Deep Learning Methods for Image Restoration and Reconstruction

by

ZAHRA ANVARI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2021

Deep Learning Methods for Image Restoration and Reconstruction

The members of the Committee approve the doctoral
dissertation of Zahra Anvari

Vassilis Athitsos

Supervising Professor _____

Christopher Conly _____

David Levine _____

Dajiang Zhu _____

Dean of the Graduate School _____

ACKNOWLEDGEMENTS

I am greatly thankful to my supervising professor Dr. Vassillis Athitsos, who I am honored to be one of his Ph.D. students. I greatly respect and appreciate his consistent help, support, guidance, patience, understanding, and insightful advice during my doctoral journey. I would also like to thank my academic committee members: Dr. Dajiang Zhu, and Dr. Chris Conly, and Dr. David Levine for their valuable feedback.

Further, I would like to extend my appreciation to all UTA-CSE faculty/staff members. I am particularly thankful to Dr. Ramez Elamsri, Dr. Bahram Khalili, Camille Costabile, Sherri Gotcher, Pamela Mcbride, and Ginger Dickens for their efforts to ease and smooth all administration process for all graduate students.

Finally, my deep and sincere gratitude to my family for their continuous love, help and support.

April 15, 2021

ABSTRACT

Deep Learning Methods for Image Restoration and Reconstruction

Zahra Anvari, Ph.D.

The University of Texas at Arlington, 2021

Supervising Professor: Vassilis Athitsos

The problem of image reconstruction and restoration refers to recovering the clean images from corrupted ones. Corruption or degradation can occur due to atmospheric conditions such as rain, fog, mist, snow, dust, and air pollution or technical drawbacks of imaging devices such as motion blurriness, compression noise, low-resolution, *etc.* Image reconstruction algorithms aim at reducing these artifacts and degradation and generate clear images. Scenes captured under bad weather conditions such as rain, fog, mist, and haze suffer from visibility issues thus introduce obstacles for computer vision applications, *e.g.* object detection, recognition, tracking, and segmentation.

In this dissertation, we focus on *single image dehazing* problem. In single image dehazing, we would like to restore the haze free image from a hazy image. Most of the recent image dehazing methods rely on paired datasets, which means for each hazy image there's a single clean/haze-free image as a ground truth. In practice, however, there is a range of clean images that can correspond to a hazy image, due to factors such as contrast or light intensity changes throughout the day. In fact, it is

infeasible to capture both ground truth/clear image and the hazy image of the same scene simultaneously. Thus there is an emerging need to develop solutions that do not rely on the ground truth images and could operate with *unpaired* supervision.

To address this, we first cast the unpaired image dehazing problem to an image-to-image translation problem and then propose a novel cycle-consistent generative adversarial network, called ECDN, that operates without paired supervision and benefits from (i) a global-local discriminator architecture to handle spatially varying haze (ii) an encoder-decoder generator architecture with residual blocks to better preserve the details (iii) skip connections in the generator to improve the performance of the network and convergence (iv) customized cyclic perceptual loss and a self-regularized color loss to generate more realistic images and mitigate the color distortion problem. Through empirical analysis we show that the proposed network can effectively remove haze and generate visually pleasing haze-free images.

In addition, most existing methods assume that haze has a uniform/homogeneous distribution and haze can have a single color, *i.e.* grayish white color similar to smoke, while in reality haze can be distributed non-uniformly with different patterns and colors. To quantify the challenges and assess the performance of these methods, we introduce a sunlight haze benchmark dataset, *Sun-Haze*, containing 107 hazy images with different types of haze created by sunlight having a variety of intensity and color. We evaluate a representative set of state-of-the-art image dehazing methods on this benchmark dataset in terms of standard metrics such as PSNR, SSIM, CIEDE2000, PI and NIQE. This uncovers the limitation of the current methods, and questions their underlying assumptions as well as their practicality.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF SYMBOLS

D	Discriminator
G	Generator
I	Hazy image
J	Haze-free image
l	Luminance
t	Transition map

LIST OF ACRONYMS

AODNet	All-in-One Dehazing Network
CNN	Convolutional Neural Network
DCP	Dark Channel Prior
DDN	Disentangled Dehazing Network
ECDN	Enhanced CyclcGAN Dehazing Network
EPDN	Enhanced Pix2pix Dehazing Network
GAN	Generative Adversarial Network
MSCNN	Multi-scale Convolutional Neural Network
PDF	Probability Density Function
PI	Perceptual Index
PSNR	Peak signal-to-noise ratio
SOTS	Synthetic Objective Testing Set
SSIM	Structural Similarity Index
VAE	Variational Autoencoder

CHAPTER 1

Introduction

Image reconstruction/restoration problem refers to recovering the clean images from corrupted ones. Corruption or degradation can occur due to atmospheric conditions such as rain, fog, mist, snow, dust, and air pollution or technical drawbacks of imaging devices such as motion blurriness, compression noise, low resolution, *etc.* Image reconstruction algorithms aim at reducing these artifacts and degradation and generate clear images. Scenes captured under bad weather conditions such as rain, fog, mist, and haze suffer from visibility issues thus introduce obstacles for computer vision applications, *e.g.* object detection, recognition, tracking, and segmentation.

In this dissertation, we focus on *single image dehazing*. In single image dehazing, we would like to restore the haze free image from a hazy image. Figure 1.1 shows a few hazing images along with their corresponding clean images. Haze is an atmospheric phenomenon that can cause visibility issues, and the quality of images captured under haze can be severely degraded. Hazy images suffer from poor visibility and low contrast, which can challenge both human visual perception and numerous intelligent systems relying on computer vision methods.

The performance of standard computer vision tasks such as object detection [6, 7], semantic segmentation [8], face detection, clustering and dataset creation [9, 10, 11, 12] can be affected significantly when images are hazy. Hence, image dehazing is an essential pre-processing task for general-purpose computer vision algorithms that are fed with hazy images. As a result, single image dehazing has received a great deal of attention over the past decade [13, 14, 15, 16, 17, 13, 14, 15, 16, 18, 17].

(a) Hazy (b) Ground Truth

Figure 1.1: An example of hazy and clean images.

However, haze removal is still a challenging and ill-posed problem and most existing methods make assumptions that do not simply hold in reality. Most of the recent image dehazing methods rely on paired datasets, which means for each hazy image there's a single clean/haze-free image as a ground truth. In practice, however, there is a range of clean images that can correspond to a hazy image, due to factors such as contrast or light intensity changes throughout the day. In fact, it is infeasible to capture both ground truth/clear image and the hazy image of the same scene simultaneously. Thus there is an emerging need to develop solutions that do not rely on the ground truth images and could operate with *unpaired* supervision.

To address this, we first cast the unpaired image dehazing problem to an image-to-image translation problem and propose a novel cycle-consistent generative adver-

sarial network, called ECDN. ECDN operates without paired supervision and benefits from the following major components:

- A global-local discriminator architecture to handle spatially varying haze.
- An encoder-decoder generator architecture with residual blocks to better preserve the details.
- Skip connections in the generator to improve the performance of the network and convergence.
- Customized cyclic perceptual loss and a self-regularized color loss to generate more realistic images and mitigate the color distortion problem.

Through ablation study we show the effectiveness of each of these components. Also through empirical analysis we show that the proposed network can effectively remove haze and generate visually pleasing haze-free images.

Furthermore, most existing methods and datasets assume that i) haze has a uniform and homogeneous distribution in the entire image, and ii) haze can only have a single color, *i.e.* grayish white similar to the color of smoke or pollution. While in reality haze density can change non-homogeneously throughout an image and it can vary in pattern and color.

To quantify the challenges and assess the performance of these methods, we introduce a sunlight haze benchmark dataset, *Sun-Haze*, containing 107 hazy images with different types of haze created by sunlight having a variety of intensity and color. We evaluate a representative set of state-of-the-art image dehazing methods on this benchmark dataset in terms of standard metrics such as PSNR, SSIM, CIEDE2000, PI and NIQE. This uncovers the limitation of the current methods, and questions their underlying assumptions as well as their practicality.

In summary, this dissertation presents the following major contributions:

- We propose a novel cycle-consistent generative adversarial network called ECDN for unpaired image dehazing. ECDN does not rely on any priors such as the physical scattering model, as opposed to many previous methods, and instead it adopts the image-to-image translation approach for unpaired image dehazing.
- We adopt a global-local discriminator structure to deal with spatially varying haze and generate better haze-free images.
- We define a self-regularized color loss and utilize it along with a customized perceptual loss to generate more visually pleasing images with vibrant colors and mitigate the color distortion problem. Self-regularization is vital to our network since in unpaired setting there is no external supervision available.
- We use an encoder-decoder generator architecture with residual blocks with skip connections to better preserve the details.
- Through empirical analysis, we show that our network outperforms the previous work in terms of PSNR and SSIM.
- We present a sunlight haze benchmark dataset, Sun-Haze, that contains 107 hazy images caused by sunlight, along with six ground truth images (five retouched by five experts and one original image before being retouched) per hazy image.
- We perform an extensive analysis to evaluate current state-of-the-art dehazing methods over Sun-Haze in terms of both reference-based and no-reference-based metrics.
- We show that existing dehazing methods can not generalize well when there is sunlight haze, in particular when we have sunlight color changes.

1.1 Dissertation Overview

In Chapter 2, we give an overview of related work in single image dehazing. This includes both traditional and deep learning based methods. We also provide related work for both paired and unpaired image dehazing.

In Chapter 3, we give an overview of GANs and image-to-image translation technique and how they work. We also describe paired vs. unpaired supervision settings.

In Chapter 4, we propose a novel cycle-consistent generative adversarial network, called ECDN, that operates without paired supervision. **Through empirical analysis, we show that the proposed network can effectively remove haze and outperforms current methods in terms of SSIM and PSNR.**

In Chapter 5, we introduce our dataset, Sun-Haze, and present our research on benchmarking current dehazing methods. **Our work reveals the limitation of the current methods, and questions their underlying assumptions as well as their practicality.**

Finally, we conclude the dissertation in Chapter 6 and provide some potential future research directions.

CHAPTER 2

Related Work

Numerous attempts have been done to solve the single image haze removal problem. These methods can be categorized into two main classes: prior-based and learning-based, that we describe them below.

2.0.1 Prior-based dehazing

Prior-based methods are mainly based on prior information and assumptions to recover the haze-free images from hazy images. They heavily depend on estimating the parameters of the physical scattering model [19, 20], *aka.* the atmospheric scattering model, which contains the transmission map and the atmospheric light to solve the haze removal problem. The physical scattering model is formulated as:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{2.1}$$

where $I(x)$ is the hazy image, $J(x)$ is the haze-free image or the scene radiance, $t(x)$ is the medium transmission map, and A is the global atmospheric light on each x pixel coordinates.

