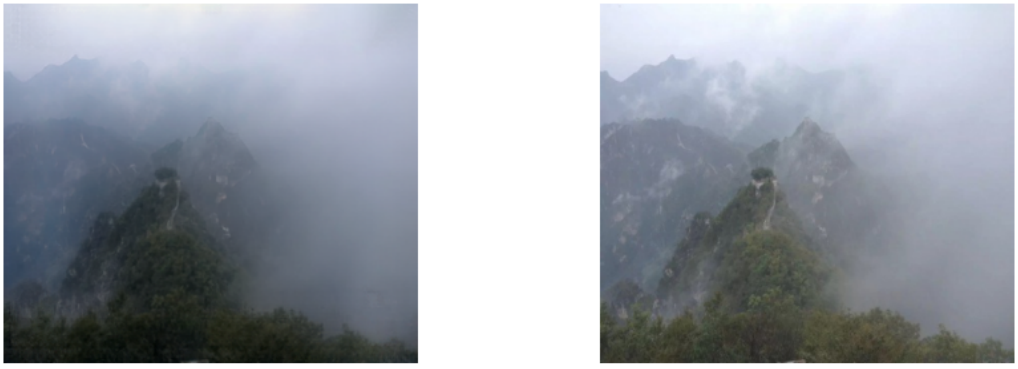

FIGURE 4.2: An Example of Synthetic to Real Haze Image

4.1.3 SOTS

The SOTS is the synthetic test set in the RESIDE[21] data set, which contains 500 indoor hazy images and 500 outdoor hazy images.

4.1.4 Middlebury

This data set is also called D-HAZY[2] and is generated from the Middelbury[32] dataset, which provides images of various scenes and their corresponding depth maps. The dataset is constructed based on the depth information and Koschmieder's light propagation model[15].

4.2 Implementation Details

The proposed Colour Balance Net is trained on the generated colored haze image dataset. $1/10$ images from the dataset are selected as the test set. Input training images are randomly cropped to 240×240 and are randomly rotated to 90, 180 or 270 degrees. Random horizontal flip is also implemented. The Adam optimizer[19] is used with $\beta 1 = 0.9$ and $\beta 2 = 0.99$ respectively. The initial learning rate is set to 0.001. The total training epochs are 160 and adjust learning rate to half of the precious one every 40 epochs.

The proposed Dehaze Net is trained on the RESIDE dataset. For training data, 6000 indoor hazy images and 6000 outdoor hazy images are selected randomly from ITS and OTS respectively. Then the SOTS dataset is used as the test set. Input training images are randomly cropped to 240×240 and are randomly rotated to 90, 180 or 270 degrees. Random horizontal flip operation is also implemented. The Adam optimizer[19] is used with $\beta 1 = 0.9$ and $\beta 2 = 0.99$ respectively. Then finetune the pretrained Dehaze Net on the Synthetic to Real dataset.

The initial learning rate is set to 0.001. The total training epochs are 200 and adjust learning rate to the half of precious one every 60 epochs. The training is carried out on the server with two NVIDIA GeForce GTX 1080Ti.

4.3 Evaluation Method

PSNR: Given an $H \times W$ clear image I and the corresponding dehazed image $\hat{I}$, the mean square error MSE is defined as:

$$MSE = \frac{1}{HW}\sum_{i=0}^{H-1}\sum_{i=0}^{W-1}(I(i,j) - \hat{I}(i,j))^2. \tag{4.1}$$

Then the Peak Signal-to-Noise Ratio(PSNR) is defined as:

$$PSNR = 10log_{10}\frac{(2^8-1)^2}{MSE}. \tag{4.2}$$

SSIM: Structural SIMilarity(SSIM)[39] is used to compare the difference on luminance l, contrast c and structure s between samples. Let gt denote ground truth clear image, dh represent the corresponding dehazed image. The SSIM can be calculated by:

$$\begin{aligned} SSIM(gt, dh) &= \frac{2\mu_{gt}\mu_{dh} + c_1}{\mu_{gt}^2 + \mu_{dh}^2 + c_1} \cdot \frac{2\sigma_{gt,dh} + c_2}{\sigma_{gt}^2 + \sigma_{dh}^2 + c_2} \\ &= l(gt, dh) * c(gt, dh)s(gt, dh) \end{aligned} \tag{4.3}$$

where μ_{gt} is the mean of ground truth clear image, μ_{dh} is the mean of dehazed image, σ_{gt}^2 is the variance of ground truth clear image, σ_{dh} is the variance of dehazed image and $\sigma_{gt,dh}$ is the covariance of ground truth and dehazed image. c_1 and c_2 are two constants.

