# DEHAZE-GLCGAN: UNPAIRED SINGLE IMAGE DEHAZING VIA ADVERSARIAL TRAINING

**Zahra Anvari**
Department of Computer Science and Engineering
University of Texas Arlington
Arlington, TX
zahra.anvari@mavs.uta.edu

**Vassilis Athitsos**
Department of Computer Science and Engineering
University of Texas Arlington
Arlington, TX
athitsos@uta.edu

August 18, 2020

## ABSTRACT

Single image dehazing is a challenging problem, and it is far from solved. Most current solutions require paired image datasets that include both hazy images and their corresponding haze-free ground-truth images. However, in reality, lighting conditions and other factors can produce a range of haze-free images that can serve as ground truth for a hazy image, and a single ground truth image cannot capture that range. This limits the scalability and practicality of paired image datasets in real-world applications. In this paper, we focus on *unpaired* single image dehazing and we do not rely on the ground truth image or physical scattering model. We reduce the image dehazing problem to an image-to-image translation problem and propose a dehazing **G**lobal-**L**ocal **C**ycle-consistent **G**enerative **A**dversarial **N**etwork (Dehaze-GLCGAN). Generator network of Dehaze-GLCGAN combines an encoder-decoder architecture with residual blocks to better recover the haze free scene. We also employ a global-local discriminator structure to deal with spatially varying haze. Through ablation study, we demonstrate the effectiveness of different factors in the performance of the proposed network. Our extensive experiments over three benchmark datasets show that our network outperforms previous work in terms of PSNR and SSIM while being trained on smaller amount of data compared to other methods.

***Keywords*** Image Dehazing · Image Reconstruction · Adverserial Training · GAN

## 1 Introduction

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. For example, autonomous driving requires the processing of images captured at different weather conditions, and being able to dehaze hazy/foggy images is a must.

The performance of standard computer vision tasks such as object detection [22, 30], semantic segmentation [23], face detection [13, 39], clustering and dataset creation [4, 20, 21, 32] can be affected significantly when images are hazy. Hence, image de-hazing can be a useful or even necessary pre-processing task before applying general-purpose computer vision algorithms to hazy images. As a result, single image dehazing has received a great deal of attention over the past ten years.

Most of the recent solutions for image dehazing heavily depend on paired datasets, which include for each hazy image a single clean (haze-free) image as 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 and 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 at the same time. Thus there is an emerging need to develop solutions that do not rely on the ground truth images and could operate with *unpaired* supervision.
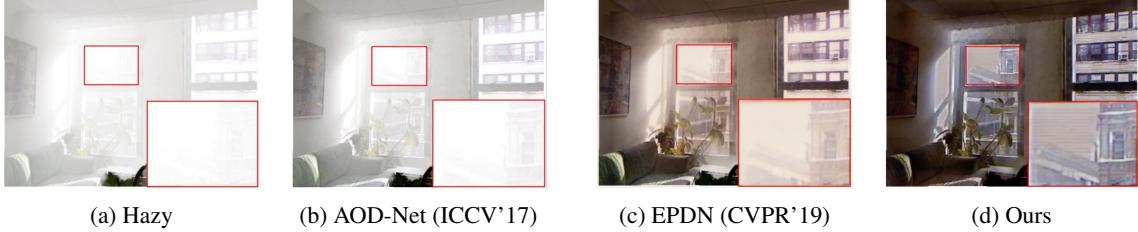
| (a) Hazy | (b) AOD-Net (ICCV'17) | (c) EPDN (CVPR'19) | (d) Ours |

Figure 1: A single image dehazing example. Our method generates an image with less haze and rich details compared with AOD-Net [18] and EPDN [29].

*Unpaired image dehazing* does not rely on the ground truth of hazy images, meaning that there is no one-to-one mapping between hazy and haze-free images in the training dataset. Thus it is more scalable and more practical. In this paper, we focus on unpaired image dehazing and propose a model to remove haze from a hazy image with spatially varying haze conditions.

Single image dehazing methods can be categorized into two main classes: prior-based [2, 3, 8, 27, 37] and learning-based [6, 31, 35, 40, 42]. Prior-based models solve the haze removal problem through estimating the physical model, *i.e.* transmission map and atmospheric light parameters. Learning-based methods mainly use CNN-based or GAN-based models to recover the haze-free/clean image. These models take advantage of large amounts of training data to learn a model that maps a hazy image to a clean image.

In this paper, we treat the unpaired image dehazing as an image-to-image translation problem. We propose a novel cycle-consistent generative adversarial network with novel generators, that operates without paired supervision and benefits from (i) a global-local discriminator architecture to handle spatially varying haze (ii) customized cyclic perceptual loss to generate more realistic and natural images. Through empirical analysis we show that the proposed network can effectively remove haze while being trained on smaller amounts of data compared to other methods.

Figure 1 shows the result of our network compared to the current state-of-the-art methods. As we can see, the proposed method (Figure 3d) removes haze more effectively and generates a more realistic clean image compared to previous work, while being trained on a smaller training dataset compared to previous work.