He *et al.* [21] proposed a dark channel prior, DCP, to estimate the transmission map effectively. DCP utilizes dark channel prior to more reliably calculate the transmission matrix. With dark channel prior, the thickness of haze is estimated and removed by the atmospheric scattering model.

Moreover, this method is proposed based on experimental statistics of experiments on haze-free images, which shows at least one color channel has some pixels

with very low intensities in most of non-haze patches. However, DCP has poor performance on dehazing the sky images and is computationally intensive.

Tan et al. [22] increase the contrast of hazy images, based on the fact that haze-free images have higher contrast than hazy images.

2.0.2 Learning-based dehazing

Recently learning based methods have been proposed that utilize CNNs and GANs for the single image dehazing problem. CNN-based methods try to recover the clean images through the atmospheric scattering model, by mainly estimating the transmission map and atmospheric light [19, 23].

MSCNN [24] is a learning-based dehazing method. The authors proposed a multi-scale deep neural network for single-image dehazing by learning the mapping between hazy images and their corresponding transmission maps.

This method contains two sub-networks called coarse-scale and fine-scale, to estimate the transmission map. The coarse-scale network estimates the transmission map based on the entire image. The results are further improved locally by the fine-scale network.

DehazeNet [25] takes advantage of both priors and the power of convolutional neural networks. DehazeNet proposed an end-to-end system for medium transmission estimation. It takes a hazy image as input, and outputs its medium transmission map that is later used to recover a haze-free image via atmospheric scattering model. In addition, DehazeNet proposed a CNN-based deep network, which its layers are specially designed to embody the established priors in image dehazing. Authors also proposed a nonlinear activation function called Bilateral Rectified Linear Unit (BReLU), to improve the quality of recovered haze-free image.

In short, DehazeNet modified the classic CNN model by adding feature extraction and non-linear regression layers. These modifications distinguish DehazeNet from other CNN-based models.

AOD-Net [26] is an end-to-end dehazing network which is based on estimating the transmission map through reformulating the atmospheric scattering model. Instead of estimating the transmission matrix and the atmospheric light separately, AODNet directly generates the clean image through a light-weight CNN. AOD-Net can be easily embedded with Faster R-CNN [27] and improve the object detection performance on hazy images with a large margin.

EPDN [28] is a recently proposed GAN-based single image dehazing method. In this work they reduced the image dehazing problem to an **paired** image-to-image translation problem and proposed an enhanced Pix2pix Dehazing network based on a generative adversarial network. This network contains generators, discriminators, and two enhancing blocks to produce a realistic dehazed image on the fine scale.

The enhancer contains two enhancing blocks based on the receptive field model, which reinforces the dehazing effect in both color and details. The GAN is jointly trained with the enhancer.

CycleDehaze [29] is an end-to-end single image dehazing method which does not require pairs of hazy and corresponding ground truth images for training, *i.e.* they train the network by feeding clean and hazy images in an unpaired manner.

This method enhances CycleGAN formulation by combining cycle-consistency and perceptual loss to improve the quality of textural information recovery and generate more visually pleasing and realistic haze-free images.

The Cycle-consistent GAN (CycleGAN) [29] method was proposed for unpaired image-to-image translation task and has gained significant attention during the past

couple of years. CycleGAN is utilized for image dehazing along with the perceptual loss to generate more visually realistic dehazed images [5].

CHAPTER 3

Background

In this section, we explain an overview of the related topics in this dissertation which includes Generative adversarial Network and paired vs unpaired supervision.

3.1 Generative Adversarial Network

Generative adversarial networks, GANs, are an exciting recent innovation in deep learning area. GANs are generative models meaning that they create new data instances that resemble the training data.

Unlike other generative models for instance Variational Autoencoders (VAEs), GANs do not estimate the Probability Density Function (PDF) of the training data and instead it learns to generate instances that are indistinguishable from the training data through a MiniMax game played by the Generator and the Discriminator.

Two main components of a generative adversarial network are the generator model and discriminator model. The generator model, learns to produce the target output, and the discriminator learns to distinguish true data from the output of the generator or fake data. The generator tries to fool the discriminator, and the discriminator tries to keep from being fooled.

In addition to generate realistic images for instance various objects, animal, human faces, *etc.*, GANs have become one of the most successful methods for image manipulation, restoration, and reconstruction. GANs have been used to super-resolve images [30], remove blurriness from images [31], remove noise [32], and haze removal to name a few.

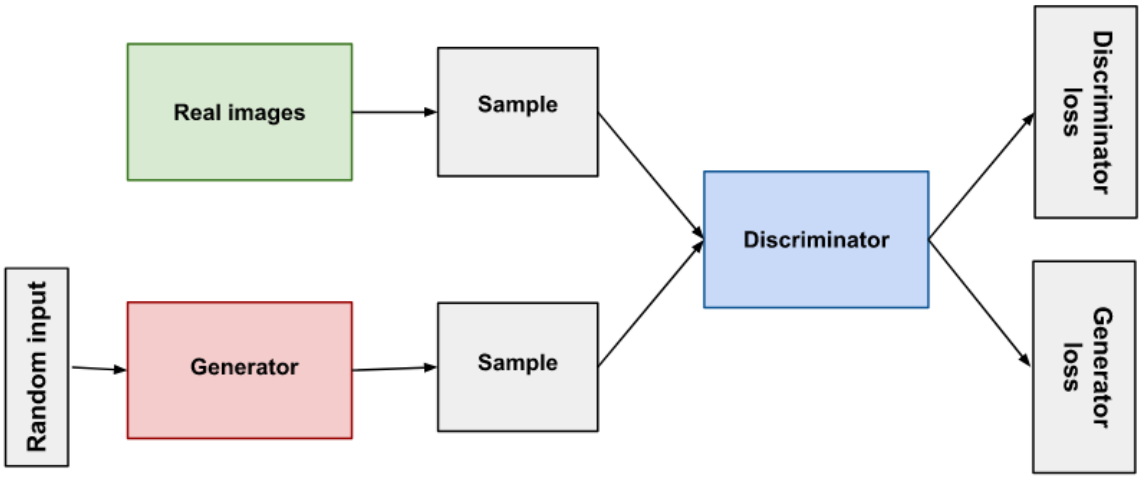


Figure 3.1: Overview of a Generative Adversarial Model. Image captured from [1].

Figure 3.1 depicts the architecture of the generative adversarial network. The input to the Generator model is a noise or a random vector, and the generator learns to generate realistic examples using the feedback given by the discriminator. The image generated by the generator and a real example from the training dataset are the inputs of the discriminator. The discriminator classifies these examples as real vs fake.

Both the generator and the discriminator are neural networks. The generator output is connected directly to the discriminator input. Through backpropagation, the discriminator's classification provides a signal that the generator uses to update its weights.

Figure 3.2 depicts the backpropagation in training the generator model and Figure 3.3 shows the backpropagation in training the discriminator model.

3.2 Paired vs. Unpaired Supervision

In this section, we explain the concept of paired vs unpaired supervision in single image dehazing problem.

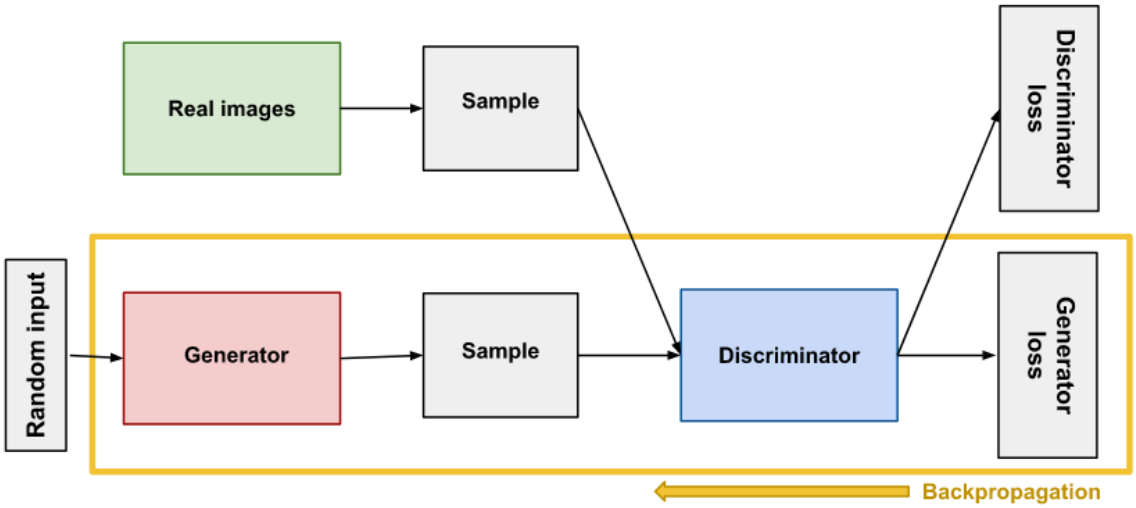


Figure 3.2: Backpropagation in generator training. Image captured from [1].

Training a model can be done in Paired or Unpaired manner. In paired supervised training, to successfully train a model we need the ground truth instance of every hazy image; these models are mostly CNN-based.

On the other hand, models that train through unpaired supervision, do not require hazy/haze-free pairs, in other words the ground truth images are not necessary to train the model. This type of training mostly belong to GAN-based models that are inherently unpaired supervised/weakly supervised. Another type of models that can be trained without pairs being present is the prior-based models, these models estimate the model parameters for instance the physical scattering model parameters in case of image dehazing problem, and do not leverage the power of data.

3.3 Evaluation Metrics

We utilize different metrics throughout this dissertation that we describe them here.

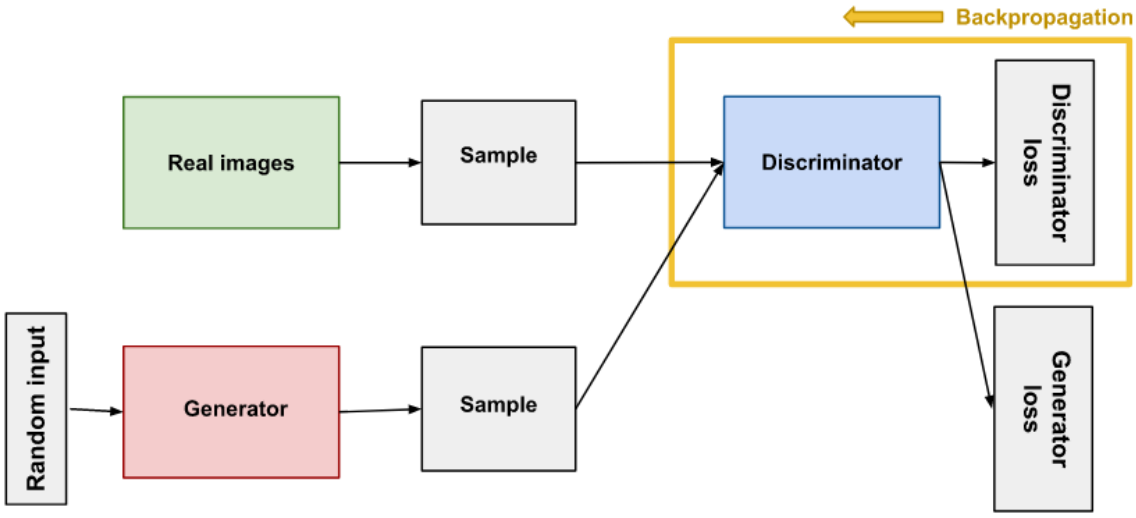


Figure 3.3: Backpropagation in Discriminator training. Image captured from [1].

