A Discrete Wavelet Transform GAN for NonHomogeneous Dehazing

A Discrete Wavelet Transform GAN for NonHomogeneous Dehazing

By Minghan Fu, B.Eng.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & 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

Master of Applied Science(2021) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE: A Discrete Wavelet Transform GAN for NonHomogeneous Dehazing

AUTHOR: Minghan Fu

B.Eng.(Automation Engineering)

Northeast Electric Power University

Jilin, China

SUPERVISOR: Dr.Jun Chen

NUMBER OF PAGES: xii, 53

To my dear parents and friends

Abstract

Hazy images are often subject to color distortion, blurring and other visible quality degradation. Some existing CNN-based methods have shown great performance on removing the homogeneous haze, but they are not robust in the non-homogeneous case. The reason is twofold. Firstly, due to the complicated haze distribution, texture details are easy to get lost during the dehazing process. Secondly, since the training pairs are hard to be collected, training on limited data can easily lead to the over-fitting problem. To tackle these two issues, we introduce a novel dehazing network using the 2D discrete wavelet transform, namely DW-GAN. Specifically, we propose a two-branch network to deal with the aforementioned problems. By utilizing the wavelet transform in the DWT branch, our proposed method can retain more high-frequency information in feature maps. To prevent over-fitting, ImageNet pre-trained Res2Net is adopted in the knowledge adaptation branch. Owing to the robust feature representations of ImageNet pre-training, the generalization ability of our network is improved dramatically. Finally, a patch-based discriminator is used to reduce artifacts of the restored images. Extensive experimental results demonstrate that the proposed method outperforms the state-of-the-art quantitatively and qualitatively.

Acknowledgements

Firstly, I want to express my deepest appreciation to my supervisor, Prof. Jun Chen. Thanks for his guidance and support for this work. During my Master program, Prof. Chen not only taught us in-depth theoretical knowledge but also taught me a rigorous and serious attitude towards research work. It is an honor for me to be his student. In future work, I will keep in mind the teachings of Prof. Chen.

Secondly, I would like to sincerely thank my committee members, Dr.Thamas and Dr.Sorina for their valuable suggestions and kind help.

Next, thank you Huan Liu, thanks for your great help in the field of Deep Learning and Image Processing. I often make many mistakes when I first started doing experiments and research work. You taught me patiently and point out my mistakes every time. It is your contribution that helps me gradually develop the ability to independently complete the design and experiments of the model. Your rigorous and serious attitude towards scientific research deeply affects me.

Furthermore, thanks Yankun and Wei. We encourage each other in study and life. Thanks Chengqi and Zhihao, the days we work together in the laboratory are unforgettable for me. Thanks Xiaohong and Linhui, your smart views have a profound influence on my research, and I am grateful for your help on this work.

Finally, thank you Mom and Dad for giving me selfless love and understanding. Every time I chat with you, I feel warm and happy.

Notation and abbreviations

GAN	Generative Adversarial Network
DW-GAN	A Discrete Wavelet Transform GAN
CNN	Convolutional Neural Network
Adam	Adaptive moment estimation algorithm
ReLu	Rectified Linear unit
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity
GPU	Graphics Processing Unit
SOTS	Synthetic Objective Testing Set
RESIDE	Realistic Single Image Dehazing Dataset
ITS	Indoor Training Set of RESIDE
NTIRE19 Dataset	Dense Haze Dataset
NTIRE20 Dataset	Non-homogeneous Haze Dataset

NTIRE21 Dataset	Non-homogeneous Haze Dataset2
DCP	Dark Channel Prior
AOD	All-in-One Dehazing Network
GFN	Gated Fusion Dehazing Network
GCA	Gated Context Aggregation Dehazing Network
FFA	Feature Fusion Attention Dehazing Network
TDN	Trident Dehazing Network
DCPDN	Densely Connected Pyramid Dehazing Network

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Problem Statement

1.1 Introduction

Hazy images are often prone to color distortion, blurring and other visible quality degradation. The varied image degradations often lower the perceptual quality of pictures and jeopardize the performance of numerous intelligent systems, such as tracking (Singh and Kumar 2019) , satellite remote sensing (Long et al. 2013; Ni et al. 2016) , and object detection (Sindagi et al. 2020; Katyal et al. 2018). Therefore, image dehazing has attracted much attention in the computer vision community. Many previous dehazing methods are based on the classical atmospheric scattering model (Middleton 2019):

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

where $I(x)$ denotes the hazy image, $J(x)$ represents the clear image, $A(x)$ is the global atmospheric light, $t(x)$ is the medium transmission and x indicates the pixels. Besides, $t(x) = e^{-\beta d(x)}$, where β and $d(x)$ are respectively the atmosphere scattering parameter and the scene depth.

Based on the atmospheric scattering formulation, some prior-based methods have been proposed (He et al. 2010; Berman, Avidan, et al. 2016; Zhu et al. 2015; Fattal 2014). These methods estimate the atmospheric light $A(x)$ and the medium transmission map $t(x)$ by hand-crafted priors, such as the dark channel prior (He et al. 2010) and the non-local prior (Berman, Avidan, et al. 2016). However, it is quite hard to accurately estimate $A(x)$ and $t(x)$. Especially in the non-homogeneous dehazing task, the haze distribution is much more complicated and the haze density is not strongly correlated to the image depth. Therefore, using prior-based method can result in huge estimation error. Such methods are no longer good choices for non-homogeneous dehazing. In recent years, with the development of deep learning techniques (LeCun et al. 2015), many deep learning based dehazing methods (Cai et al. 2016; Li et al. 2017b; Zhang and Patel 2018; Chen et al. 2019; Qin et al. 2020) have also been proposed. These methods use convolutional neural networks (CNNs) to extract features and learn the mappings directly between hazy and haze-free image pairs. However, these methods usually require a large number of image pairs during the training process. With insufficient amount of training data, many deep learning based methods are harder to succeed. In addition, the high-frequency components in the clear images, such as edges and fine textures, are often degraded significantly by non-homogeneous haze. Therefore, restoring clear texture details and sharp edges from hazy images are essential for achieving

good perceptual quality.

In summary, difficulties mainly come from two aspects in non-homogeneous dehazing. Firstly, due to the complex haze distribution, texture and color details are easy to get lost during restoration. Secondly, the training image pairs are hard to be collected. Using limited data to train for a robust non-homogeneous dehazing model is quite challenging. To address the above two problems, we propose a two-branch generative adversarial network. In our first branch, we use the designed wavelet down-sampling modules to replace parts of the convolution layers. By doing this, the number of parameters can be reduced. The lightweight model can achieve better performance on small training datasets and avoids over-fitting problems caused by model redundancy. In addition, the discrete wavelet transform (Mallat 1989) retains the frequency domain information in the images and feature maps. Such information is more conducive to the restoration of texture details. In our second branch, we employ the pre-trained Res2Net (Gao et al. 2019) as the backbone to extract multilevel features. This pre-trained encoder can bring substantial prior knowledge for small training datasets (He et al. 2019). By leveraging the prior knowledge, we can observe significant improvements on small-scale datasets in terms of testing accuracy. Moreover, we further employ an attention mechanism in our pipeline. Pixel-wise attention module and channel-wise attention module enable the network to focus on the hazy zones and more critical channel information. Finally, the discriminator is used to introduce an adversarial loss in the training stage. By adopting the adversarial loss, our network is guided to learn for natural and photo-realistic solutions.

1.2 Contributions

In this thesis, we propose a fully end-to-end trainable Generative Adversarial Network, named DW-GAN, for non-homogeneous image dehazing, which takes the hazy image as input and directly generates the haze-free image without the estimation of intermediate parameters. The proposed method utilizes the advance of deep learning techniques and leverages the power of 2D discrete wavelet transform. Compared with the prior-based dehazing model, the proposed method avoids the pitfall of inaccurate parameter estimation. The inaccurate estimation of the transmission map and the global atmospheric light may lead to unsatisfactory dehazing results. As for some exiting CNN-based methods, they do not deliver visually pleasing results on non-homogeneous dehazing tasks. Our proposed model surpasses these SOTA methods by a large margin in both quantitative results and qualitative visual results. Figure 1.1 visually demonstrates examples of our dehazing results in NH-HAZE2 test dataset.

In summary, the main innovations and contributions of this work are as follows:

1. We propose a two-branch end-to-end trainable GAN to address the non-homogeneous dehazing problem.

2. We introduce a novel way to embed 2D discrete wavelet transform in our proposed network, aiming at preserving sufficient high-frequency information and restoring clear texture details. To perform well in small-scale datasets, we leverage the prior feature knowledge by using ImageNet pre-trained weights as initialization.

Figure 1.1: Part of our dehazing results on NH-HAZE2 test set.

3. We show extensive experimental results and comprehensive ablation analysis to illustrate the effectiveness of our proposed method.

1.3 Thesis Structure

To elucidate the advantages of the proposed DW-GAN clearly, the outline of this thesis is organized as follows: In Chapter 2, we will review the existing dehazing methods, frequency domain learning techniques and generative adversarial network structure. Then, Chapter 3 will introduce the proposed DW-GAN in detail including the network architecture, discrete wavelet transform, and the loss functions. Furthermore, Chapter 4 will introduce our experimental settings, conduct

ablation studies and compare the performance of our proposed method with other state-of-the-art methods qualitatively and quantitatively. In Chapter 5, we will discuss our data pre-processing method and dehazing results in the NTIRE2021 NonHomogeneous Dehazing Challenge. Finally, Chapter 6 concludes our work.

Chapter 2

Related Works

2.1 Homogeneous Dehazing

In the homogeneous dehazing task, the haze distribution is uniform and training image pairs are easy to get. We can use clean images to generate synthesized corresponding haze images via using Equation1.1. Researchers have proposed many methods for homogeneous dehazing tasks. These methods can be roughly divided into two categories: prior-based methods and learning-based methods.

2.1.1 Prior-based methods

Prior-based methods utilize prior statistical knowledge and hand-crafted features for image dehazing. There are some simple but powerful priors, such as dark channel prior (DCP) (He et al. 2010), non-local prior(Berman, Avidan, et al. 2016) and color attenuation prior(Zhu et al. 2015).