In summary, this paper presents the following contributions:

- We propose a novel cycle-consistent generative adversarial network called Dehaze-GLCGAN for unpaired image dehazing. Dehaze-GLCGAN does not rely on the physical scattering model, as opposed to many previous methods, and instead it adopts the image-to-image translation approach for unpaired image dehazing [9, 43].

- We propose a novel dehazing generator that combines an encoder decoder structure with embedded Residual blocks to better preserve the details of images.

- We employ color loss along with perceptual loss to generate more visually pleasing images.

- We adopt a global-local discriminator structure to deal with spatially varying haze and generate cleaner images.

- Through empirical analysis, we show that our network outperforms the previous work in terms of PSNR and SSIM with a large margin while trained on smaller amount of data compared to previous work.

This paper is organized as follows. Section 2 describes background information as well as previous work. Section 3 describes our proposed method. Section 4 presents our experiments and results.

## 2   Related Work

Single image dehazing and generative adversarial networks are the two topics that are the most relevant for our method. We briefly describe them and their related work in this section.

### 2.1   Single image dehazing

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.

### 2.1.1 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 the parameters of the physical scattering model [26, 34], *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:

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

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.* [10] proposed a dark channel prior to estimate the transmission map effectively. Tan et al. [36] increase the contrast of hazy images, based on the fact that haze-free images have higher contrast than hazy images.

### 2.1.2 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 [26, 28].

MSCNN [31] contains two sub-networks called coarse-scale and fine-scale, to estimate the transmission map. The coarse-scale network estimates the transmission map and is further improved locally by the fine-scale network. In DehazeNet [6], authors modified the classic CNN model by adding feature extraction and non-linear regression layers. These modifications distinguish DehazeNet from other CNN-based models. The All-In-One Dehazing Network (AOD-net) [18] proposed an end-to-end network that produces the haze-free/clean images through reformulating the atmospheric scattering model. DCPDN [41] uses an end-to-end single image dehazing network to jointly learn the transmission map, atmospheric light and image dehazing. EPDN [29] employs a three-part network which consists of generator, discriminator, and two enhancer blocks.

## 2.2 Generative Adversarial Networks

GANs have become one of the most successful methods for image generation, manipulation, restoration, and reconstruction. GANs have been used to super-resolve images [17], remove motion blurriness from images [16], and remove noise [7], to name a few applications. GANs are also utilized in image dehazing [9, 40]. Yang et al. [40] proposed the disentangled dehazing network without paired supervision.The GAN network they proposed consists of three generators: a generator for generating haze-free image, a generator for the atmospheric light, and the third generator for transmission map.

The Cycle-consistency GAN (CycleGAN) [43] method was proposed for unpaired image-to-image translation task and has gained significant attention during the past couple of years. Engin et al. [9] utilized CycleGAN for image dehazing along with the perceptual loss to generate more visually realistic dehazed images.
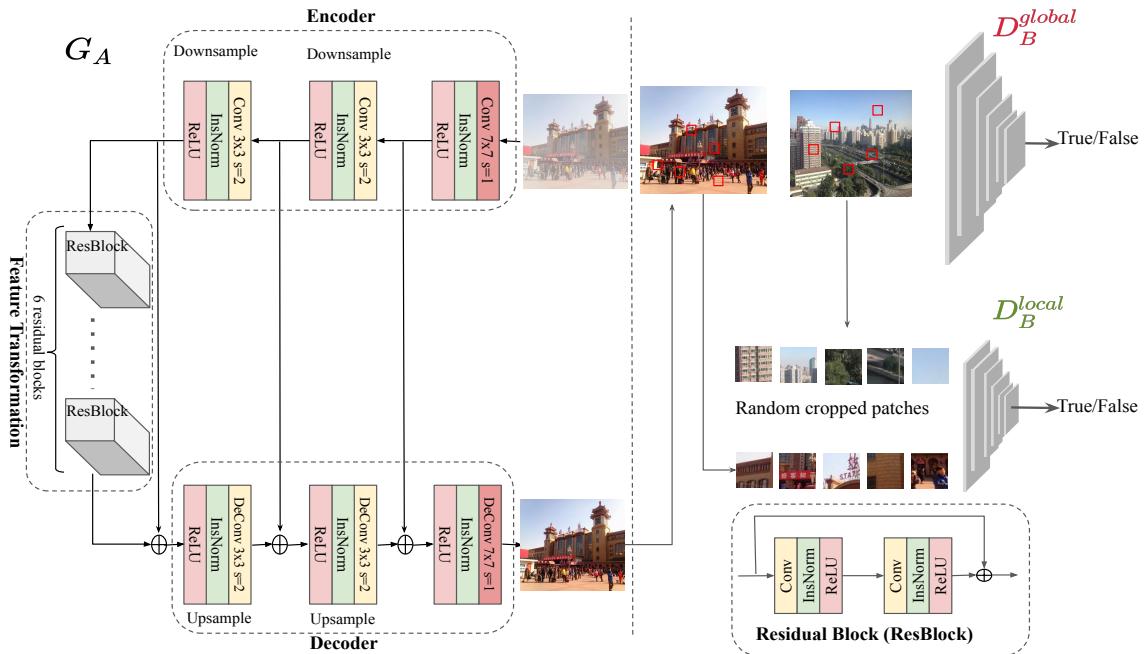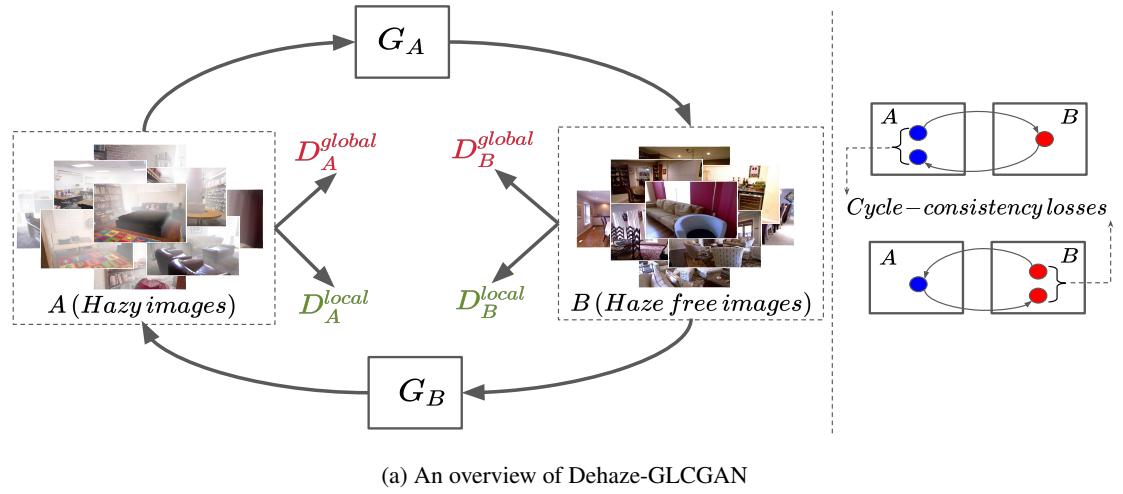
## 3 Proposed Method