- **PSNR (Peak Signal to Noise Ratio)**: It measures the ratio between the maximum possible value of a signal and the power of distorting noise that affects the quality of its representation. The higher the PSNR, the better the reconstructive method. It is formulated as follows:

$$PSNR = 10\log_{10}(peakval^2)/MSE \tag{3.1}$$

  where peakval is the maximal in the image data and MSE is the Mean Squared Error between hazy and de-hazed image.

- **SSIM (Structural Similarity Index)**: In this measurement, image degradation is considered as the change of perception in structural information. It also collaborates some other important perception based fact such as luminance masking, contrast masking, *etc.* It is a measure that is consistent with human perception, and is calculated based on the computation of three major aspects that are: luminance, contrast, and structural.

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)^{\beta}.[s(x,y)]^{\gamma} \tag{3.2}$$

where, l is the luminance, c is the contrast and s is the structure and α, β and γ are the positive constants [33].

- **PI (Perceptual Index)**: It measures the quality of restored and reconstructed images based on human perception [34]. A lower perceptual index indicates better perceptual quality. The perceptual quality of an image is the degree to which the image looks like a natural image. This metric is a no-reference quality measurement metric which means that it does not require a ground truth image. The equation below shows the PI formula:

$$PI = ((10 - Ma) + NIQE) \tag{3.3}$$

  which Ma and NIQE are two image qualification indexes detailed in [35, 36].

- **CIEDE2000**: It measures the color difference between hazy and dehazed images; smaller values indicate better color preservation, thus better dehazing and perceptual quality [37].
- **NIQE**: It is a well-known no-reference image quality assessment metric for evaluating real image restoration without requiring the ground-truth.

The larger values of PSNR, SSIM and the smaller values of CIEDE2000, NIQE, and PI indicate better dehazing and perceptual quality.

CHAPTER 4

Enhanced CycleGAN Dehazing Network

Haze is an atmospheric phenomenon that can cause visibility issues, and the quality of images captured under haze can be severely degraded. Hazy images suffer from poor visibility and low contrast, which can challenge both human visual perception and numerous intelligent systems relying on computer vision methods.

The performance of standard computer vision tasks such as object detection [6, 7], semantic segmentation [8], face detection, clustering and dataset creation [9, 10, 11, 12] can be affected significantly when images are hazy. Hence, image dehazing is an essential pre-processing task for general-purpose computer vision algorithms that are fed with hazy images. As a result, single image dehazing has received a great deal of attention over the past decade [13, 14, 15, 16, 17, 13, 14, 15, 16, 18, 17].

Most of the recent image dehazing methods rely on paired datasets, which means for each hazy image there's a single clean/haze-free image as a ground truth. In practice, however, there is a range of clean images that can correspond to a hazy image, due to factors such as contrast or light intensity changes throughout the day. In fact, it is infeasible to capture both ground truth/clear image and the hazy image of the same scene simultaneously. Thus there is an emerging need to develop solutions that do not rely on the ground truth images and could operate with *unpaired* supervision.

Single image dehazing methods can be categorized into two main classes: prior-based methods and learning-based methods. Prior-based models solve the haze removal problem through estimating the physical model, *i.e.* transmission map and

(a) Hazy (b) MSCNN

(c) DehazeNet (d) Ours

Figure 4.1: A single image dehazing example. Our method generates an image with less haze and rich details compared with MSCNN and DehazeNet.

atmospheric light parameters. Learning-based methods mainly use CNN-based or GAN-based models to recover the haze-free images. These models take advantage of large amount of training data to learn a model that recovers the haze-free image of a hazy image.

In this Chapter, we focus on *unpaired* image dehazing and first cast the unpaired image dehazing problem to an image-to-image translation problem and then propose a novel cycle-consistent generative adversarial network, called ECDN, that operates without paired supervision and benefits from (i) a global-local discriminator

architecture to handle spatially varying haze (ii) an encoder-decoder generator architecture with residual blocks to better preserve the details (iii) skip connections in the generator to improve the performance of the network and convergence (iv) customized cyclic perceptual loss and a self-regularized color loss to generate more realistic images and mitigate the color distortion problem. Through empirical analysis we show that the proposed network can effectively remove haze and generate visually pleasing haze-free images.

Figure 4.1 shows the result of our method compared to the current state-of-the-art methods. Our proposed method removes haze more effectively and generates a more realistic clean image compared to previous work.

In summary, this Chapter presents the following contributions:

- We propose a novel cycle-consistent generative adversarial network called ECDN for unpaired image dehazing. ECDN does not rely on any priors such as the physical scattering model, as opposed to many previous methods, and instead it adopts the image-to-image translation approach for unpaired image dehazing.
- We adopt a global-local discriminator structure to deal with spatially varying haze and generate better haze-free images.
- We define a self-regularized color loss and utilize it along with a customized perceptual loss to generate more visually pleasing images with vibrant colors and mitigate the color distortion problem. Self-regularization is vital to our network since in unpaired setting there is no external supervision available.
- We use an encoder-decoder generator architecture with residual blocks with skip connections to better preserve the details.
- Through empirical analysis, we show that our network outperforms the previous work in terms of PSNR and SSIM.

4.1 Proposed Method

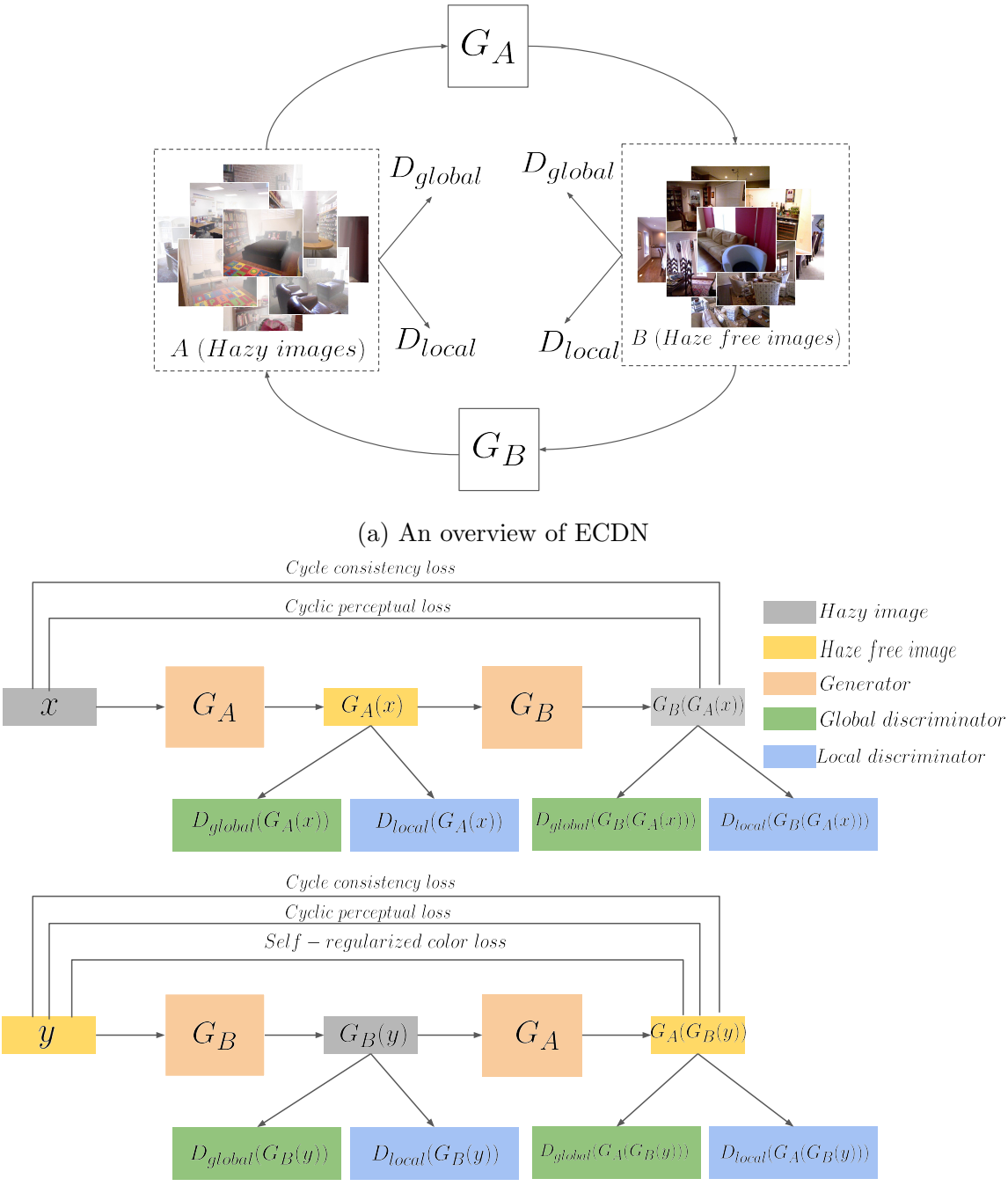
First we reduce the unpaired image dehazing problem to an image-to-image translation problem, and then propose an **E**nhanced **C**ycleGAN **D**ehazing **N**etwork (ECDN) to translate a hazy image to a haze-free one. Next we describe our network in details.

4.1.1 Overview of ECDN

Figure 4.2a demonstrates an overview of our proposed network. We have two domains *i.e.* hazy and haze-free, and the generator G_A which generates haze-free image of a hazy image and G_B which does the backward translation from haze-free to hazy. We need these forward and backward translations to ensure the cycle consistency. At each direction we have two discriminators *i.e.* D_{global} and D_{local} for each generator to enforce them to generate more realistic and better haze-free images.

Figure 4.2b illustrates our proposed network in forward and backward cycles. Top row depicts the hazy to haze-free translation cycle and how the components interact. x is the hazy image and $G_B(G_A(x))$ is the reconstructed hazy image that is used to calculate loss values *i.e.* cycle consistency loss and cyclic perceptual loss. The bottom row shows the backward cycle *i.e.* how the haze-free image is reconstructed through the backward cycle. y is the haze-free image and $G_A(G_B(y))$ is the reconstructed haze-free image that is used to calculate different loss values *i.e.* cycle consistency loss, cyclic perceptual loss, and also self-regularized color loss. We only use self-regularized color loss in the backward cycle, since we want to make the haze free and the reconstructed haze free images closer in terms of color, and prevent color shifting and distortion.