DCP assumes that in most of the non-sky and non-haze patches, the hazy image may have extremely low intensities (value tend to be 0) in at least one color

channel. The DCP of a hazy image J can be defined as:

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

where the J^{dark} denotes the low intensity channel of a hazy image J, J^c is the color channel of J and $\Omega(x)$ is a local patch centered at pixel x. Then, the transmission map $t(x)$ can be estimated by mathematical deduction. Choosing the min value of each small local patch to represent the value of each local zone based on Equation1.1, we can obtain:

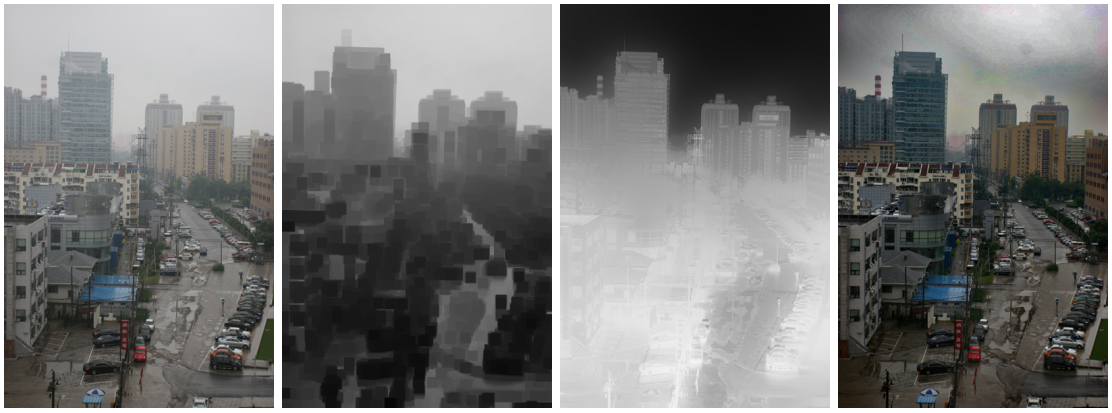
$$\min_{y\in\Omega(x)} (I^c(y)) = \hat{t}(x) \min_{y\in\Omega(x)} (J^c(y)) + (1-\hat{t}(x))A^c, \tag{2.2}$$

where $\hat{t}(x)$ is the transmission map based on each small local patch, $I^c(y)$ is the color channel of a clear image, and A^c is the color channel of A and is always supposed to be a positive number. Dividing both sides of the equation by the same value A^c , we can get:

$$\min_{y\in\Omega(x)} \left(\frac{I^c(y)}{A^c} \right) = \hat{t}(x) \min_{y\in\Omega(x)} \left(\frac{J^c(y)}{A^c} \right) + (1-\hat{t}(x)). \tag{2.3}$$

Performing the min operation on both sides of the Equation 2.3, we can obtain:

$$\min_{c} \left(\min_{y\in\Omega(x)} \left(\frac{I^c(y)}{A^c} \right) \right) = \hat{t}(x) \min_{c} \left(\min_{y\in\Omega(x)} \left(\frac{J^c(y)}{A^c} \right) \right) + (1-\hat{t}(x)). \tag{2.4}$$



Hazy input Dark channel Transmission map Dehazed image

Figure 2.1: Visualized dehazing results of DCP

According to the dark channel assumption and in view of Equation2.1, the dark channel value of haze-free zones in J tends to be zero; as a consequence, we get:

$$\hat{t}(x) = 1 - \min_{c} \left(\min_{y \in \Omega(x)} \left(\frac{I^c(y)}{A^c} \right) \right). \tag{2.5}$$

To estimate A, the top 0.1% brightest pixels are chosen and the highest intensity among their corresponding pixel of the hazy image tends to be atmospheric light A. Thus, the hazy image can be easily restored by using the estimation of A, $t(x)$ and invoking Equation1.1. The dehazing results by DCP are shown in Figure 2.1.

Although DCP shows a good effect on haze removal, there still exist some limitations of DCP. For example, DCP can not distinguish between large white regions, like the sky, huge walls and real hazy zones. To tackle such problems, color attenuation prior (CAP) (Zhu et al. 2015) has been proposed. CAP creates a linear model to estimate scene depth as strong prior knowledge. The author observed that hazy zones always have the highest brightness and lowest saturation, so the difference between the brightness and the saturation can indicate hazy patches in

a hazy image. Besides, CAP assumes that scene depths are linearly related to depth map $d(x)$ of target image as below:

$$d(x) = \theta_0 + \theta_1 v(x) + \theta_2 s(x) + \epsilon(x), \tag{2.6}$$

where v is the brightness component of the hazy image, s is the saturation component. $\theta_0, \theta_1, \theta_2$ are learnable parameters, $\varepsilon(x)$ is the random noise that follows the Gaussian distribution $N(0, \sigma^2)$. By using the training dataset, the linear regression model updates all the learnable parameters. The author provides the proper value of $\theta_0, \theta_1, \theta_2$, which are 0.1218, 0.9597 and -0.7802 respectively. Because the transmission map is related to the $t(x) = e^{-\beta d(x)}$, $t(x)$ can be easily calculated. $A(x)$ can be estimated using haze-lines prior (Berman, Avidan, et al. 2016). Then we can dehaze images by using Equation 1.1.

Non-local dehazing method(Berman, Avidan, et al. 2016) makes use of the observation that the RGB color value of an image can be represented by hundreds of RGB values. In a haze-free image, these color values can form tight clusters, and the pixel distribution in each cluster is non-local, spreading across the entire image. But for hazy images, these RGB values are distributed along lines in RGB space. This suggests the approach of utilizing haze lines to estimate the transmission map and then performing image dehazing.

The strong assumption and inaccurate estimation impede these prior methods to achieve satisfying performance. With the development of deep learning techniques, some deep learning based methods have achieved great performance and become much more popular in recent years.

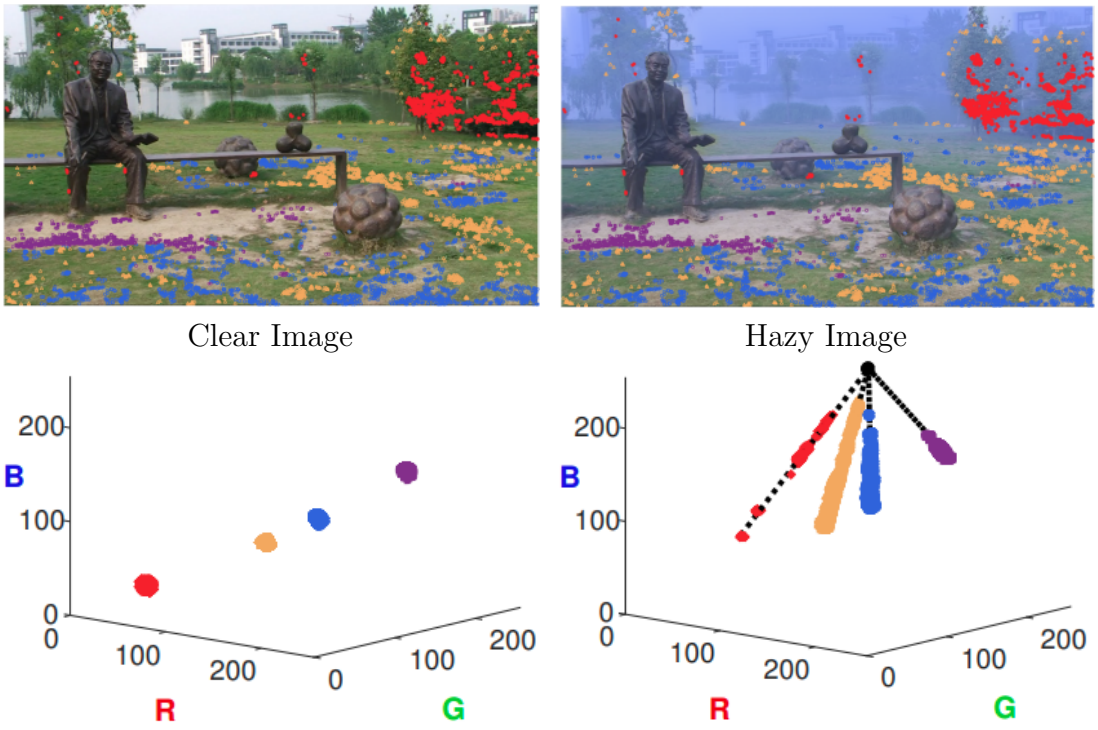


Figure 2.2: Non-local image deahzing method. Images are adopted from original paper of (Berman, Avidan, et al. 2016)

2.1.2 Deep Learning-based methods

Recently, with the rapid progress of the deep learning approach (LeCun et al. 2015), deep learning shows its remarkable ability in solving single image dehazing problems. Some of the deep learning based methods still rely on the atmospheric scattering model. For example, (Cai et al. 2016) proposed DehazeNet as the first end-to-end CNN to learn the transmission map. Specifically, it follows the traditional procedure and atmospheric scattering formulation but uses CNN to estimate the transmission map. Similarly, a novel multi-scale neural network (MSCNN) (Ren et al. 2016) is then proposed to estimate the transmission map from the hazy image.

Different from (Cai et al. 2016; Ren et al. 2016), AOD-Net (Li et al. 2017b) jointly estimates the transmittance and atmospheric light through a lightweight neural network. Besides, AOD-Net is easily embedded with Faster R-CNN (Ren et al. 2015) and improves the object detection performance on hazy images with a large margin. The network structure of AOD-Net is shown in Figure 2.3.

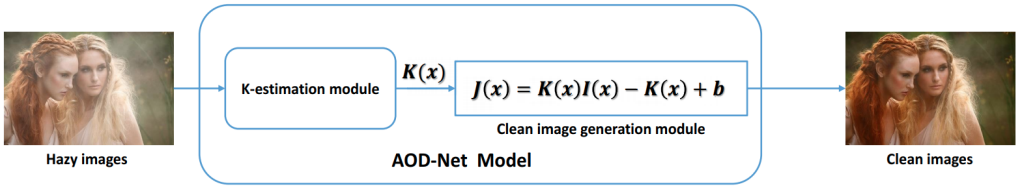


Figure 2.3: The network structure of AOD-Net. Image originally used in (Li et al. 2017b)

DCPDN (Zhang and Patel 2018) employs pyramid densely connected encoder-decoder to estimate the transmission map and uses a U-Net (Ronneberger et al. 2015) structure to estimate atmospheric light.