We first reduce the unpaired image dehazing problem to an image-to-image translation problem, and then propose a dehazing **G**lobal-**L**ocal **C**ycle-consistent **G**enerative **A**dversarial **N**etwork (Dehaze-GLCGAN) with discriminators that operate at global and local levels, to translate a hazy image to a haze-free image. Next we describe our network in detail.

### 3.1 Overview of Dehaze-GLCGAN

Figure 2a demonstrates an overview of the proposed network, and Figure 2b depicts the architecture of the generators and discriminators. The bottom right part of Figure 2b shows in more detail the structure of the residual blocks that are used in the network.

This network is designed to translate images from domain $A$, *i.e.* hazy, to domain $B$, *i.e.* clean/haze-free, in a cycle-consistent manner. The components of Dehaze-GLCGAN are Generator $A$ ($G_A$), Generator $B$ ($G_B$), Global Discriminator $A$ ($D_A^{Global}$), Global Discriminator $B$ ($D_B^{Global}$), Local Discriminator $A$ ($D_A^{Local}$), and Local Discriminator $B$ ($D_B^{Local}$).

Figure 2a demonstrates how the generators and discriminators interact. $G_A$ translates images from domain $A$ (hazy) to domain $B$ (haze-free). Global discriminator $B$ ($D_B^{Global}$) classifies if a haze-free image generated by $G_A$ is real or fake,

(a) An overview of Dehaze-GLCGAN



(b) The architecture of Dehaze-GLCGAN. 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, *e.g.,* the input to $G_B$ is a clean image and the input to $G_A$ is a hazy image.

Figure 2: The overview and architecture of Dehaze-GLCGAN

**based on the entire image**. Local discriminator $B$ ($D_B^{Local}$) classifies if a haze-free image generated by $G_A$ is real or fake, **based on 5 randomly cropped image patches of size** $64 \times 64$ **pixels from that image** (inspired by [14]).

Similarly, $G_B$ translates images from domain $B$ (haze-free) to domain $A$ (hazy). And $D_A^{Global}$ and $D_A^{Local}$ classify if a hazy image generated by $G_B$ is real or fake based on the entire image and 5 randomly cropped patches respectively.

$G_A$ and $G_B$ utilize the same network architecture (Figure 2). Also all discriminators share the same network architecture, however operate on different scales. Through ablation study we show that utilizing global-local discriminators effectively improves the network performance.

## 3.2 Generator

Figure 2b presents the architecture of Dehaze-GLCGAN. The architecture of generator $A$ ($G_A$) is depicted on the left. Note that $G_B$ has the same architecture. 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. To do so, we designed a network with three modules: encoder, feature transformation, and decoder.