Figure 4.3 depicts the network architecture of the generator G_A and the global and local discriminators. G_A and G_B utilize the same network architecture. Similarly



(a) An overview of ECDN

(b) Overall architecture of ECDN

Figure 4.2: The architecture of ECDN

all discriminators share the same network architecture, however operate on different scales.

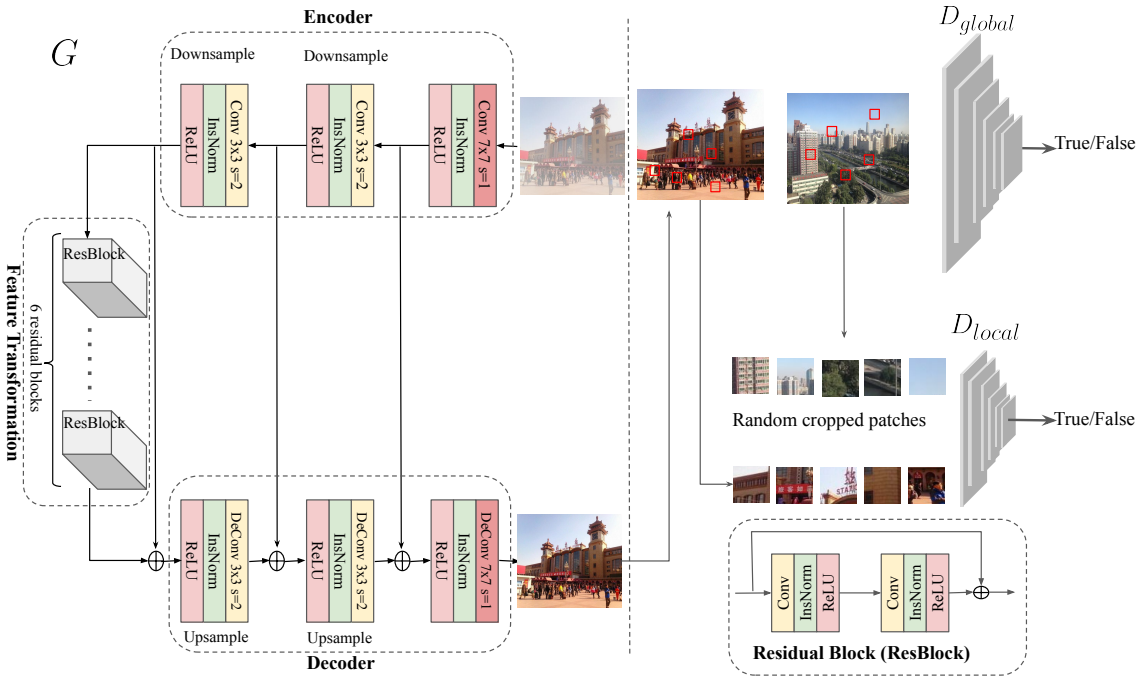


Figure 4.3: The architecture of Generators and Discriminators of ECDN. This figure shows the architecture of G_A, D_B^{Global} and D_B^{Local}. G_B, D_A^{Global} and D_A^{Local} have the same architecture as G_A, D_B^{Global}, D_B^{Local} respectively, except that they work on different inputs, *i.e.,* the input to G_B is a clean image and the input to G_A is a hazy image.

4.1.2 Generator

Figure 4.3 presents the architecture of ECDN model. The architecture of generator G_A is depicted on the left. Note that G_B has the same architecture as G_A. In order to generate a haze-free image without paired supervision in a cycle-consistent manner, we require a generator network that can preserve the images' texture, structure and details while removing haze. Therefore, we designed a network with three parts: encoder, feature transformation, and decoder.

Table 4.1: Generator Network Details

Component	Layers	#Filter	Kernel Size	Stride	Padding
Encoder	Conv, InstanceNorm, Relu	64	7x7	1	0
	Conv, InstanceNorm, Relu	128	3x3	2	1
	Conv, InstanceNorm, Relu	256	3x3	2	1
Feature Transformation	Conv, InstanceNorm, Relu	256	3x3	1	1
	Conv, InstanceNorm, Relu	256	3x3	1	1
	Conv, InstanceNorm, Relu	256	3x3	1	1
	Conv, InstanceNorm, Relu	256	3x3	1	1
	Conv, InstanceNorm, Relu	256	3x3	1	1
Decoder	ConvTranspose2d, InstanceNorm, Relu	128	3x3	2	1
	ConvTranspose2d, InstanceNorm, Relu	64	3x3	2	1
	ConvTranspose2d, Tanh	3	7x7	1	0

The encoder module starts with a convolution layer followed by an Instance Normalization and Relu non-linearity and two downsampling blocks. Feature transformation, has six Residual Blocks to extract complex and deep features whilst removing haze. Going deeper in network helps it to become capable of representing complex functions and also learn features at many different levels of abstraction. Decoder consists of two upsampling blocks which are deconvolution layers, followed by Instance Normalization and Relu. The deconvolution layers are used to recover image structural details and convert the feature maps to a haze-free RGB image. The upsampling operations are performed through the deconvolution layer to obtain intermediate feature mappings with double spatial size and half channels than its previous counterpart.

We use skip links between corresponding layers of different levels from encoder and decoder to guarantee better convergence. A skip connection before downsampling, is also applied between input and output of the feature transformation module, as shown in Figure 4.3. Table 4.1 depicts the network details of the generators.

4.1.3 Discriminator

The right side of Figure 4.3 shows D_B^{Global} and D_B^{Local}. Note that D_A^{Global} and D_A^{Local} have the same architecture as D_B^{Global} and D_B^{Local} respectively. We have two types of discriminators, global and local, each performing a particular operation to classify real vs. fake images.

Initially our model contained only global discriminators. However, we have observed that global discriminators often fail on spatially-varying hazy images, *i.e.*, in cases where haze density variation exists in an image.Thus we decided that different image parts need to be enhanced differently. In order to enhance each region of an image appropriately, in addition to improving the haze removal globally, we utilized a global-local discriminator scheme inspired by [38] in a cycle-consistent manner.

Global discriminator D_B^{Global} classifies if a haze-free image generated by G_A is real or fake, **based on the entire image**. Local discriminator D_B^{Local} classifies if a haze-free image generated by G_A is real or fake, **based on 5 randomly cropped image patches of size** 64×64 **pixels from that image**. Table 4.2 depicts the network details of the discriminators. Global and local discriminators share the same network architecture, *i.e.* PatchGAN architecture.

Table 4.2: Discriminator Network Details

Layers	#Filters	Kernel Size	Stride	Padding
Conv layer, InstanceNorm, Leaky Relu	64	4x4	2	1
Conv layer, InstanceNorm, Leaky Relu	128	4x4	2	1
Conv layer, InstanceNorm, Leaky Relu	256	4x4	2	1
Conv layer, InstanceNorm, Leaky Relu	512	4x4	1	1
Conv layer, InstanceNorm, Leaky Relu	1	4x4	1	1

4.1.4 Loss functions

Our objective loss function contains:

- **Adversarial loss** for matching the distribution of generated images to the data distribution in the target domain.
- **Cycle consistency loss** to prevent the learned mappings G_A and G_B from contradicting each other.
- **Cyclic perceptual loss** to help the generators generate more realistic and visually pleasing images.
- **Self-regularized color loss** to avoid color shifting and artifacts in generated haze-free images and also guide the generator to generate images with vibrant colors.

The overall loss function for training ECDN is defined as follows:

$$\begin{aligned} Loss_{total} = L_{global}^{GAN} + L_{local}^{GAN} + L_{global}^{Cycle} + \\ L_{local}^{Cycle} + L_{global}^{CP} + L_{local}^{CP} + L_{global}^{SRColor} \end{aligned} \tag{4.1}$$

Next we describe these loss functions in details.

4.1.4.1 Adversarial loss

We adopted Least Squares GAN to calculate the adversarial loss. Equations 4.2 and 4.3 show how we calculate the adversarial loss for the global discriminators and the global generators respectively.

$$\begin{aligned} L_D^{Global} = E_{x_r \sim P_{real}}[(D(x_r) - 1)^2] + \\ E_{x_f \sim P_{fake}}[(D(x_f) - 0])^2] \end{aligned} \tag{4.2}$$

$$L_G^{Global} = E_{x_r \sim P_{fake}}[(D(x_f) - 1)^2] \tag{4.3}$$

where D denotes the discriminator, and x_r and x_f are sampled from the real and fake distribution respectively.

We introduced the local discriminator to further enhance hazy image and deal with spatially-varying hazy images. Equations 4.4 and 4.5 depicts the corresponding loss functions:

$$\begin{aligned} L_D^{Local} = E_{x_r \sim P_{real-patches}}[(D(x_r)-1)^2] + \\ E_{x_f \sim P_{fake-patches}}[(D(x_f)-0)^2] \end{aligned} \tag{4.4}$$

$$L_G^{Local} = E_{x_f \sim P_{fake-patches}}[(D(x_f)-1)^2] \tag{4.5}$$

where D denotes the discriminator, x_r and x_f are sampled from **patches** taken from real and fake distributions.

4.1.4.2 Cycle consistency loss

Adversarial loss can not guarantee that the learned function can map an individual input x_i to desired output y_i. Thus a cycle-consistency loss is proposed by CycleGAN to reduce the space of possible mapping functions. Cycle-consistency loss function ($L1-norm$) compares the cyclic image and the original image in an unpaired image-to-image translation process [29]. Cycle consistency loss is defined as:

$$\begin{aligned} L_{cycle}(G_A, G_B) = E_{x \sim p_{data(x)}}[\|(G_B(G_A(x)) - x)\|]_1 + \\ E_{y \sim p_{data(y)}}[\|(G_A(G_B(y)) - y)\|]_1 \end{aligned} \tag{4.6}$$

where G_A and G_B are forward and backward generators, x belongs to domain X (*i.e.* the original domain, hazy images here) and y belongs to domain Y (*i.e.* the haze-free images). $G_B(G_A(x))$ and $G_A(G_B(y))$ are the reconstructed images.

If cycle-consistency loss's goal is met, the reconstructed images $G_B(G_A(x))$ will match closely to the input image x and also the reconstructed images $G_A(G_B(y))$ will match closely to the input image y.

4.1.4.3 Self-regularized color loss

Hazy images usually lack brightness and contrast, to improve these lacking features we define a self-regularized color loss, inspired by [39] to measure color difference between the haze-free images and the reconstructed images. We call it *self-regularized* because we do not rely on the ground truth image.