Other than the above-listed methods that depend on the atmospheric scattering model, most recent dehazing methods pay great attention to discover a good mapping from hazy images to clear counterparts. GFN (Ren et al. 2018) uses the white balance, the contrast enhancement and the gamma correction methods to pre-process the hazy input. And the dehazed output is then generated by fusing the features of the three derived inputs.

GCANet (Chen et al. 2019) adopts the smoothed dilation technique into the pipeline for removing the grid artifacts caused by the dilated convolution and uses a gated subnetwork to fuse the features of different levels. The network structure of GCANet is shown in Figure 2.4

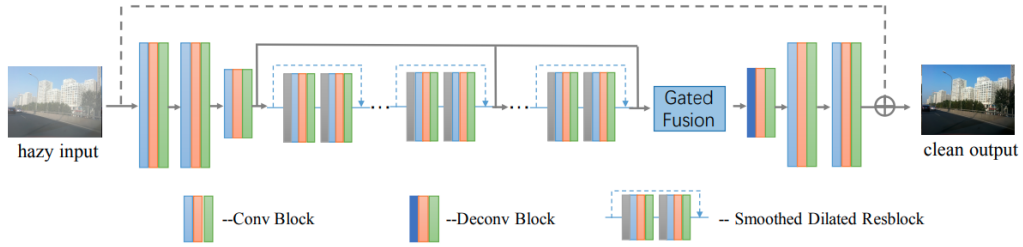


Figure 2.4: The network structure of GCANet. Image originally used in (Chen et al. 2019)

FFA (Qin et al. 2020) proposes a novel feature attention module, which combines the channel attention and the pixel attention mechanism. The proposed feature attention module provides additional flexibility in dealing with different types of haze patterns, directs the network to pay more attention to the hazy zones and more important channel information. The network structure of FFA is shown in Figure 2.5

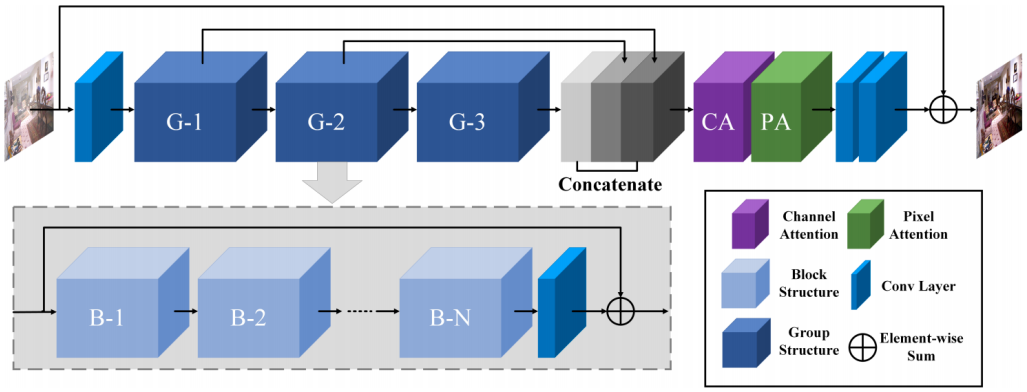


Figure 2.5: The network structure of FFA. Image originally used in (Qin et al. 2020)

FD-GAN (Dong et al. 2020) leverages an end-to-end GAN (Goodfellow et al. 2014) with a fusion discriminator. The fusion discriminator integrates the high-frequency and low-frequency information as additional priors during the training

progress. Specifically, they concatenate the generated image G(I) and its corresponding LF and HF components together as input information and feed them into the discriminator. Different from traditional discriminator that only learns to distinguish between the dehazed image and haze-free image, the proposed FD-GAN additionally takes LF and HF information into account. The network structure is shown in Figure 2.6. Unlike the FD-GAN(Dong et al. 2020) that discriminates the

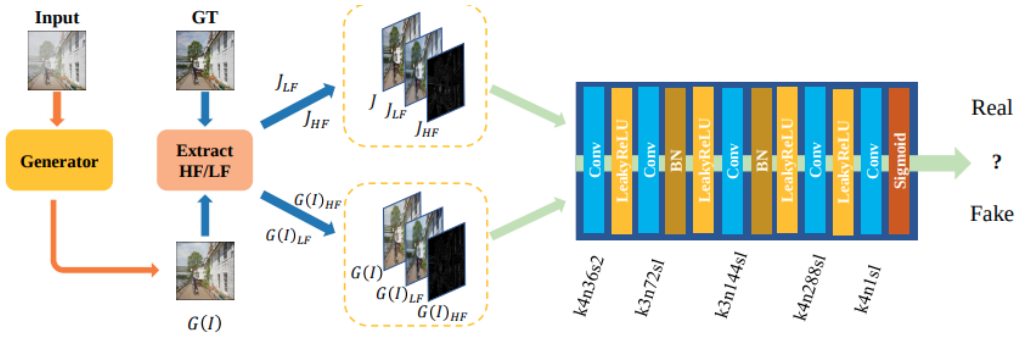


Figure 2.6: The hazy image processing procedure of FD-GAN. Image originally used in (Dong et al. 2020)

clear image and the generated image in the frequency domain, our method focuses on directly fusing the high-frequency domain features into the generator.

2.2 NonHomogeneous Dehazing

In the non-homogeneous dehazing task, the haze distribution is un-uniform and training image pairs are hard to collect. Due to the complex haze distribution, researchers tend to design the convolution neural network(CNN) to directly learn the color mapping from hazy images to clear counterparts. TDN is one of the outstanding methods for non-homogeneous dehazing.

TDN (Liu et al. 2020)proposes a novel coarse to fine model, which consists of three subnets. The Details Refinement sub-Net focuses on reconstructing high frequency image details, the Haze Density Map Generation sub-Net can automatically generate the haze density map and the Encoder-Decoder sub-Net extracts features and learns the color mapping from hazy images to clear counterparts. TDN shows great performance on non-homogeneous dehazing and won the winner award on NTIRE2020 NonHomogeneous Dehazing Challenge. The network structure of TDN is shown in Figure 2.7.

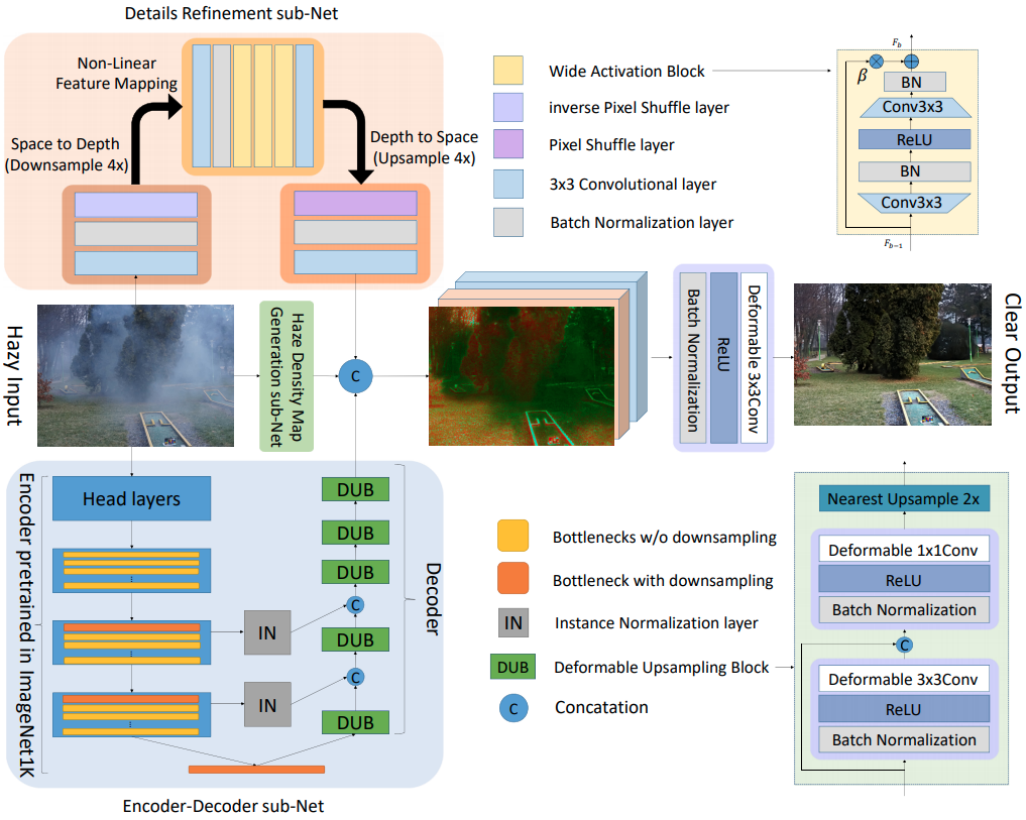


Figure 2.7: The network structure of TDN. Image originally used in (Liu et al. 2020)

2.3 Frequency Domain Learning

Frequency analysis has always been a powerful tool in image processing. Effective usage of image frequency domain information can greatly improve the performance of the methods in image restoration. Recently, some approaches embedded frequency information into the network structure and exploited the effectiveness of frequency domain information. A wavelet residual network (Bae et al. 2017) is proposed with the discovery that neural networks can benefit from learning on wavelet subbands. DWSR (Guo et al. 2017) designed a deep wavelet network that can recover missing details in subbands. MWCNN(Liu et al. 2018) considered multi-level wavelet transform to enlarge receptive field. AWNet (Dai et al. 2020) further utilizes the frequency domain knowledge for image ISP. These methods take advantage of discrete wavelet transform and use it in designing deep learning network architectures.