The encoder module starts with a convolution layer followed by an Instance Normalization and Relu non-linearity and then two downsampling blocks. Our next module, feature transformation, uses 6 Residual Blocks to extract more complex and deeper features whilst removing haze. The main benefit of going deeper in network is that it becomes capable of representing very complex functions and also learns features at many different levels of abstraction.

The third module, which is the decoder, consists of two upsampling blocks which are deconvolution layers, followed by Instance Normalization and Relu. The deconvolution layers of the decoder are used to recover image structural details and convert the feature maps to a haze-free RGB image. Through the deconvolution layer also known as transposed layer, the upsampling operations are performed to obtain intermediate feature mappings with double spatial size and half channels than its previous counterpart.

## 3.3 Discriminator

The right side of Figure 2b shows $D_B^{Global}$ and $D_A^{Local}$. 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 haze images, *i.e.*, in cases where haze density variation exists in an image, and thus different image parts need to be enhanced differently from other parts. In order to enhance each local region appropriately, in addition to improving the haze removal globally, we utilized a global-local discriminator scheme inspired by [14] in a cycle-consistent manner, both using PatchGAN discriminators.

PatchGAN discriminators classify **individual (N x N) patches** in the image as real vs. fake, as opposed to classifying the **entire image** as real vs. fake.

PatchGAN imposes more constraints which help with sharp high-frequency details. Furthermore, PatchGAN has fewer number of parameters thus performs faster compared to classifying the entire image [12]. It is shown that $70 \times 70$ patches produce the best results [12], thus we employed the same patch size in our network.

## 3.4 Loss functions

Our objective loss function contains: i) adversarial losses for matching the distribution of generated images to the data distribution in the target domain; ii) cycle consistency losses to prevent the learned mappings $G_A$ and $G_B$, as mentioned in Fig. 2a, from contradicting each other, and iii) cyclic perceptual loss to help the generators generate more visually pleasing images. Next we describe these loss functions in detail.

### 3.4.1 Adversarial loss

Here we adopted Least Squares GAN (LSGAN) [25] to calculate the adversarial loss. It has been shown that LSGAN is able to generate higher quality images than regular GANs (*aka.* Vanilla GANs) [25]. In addition, it is more stable during the learning process. Equations 2 and 3 show how we calculate the adversarial loss for the global discriminator and the global generator respectively.

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

5

$$L_G^{Global} = E_{x_r \sim P_{fake}}[(D(x_f) - 1)^2] \tag{3}$$

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 enhance small regions of the hazy image. Equations 4 and 5 depicts the corresponding loss functions:

$$L_D^{Local} = E_{x_r \sim P_{real-patches}}[(D(x_r) - 1)^2] + E_{x_f \sim P_{fake-patches}}[(D(x_f) - 0)^2] \tag{4}$$

$$L_G^{Local} = E_{x_f \sim P_{fake-patches}}[(D(x_f) - 1)^2] \tag{5}$$

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

### 3.4.2 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 unpaired image-to-image translation process [43]. Cycle consistency loss can be written as:

$$L_{cycle}(G_A, G_B) = E_{x \sim p_{data(x)}}[\|(G_B(G_A(x)) - x)\|]_1 + E_{y \sim p_{data(y)}}[\|(G_A(G_B(y)) - y)\|]_1 \tag{6}$$

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 $y$.

### 3.4.3 Color loss

Hazy images usually lack brightness and contrast, to improve these lacking features we employed color loss [11] to measure the color difference between the enhanced images and haze-free images. This loss function forces the generator to generate images with the same color distribution as the haze-free images.

To implement this loss function we apply a Gaussian blur to the enhanced image and the haze-free image and then we compute the Euclidean distance between the blurred images.

$$L_{color}(A, B) = \|(A_{blurred} - B_{blurred})\|_2^2 \tag{7}$$

$A_{blurred}$ and $B_{blurred}$ are the blurred images of A and B images.

### 3.4.4 Cyclic perceptual loss

Adversarial and cycle consistency loss are not able to preserve the textual and perceptual information of corrupted hazy images. Therefore, to achieve the perceptual similarity we employed a cyclic perceptual loss inspired by cycle consistency loss, to make sure that this information is recovered as much as possible.

This loss can be calculated using a pre-trained VGG16 [33] model to calculate the feature space distance between images. This loss function has been employed by other vision-based tasks [16, 17]. It constrains the extracted feature distance between the output image and its ground truth.

Since our method is designed using unpaired supervision, *i.e.* the ground truth of images is not available, perceptual loss cannot be directly applied. Thus a modified version of perceptual loss is adopted when the ground truth is unavailable, originally proposed by [14]. They 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.

The goal of this loss function is to preserve the image structure and content features during dehazing. To calculate this loss, we focused on feature maps extracted from the $2^{nd}$ and $5^{th}$ pooling layers of VGG-16 pre-trained model.