This loss function forces the generator to generate images with the same color distribution as the haze-free images. In addition, we observed that some of the reconstructed images have color artifacts which is an inherent problem of CycleGAN, this loss function was employed to deal with this problem as well. Equation 4.7 shows color loss function.

$$L_{SRColor} = \sum_{p} ANGLE(G_A(G_B(y))_p, y_p) \tag{4.7}$$

Where $()_p$ denotes a pixel; ANGLE is a function that calculates the angle between two colors regarding the RGB color as a 3D vector. y belongs to domain Y (*i.e.* faze-free images) and $G_A(G_B(y))$ the reconstructed haze-free image.

Eq. 4.7 sums the angles between the color vectors for every pixel pair in $G_A(G_B(y))$ and image y. The reason that we use this color loss calculation instead of an L2 distance in other color space is that the L2 metric only numerically measures the color difference, it cannot ensure that the color vectors have the same direction and the formulation is simple and fast for network computation.

4.1.4.4 Cyclic perceptual loss

Adversarial and cycle consistency losses are not able to preserve the textual and perceptual information of corrupted hazy images. Therefore, to achieve the perceptual quality we employed a cyclic perceptual loss. The goal of this loss function is to preserve the image structure and content features during dehazing and generate more realistic images.

However, since our method is designed using unpaired supervision, *i.e.*, perceptual loss cannot be directly applied between hazy and its ground truth counterpart. Thus a modified version of perceptual loss is adopted.

Authors in [38] adopted a modified version of perceptual loss when the ground truth is unavailable. They technically calculate the loss function between the original image and its enhanced version and call it self preserving perceptual loss and show that it helps with preserving the image structure and details. We adopted this loss and utilized it in a cycle-consistent manner, and we call it Cyclic Perceptual Loss.

To calculate this loss, we focus on feature maps extracted from the 2^{nd} and 5^{th} pooling layers of VGG-16 pre-trained model. Equation 4.8 shows how this loss is calculated:

$$\begin{aligned} Loss_{CP} = \|(Vgg(G_B(G_A(x))) - Vgg(x))\|_2 + \\ \|(Vgg(G_A(G_B(y))) - Vgg(y))\|_2 \end{aligned} \tag{4.8}$$

where G_A and G_B are forward and backward generators, x belongs to domain X (*i.e.* the original domain, hazy images here) and y belongs to domain Y (*i.e.* the haze-free images). $G_B(G_A(x))$ and $G_A(G_B(y))$ are the reconstructed images. Vgg is a VGG16 feature extractor from the second and fifth pooling layers.

To calculate the L_{CP}^{Local} for the local discriminator we used the cropped local patches of input and output images and used the same equation 4.8.

(a) w/o color loss (b) with color loss

(c) w/o color loss (d) with color loss

Figure 4.4: Examples showing the importance of color loss in our model ECDN.

Table 4.3: Ablation study over NYU dataset. The larger values of PSNR, SSIM and the smaller value of CIEDE2000 indicate better dehazing and perceptual quality.

Setting	↑ PSNR	↑SSIM	↓ CIEDE2000
CycleGAN	13.3879	0.5223	17.6113
ECDN w/o color loss	14.5402	0.7407	15.6401
ECDN w/o perceptual loss	14.6582	0.7312	15.6348
EDCN w/o residual blocks	14.1092	0.6923	16.4344
EDCN w/o local discriminator	14.0681	0.7111	19.9466
ECDN	**16.0531**	**0.8244**	**14.9436**

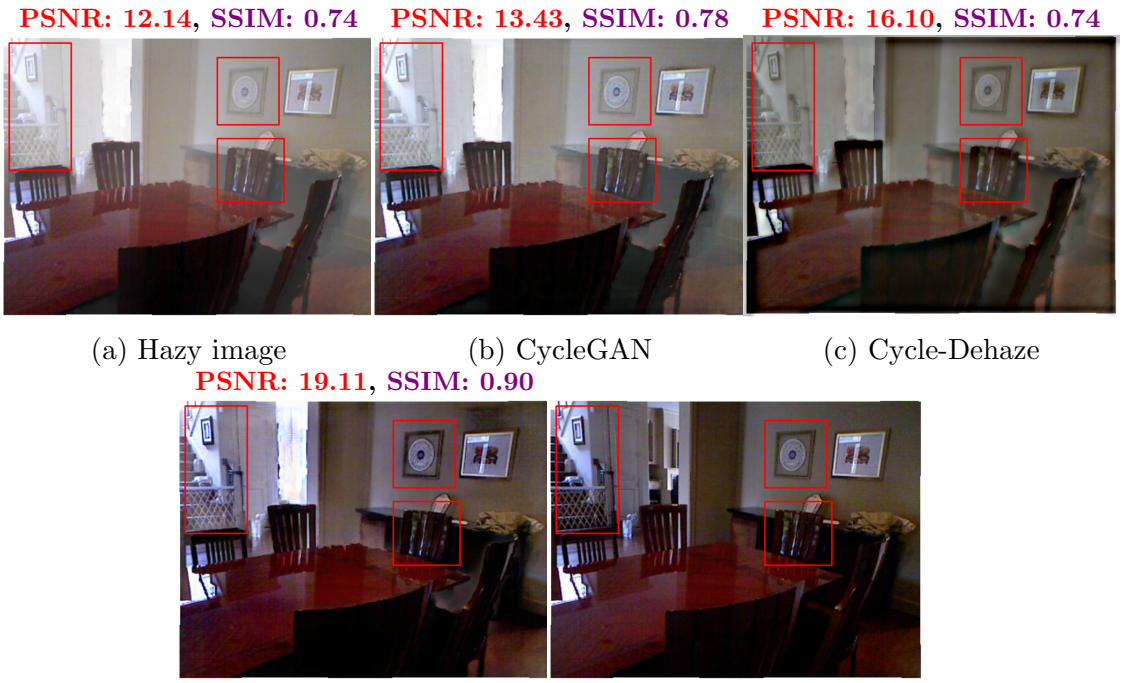


Figure 4.5: Comparison between CycleGAN, Cycle-Dehaze and the proposed method.

4.2 Experiments and Results

To evaluate the performance of our method compared to previous *paired and unpaired* methods, we train a model on NYU dataset [40] and test it on NYU dataset and also Middlebury dataset [41] as a cross-dataset to show how our model generalizes. NYU contains 1449 hazy images paired with their ground truth images and Middlebury contains 23 high-resolution(2k) hazy images with their ground truth. Since our method uses unpaired supervision, the training process received no information about which haze-free image corresponds to each hazy image.

4.2.1 Training

For training we need two sets of training datasets: trainA includes hazy images and trainB includes ground truth images (shuffled to simulate the unpaired super-

Table 4.4: Results on NYU dataset. Some of the numbers for the previous work are taken from [4, 5].

Method	↑PSNR	↑SSIM	↓CIEDE2000
DCP	10.9803	0.6458	18.9781
CycleGAN	13.3879	0.5223	17.6113
Cycle-Dehaze	15.41	0.66	19.04432
DDN	15.5456	0.7726	**11.8414**
DehazeNet	12.8426	0.7175	15.8782
MSCNN	12.2669	0.7000	17.4497
Ours	**16.0531**	**0.8244**	14.9436

vision similar to other unpaired methods [4]). The generators network layers are depicted in Table 4.1, and the local and global discriminators network layers and details are shown in Table 4.2. We opted for Adam optimizer ($momentum = 0.5$) with batch size of 1. Our initial learning rate was 0.0002 for the first 100 epochs, with linear decay to zero over the next 100 epochs. We set the slope of Leaky Relu to 0.2. We implemented our model in PyTorch using two NVIDIA Tesla P100 GPUs and trained our network for 200 epochs.

4.2.2 Quality Measures

We used PSNR, SSIM, and CIEDE2000 metrics, as described in Section 3.3.

4.2.3 Ablation Study

To demonstrate the effectiveness of the local discriminator, cyclic perceptual loss, and self-regularized color loss, we perform several ablation experiments.

Figure 4.4 depicts a couple of examples on how color loss helps with color artifacts removal. Employing color loss has enabled the network to remove artifacts effectively.

Table 4.5: Results on Middlebury dataset. The numbers for the previous work are taken from [4, 5].

Method	↑PSNR	↑SSIM
DCP	12.0234	0.6902
CycleGAN	11.3037	0.3367
Cycle-Dehaze	15.6016	0.8532
DDN	14.9539	0.7741
DehazeNet	13.5959	0.7502
MSCNN	13.5501	0.7365
Ours	**15.8747**	**0.8601**

We compared our method with other cycle-consistent unpaired image-to-image translation methods. Figure 4.5 shows the comparison between CycleGAN, Cycle-dehaze and our method using an example image from NYU dataset. As one can observe our method removed more haze and the generated haze-free images is closer to the ground truth image. The red bounding boxes signify some parts of the image with different amount of haze removed by these methods.

Table 4.3 depicts the results of our ablation study in terms of PSNR, SSIM and CIEDE2000. One can observe that incorporating local discriminators can help achieve better PSNR, SSIM and CIEDE2000, meaning better restoration and generation of more visually pleasing results. The best results in terms of PSNR, SSIM, CIEDE2000 are achieved when the local discriminators, cyclic perceptual loss, and self-regularized color loss are incorporated.

4.2.4 Quantitative and Qualitative Analysis

We compare our model with both *paired* and *unpaired* methods, on the NYU and Middlebury datasets. Our method as well as the competitors are trained on the NYU dataset, and tested on NYU dataset and Middlebury dataset as a cross-datase.