4.4 Quantitative Results

The quantitative evaluations are conducted on SOTS and Middlebury datasets. It is clear that the proposed method have improvement on both SOTS and Middlebury dataset, compared with DANet and other state-of-the-art methods.

Method	DCP	AOD-Net	EPDN	GDN	DANet	Proposed
SOTS	15.49/0.64	19.06/0.85	23.82/0.89	31.51/0.98	27.76/0.93	31.88/0.98
Middlebury	15.91/0.81	13.86/0.79	15.11/0.83	16.70/0.85	15.93/0.70	16.33/0.85

TABLE 4.1: Quantitative Evaluations on SOTS and Middlebury. PSNR/SSIM Evaluation Methods are Performed.

4.5 Qualitative Results

The quantitative evaluations are conducted on two synthetic hazy image datasets and two real hazy image datasets. They are SOTS[21], Middlebury[32], URHI[21] and Fattal's real world dataset[9] respectively.

Note that in Table.4.3, Table.4.2 and Table.4.4, the proposed method have great improvement on the two synthetic hazy image datasets.

In Table.4.5 and Table.4.6, it is clear that the proposed method performs better not only on real white haze images, but also have large improvement in the visual effects of real colored haze images.

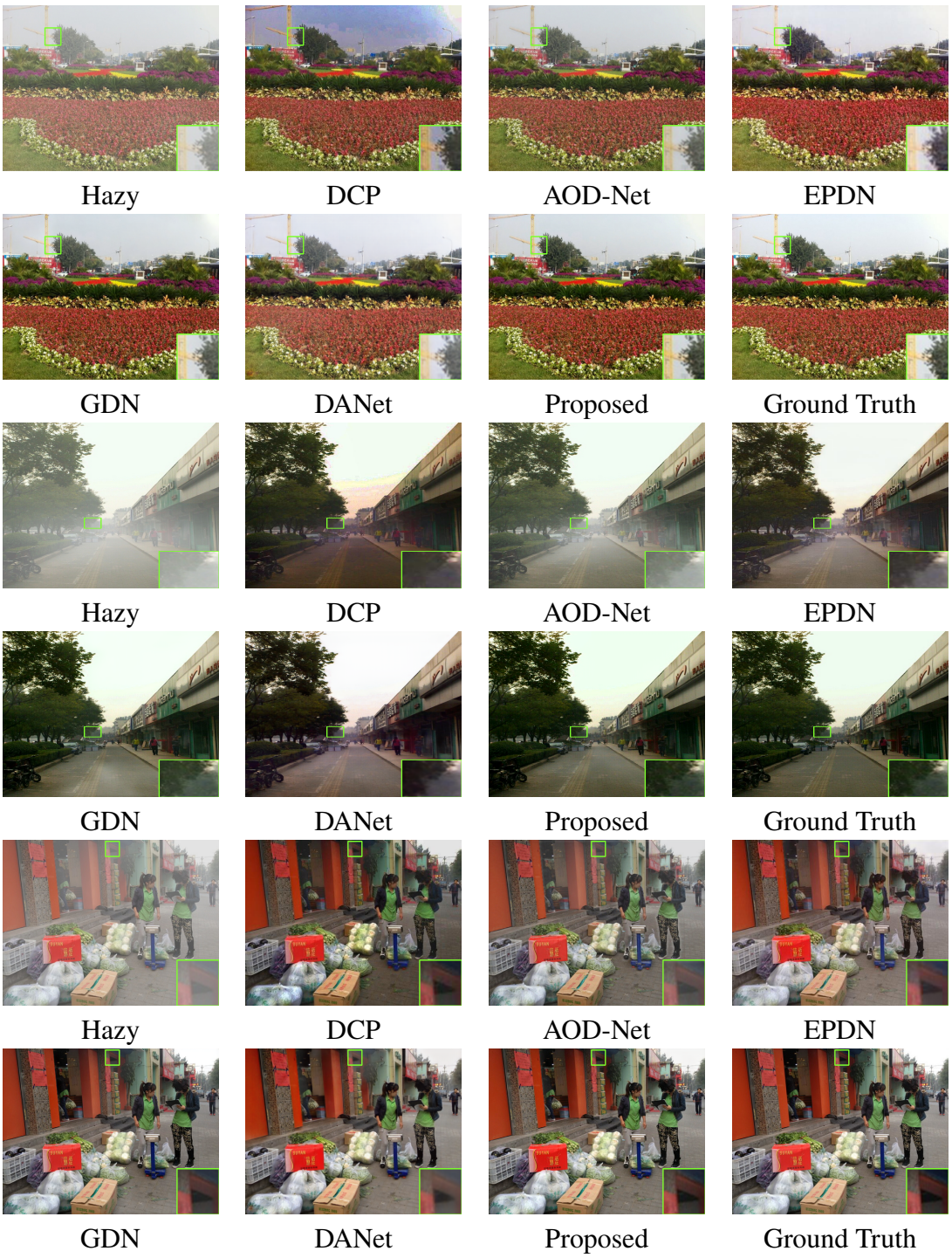

TABLE 4.2: Qualitative Comparisons on SOTS Outdoor Dataset

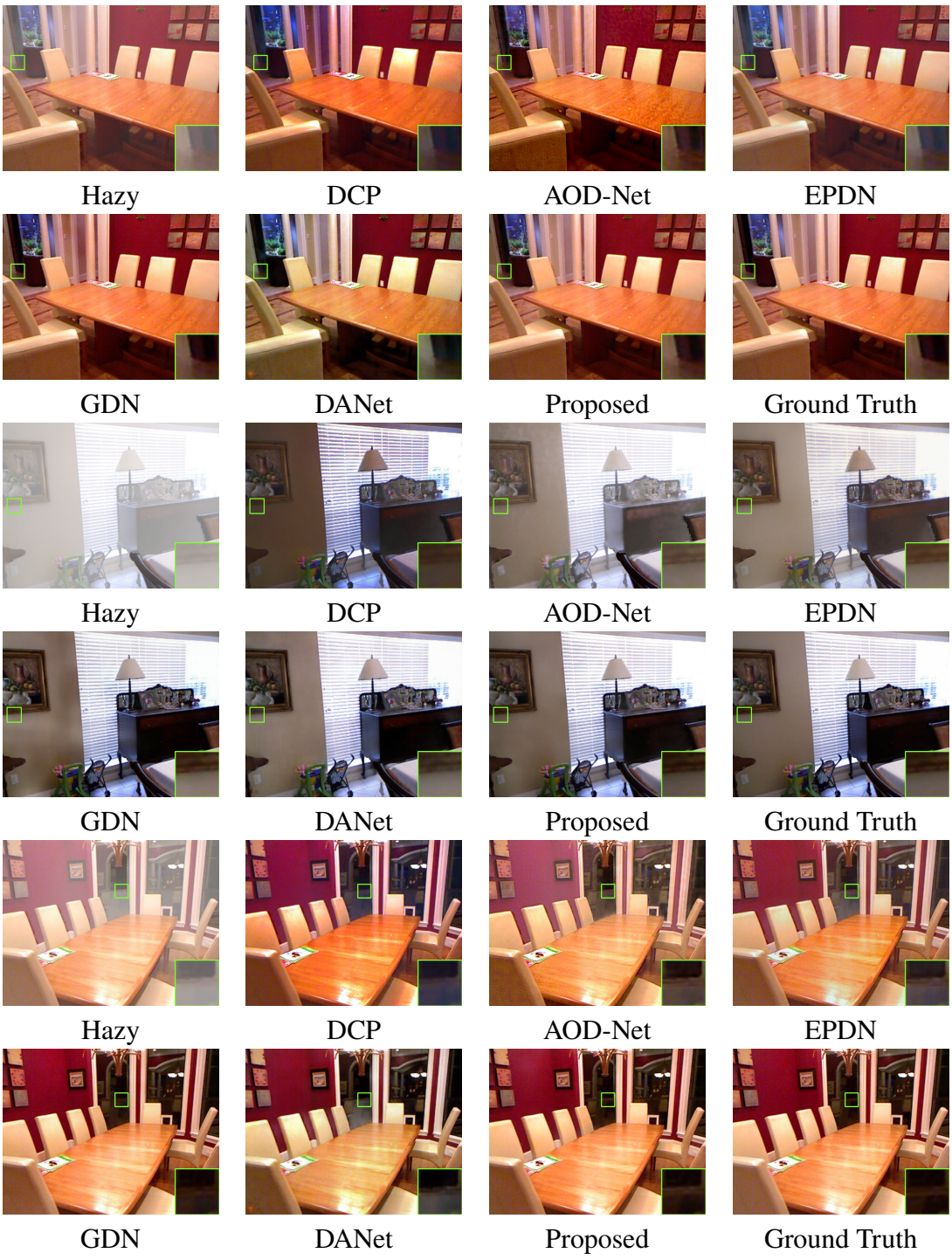

TABLE 4.3: Qualitative Comparisons on SOTS Indoor Dataset

TABLE 4.4: Qualitative Comparisons on Middlebury Dataset

TABLE 4.5: Qualitative Comparisons on Fattal's Real World Dataset[9]