2.4 Generative Adversarial Network

Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) consist of two parts: Generator and Discriminator. This is a min-max optimization problem, where two players contest with each other and boost the performance of each other. GANs have achieved great performance in synthesizing realistic images. In addition, many researchers utilize the output value of discriminator as adversarial loss during the training stage for various low-level vision tasks, such as image to image translation (Isola et al. 2017; Liu and Tuzel 2016), super-resolution (Ledig et al. 2017), single image dehazing (Deng et al. 2020) and image deraining (Zhang et al. 2019).

Chapter 3

Proposed Method

In this section, we first describe the overall network architecture (shown in Figure 3.1) and explain the sense of two-branch designing. Then, we introduce the concept of the discrete wavelet transform (DWT) and analyze the benefits of using DWT in our pipeline. In the end, we further demonstrate the loss functions adopted in the training stage.

3.1 Network Architecture

The two-branch designed network has been successfully applied in various computer vision tasks (Wang et al. 2018; Li et al. 2017a). By using this architecture, each network branch can have its own information processing procedures and extract different representations from the same input. In observing that, if we can use such distinct information wisely and make them complement each other by proper fusion strategies, sufficient and comprehensive information from two branches can greatly boost the performance of image dehazing. Based on this idea, we design a two-branch neural network.

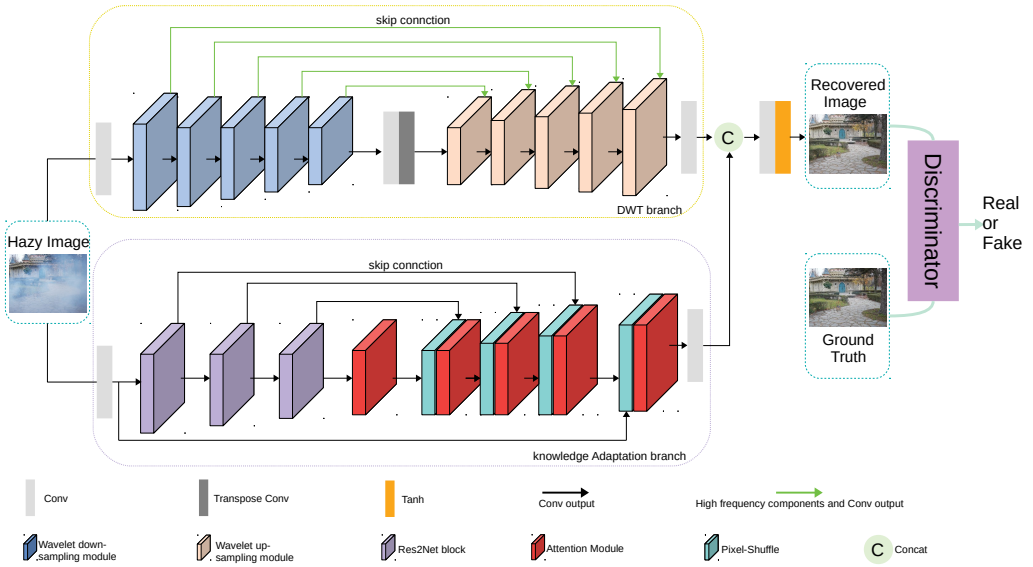


Figure 3.1: The network structure of the proposed method. The generator is a two-branch network, which consists of the DWT branch and the knowledge adaptation branch. The same color used in the cubic denotes the same operation.

3.1.1 DWT Branch

Our first branch, i.e., the DWT branch (shown in Figure 3.1), is designed to directly learn the color mapping from hazy to haze-free images. To achieve this, we follow the U-Net (Ronneberger et al. 2015) to construct our DWT branch. It has an encoder, a decoder and massive skip connections at each feature scale. To meet our requirements for preserving more texture details and reduce the redundant convolution operation during the dehazing process, we propose to use the discrete wavelet transform (DWT) in the feature extraction stage. Since the input feature maps can be decomposed into the low-frequency and high-frequency components by the DWT (detailed explanation can be seen in Section3.2), our network can be guided to learn from both high-frequency and low-frequency information. As shown in Figure 3.2, low-frequency components are concatenated with the convolution output as down-sampling features and high-frequency components are

Figure 3.2: Left:DWT down-sampling module. Right:DWT up-sampling module.

added to the DWT up-sampling module by the skip connection. By doing this, our network not only learns abundant information from both spatial domain and frequency domain but also retains favorable image details by the high-frequency skip connection. Besides, using the DWT operation to enlarge the receptive field makes the model become lightweight, which boosts the model performance on a small dataset to a great extent.

However, due to the limited data in non-homogeneous dehazing tasks, it is hard to achieve plausible performance solely relying on the DWT branch. Towards better performance on small-scale datasets, we introduce our second branch to utilize the additional knowledge further.

3.1.2 Knowledge Adaptation Branch

Our second branch, i.e., the knowledge adaptation branch (shown in Figure 3.1), focuses on adopting the prior knowledge gained from the image classification task to the current dehazing task. It leverages the power of transfer learning (Donahue et al. 2014; Zeiler and Fergus 2014) and brings extra information to the small dataset. To achieve this, we use the ImageNet (Deng et al. 2009) pre-trained Res2Net (Gao et al. 2019) as the backbone of our encoder.

The architecture of the Res2Net block is shown in Figure 3.3. Input features are divided into four components, we denote these components as feature1, feature2, feature3 and feature4 respectively. Feature1 avoids using the convolution operation, which aims to preserve the original input information. Feature maps in feature2 not only doing the convolution operation but also concat to feature3 and continue extracting features. The operation of feature3 and feature4 is quite similar to that in feature2. By doing this, the Res2Net block can get multiple receptive fields for input features. Such a multi-scale feature extraction structure boosts the feature extraction ability of our proposed knowledge adaptation branch.

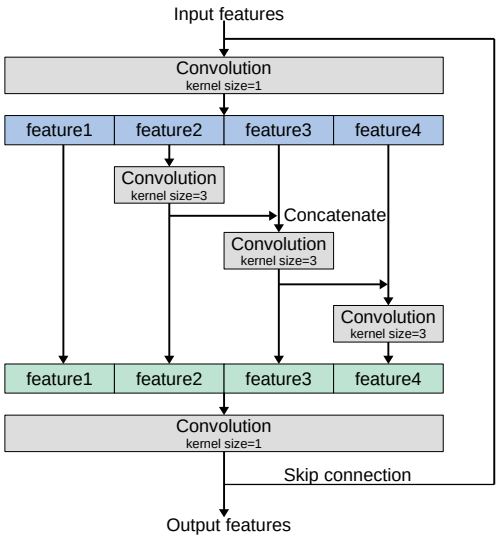


Figure 3.3: The network structure of Res2Net block. Input features are divided into four components to do convolution operation and get multiple receptive fields.

In the decoder module, we use pixel-shuffle layers for up-sampling, which reduces the computational overload (Shi et al. 2016) and makes the size of feature maps gradually recovered to the original resolution.

Besides, inspired by (Qin et al. 2020), attention blocks are employed after each pixel-shuffle layer to identify the dynamic haze patterns. The attention block is composed of the channel attention block, the pixel attention block and convolution layers. Using global average pooling to calculate represent values G_c for each channel of input feature maps F_c:

$$G_c = H_p\left(F_c\right) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i, j), \tag{3.1}$$

where $X_c(i, j)$ stands for the value of c -th channel X_c at position (i, j), H_p is the global average pooling operation. After the pooling operation, the shape of the input feature map changes from $C \times H \times W$ to $C \times 1 \times 1$, which means each channel can use 1×1 value to represent. Using the sigmoid function to activate each representation, we can get different values to represent the importance of each channel. Similarly, the pixel attention block uses simple convolution layers to compress the input feature maps from $C \times H \times W$ to $1 \times H \times W$, which means we can use one compressed feature map to represent all input feature maps after the convolution operation. The rest operations of the pixel-wise attention block are same as that in the channel attention block. Finally, the attention module is designed to combine the channel-wise attention block and the pixel-wise attention block, which can treat feature maps and pixels unequally. By using the attention module block, additional flexibility is provided to our knowledge adaptation branch in dealing with different types of information. The network structure of the attention module is shown in Figure 3.4.

In the end, multiple skip connections are added between the encoder and the decoder as shown in Figure 3.1. In this way, our DW-GAN becomes much more

robust and has better generalization ability. Finally, we add a simple 7×7 convo-

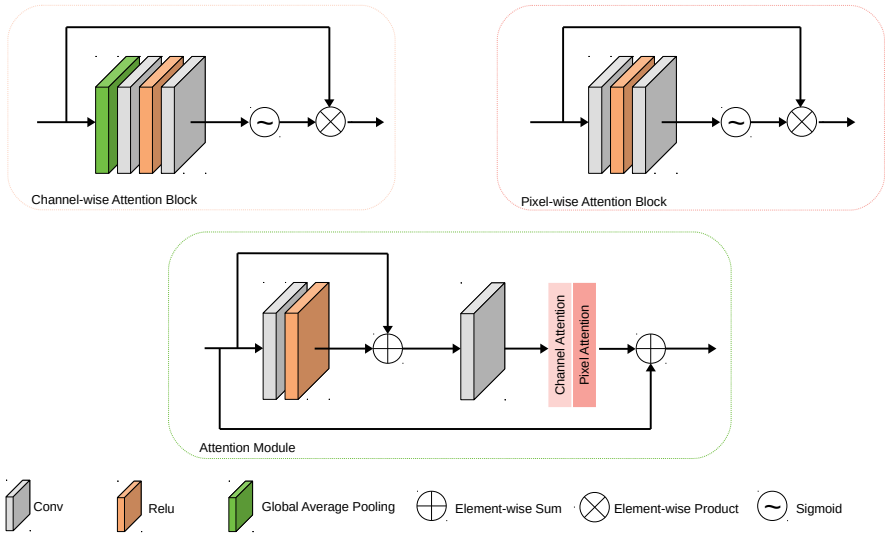


Figure 3.4: Attention module.

lution layer as a fusion operation to map the combined features from two branches to clear images.

3.1.3 Discriminator

We use the discriminator (Goodfellow et al. 2014) in our network pipeline. The discriminator follows the architectural guidelines summarized by (Ledig et al. 2017) and uses LeakyReLU as the activation function. Instead of using max-pooling, the discriminator uses global average pooling to preserve more global information. The trained discriminator architecture is used to differentiate between the dehazed images and clear counterparts. With this approach, our generator can learn to generate dehazed images that are highly similar to real images. The network structure of the discriminator is shown in Figure 3.5.