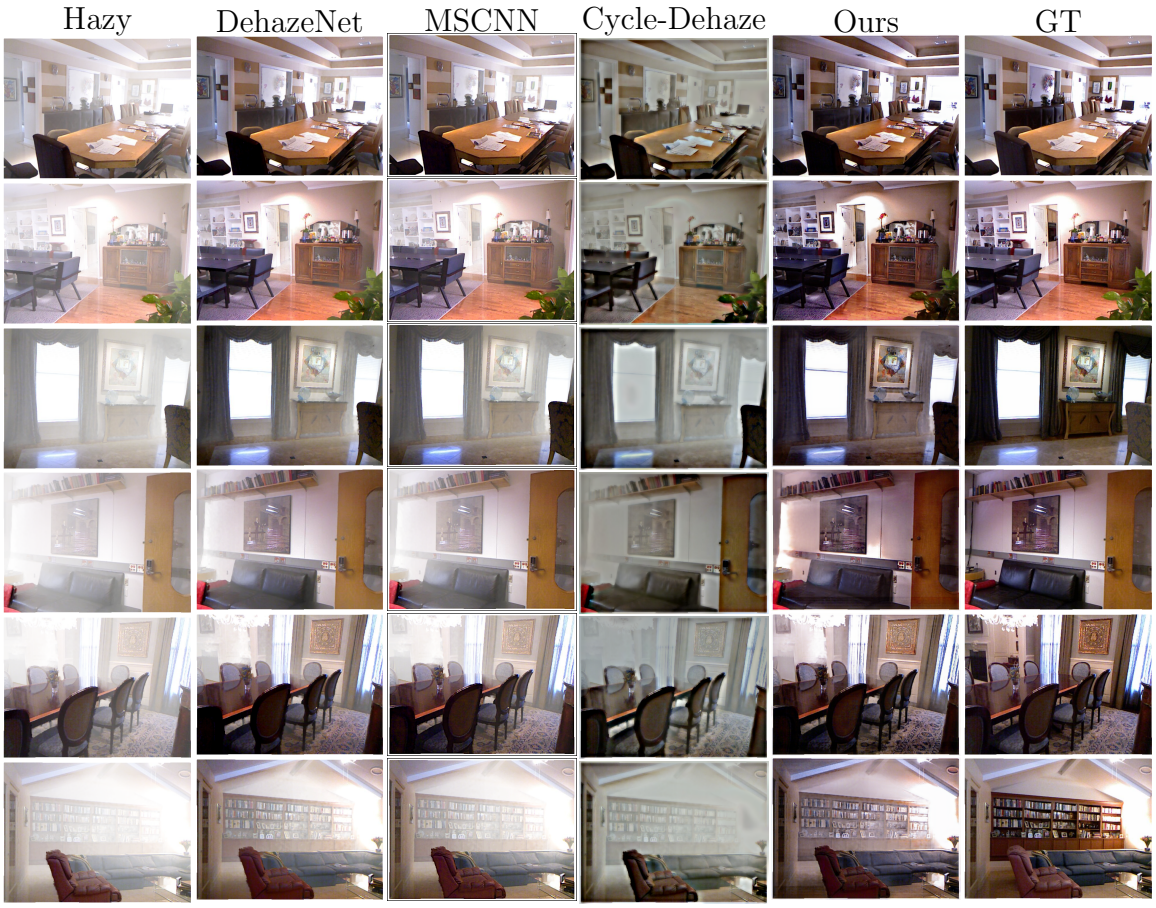


Figure 4.6: Comparison of the state-of-the-art dehazing methods on NYU dataset.

Our method outperforms other methods in terms of SSIM and PSNR on both NYU and Middlebury datasets.

Table 4.4 and 4.5 and show the results on NYU and Middlebury datasets respectively. Our method outperforms the other methods in terms of SSIM and PSNR, and is the second best in terms of CIEDE2000.

Figure 4.6 shows the results of our method compared with other methods. DCP suffers from color distortion and over-exposure. CycleGAN introduces color artifacts and color shifting, and fails to remove much haze especially from dense hazy images. MSCNN and DehazeNet similarly fail to remove much haze from hazy images as well.

Our method, on the other hand is able to generate more natural haze-free images which are much closer to the ground truth image. Moreover, one can observe that our model outperforms the above-mentioned methods in recovery of details, and generates more natural images with least color artifacts.

4.3 Conclusion

In this Chapter, we treated the image dehazing problem as an image-to-image translation problem, and proposed a cycle-consistent generative adversarial network, called ECDN, for unpaired image dehazing. ECDN utilizes discriminators with a local-global structure and generators with an encoder-decoder architecture with residual blocks and skip links to remove haze effectively. It also leverages different loss functions to generate realistic clean images. Using two benchmark test datasets, we showed the effectiveness of the proposed method. Our method outperforms other methods in terms of PSNR and SSIM.

We showed that the global-local discriminators structure can be effectively applied to unpaired single image dehazing through adversarial training. We speculate that this structure can be generalized to other image restoration and reconstruction applications such as single image de-raining.

CHAPTER 5

Evaluating Single Image Dehazing Methods Under Realistic Sunlight Haze

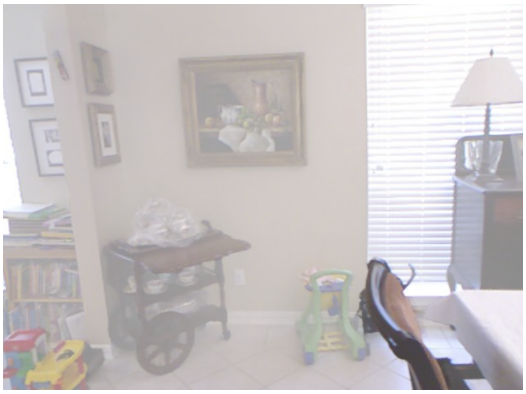
Haze removal is still a challenging and ill-posed problem and most existing methods make assumptions that do not simply hold in reality. For example, most existing methods and datasets assume that i) haze has a uniform and homogeneous distribution in the entire image, and ii) haze can only have a single color, *i.e.* grayish white similar to the color of smoke or pollution. While in reality haze density can change non-homogeneously throughout an image and it can vary in pattern and color.

Figure 5.1 shows a few sample images of different haze datasets that are widely used to test image dehazing methods. As you can see, haze is monochromatic and homogeneous in all these images. These datasets are created synthetically and do not look realistic, which may limit the practicality of dehazing methods.

In this Chapter, we focus on haze created by sunlight which present a unique challenge to dehazing methods. The reason we focus on sunlight is that haze created by sunlight is one of the most prevalent type of haze for the outdoor and indoor settings, and yet it has not received enough attention. To the best of our knowledge, our work is the first work that focuses on sunlight haze.

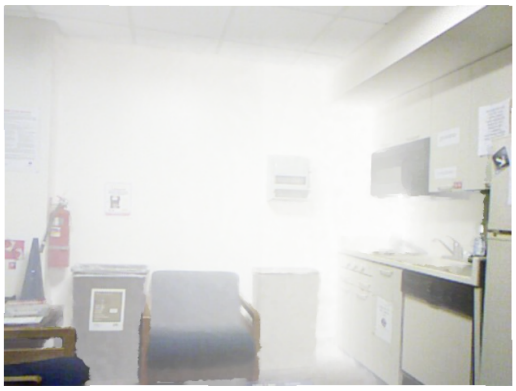
Haze created by sunlight has multiple unique features:

- It can drastically vary in between sun rays throughout an image.
- It can corrupt some parts of an image more than another, meaning haze density drastically varies.
- It can have an spectrum of colors, due to the sunlight color changes during the day.



(a) SOTS indoor

(b) SOTS outdoor



(c) NYU

(d) Middlebury

Figure 5.1: Sample hazy images of the datasets widely used to test image dehazing methods, SOTS test dataset [2], NYU dataset [3], and Middlebury [3].

- It has a unique gradually diminishing pattern.

To quantify the challenges and assess the performance of state-of-the-art dehazing methods, we present a sunlight haze benchmark dataset, Sun-Haze, containing 107 hazy images with different haze density, coverage, and color, caused by sunlight. Figure 5.2 presents some sample hazy images of Sun-Haze along with their corresponding ground truth images. We describe our dataset in the next section in details.

Figure 5.2: Sample images of Sun-Haze

Since the ground truth/haze-free image of a hazy image can be a variety of clean images, for instance images with different contrast or lighting, having only a single image as the ground truth might not be a fair representative and it lacks flexibility and practicality. Thus we build our dataset on top of MIT-Adobe FiveK dataset [42] which includes images retouched by five experts. The retouched images are clean images that we can employ as ground truth. Therefore, our dataset contains six ground truth images per hazy image, including the original one before adding haze. This provides us with the opportunity to compare existing methods more widely, fairly and more importantly in a more practical way.

Our evaluation of the current state-of-the-art methods shows that there is no clear winner and all these dehazing methods suffer to generalize well to the haze created by sunlight, specially when dealing with haze with a different color.

In summary, this Chapter presents the following contributions:

- We present a sunlight haze benchmark dataset, Sun-Haze, that contains 107 hazy images caused by sunlight, along with six ground truth images (five retouched by five experts and one original image before being retouched) per hazy image.
- We perform an extensive analysis to evaluate current state-of-the-art dehazing methods over Sun-Haze in terms of both reference-based and no-reference-based metrics.
- We show that existing dehazing methods can not generalize well when there is sunlight haze, in particular when we have sunlight color changes.

5.1 Sun-Haze dataset

In this section, we describe how we created our hazy dataset, called Sun-Haze. This dataset is built on top of MIT-Adobe FiveK dataset [42]. MIT-Adobe FiveK

dataset contains 5,000 photos taken by photographers with SLR cameras. These photos are all in RAW format, meaning that all the information recorded by the camera sensors, *i.e.* metadata, is preserved. These photos are captured from different scenes, subjects, and during various lighting conditions. These photographs are retouched to obtain a visually pleasing renditions by five photography experts using *Adobe Lightroom* [43].

To create our own dataset, we carefully selected a subset of 107 images from MIT-Adobe FiveK dataset, and added sunlight haze to them. We selected images that would create a realistic image after adding the sunlight haze. For example, we did not select night time images, or indoor images with no windows.

To add sunlight haze and mimic the real sunlight haze effect, we utilized *Adobe Photoshop* [44] and *Luminar 4* [45] which are photo editing applications. They enable us to add realistic sunlight haze. To produce realistic sunlight haze effect, we carefully used different parameters in Luminar 4, such as sunlight length, number of sunlight rays, intensity, penetration, and warmth. Increasing intensity would create a thicker and more dense haze effect. Increasing penetration would expand the sunlight haze effect to a broader region of the image. Increasing sunlight warmth creates a golden yellow type of haze, which enables us to create realistic sunlight color changes during the day. We also added sunlight haze from different angles to further diversify our dataset. To create sunset/sunrise haze effect, we used Adobe Photoshop and professionally added a gradient sunlight haze effect.

Sun-Haze dataset includes 107 outdoor and indoor images with sunlight haze professionally added. It also includes the original image (before retouch) as well as five retouched images (retouched by five experts) per hazy image, that serve as the ground truth images for each hazy image. Therefore each hazy image in our dataset has six ground truth images, enabling us to evaluate dehazing methods more widely

and in a more practical way (since the ground truth of a hazy image could be a variety of clean/haze-free images). We will make our dataset public, and we hope that this dataset can help other researchers to test dehazing methods in a more practical and realistic way.

5.2 Dehazing Methods

In this section, we select a representative of existing dehazing methods from the earliest ones to the current state-of-the-art methods. We can categorize the dehazing methods into the following categories (Note that some of the methods might fall into multiple categories):

- **Prior-based:** Prior-based methods also known as prior information-based methods are mainly based on the parameter estimation of atmospheric scattering model by utilizing the priors, such as dark channel priors [21], color attenuation prior [46], haze-line prior [47, 48]. The physical scattering model consists of the transmission map and the atmospheric light, and it is formulated as follows:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{5.1}$$

  where $I(x)$ is the hazy image, $J(x)$ is the haze-free image or the scene radiance, $t(x)$ is the medium transmission map, and A is the global atmospheric light on each x pixel coordinates.