Equation 8 shows how this loss is calculated:

$$Loss_{CP}(I_h) = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\sigma_{i,j}(I_h) - \sigma_{i,j}(G(I_h)))^2 \tag{8}$$

where $I_h$ represents the hazy input image and $G(I_h)$ represents the dehazed output. Symbol $\sigma_{i,j}$ denotes the features extracted from the pre-trained (on ImageNet dataset) VGG16 model. $W_{i,j}$ and $H_{i,j}$ are the dimensions of the extracted feature. $i$ shows the $i_{th}$ max pooling and $j$ denotes the $j_{th}$ conv layer after $i_{th}$ max pooling layer.

We also add an instance normalization layer [38] after the VGG feature maps before feeding into $L_{CP}$ and $L_{CP}^{Local}$ in order to stabilize training. 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 8.

### 3.5 Final objective function

The overall loss function for training Dehaze-GLCGAN is defined as follows:

$$Loss_{total} = L_{global}^{GAN} + L_{local}^{GAN} + L_{global}^{Cycle} + L_{local}^{Cycle} + L_{global}^{CP} + L_{local}^{CP} + L_{global}^{color} + L_{local}^{color} \tag{9}$$

## 4 Experiments and Results

Previous *paired* methods have primarily reported results on the ITS dataset [19]. To compare with those methods (even though our method is unpaired), we evaluated our method on the ITS dataset. However, we should stress that, due to lack of computing power, **we only train our model over 20% of the ITS training set, while previous methods are trained over the entire ITS dataset.** The ITS dataset contains 100,000 indoor hazy images. This dataset is part of the RESIDE dataset [19], which is a standard dataset of synthetic hazy images. RESIDE also has a test dataset called SOTS which contains 500 indoor hazy images and 500 outdoor hazy images with their clear ground truth. We report results on the SOTS test dataset.

To evaluate the performance of our method compared to previous *unpaired* methods, we trained a model over NYU-Depth dataset [1], and test it on the NYU-Depth and Middlebury datasets. We should note that **previous unpaired methods were trained over the augmented NYU-Depth dataset, while we do not use the augmented data**. The NYU-Depth and Middlebury datasets belong to the D-hazy dataset [1], which is a standard synthetic hazy dataset for image dehazing. NYU-Depth contains 1449 hazy images couples with their ground truth images and Middlebury has 23 high-resolution (2k images) hazy images with their ground truth. NYU-Depth has a ground truth image for each hazy image. Since our method uses unpaired supervision, the training process received no information about which haze-free image corresponds to which hazy image.

### 4.1 Training

For training we need two sets of training data: trainA includes hazy images and trainB includes ground truth images (shuffled to simulate the unpaired supervision like other unpaired methods [40]). We opted for Adam optimizer ($momentum = 0.5$) with batch size of 1. Our initial learning rate ($\alpha$) was set to 0.0002 for 100 epochs, with linear decay to zero over the next 100 epochs. We implemented our model in PyTorch using eight NVIDIA Tesla P100 GPUs and trained our network for 200 epochs.

### 4.2 Quality Measures

We used the following measures, to analyze the performance of our proposed method:

- **PSNR (Peak Signal to Noise Ratio)**: 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.
- **SSIM (Structural Similarity Index)**: is a perceptual metric that quantifies image quality degradation caused by processing. In this measurement, image degradation is considered as the change of perception in structural information [15].
- **PI (Perceptual Index)**: measures the quality of restored and reconstructed images based on human perception [5]. A lower perceptual index indicates better perceptual quality.
- **CIEDE2000**: This measures the color difference between hazy and dehazed images; smaller values indicate better color preservation, thus better dehazing and perceptual quality [24].

**PSNR: 12.14, SSIM: 0.74**   **PSNR: 13.43, SSIM: 0.78**   **PSNR: 19.11, SSIM: 0.90**

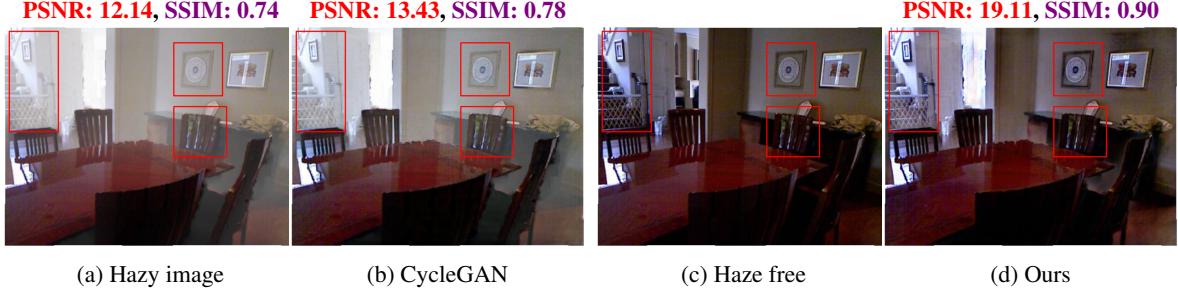(a) Hazy image          (b) CycleGAN          (c) Haze free          (d) Ours

Figure 3: Comparison between CycleGAN and the proposed method.