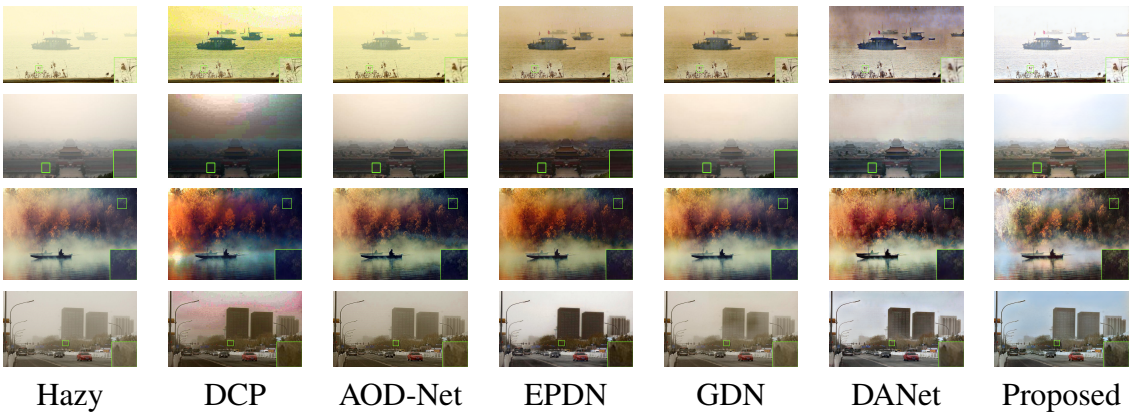

TABLE 4.6: Qualitative Comparisons on the URHI Real Hazy Image Dataset

4.6 Analysis and Ablation Study

4.6.1 Effect of Colour Balance Net

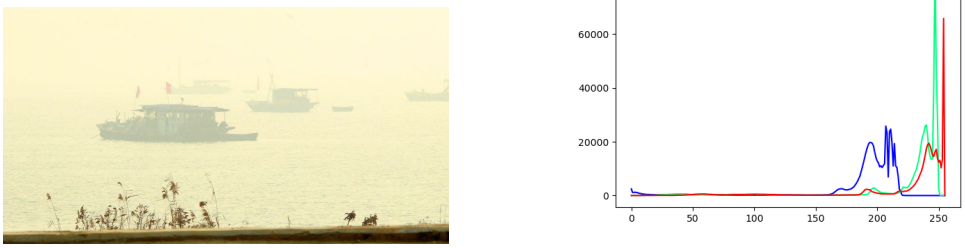

FIGURE 4.3: Histogram of Real Color Haze Image

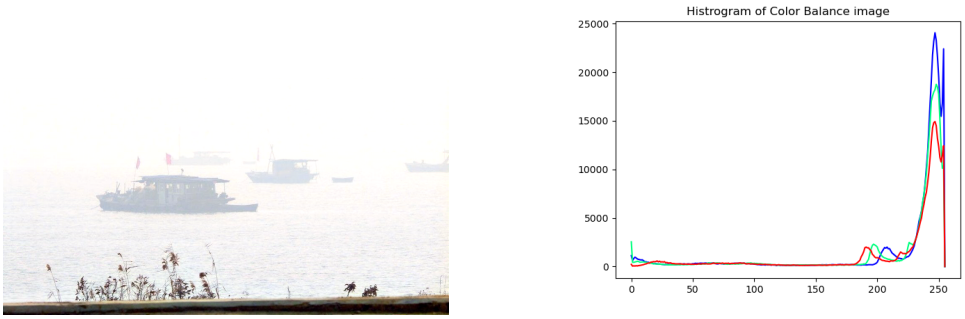

FIGURE 4.4: Histogram of Real Color Balance Image

Fig.4.3 and Fig.4.4 show the changes in color histograms between original image and the color corrected image. It is evident that the deviation value of blue channel tends to be normal. Besides, the visual effect is also improved.

In addition, the proposed Colour Balance Net also can help in improving the dehazed results of different backbones. Taking the Colour Balance Net as an image preprocessing module, the results of selected backbones are illustrated in Table.4.7. Note that all the backbones performs much better after using the result of Colour Balance Net. The new dehazed images are more clear and have better visual effects.

4.6.2 Ablation Study

First we do experiments to verify the effectiveness of attention module in Colour Balance Net and adversarial loss, pixelshuffle in Dehaze Net. Quantitative results are reported in Table.4.8 and Table.4.9

We also did ablation studies to verify the effectiveness of deformable feature block and attention based skip connection. Experiments showed that the proposed Dehaze Net failed to generate dehazed images without any of these two blocks.

Then we will study the effectiveness of Synthetic to Real(denote as s2r) Image data and the Colour Balance Net in the proposed method. Qualitative results are illustrated in Table.4.10.

Note that Table.4.10 shows model trained on the synthetic to real hazy image database can have great improvement in real hazy image dehazing task. The use of Colour Balance net can improve the visual effects in colored haze image dehazing tasks.