- **Learning-based:** On the other hand, some methods utilize the deep convolutional neural networks to estimate the transmission map indirectly [25, 24]. Some work employ deep convolutional neural networks to jointly estimate the parameters of the physical scattering model, *i.e.* atmospheric light and the transmission map [4, 49].

Table 5.1: Description of the evaluated existing methods.

Method	Paired vs. Unpaired	Prior-based	Learning-based	Adversarial-based
DCP [21]	NA	✓		
MSCNN [24]	Paired	✓	✓	
DehazeNet [25]	Paired	✓	✓	
AODNet [26]	Paired	✓	✓	
EPDN [28]	Paired		✓	✓
ECDN [50]	Unpaired		✓	✓
CycleDehaze [5]	Unpaired		✓	✓

- **Paired/Unpaired Supervision:** Paired single image dehazing methods need the haze-free/ground truth of each hazy image for training [28, 26], while unpaired dehazing methods do not require the haze-free pair of the hazy images [50, 5].
- **Adversarial-based:** Some image dehazing methods utilize generative adversarial networks for image dehazing and learn transmission map and atmospheric light simultaneously in the generators. Some data-driven methods use adversarial training to solve the dehazing problem without using priors. Some recently proposed work use image-to-image translation techniques to tackle the image dehazing problem through adversarial training [50, 28, 5]. Xitong *et. al* proposed a joint model that learns to perform physical-model based disentanglement by adversarial training [4].

Table 5.1 represents a description of the methods that we evaluated. As you can observe, we selected a variety of methods from different categories. Next, we will describe these methods in more details.

5.3 Results and Discussion

In this section, we present the quantitative and qualitative evaluation results and then discuss the performance of the dehazing methods over Sun-Haze dataset.

5.3.1 Quantitative Evaluation

In this section, the Sun-Haze dataset is used to perform a comprehensive quantitative evaluation of several state-of-the-art single image dehazing methods, as described in Section 5.2.

Table 5.2 shows the quantitative results of evaluating the dehazing methods in terms of PSNR and SSIM, CIEDE2000, NIQE, and PI. We conducted multiple experiments to evaluate the dehazing and generalization capability of the dehazing techniques.

We used our Sun-Haze dataset as the hazy images and the retouched images by five experts as well as the original image as the ground truth for each experiment. The best results are depicted in bold. We also highlighted the top three best results. The green highlights represent the best results, yellow the second best and pink the third best results.

As one can observe, on average MSCNN achieved best PSNR of 17.51 and the second best SSIM and CIEDE2000. ECDN outperformed other methods in terms of SSIM, CIEDE2000, and PI by a very small margin. AOD-Net achieved the best NIQE of 3.93 which is very comparable with the second and third best results.

As you can see, no method is superior to other methods in all five measurements and the best results are only slightly better than the second and third best results. This suggests that current methods can not generalize well to remove haze caused by sunlight.

Table 5.2: Results over Sun-Haze dataset. We performed separate analysis for different ground truth images. The images retouched by 5 experts and the original image before retouch are considered as ground truth/haze free for each experiment. We also present results for the no-reference metrics that do not require a ground truth image.

Ground truth	Metric	DCP	MSCNN	Dehazenet	AOD-Net	EPDN	ECDN	CycleDehaze
Expert A	PSNR	11.01	**16.48**	15.62	14.83	15.88	14.38	15.53
	SSIM	0.641	0.773	0.733	0.698	0.784	**0.789**	0.778
	CIEDE	34.76	**23.46**	27.55	30.26	26.29	24.37	26.62
Expert B	PSNR	11.28	16.33	15.13	14.15	14.96	15.12	15.38
	SSIM	0.655	0.763	0.709	0.676	0.761	**0.801**	0.763
	CIEDE	32.66	25.43	30.90	34.33	31.64	**22.03**	26.49
Expert C	PSNR	11.32	**16.57**	15.49	14.49	15.44	14.74	15.23
	SSIM	0.643	0.746	0.703	0.670	0.756	**0.782**	0.737
	CIEDE	33.22	24.57	28.84	31.93	28.73	**23.98**	28.98
Expert D	PSNR	11.43	14.91	13.75	12.82	13.55	**14.93**	14.35
	SSIM	0.649	0.722	0.667	0.632	0.713	**0.781**	0.728
	CIEDE	30.87	29.10	34.76	38.87	36.28	**22.36**	28.67
Expert E	PSNR	11.32	15.27	13.86	12.99	13.56	**15.32**	14.84
	SSIM	0.640	0.719	0.660	0.626	0.704	**0.780**	0.733
	CIEDE	33.07	28.56	34.38	38.11	35.88	**22.30**	28.13
Original image	PSNR	11.40	**19.10**	17.78	16.89	17.96	14.57	16.39
	SSIM	0.686	**0.867**	0.814	0.782	0.857	0.810	0.834
	CIEDE	36.10	**18.44**	23.651	26.63	22.72	24.46	23.46
Average	PSNR	11.34	**17.51**	16.28	15.37	16.32	14.73	15.73
	SSIM	0.651	0.765	0.721	0.680	0.763	**0.799**	0.762
	CIEDE	33.45	24.93	30.06	33.36	30.26	**23.25**	27.06
No reference	NIQE	5.35	4.06	4.08	**3.93**	4.13	4.09	4.60
No reference	PI	3.71	3.25	3.25	3.02	3.04	**2.93**	4.08

Hazy image

AOD-Net [26]

ECDN [50]

DehazeNet [25]

MSCNN [24]

EPDN [28]