Table 1: Ablation study over NYU-Depth dataset. The larger values of PSNR, SSIM and the smaller value of CIEDE2000 indicate the better dehazing and perceptual quality.

| Setting | ↑ PSNR | ↑SSIM | ↓ CIEDE2000 |
|---|---|---|---|
| CycleGAN | 13.3879 | 0.5223 | 17.6113 |
| Dehaze-GLCGAN w/o color loss | 14.5402 | 0.7407 | 15.6401 |
| Dehaze-GLCGAN w/o perceptual loss | 14.6582 | 0.7312 | 15.6348 |
| Dehaze-GLCGAN w/o residual blocks | 14.1092 | 0.6923 | 16.4344 |
| Dehaze-GLCGAN w/o local discriminator | 14.0681 | 0.7111 | 19.9466 |
| Dehaze-GLCGAN | **15.4780** | **0.7808** | **14.8877** |

## 4.3 Ablation Study

To demonstrate the effectiveness of the local discriminator and cyclic perceptual loss, we conduct several ablation experiments. In particular, to evaluate the effectiveness of local discriminators, we trained our model without the local discriminators and tested the model on NYU-Depth dataset. Afterwards, to demonstrate the effect of the perceptual loss, we trained a model without this loss function and evaluated the model on NYU-Depth dataset.

Table 1 depicts the results of our ablation study in terms of PSNR, SSIM, and CIEDE2000. It is observed that incorporating local discriminators can help achieve better PSNR, SSIM and CIEDE2000, meaning better restoration and generation of more visually pleasing results.

Having removed cyclic perceptual and color loss, our model achieved lower PSNR, SSIM, and CIEDE2000, compared to its counterpart, *i.e.* perceptual loss and color incorporated. This means our generator can generate more natural and perceptually pleasing clean images when the cyclic perceptual and color loss function is utilized. The best results in terms of PSNR, SSIM, and CIEDE2000 are achieved when both local discriminators and cyclic perceptual loss are incorporated.

Figure 3 also shows the comparison of the proposed method with CycleGAN. As we can see, the proposed method can achieve higher PSNR and SSIM compared to CycleGAN and removes more hazy.

## 4.4 Quantitative and Qualitative Analysis

**Comparison with paired methods:** To evaluate the performance of our proposed method quantitatively, we first compare our model with *paired* methods on the SOTS test dataset, since that is the dataset where existing paired methods have reported results. As noted earlier, for these experiments, our method as well as the competitors are trained on the ITS dataset, however our method was only trained on 20% of that dataset.

Table 2 shows the results on SOTS test dataset. It demonstrates that Dehaze-GLCGAN achieves the best performance in single image dehazing in terms of both PSNR and SSIM. These results were achieved despite the fact that our method is unpaired, and the competitors use paired supervision.

**Comparison with unpaired methods:** We also compare our model with *unpaired* methods, on the NYU-Depth and Middlebury datasets. As noted earlier, for these experiments, our method as well as the competitors are trained on the NYU-Depth dataset, however the competitors used an augmented version of that dataset, whereas we used the original, unaugmented version. Table 3 and 4 show the results on NYU-Depth and Middle-berry datasets respectively. Our method outperforms the other methods in terms of PASNR and SSIM.

Table 2: Results on SOTS outdoor dataset. Most of the numbers for the previous work are taken from [29].

| Method | ↑PSNR | ↑SSIM | ↓PI |
|---|---|---|---|
| DCP [10] | 19.13 | 0.8148 | 2.5061 |
| DehazeNet [6] | 22.46 | 0.8514 | 2.4346 |
| AOD-NET [18] | 20.29 | 0.8765 | 2.4280 |
| DCPDN [41] | 19.93 | 0.8449 | 2.7269 |
| EPDN [29] | 22.57 | 0.8630 | **2.3858** |
| **Ours** | **23.0276** | **0.9165** | 2.4051 |

Table 3: Results on NYU-Depth dataset. The numbers for the previous work are taken from [9, 40].

| Method | ↑PSNR | ↑SSIM | ↓CIEDE2000 |
|---|---|---|---|
| CycleGAN [6] | 13.3879 | 0.5223 | 17.6113 |
| Cycle-Dehaze [18] | 15.41 | 0.66 | NA |
| DDN [10] | 15.5456 | 0.7726 | 11.8414 |
| **Ours** | **15.9780** | **0.8208** | **10.3862** |

Table 4: Results on Middlebury dataset. The numbers for the previous work are taken from [9, 40].

| Method | ↑PSNR | ↑SSIM | ↓CIEDE2000 |
|---|---|---|---|
| CycleGAN [6] | 11.3037 | 0.3367 | 26.3118 |
| Cycle-Dehaze [18] | 12.7634 | 0.5812 | NA |
| DDN [10] | 14.9539 | 0.7741 | **13.4826** |
| **Ours** | **15.6802** | **0.8482** | 16.6745 |



| (a) Hazy | (b) DehazeNet | (c) AOD-Net | (d) EPDN | (e) Ours | (f) Haze free |

Figure 4: Comparison of the state-of-the-art dehazing methods on NYU-Depth dataset. Please magnify to see the details. For high resolution version of these images, please refer to the supplementary material.