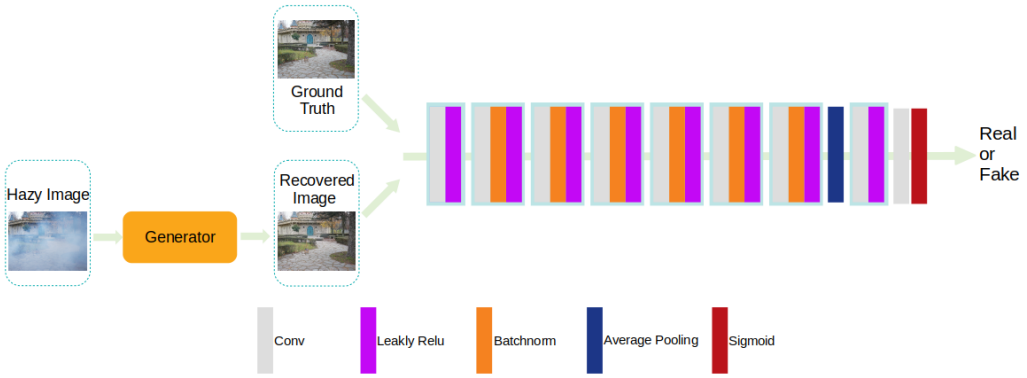


Figure 3.5: The architecture of the discriminator in DW-GAN.

3.2 Discrete Wavelet Transform

In 2D discrete wavelet transform, there are four filters, i.e., low-pass filter f_{LL} and high-pass filters f_{LH}, f_{HL}, f_{HH}. These filters have fixed parameters and can be written as:

$$\mathbf{f}_{LH} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathbf{f}_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{f}_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \mathbf{f}_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{3.2}$$

By convolving with these four filters, each input feature map and image can be decomposed into four subbands, i.e., x_{LL}, x_{LH}, x_{HL} and x_{HH}. We can define x_{LL} as $(f_{LL} \otimes x) \downarrow_2$, where $\otimes$ represents the convolution operation, x is the input feature map and $\downarrow_2$ indicates down-sampling by the scale factor of 2. Use low-pass filter as an example, the value of pixel (i, j) of x_{LL} after 2D Haar wavelet transform can be defined as:

$$x_{LL}(i,j) = x(2i-1, 2j-1) + x(2i-1, 2j) + x(2i, 2j-1) + x(2i, 2j), \tag{3.3}$$

where x is the original input feature maps or hazy images.The expressions of x_{LH}, x_{HL} and x_{HH} are similar to that of x_{LL}. By using DWT, we can obtain the frequency domain information that retains hazy image details, especially from x_{LH}, x_{HL} and x_{HH}. Also, we can reduce the network redundancy by using low-pass filters instead of convolution operations. However, only using the DWT to do image dehazing is not enough. We thus combine these frequency domain operations with convolution operations so that the network can learn from both spatial and frequency information. Experiment results show the improvement of using the DWT method (see details in Section4.3).

3.3 Loss Functions

We denote our dehazed image as $\hat{I}$. I^{gt} and I^{hazy} are respectively the ground truth image and the hazy image. We use G and D to represent the two-branch dehazing network and the discriminator respectively.

Smooth L1 Loss. $\hat{I}_c(i)$ and $I_c^{gt}(i)$ denote the intensity of the c-th channel of pixel i in the dehazed image and the ground truth image respectively. N denotes the total number of pixels. The smooth L1 loss can be defined as:

$$\mathcal{L}_{\text{smooth-L1}} = \frac{1}{3N} \sum_{i=1}^{N} \sum_{c=1}^{3} \alpha \left(\hat{I}_c(i) - I_c^{gt}(i) \right), \tag{3.4}$$

where

$$\alpha(e) = \begin{cases} 0.5e^2, \text{ if } |e| < 1 \\ |e| - 0.5, \text{ otherwise .} \end{cases} \tag{3.5}$$

Perceptual Loss. Besides the pixel-wise supervision, we use the VGG16 (Simonyan and Zisserman 2014) pre-trained on ImageNet (Deng et al. 2009) as the loss network to measure perceptual similarity. The loss function is defined as:

$$\mathcal{L}_{\text{per}} = \sum_{j=1}^{3} \frac{1}{C_j H_j W_j} \left\| \phi_j(I^{gt}) - \phi_j(\hat{I}) \right\|_2^2, \tag{3.6}$$

where H_j , W_j , and C_j denote the height, the width, and the channel of the feature map in the j-th layer of the backbone network, ϕ_j is the activation of the j-th layer. I^{gt} and $\hat{I}$ are respectively the ground truth image and our dehazed result.

MS-SSIM Loss. Let O and G denote two windows of common size centered at pixel i in the dehazed image and the haze-free image, respectively. Use a Gaussian filter to O and G. Then compute the resulting means μ_O, μ_G, standard deviations σ_O, σ_G and covariance σ_{OG}. The SSIM for pixel i is defined as:

$$\text{SSIM}(i) = \frac{2\mu_O\mu_C + C_1}{\mu_O^2 + \mu_G^2 + C_1} \cdot \frac{2\sigma_{OG} + C_2}{\sigma_O^2 + \sigma_G^2 + C_2} = l(i) \cdot cs(i), \tag{3.7}$$

where $l(i)$ represents luminance. $cs(i)$ represents contract and structure measures. C_1, C_2 are two variables to stabilize the division with weak denominator. The MS-SSIM loss is computed using M levels of SSIM. Specifically, we have

$$\mathcal{L}_{\text{MS-SSIM}} = 1 - \text{MS-SSIM}, \tag{3.8}$$

where

$$\text{MS-SSIM} = l_M^{\alpha_m}(i) \cdot \prod_{m=1}^{M} cs_m^{\beta_m}(i), \tag{3.9}$$

with α_m and β_m being default parameters. MS-SSIM loss considers multiple scales of similarity. The image size of each scale is reduced to half of the previous scale. Different scales have different α_m and β_m value, where $\beta_1 = \gamma_1 = 0.0448$, $\beta_2 = \gamma_2 = 0.2856$, $\beta_3 = \gamma_3 = 0.3001$, $\beta_4 = \gamma_4 = 0.2363$ and $\beta_5 = \gamma_5 = 0.1333$.

Adversarial Loss. The adversarial loss l_{adv} is defined based on the probabilities of the discriminator $D(G(I^{hazy}))$ over all training samples as:

$$\mathcal{L}_{adv} = \sum_{n=1}^{N} -\log D(G(I^{hazy})). \tag{3.10}$$

Here, $D(G(I^{hazy}))$ is the probability of reconstructed image $G(I^{hazy})$ to be a haze-free image.

Discriminator Loss. The discriminator loss l_{dis} is used for update the parameters in the discriminator over all training samples. It can be defined as:

$$\mathcal{L}_{dis} = \sum_{n=1}^{N}(1 - D(I^{gt}) + D(G(I^{hazy}))), \tag{3.11}$$

where, $D(I^{gt})$ is the probability of the real image and $D(G(I^{hazy}))$ is the probability of the generated image.

Total Loss. We combine the smooth L1 Loss, perceptual loss, MS-SSIM loss and adversarial loss to supervise the training of our dehazing network.

$$\mathcal{L}_{total} = \mathcal{L}_{\text{smooth-L1}} + \alpha\mathcal{L}_{\text{MS-SSIM}} + \beta\mathcal{L}_{per} + \gamma\mathcal{L}_{adv}, \tag{3.12}$$

where $\alpha = 0.2$, $\beta = 0.001$ and $\gamma = 0.005$ are the hyperparameters weighting for each loss function.

Chapter 4

Experiments

In this section, we firstly describe the datasets that are used for evaluating the effectiveness of our proposed method. Secondly, we introduce our experimental settings, i.e., implementation details and evaluation metrics. Then, we conduct ablation studies to illustrate the benefits of each component in DW-GAN. After that, we compare the performance of our proposed method with other state-of-the-art methods qualitatively and quantitatively.

4.1 Datasets

We trained and evaluated our model on various datasets. These datasets contain different scenes and haze patterns. Indoor hazy images, outdoor hazy images, synthesized hazy images and real-word hazy images are all included in our experiments.

4.1.1 Synthetic Dataset

The Indoor Training Set (ITS) of RESIDE(Li et al. 2018) contains 1399 clean images and 13990 hazy images, generated by corresponding clean images with the medium extinction coefficient β chosen uniformly from $[0.6, 1.8]$ and the global atmospheric light A chosen uniformly from $[0.7, 1.0]$. These images are all indoor scenes. We use the ITS dataset to train our network. For testing, the Synthetic Objective Testing Set (SOTS) is adopted, which contains 500 indoor image pairs.

4.1.2 Real-world Dataset

We further evaluate our performance on three small-scale real-world datasets: DENSE-HAZE (Ancuti et al. 2019), NH-HAZE (Ancuti et al. 2020b; Ancuti et al. 2020a) and NH-HAZE2 (Ancuti et al. 2021). DENSE-HAZE is characterized by dense and homogeneous hazy scenes. It contains 45 training data, 5 validation data and 5 testing data. In our work, we use the official testing data for evaluation and combine the official training set and evaluation set as our training set. NH-HAZE contains 45 training data, 5 validation data and 5 testing data. The haze pattern in this dataset is un-uniformly distributed. We use 45 training pairs and 5 validation pairs as the training set, and use 5 testing pairs as the testing set. NH-HAZE2 is introduced in the NTIRE2021 Dehazing Challenge. It only contains 25 training data, 5 validation data and 5 testing data. Because the validation and testing set is not public by far, we use images 1-20 as the training set and 21-25 as the testing set.

4.2 Implementation

To quantitatively evaluate the performance of our method, we adopt two common metrics: the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) as our evaluation criteria. During the experiment process, we constantly adjust our hyperparameters and find out proper values.