Original

Expert A

Expert B

Expert C

Expert D

Expert E

Hazy image

AOD-Net

ECDN

DehazeNet

MSCNN

EPDN

Original

Expert A

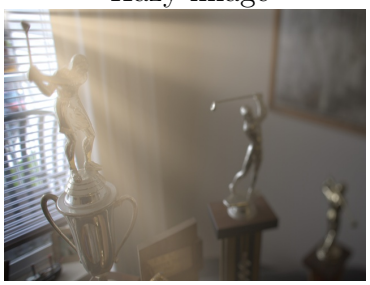
Expert B

Expert C

Expert D

Expert E

Hazy image

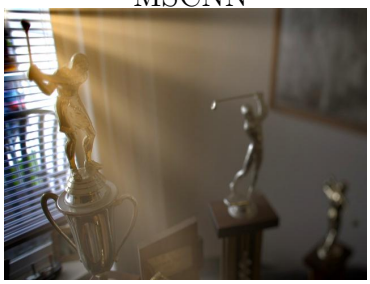


AOD-Net

ECDN

DehazeNet

MSCNN

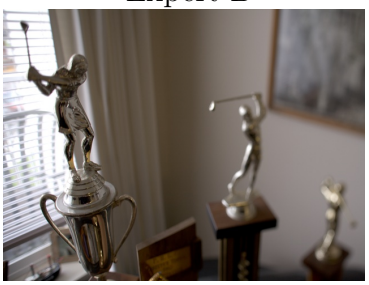


EPDN

Original

Expert A

Expert B

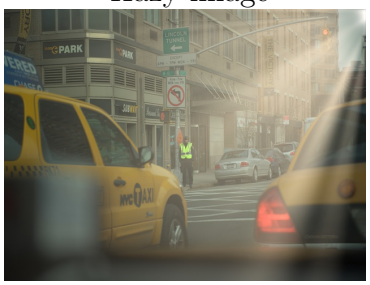


Expert C

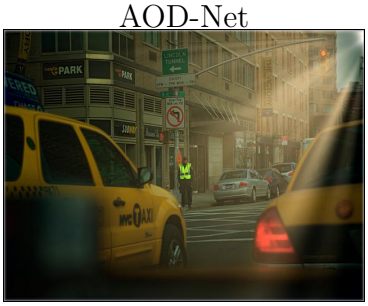
Expert D

Expert E

Hazy image



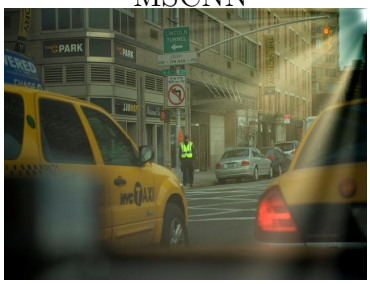
AOD-Net



ECDN

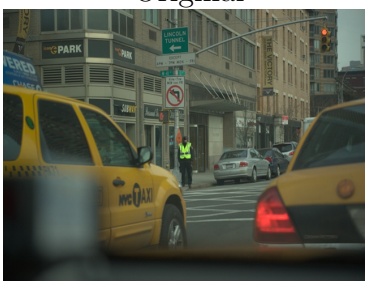
DehazeNet

MSCNN

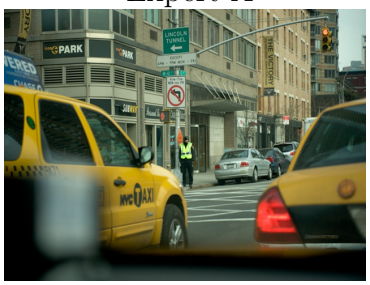


EPDN

Original



Expert A



Expert B

Expert C

Expert D

Expert E

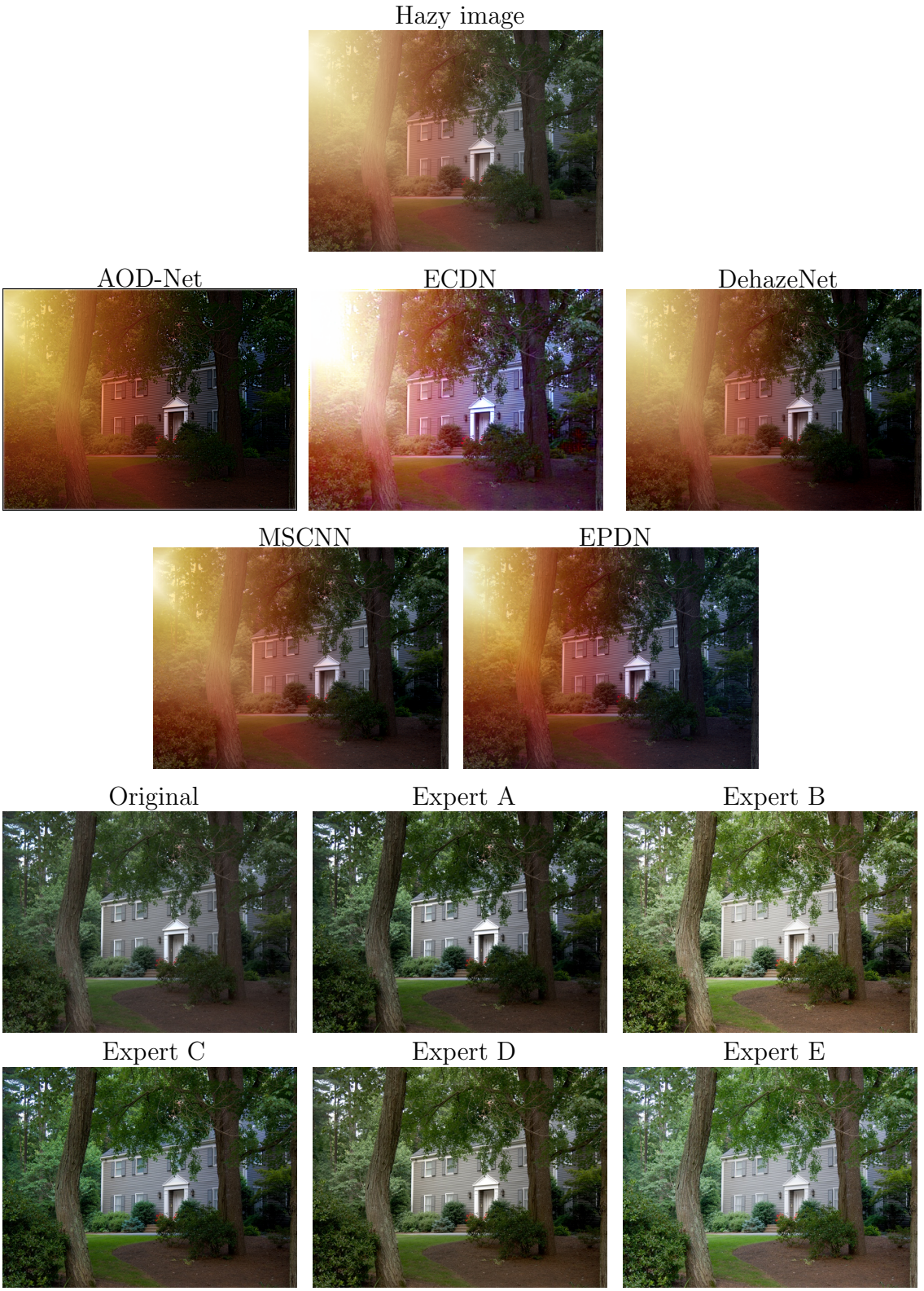


Figure 5.3: Comparison of the dehazing methods on Sun-Haze dataset.

5.3.2 Qualitative Evaluation

Fig. 5.3 depicts five hazy images from Sun-Haze dataset, and the dehazing results yielded by AOD-Net [26], ECDN [50], DehazeNet [25], MSCNN [24], and EPDN [28]. In this figure we have five pairs of rows. In each pair the top row shows a hazy image followed by the results of five dehazing methods mentioned above and the second row shows the five experts' retouched images and the untouched original image to compare with.

Qualitatively, most methods were unable to remove the sunlight haze without introducing color shifting or artifacts. EPDN which is a paired image-to-image translation technique, has mainly learned to remove haze through increasing the color intensity at different channels, thus the dehazed images look visibly darker and the sunlight haze more yellow or orange than the original hazy image. EPDN also introduced artifacts to some of the generated images while partially removing haze. It also created the halo effect near edges.

Other methods also were unable to generalize to remove sunlight haze and they removed haze partially and improved the visibility to a small extent. For instance, even though AOD-Net, ECDN, Dehazenet, and MSCNN could recover the image structure similar to EPDN, they introduced color shifting and artifacts which are not visually pleasing and made the haze even more prominent.

ECDN achieved the best/highest structural similarity index and recovered the image structure well but introduced artifacts and overexposure in particular where the sun rays lie.

In conclusion, these methods were unable to generalize well to remove sunlight haze and performed even more poorly removing sunset hazy images which embody more varicolored haze. This questions the underlying assumptions of these methods and their practicality in the real-world scenarios. Therefore, for dehazing methods

to be practical they need to be trained and tested using more realistic and practical datasets which include variety of realistic haze patterns and colors.

CHAPTER 6

Conclusion and Future Work

In this dissertation, we focused on a very challenging image restoration task, *i.e.,* single image dehazing. In single image dehazing, we need to remove haze from an image while restoring the details of the image which might have been corrupted due to the haze. Most of the recent image dehazing methods rely on paired datasets, which means for each hazy image there is a single clean/haze-free image as a ground truth. In practice, however, there is a range of haze-free/clean images that could correspond to a hazy image, due to factors such as contrast or light intensity changes throughout the day. Thus methods developed by paired supervision are not practical.

Therefore we focused on *unpaired* image dehazing, and reduced it to an image-to-image translation problem. We then proposed a novel cycle-consistent adversarial network, called *ECDN*, that operates without paired supervision and benefits from (i) a global-local discriminator architecture to handle spatially varying haze (ii) an encoder-decoder generator architecture with residual blocks to better preserve the details (iii) skip connections in the generator to improve the performance of the network and convergence (iv) customized cyclic perceptual loss and a self-regularized color loss to generate more realistic images and mitigate the color distortion problem. Through ablation study we showed the effectiveness of each part of our network. In addition, through empirical analysis we showed that this network outperforms previous work in terms of SSIM and PSNR metrics.

Next, we investigated one of the fundamental assumption that current dehazing methods undertake, *i.e.,* haze has a uniform and homogeneous distribution in the

entire image and can only have a single color, *i.e.,* smoke like color. While in reality haze density can change non-homogeneously throughout an image and it can vary in pattern and color.

To investigate this, we focused on haze created by sun, which is one of the most prevalent type of haze in the wild. Sun haze can have varying patterns due to sun rays with different color spectrum due to sunlight warmth changes throughout the day. We introduced a new dataset, called *Sun-Haze*, which includes six ground truth images (one original image and five images retouched by five experts) per each hazy image. Since each hazy image could potentially correspond with multiple clean images, Sun-Haze includes five grounds truth images that are touched with different experts. In this way, we can have a more fair and practical comparison over the previous work. We evaluated a representative set of previous work over Sun-Haze dataset in terms of different metrics like PSNR, SSIM, and CIEDE2000, NIQE.

We concluded that there is no clear winner among these methods. In fact, all methods were unable to generalize to remove sunlight haze effectively and they removed haze partially and improved the visibility to a small extent. Some methods even introduced artifacts and halo effect. This questions the underlying assumptions of these methods and their practicality in the real-world scenarios.

6.1 Future Work

The work in this dissertation paves the way for future research into single image dehazing in terms of realistic solutions and datasets. We hope that our Sun-Haze dataset can help other researchers to develop more practical methods for single image dehazing problem. Through our analysis we showed that current methods can not generalize well when haze is non-homogeneous and/or haze has a spectrum of colors. That opens up a new avenue of research. Thus, one direction for future work is

to develop methods that can remove non-homogeneous and/or vary colored haze. Sun-haze can serve as a test dataset for this purpose.

Another future work could be to collect a more comprehensive dataset for training and testing deep learning based methods for image dehazing. While Sun-Haze dataset took the first step to create a non-homogenous and vary colored haze test dataset, this dataset is limited to haze created by sunlight only. We can extend this dataset to include more realistic, non-homogenous, vary colored and diverse hazy images, such as haze created by lights at night, or haze created by smoke or pollution. Such datasets would be critical in developing practical solutions for image dehazing.

REFERENCES

[1] "gan," https://developers.google.com/machine-learning/gan.

[2] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[3] 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, pp. 2226–2230.

[4] X. Yang, Z. Xu, and J. Luo, "Towards perceptual image dehazing by physics-based disentanglement and adversarial training," in *Thirty-second AAAI conference on artificial intelligence*, 2018.

[5] D. Engin, A. Genç, and H. Kemal Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 825–833.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[9] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5525–5533.

[10] Z. Anvari and V. Athitsos, "A pipeline for automated face dataset creation from unlabeled images," in *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2019, pp. 227–235.

[11] W.-A. Lin, J.-C. Chen, C. D. Castillo, and R. Chellappa, "Deep density clustering of unconstrained faces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8128–8137.

[12] W.-A. Lin, J.-C. Chen, and R. Chellappa, "A proximity-aware hierarchical clustering of faces," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 2017, pp. 294–301.

[13] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Night-time dehazing by fusion," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 2256–2260.

[14] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 501–514.

[15] S. Emberton, L. Chittka, and A. Cavallaro, "Hierarchical rank-based veiling light estimation for underwater dehazing," 2015.

[16] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 617–624.

[17] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2201–2208.

[18] Z. Anvari and V. Athitsos, "Evaluating single image dehazing methods under realistic sunlight haze," *arXiv preprint arXiv:2008.13377*, 2020.

[19] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[20] G. Srinivasa and K. Shree, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[21] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

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

[23] S. G. Narasimhan and S. K. 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. IEEE, 2000, pp. 598–605.

[24] 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, pp. 154–169.

[25] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[26] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4770–4778.

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

[28] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8160–8168.

[29] 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, pp. 2223–2232.

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

[31] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183–8192.

[32] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3155–3164.

[33] R. Kumar and V. Moyal, "Visual image quality assessment technique using fsim," *International Journal of Computer Applications Technology and Research*, vol. 2, no. 3, pp. 250–254, 2013.

[34] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 pirm challenge on perceptual image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[35] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Computer Vision and Image Understanding*, vol. 158, pp. 1–16, 2017.

[36] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[37] M. R. Luo, G. Cui, and B. Rigg, "The development of the cie 2000 colour-difference formula: Ciede2000," *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, vol. 26, no. 5, pp. 340–350, 2001.

[38] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *arXiv preprint arXiv:1906.06972*, 2019.

[39] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6849–6857.

[40] 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, pp. 746–760.

[41] 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, pp. 31–42.

[42] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *CVPR 2011*. IEEE, 2011, pp. 97–104.

[43] "Adobe lightroom," https://www.adobe.com/products/photoshop-lightroom.html.

[44] "Adobe photoshop," https://www.adobe.com/products/photoshop.html.

[45] "Luminar 4," https://skylum.com/luminar.

[46] T. Zhang, C. Shao, and X. Wang, "Atmospheric scattering-based multiple images fog removal," in *2011 4th International Congress on Image and Signal Processing*, vol. 1. IEEE, 2011, pp. 108–112.

[47] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1674–1682.

[48] D. Berman, T. Treibitz, and S. Avidan, "Air-light estimation using haze-lines," in *2017 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2017, pp. 1–9.

[49] H. Zhu, X. Peng, V. Chandrasekhar, L. Li, and J.-H. Lim, "Dehazegan: When image dehazing meets differential programming." in *IJCAI*, 2018, pp. 1234–1240.

[50] Z. Anvari and V. Athitsos, "Enhanced cyclegan dehazing network," in *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, vol. 4, February 2021, pp. 193–202.

BIOGRAPHICAL STATEMENT

Zahra Anvari received her Bachelor's degree in Information Technology Engineering from Shahid Rajaee University in 2008, and her Master's degree in Computer Engineering from Urmia University in 2013. In 2015, she started her studies to pursue her Ph.D in Computer Science at the University of Texas at Arlington. Her current research interest is in the area of Computer Vision, Deep Learning and Machine Learning.