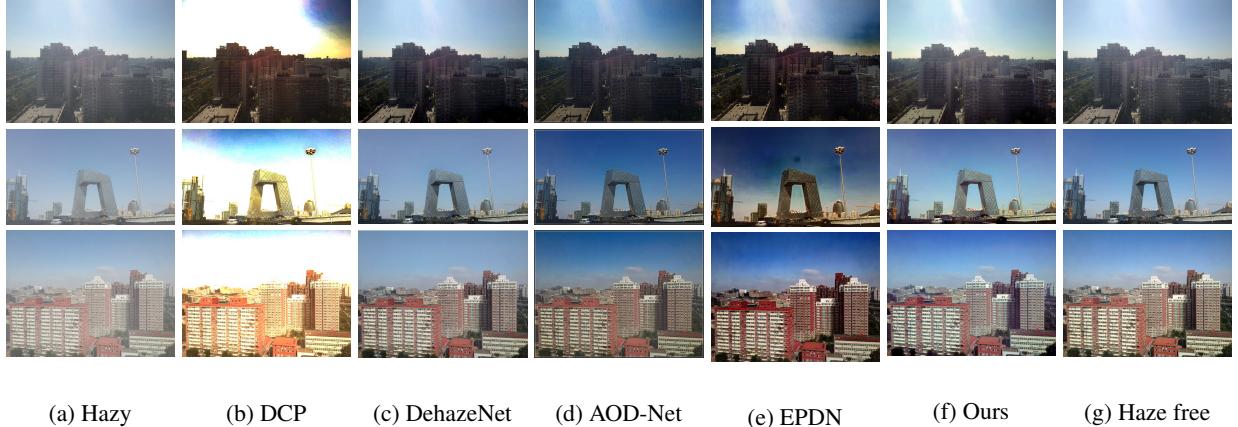| (a) Hazy | (b) DCP | (c) DehazeNet | (d) AOD-Net | (e) EPDN | (f) Ours | (g) Haze free |

Figure 5: Comparison of the state-of-the-art dehazing methods on SOTS dataset. For high resolution version of these images, please refer to the supplementary material.

Figure 4 depicts the qualitative results of our proposed method vs. other single image dehazing methods on NYU-Depth images. As one can see, our method removes more haze as shown in the red boxes. In addition, Figure 5 depicts the qualitative results of our proposed method vs. other single image dehazing methods on SOTS images. Our method generates more visually pleasing images while effectively removing haze compared to to other methods.

DCP which is one of the earliest prior-based methods suffers from color distortion and over-exposure. Cycle-Dehaze and CycleGAN fail to remove much haze from dense hazy images. AOD-Net and DehazeNet fail to remove much haze from dense hazy images. EPDN adds generates much darker images compared to the ground truth. DCP suffers from color distortion and over-exposure. DCPDN also suffers from over-exposure.

Our method, Dehaze-GLCGAN, on the other hand generates more natural and visually pleasant haze-free images and much closer to the ground truth image. Moreover, our method outperforms the above-mentioned methods in recovery of details, and generates more natural images.

## 5 Conclusion

In this paper, we treated the image dehazing problem as an image-to-image translation problem, and proposed a cycle-consistent generative adversarial network, called Dehaze-GLCGAN, for unpaired image dehazing. Dehaze-GLCGAN utilizes discriminators with a local-global structure to remove haze effectively and perceptual and color loss to generate realistic clean images. Using three benchmark test datasets, we showed the effectiveness of the proposed method.

We showed that the global-local discriminators structure as well as perceptual and color loss 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.

## References

[1] ANCUTI, C., ANCUTI, C. O., AND DE VLEESCHOUWER, C. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In *2016 IEEE International Conference on Image Processing (ICIP)* (2016), IEEE, pp. 2226–2230.

[2] ANCUTI, C., ANCUTI, C. O., DE VLEESCHOUWER, C., AND BOVIK, A. C. Night-time dehazing by fusion. In *2016 IEEE International Conference on Image Processing (ICIP)* (2016), IEEE, pp. 2256–2260.

[3] ANCUTI, C. O., ANCUTI, C., HERMANS, C., AND BEKAERT, P. A fast semi-inverse approach to detect and remove the haze from a single image. In *Asian Conference on Computer Vision* (2010), Springer, pp. 501–514.

[4] ANVARI, Z., AND ATHITSOS, V. 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.

[5] BLAU, Y., MECHREZ, R., TIMOFTE, R., MICHAELI, T., AND ZELNIK-MANOR, L. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 0–0.

[6] CAI, B., XU, X., JIA, K., QING, C., AND TAO, D. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing 25*, 11 (2016), 5187–5198.

[7] CHEN, J., CHEN, J., CHAO, H., AND YANG, M. 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.

[8] EMBERTON, S., CHITTKA, L., AND CAVALLARO, A. Hierarchical rank-based veiling light estimation for underwater dehazing.