4.2.1 Experiment settings

Despite the varied characteristics of each dataset, we adopt the same training strategy for all datasets. Specifically, we randomly crop patches with a size of 256 $\times$ 256. To augment training data, we implement random rotation (90, 180 or 270 degrees) and random horizontal flip. We train DW-NET with the batch size of 16 and utilize the Adam optimizer (Kingma and Ba 2014) (β_1=0.9, β_2=0.999). In the training process, a specific decay strategy is used, where the initial learning rate is set to 1e-4 and decays 0.5 times at 3000, 5000, 6000 epoch for a total of 10000 epochs. The discriminator uses the same optimizer and training strategies. All the experiments are conducted on two NVIDIA 1080Ti GPUs.

4.2.2 Model convergence

Taking the model training process on NH-HAZE2 as an example, as shown in Figure 4.1, the loss of the model dropped to 0.0004 after 20,000 iterations, which is a good indication of convergence. Similarly, the PSNR and SSIM are gradually rising during the 10,000 epoch, which indicates that the parameters in the model have been well optimized and the model does not achieve overfitting. PSNR and SSIM curves are shown in Figure 4.2.

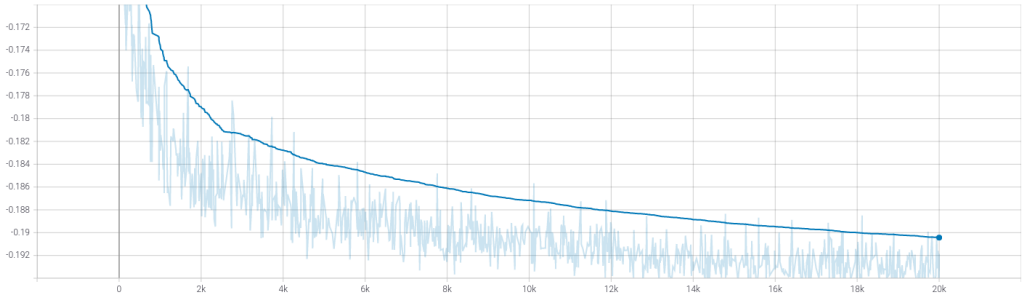


Figure 4.1: The loss curve.

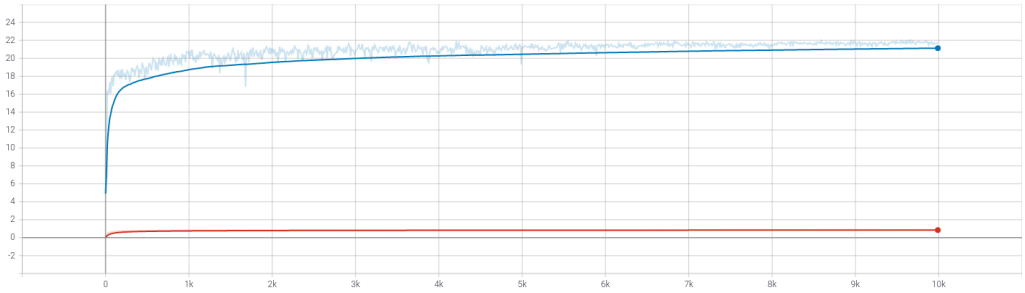


Figure 4.2: PSNR and SSIM cureves. The blue curve denotes PSNR values and the red curve denotes SSIM values.

4.3 Abalation Study

Firstly, we conduct comprehensive ablation studies to demonstrate the necessity of each component in our proposed method. According to the ablation principle, we construct four different networks to illustrate the importance of each module. (1) vanilla DWT branch: only use the vanilla DWT branch without DWT down-sampling modules and high-frequency skip connections. (2) knowledge adaptation branch: only use the knowledge adaptation branch to restore hazy images. (3) Two-branch: use a two-branch structure, which consists of vanilla DWT branch and knowledge adaptation branch. (4) Two-branch+DWT: use a two-branch structure, where DWT down-sampling and up-sampling modules are embedded into the

Methods	l_1	l_p	L_{SSIM}	l_{adv}	PSNR	SSIM
(1)vanilla DWT branch	√	×	×	×	18.15	0.7483
(2)knowledge adaptation branch	√	×	×	×	20.15	0.8156
(3)Two-branch	√	×	×	×	<u>21.35</u>	<u>0.8273</u>
(4)Two-branch+DWT	√	×	×	×	**21.52**	**0.8403**
(5)Two-branch+DWT	√	√	×	×	21.67	0.852
(6)Two-branch+DWT	√	√	√	×	<u>21.86</u>	<u>0.8555</u>
(7)Two-branch+DWT	√	√	√	√	**21.99**	**0.856**

Table 4.1: Ablation Studies for architectures and loss functions. It can be observed that the model with all components and supervised by all loss functions performs the best in terms of PSNR and SSIM.

vanilla DWT branch. From the top of Table 4.1, we can observe that using a two-branch structure can significantly improve our performance in terms of PSNR and SSIM (by comparing (1), (2) and (3)). The reason is that the two-branch network can not only learn the mapping directly between hazy and haze-free image pairs via the vanilla DWT branch but also adapt the pre-learned knowledge to the current task by the knowledge adaptation branch. To demonstrate the effectiveness of the discrete wavelet transform, we compare cases that the two-branch network adopts DWT or not. By observing the performance of (3) and (4), we can conclude that DWT plays an important role in improving PSNR and SSIM. The increased SSIM also indicates that the frequency domain information is essential for restoring texture details. Besides, we further illustrate the importance of the loss functions adopted in this work. In observing the fourth to seventh rows of Table 4.1, each loss is effective and vital to raising PSNR and SSIM. Smooth L1 loss provides pixel-wise supervision, perceptual loss lets the outputs tend to be consistent with ground truth in deep feature space, MS-SSIM loss is employed for minimizing the structural similarity error and GAN loss further improves the output results. From the loss curve in Figure 4.3, we can see that the discriminator

in the initial stage can not discriminate well, fake images and real images have the same output value. However, with the continuous optimization and update of the parameters in the discriminator, the discriminator has good performance on distinguishing between dehazed images and real images. When using real images as input, the output value of the discriminator is close to 1. When using fake images as input, the output value of the discriminator is close to 0. The trained discriminator provides an effective reference to optimize the parameters in the generator.

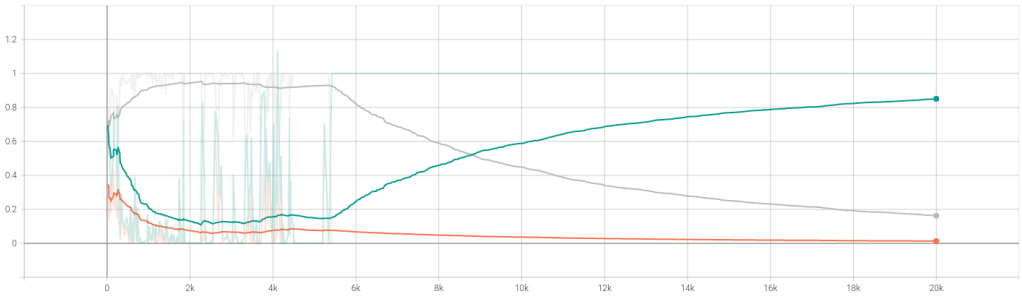


Figure 4.3: The gray curve denotes the output values of dehazed images. The green curve denotes the output values of real images. The red curve denotes the adversarial loss.

By integrating all the losses on the training stage, our model acquired the best performance (see the last row in Table 4.1).

4.4 Comparisons with State-of-the-art Methods

We compare the proposed method with state-of-the-art methods on a synthetic dataset and real-world datasets. These SOTA methods include DCP (He et al. 2010), AOD-Net (Li et al. 2017b), GCANet (Chen et al. 2019), FFA (Qin et al.

2020) and TDN (Liu et al. 2020). TDN is the winner method in NTIRE 2020 NonHomogeneous Dehazing Challenge.

4.4.1 Inference Time Comparison.

We compare the inference time with these SOTA methods for processing one 1600 × 1200 image by an NVIDIA 1080Ti GPU. As shown in Figure 4.4, AOD-Net and DCP take less time to complete dehazing processing. However, these two methods cannot remove haze effectively (details have been discussed in Section4.4.2 and Section4.4.3). It is a decisive fact that our proposed method takes less running time than GCANet, FFA and TDN. Meanwhile, our approach has better performance both qualitatively and quantitatively.

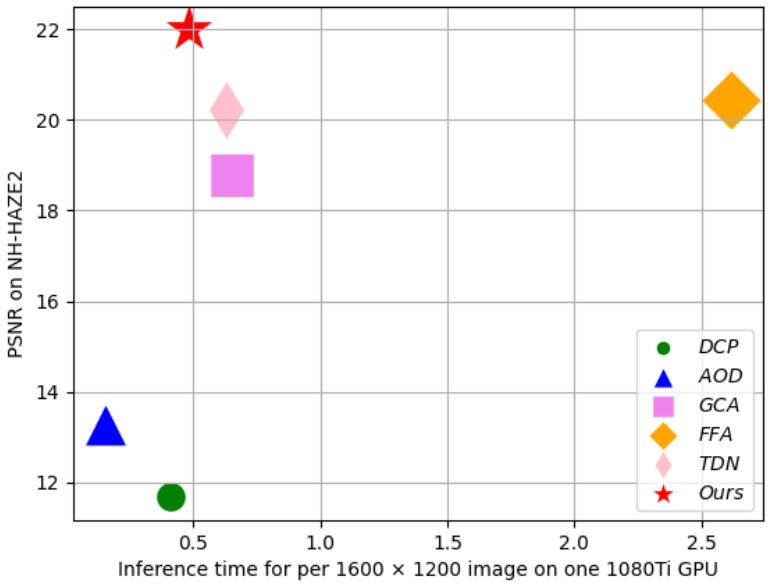


Figure 4.4: We compare the performance of our model with DCP, AOD, GCA, FFA and TDN on NH-HAZE2 dataset.

Methods	DCP	AOD-Net	GCANet	FFA	TDN	Ours
PSNR	19.63	19.06	30.23	**36.39**	34.59	<u>35.94</u>
SSIM	0.8823	0.8524	0.9814	**0.9886**	0.9754	<u>0.9860</u>

Table 4.2: Quantitative comparisons of SOTA methods over the SOTS dataset. The best results are in **bold**, and the second best are with <u>underline</u>.