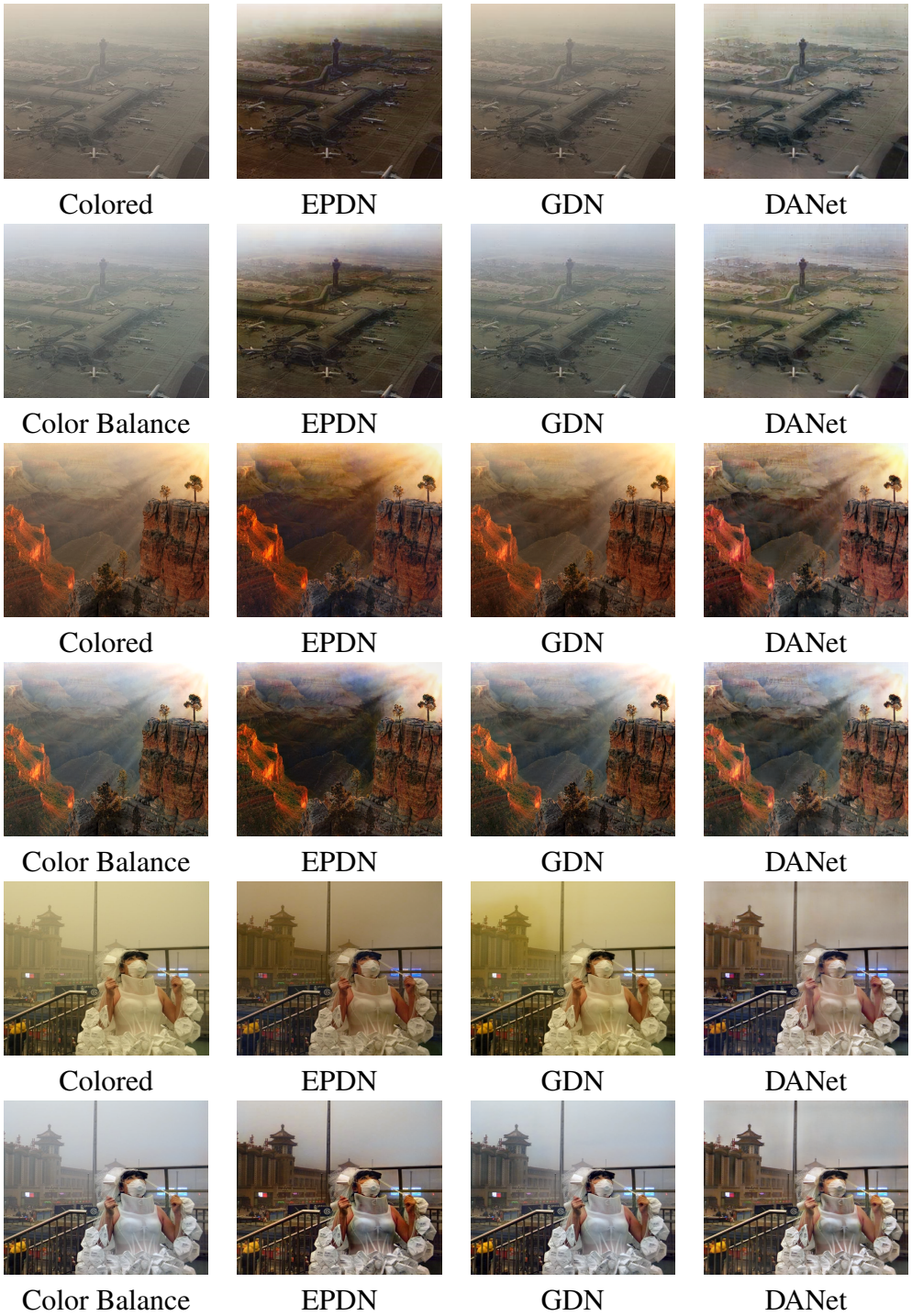

TABLE 4.7: Qualitative Comparisons on Dehazing Result of the Color-Balance Hazy Image for Different Backbones

Table 4.8: Colour Balance Net Performance of w/ Attention and w/o Attention

Methods	w/ attention	w/o attention
Generated dataset	31.21/0.97	29.48/0.97

Table 4.9: Dehaze Net Performance of w/ Adversarial Loss and w/o Adversarial Loss

Methods	w/ adversarial loss	w/o adversarial loss	adversarial loss + Convtranspose
SOTS	31.88/0.98	31.58/0.97	31.46/0.98

w/o s2r w/o CBNet w/o s2r and CBNet proposed

Table 4.10: Qualitative Results of w/o s2r, w/o Colour Balance, w/o s2r and Colour Balance, Proposed

Chapter 5

Conclusion

In this paper, we propose a two step method to deal with the colored haze. A colored haze image database is also constructed. The Colour Balance Module learns the transfer from colored haze distribution to white haze distribution. Pixel attention is applied to get the enhanced information of every pixel. The Dehazing Module firstly learns the mapping from hazy images to clear images. In order to improve performance, the synthetic to real hazy image dataset is applied to alleviate the domain gap between synthetic hazy images and real hazy images. Channel and spatial attention blocks are introduced in the dehazing net. Deformable fexture extractor and pixeshuffle method are also used. Experiments and comparisons illustrate that the proposed method have competitive performance in dehazing task, especially in colored haze image dehazing.

Bibliography

[1] C. O. Ancuti and C. Ancuti. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing* 22(8) (2013), 3271–3282.

[2] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In: *2016 IEEE international conference on image processing (ICIP)*. IEEE. 2016, 2226–2230.

[3] D. Berman, T. Treibitz, and S. Avidan. Non-local Image Dehazing. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, 1674–1682.

[4] D. Berman, T. Treibitz, and S. Avidan. Single Image Dehazing Using Haze-Lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(3) (2020), 720–734.

[5] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Transactions on Image Processing* 25(11) (2016), 5187–5198.

[6] A. Cantor. Optics of the atmosphere–Scattering by molecules and particles. *IEEE Journal of Quantum Electronics* 14(9) (1978), 698–699.