[9] ENGIN, D., GENÇ, A., AND KEMAL EKENEL, H. 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.

[10] HE, K., SUN, J., AND TANG, X. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence 33*, 12 (2010), 2341–2353.

[11] IGNATOV, A., KOBYSHEV, N., TIMOFTE, R., VANHOEY, K., AND VAN GOOL, L. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 3277–3285.

[12] ISOLA, P., ZHU, J.-Y., ZHOU, T., AND EFROS, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134.

[13] JIANG, H., AND LEARNED-MILLER, E. Face detection with the faster r-cnn. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (2017), IEEE, pp. 650–657.

[14] JIANG, Y., GONG, X., LIU, D., CHENG, Y., FANG, C., SHEN, X., YANG, J., ZHOU, P., AND WANG, Z. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972* (2019).

[15] KUMAR, R., AND MOYAL, V. Visual image quality assessment technique using fsim. *International Journal of Computer Applications Technology and Research 2*, 3 (2013), 250–254.

[16] KUPYN, O., BUDZAN, V., MYKHAILYCH, M., MISHKIN, D., AND MATAS, J. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 8183–8192.

[17] LEDIG, C., THEIS, L., HUSZÁR, F., CABALLERO, J., CUNNINGHAM, A., ACOSTA, A., AITKEN, A., TEJANI, A., TOTZ, J., WANG, Z., 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.

[18] LI, B., PENG, X., WANG, Z., XU, J., AND FENG, D. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 4770–4778.

[19] LI, B., REN, W., FU, D., TAO, D., FENG, D., ZENG, W., AND WANG, Z. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing 28*, 1 (2018), 492–505.

[20] LIN, W.-A., CHEN, J.-C., CASTILLO, C. D., AND CHELLAPPA, R. Deep density clustering of unconstrained faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 8128–8137.

[21] LIN, W.-A., CHEN, J.-C., AND CHELLAPPA, R. A proximity-aware hierarchical clustering of faces. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (2017), IEEE, pp. 294–301.

[22] LIU, W., ANGUELOV, D., ERHAN, D., SZEGEDY, C., REED, S., FU, C.-Y., AND BERG, A. C. Ssd: Single shot multibox detector. In *European conference on computer vision* (2016), Springer, pp. 21–37.

[23] LONG, J., SHELHAMER, E., AND DARRELL, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.

[24] LUO, M. R., CUI, G., AND RIGG, B. 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 26*, 5 (2001), 340–350.

[25] MAO, X., LI, Q., XIE, H., LAU, R. Y., WANG, Z., AND PAUL SMOLLEY, S. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 2794–2802.

[26] MCCARTNEY, E. J. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons, Inc., 1976. 421 p.* (1976).

[27] MENG, G., WANG, Y., DUAN, J., XIANG, S., AND PAN, C. Efficient image dehazing with boundary constraint and contextual regularization. In *Proceedings of the IEEE international conference on computer vision* (2013), pp. 617–624.

[28] NARASIMHAN, S. G., AND NAYAR, S. K. Chromatic framework for vision in bad weather. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)* (2000), vol. 1, IEEE, pp. 598–605.

[29] QU, Y., CHEN, Y., HUANG, J., AND XIE, Y. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 8160–8168.

[30] REDMON, J., DIVVALA, S., GIRSHICK, R., AND FARHADI, A. 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.

[31] REN, W., LIU, S., ZHANG, H., PAN, J., CAO, X., AND YANG, M.-H. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision* (2016), Springer, pp. 154–169.

[32] SCHROFF, F., KALENICHENKO, D., AND PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 815–823.

[33] SIMONYAN, K., AND ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[34] SRINIVASA, G., AND SHREE, K. Vision and the atmosphere. *International Journal of Computer Vision 48*, 3 (2002), 233–254.

[35] SWAMI, K., AND DAS, S. K. Candy: Conditional adversarial networks based fully end-to-end system for single image haze removal. *arXiv preprint arXiv:1801.02892* (2018).

[36] TAN, R. T. Visibility in bad weather from a single image. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (2008), IEEE, pp. 1–8.

[37] TAREL, J.-P., AND HAUTIERE, N. Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th International Conference on Computer Vision* (2009), IEEE, pp. 2201–2208.

[38] ULYANOV, D., VEDALDI, A., AND LEMPITSKY, V. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 6924–6932.

[39] YANG, S., LUO, P., LOY, C.-C., AND TANG, X. Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 5525–5533.

[40] YANG, X., XU, Z., AND LUO, J. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *Thirty-second AAAI conference on artificial intelligence* (2018).

[41] ZHANG, H., AND PATEL, V. M. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 3194–3203.

[42] ZHANG, H., SINDAGI, V., AND PATEL, V. M. Joint transmission map estimation and dehazing using deep networks. *arXiv preprint arXiv:1708.00581* (2017).

[43] ZHU, J.-Y., PARK, T., ISOLA, P., AND EFROS, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2223–2232.