4.4.2 Experiment results on Synthetic Dataset

Quantitative Results Comparison. In the synthetic dataset, the performance of our model is slightly lower than that of FFA. The success on the large-scale benchmarks often requires heavy network designing. FFA does all full resolution feature mapping, which results in a huge amount of computation. In contrast, we aim to build a suitable model to balance between good mapping capability and over-fitting. But surprisingly, we still perform the second best and approach to FFA. The experiment results are shown in Table 4.2.

Qualitative Visual Effect Comparison. The dehazing results of DCP and AOD are relatively poor. The dehazed images have many artifacts and the haze removal is incomplete. GCANet performs better than the above two methods but still suffers from severe color distortion. FFA, TDN and our proposed method have better dehazing effects. Our proposed method has the best visual effect in part of the dehazing results in the SOTS dataset. The dehazed images are shown in Figure 4.5.

4.4.3 Experiment results on Real-world Dataset

Quantitative Results Comparison. For three different real-world datasets, our method has outstanding performance and achieves the best in terms of PSNR

	NTIRE19		NTIRE20		NTIRE21	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP	11.06	0.4368	13.28	0.4954	11.68	0.7090
AOD-Net	13.21	0.4694	13.44	0.4136	13.30	0.4693
GCANet	12.46	0.4712	17.49	0.5918	18.79	0.7729
FFA	<u>16.31</u>	<u>0.5362</u>	18.60	0.6374	<u>20.45</u>	<u>0.8043</u>
TDN	15.50	0.5081	<u>20.44</u>	<u>0.6683</u>	20.23	0.7622
Ours	**16.49**	**0.5911**	**21.51**	**0.7111**	**21.99**	**0.8560**

Table 4.3: Quantitative comparisons of SOTA methods over DENSE-HAZE, NH-HAZE and NH-HAZE2. The best results are in **bold**, and the second best are with <u>underline</u>.

and SSIM. It is worth noticing that our model has first-class performance on non-homogeneous dehazing and surpasses the second-ranked model by a large margin (1.07dB and 1.54dB higher on NH-HAZE and NH-HAZE2, respectively).

Qualitative Visual Effect Comparison. DCP gets much bluer results on real-world datasets. The output results of AOD-Net can only remove thick haze. GCANet and FFA perform much better than the above two methods, they still fail to handle the hazy zones. GCANet tends to generate blurry and color distorted images, and it is unable to remove haze in DENSE-HAZE dataset. FFA performs comparably worse in the non-homogeneous dehazing task. For example, in NH-HAZE and NH-HAZE2, FFA cannot remove haze effectively and produce unpleasant artifacts. Surprisingly, TDN shows unsatisfied results in DENSE-HAZE and NH-HAZE2. For example, a considerable color deviation between dehazed images and ground truths can be observed in the DENSE-HAZE dataset. The brightness of dehazed images is much darker, and image details are not restored well on NH-HAZE2. It is worth pointing out that our proposed method performs well on all the datasets, which further reveals the robustness of our model. It can

be seen that our dehazed images are visually pleasing and closest to the ground truths. The dehazed images are shown in Figure 4.6, Figure 4.7 and Figure 4.8.

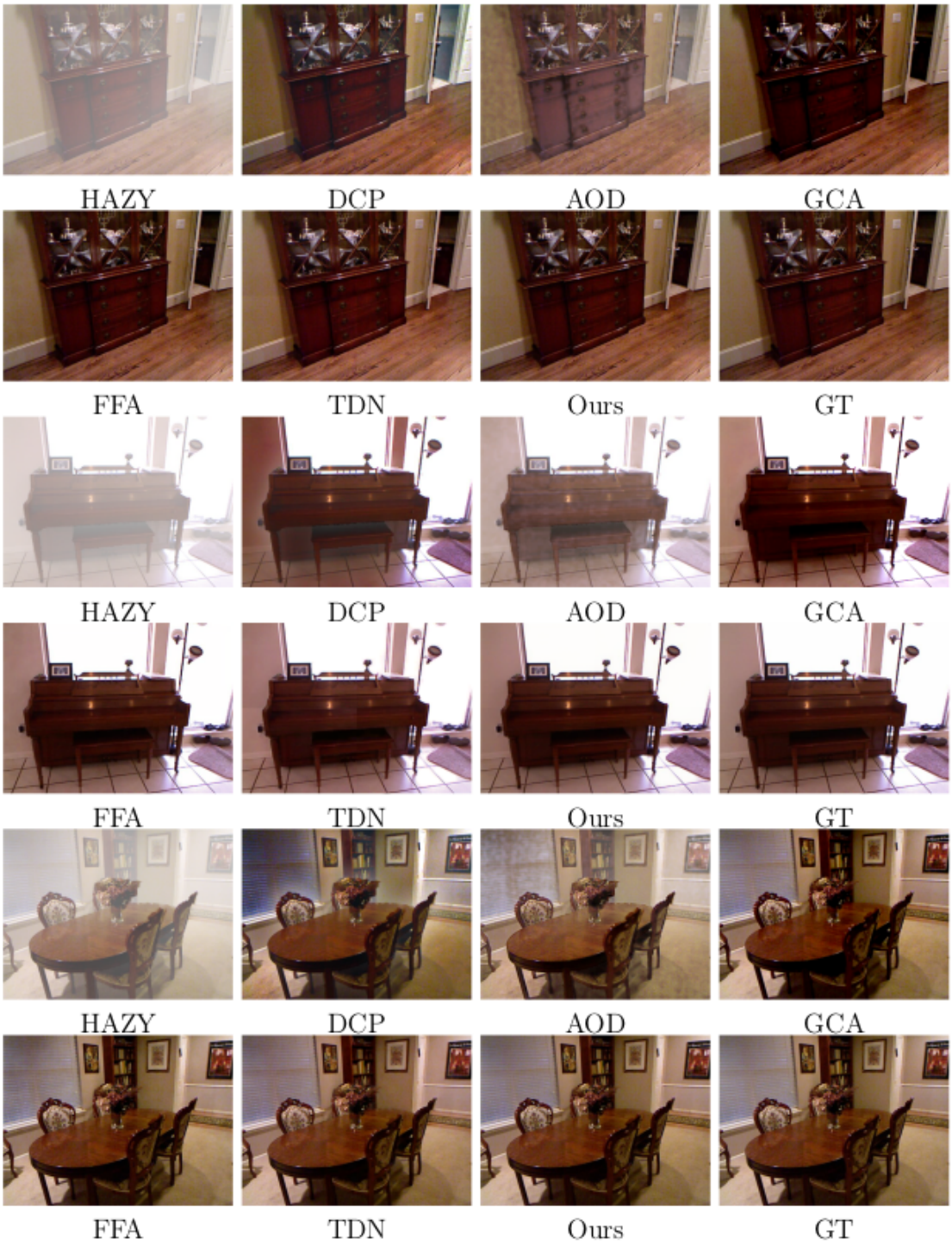


Figure 4.5: Qualitative evaluation examples of the RESIDE SOTS indoor testing dataset.

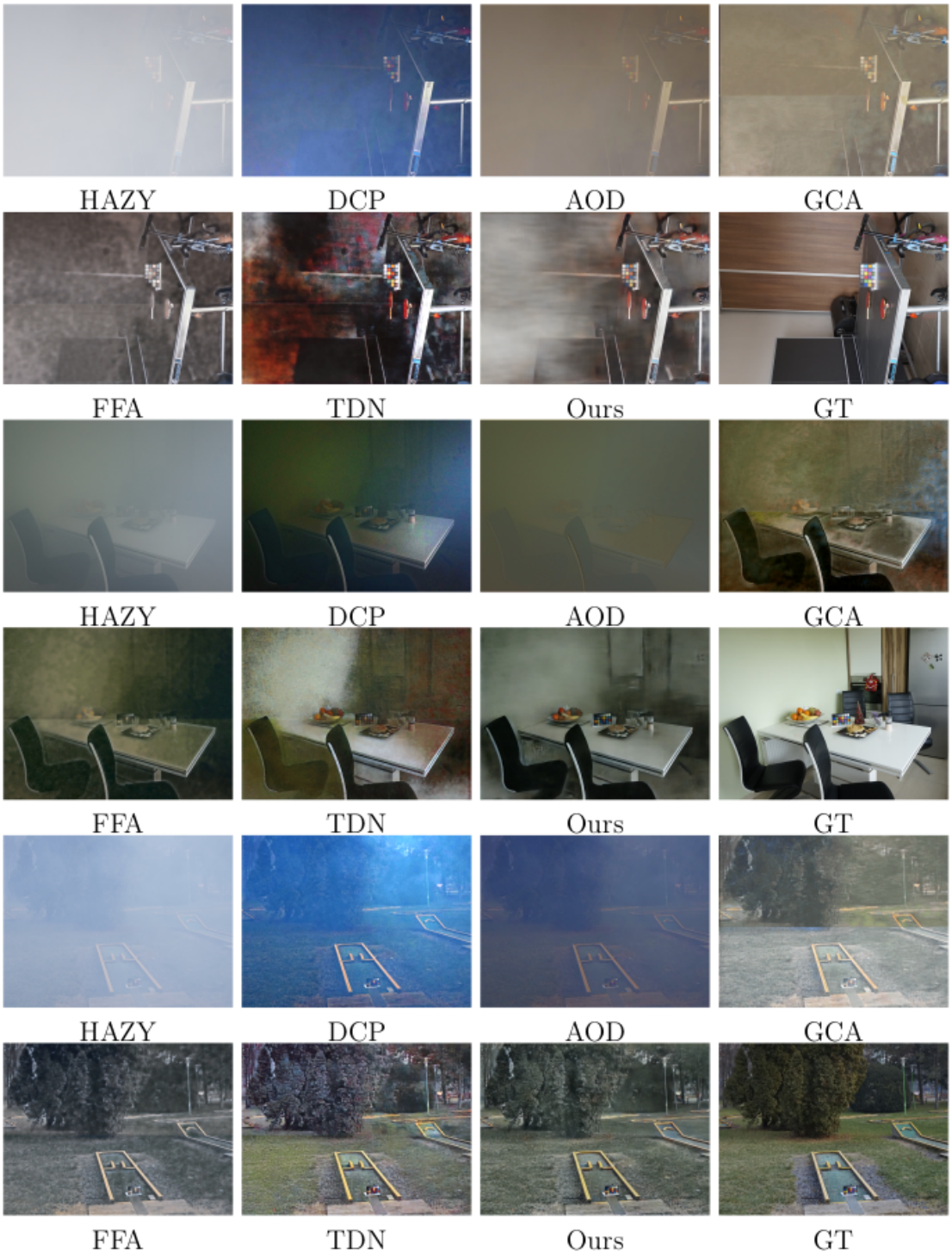


Figure 4.6: Qualitative evaluation examples of our method with others on the Dense Haze dataset.