[7] 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)*. 2019, 1375–1383.

[8] A. Dudhane, K. M. Biradar, P. W. Patil, P. Hambarde, and S. Murala. Varicolored Image De-Hazing. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, 4563–4572.

[9] R. Fattal. Dehazing Using Color-Lines. *ACM Transactions on Graphics (TOG)* 34 (2014), 1–14.

[10] R. Fattal. Dehazing using color-lines. *ACM transactions on graphics (TOG)* 34(1) (2014), 1–14.

[11] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Tremeau, and C. Wolf. Residual conv-deconv grid network for semantic segmentation. *arXiv preprint arXiv:1707.07958* (2017).

[12] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmio. Fusion-based variational image dehazing. *IEEE Signal Processing Letters* 24(2) (2016), 151–155.

[13] K. He, J. Sun, and X. Tang. Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(12) (2011), 2341–2353.

[14] K. He, X. Zhang, S. Ren, and J. Sun. *Deep Residual Learning for Image Recognition*. 2015.

[15] H. Israël and F. Kasten. Koschmieders theorie der horizontalen sichtweite. In: *Die Sichtweite im Nebel und die Möglichkeiten ihrer künstlichen Beeinflussung*. Springer, 1959, 7–10.

[16] X. Ji, Y. Feng, G. Liu, M. Dai, and C. Yin. Real-Time Defogging Processing of Aerial Images. In: *2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*. 2010, 1–4.

[17] J. Johnson, A. Alahi, and L. Fei-Fei. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*. 2016.

[18] J.-Y. Kim, L.-S. Kim, and S.-H. Hwang. An advanced contrast enhancement using partially overlapped sub-block histogram equalization. *IEEE transactions on circuits and systems for video technology* 11(4) (2001), 475–484.

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

[20] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. AOD-Net: All-in-One Dehazing Network. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, 4780–4788.

[21] 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* 28(1) (2019), 492–505.

[22] R. Li, J. Pan, Z. Li, and J. Tang. Single Image Dehazing via Conditional Generative Adversarial Network. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, 8202–8211.

[23] X. Liu, Y. Ma, Z. Shi, and J. Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, 7314–7323.

[24] X. Liu, Z. Shi, Z. Wu, and J. Chen. *GridDehazeNet+: An Enhanced Multi-Scale Network with Intra-Task Knowledge Transfer for Single Image Dehazing*. 2021.

[25] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley. *Least Squares Generative Adversarial Networks*. 2017.

[26] S. Narasimhan and S. Nayar. Chromatic framework for vision in bad weather. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*. Vol. 1. 2000, 598–605 vol.1.

[27] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia. FFA-Net: Feature fusion attention network for single image dehazing. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07. 2020, 11908–11915.

[28] 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)*. 2019, 8152–8160.

[29] Y. Qu, Y. Chen, J. Huang, and Y. Xie. Enhanced pix2pix dehazing network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, 8160–8168.

[30] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In: *European conference on computer vision*. Springer. 2016, 154–169.

[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*. 2018, 3253–3261.

[32] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In: *German conference on pattern recognition*. Springer. 2014, 31–42.

[33] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang. Domain adaptation for image dehazing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, 2808–2817.

[34] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 1874–1883.

[35] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In: *European conference on computer vision*. Springer. 2012, 746–760.

[36] K. Simonyan and A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015.

[37] J. A. Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on image processing* 9(5) (2000), 889–896.

[38] R. T. Tan. Visibility in bad weather from a single image. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008, 1–8.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13(4) (2004), 600–612.

[40] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, 3–19.

[41] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma. Contrastive Learning for Compact Single Image Dehazing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, 10551–10560.

[42] H. Zhang and V. M. Patel. Densely Connected Pyramid Dehazing Network. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, 3194–3203.

[43] H. Zhang, J. Li, L. Li, Y. Li, Q. Zhao, and Y. You. Single image dehazing based on detail loss compensation and degradation. In: *2011 4th International Congress on Image and Signal Processing*. Vol. 2. 2011, 807–811.

[44] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. 2017, 2223–2232.

[45] K. Zuiderveld. Contrast Limited Adaptive Histogram Equalization. In: *Graphics Gems*. 1994.