Figure 4.7: Qualitative evaluation examples of our method with others on the NH-HAZE dataset.

Figure 4.8: Qualitative evaluation examples of our method with others on the NH-HAZE2 dataset.

Chapter 5

Ntire 2021 Dehazing Challenge

5.1 Discussion of Data Pre-processing

The NTIRE2021 NonHomogeneous Dehazing Challenge provides limited data (only 25 training image pairs). To augment training data, we mix image pairs from NH-HAZE, which consists of 55 non-homogeneous hazy images and clear counterparts.

These two datasets only contain image pairs of outdoor scenes and these hazy images are all non-homogeneous haze patterns. Compared with other homogeneous haze datasets, the tonal styles of NH-HAZE image pairs are much more closer to NH-HAZE2 image pairs, which indicates the color mapping in the two datasets is similar. So it is a good choice to use NH-HAZE as extra data to enlarge the training dataset and help boost the performance of DW-GAN. However, directly using NH-HAZE as extra data may also lead to problems, because images in NH-HAZE and NH-HAZE2 have huge differences in terms of brightness. The visual effect of images in NH-HAZE is much darker, while that of NH-HAZE2 is brighter.

Figure 5.1: Illustration of the gamma corrected the clear image in the NH-HAZE dataset. Left:before the gamma correction. Right:after the gamma correction

To further verify our observation, we quantitatively analyzed the grayscale distribution of the two datasets (see in Figure 5.2 (left)). The average gray value of all haze-free images in NH-HAZE is 102.30 and the variance is 62.42, while in NH-HAZE2, the statistics are 131.45 and 57.45 separately. Due to the difference in brightness, if we simply adopt the model trained with NH-HAZE as extra data to restore hazy images in NH-HAZE2, the average brightness of these dehazed images should be lower than 131.45 and higher than 102.3. The inaccurate brightness estimation may result in unsatisfied performance.

We aim to boost the performance on NH-HAZE2 test dataset, so to reduce the brightness discrepancy between the two datasets, we use the gamma correction on NH-HAZE. When the gamma value is set to 0.65, the average gray value of NH-HAZE is shifted to 133.30 and the variance is changed to 57.78. With the preprocessing, images in NH-HAZE become brighter(shown in Figure 5.1) and the grayscale distribution of NH-HAZE is much more similar to that of NH-HAZE2 (see in Figure 5.2 (right)).

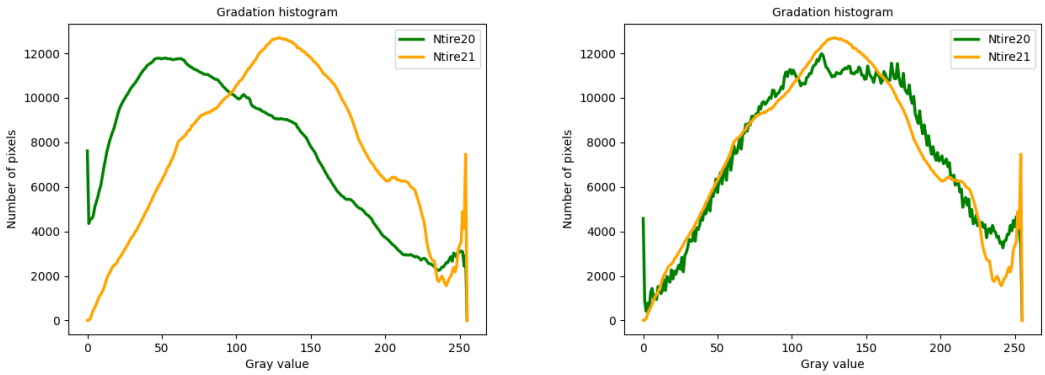


Figure 5.2: Left:comparison of data distribution without the gamma correction. Right:comparison of data distribution with the gamma correction.

5.2 Performance on Dehazing Challenge

From the reported results (Ancuti et al. 2021), our DW-GAN is among the top performed methods in terms of PSNR and SSIM. To be specific, our dehazed results achieve plausible PSNR (21.6815dB), SSIM (0.8344dB) during the validation phase and PSNR (21.08dB), SSIM (0.8393) during the testing phase. To visually demonstrate our performance, we show the dehazing results of our DW-GAN in Figure 5.3 and Figure 5.4. It can be observed that our DW-GAN can remove most of the haze and generate visually pleasing results.

Figure 5.3: Our dehaazing results on NH-HAZE2 validation set.

Figure 5.4: Our dehaazing results on NH-HAZE2 test set.

Chapter 6

Conclusion

In this paper, we propose a novel generative adversarial network for single image dehazing, namely DW-GAN. The DWT-branch directly learns the image mapping from hazy to haze-free images and leverages the power of discrete wavelet transform in helping the network acquire more frequency domain information. The knowledge adaptation branch exploits the prior knowledge by using the pre-trained Res2Net as an encoder. Extra information from the heterogeneous task, i.e., image classification, is introduced to complement the small-scale datasets, which allows our DW-GAN to be much more robust in dealing with limited real-world data. Extensive experimental results illustrate that DW-GAN has great performance in synthetic datasets, real-world scenes with dense haze and non-homogeneous haze.

Bibliography

Ancuti, C. O., Ancuti, C., Sbert, M., and Timofte, R. (2019). Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In: *2019 IEEE international conference on image processing (ICIP)*. IEEE, 1014–1018.

Ancuti, C. O., Ancuti, C., and Timofte, R. (2020a). NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 444–445.

Ancuti, C. O., Ancuti, C., Vasluianu, F.-A., Timofte, R., et al. (2021). NTIRE 2021 NonHomogeneous Dehazing Challenge Report. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Ancuti, C. O., Ancuti, C., Vasluianu, F.-A., and Timofte, R. (June 2020b). NTIRE 2020 Challenge on NonHomogeneous Dehazing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Bae, W., Yoo, J., and Chul Ye, J. (2017). Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 145–153.

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

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

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

Dai, L., Liu, X., Li, C., and Chen, J. (2020). AWNet: Attentive Wavelet Network for Image ISP. In: *Computer Vision – ECCV 2020 Workshops*. Ed. by A. Bartoli and A. Fusiello. Cham: Springer International Publishing, 185–201. ISBN: 978-3-030-67070-2.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

Deng, Q., Huang, Z., Tsai, C.-C., and Lin, C.-W. (2020). HardGAN: A Haze-Aware Representation Distillation GAN for Single Image Dehazing. In: *European Conference on Computer Vision*. Springer, 722–738.

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

Dong, Y., Liu, Y., Zhang, H., Chen, S., and Qiao, Y. (2020). FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07, 10729–10736.

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

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

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.

Guo, T., Seyed Mousavi, H., Huu Vu, T., and Monga, V. (2017). Deep wavelet prediction for image super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 104–113.

He, K., Girshick, R., and Dollár, P. (2019). Rethinking imagenet pre-training. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4918–4927.

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

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Katyal, S., Kumar, S., Sakhuja, R., and Gupta, S. (2018). Object detection in foggy conditions by fusion of saliency map and YOLO. In: *2018 12th International Conference on Sensing Technology (ICST)*. IEEE, 154–159.

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

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature* 521(7553), 436–444.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690.

Li, B., Luo, H., Zhang, H., Tan, S., and Ji, Z. (2017a). A multi-branch convolutional neural network for detecting double JPEG compression. *arXiv preprint arXiv:1710.05477*.

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

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., and Wang, Z. (2018). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing* 28(1), 492–505.

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

Liu, M.-Y. and Tuzel, O. (2016). Coupled generative adversarial networks. *arXiv preprint arXiv:1606.07536*.

Liu, P., Zhang, H., Zhang, K., Lin, L., and Zuo, W. (2018). Multi-level wavelet-CNN for image restoration. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 773–782.

Long, J., Shi, Z., Tang, W., and Zhang, C. (2013). Single remote sensing image dehazing. *IEEE Geoscience and Remote Sensing Letters* 11(1), 59–63.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence* 11(7), 674–693.

Middleton, W. E. K. (2019). *Vision through the atmosphere*. University of Toronto Press.

Ni, W., Gao, X., and Wang, Y. (2016). Single satellite image dehazing via linear intensity transformation and local property analysis. *Neurocomputing* 175, 25–39.

Qin, X., Wang, Z., Bai, Y., Xie, X., and Jia, H. (2020). FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07, 11908–11915.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28, 91–99.

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

Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., and Yang, M.-H. (2018). Gated fusion network for single image dehazing. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3253–3261.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). 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*, 1874–1883.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sindagi, V. A., Oza, P., Yasarla, R., and Patel, V. M. (2020). Prior-based domain adaptive object detection for hazy and rainy conditions. In: *European Conference on Computer Vision*. Springer, 763–780.

Singh, D. and Kumar, V. (2019). A comprehensive review of computational dehazing techniques. *Archives of Computational Methods in Engineering* 26(5), 1395–1413.

Wang, L., Li, Y., Huang, J., and Lazebnik, S. (2018). Learning two-branch neural networks for image-text matching tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41(2), 394–407.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In: *European conference on computer vision*. Springer, 818–833.

Zhang, H. and Patel, V. M. (2018). Densely connected pyramid dehazing network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3194–3203.

Zhang, H., Sindagi, V., and Patel, V. M. (2019). Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology* 30(11), 3943–3956.

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