Generative Adversarial Networks for Effective Single Image Dehazing

Jaehyeok Kim Hyeonjae Kim
The Hong Kong University of Science and Technology
{jkimbf, hkimar}@connect.ust.hk

Abstract

Single image dehazing is a challenging task which is critical for success in downstream computer vision tasks. Recently, generative adversarial networks (GANs) have achieved significant advancement in single image dehazing and received great attention in research. However, a great number of existing learning-based dehazing models are still not fully end-to-end, following the conventional dehazing methods. Yet, due to the ill-posed natures of the problem, it is intricate to accurately estimate intermediate parameters in the conventional methods. On top of that, most existing methods train the model only on synthetic hazy images, thereby making the model hard to generalize well on real hazy images. To address this problem, we propose a novel GAN-based end-to-end model that incorporates the strength of two prior GAN-based works. Our model shows satisfactory performance even with the limited amount of high-quality real-world hazy images. At the same time, our model is able to generate more authentic and natural dehazed images with lessened color distortion and fewer artifacts. Experimental results on real-world images show that our model shows comparable performance with the state-of-the-art dehazing algorithms and improved qualitative aspects in terms of color casting and color constancy. The code is released in https://github.com/jkimbf/DoubleGANDehaze.

1. Introduction

The particles suspended in the atmosphere, such as dust, mist and smog, significantly degrade the image qualities in our daily life. Such turbid mediums often corrupt outdoor images, resulting in undesirable image qualities such as lowered contrast and limited visibility. Due to the negative effects on the performance of ensuing high-level computer vision applications, such as object localization and image segmentation, increasing attention has been drawn to the image dehazing task. In order to address this issue, the physical scattering model [14] has been widely adopted

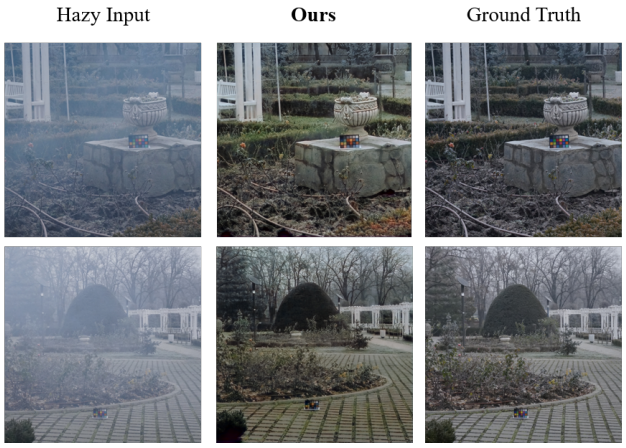

Figure 1: Dehazing examples on real-world hazy images. Our model is able to remove the haze effect in a thorough manner and therefore generate natural and visually satisfying results with low color distortion.

as a conventional method to model a hazy input:

$$I(x) = J(x)t(x) + A(1–t(x)) \quad (1)$$

where $I(x)$, $J(x)$ denote the observed hazy image and the scene radiance of pixel x, respectively. In addition, A is the global atmospheric light and $t(x)$ is the transmission map. Since only $I(x)$ is known, the single image dehazing problem that aims to estimate the $J(x)$ is ill-posed.

Recently, various approaches such as conventional physics-based models [9, 7, 13] and deep learning-based methods [5, 12, 15, 18, 6, 17] have been proposed to solve this problem. In physics-based models, intermediate variables such as A and $t(x)$ are estimated using hand-crafted priors. On the other hand, deep learning-based models employ convolutional neural networks (CNNs) [5, 12, 15] or generative adversarial networks (GANs) [18, 6, 17] to extract features and learn the mappings between hazy and ground-truth images. Currently, most learning-based models still adopt the conventional procedure of estimating intermediate variables in the equation (1). Yet, owing to the inadequacy of real-world hazy images and effective priors,

it is demanding to accurately estimate these intermediate parameters. In order to handle this issue, fully end-to-end learning-based models [5, 12, 15] have been proposed. Nevertheless, a number of models suffer from color reconstruction issues such as color casting and color distortion.

To tackle the aforementioned problems simultaneously, we propose a fully end-to-end trainable dehazing framework that utilizes a Generative Adversarial Networks (GAN) [8] architecture. To be specific, to handle color distortion issues and improve color constancy, we adopted the fusion-discriminator (FD) architecture from prior work [6]. At the same time, inspired by the work by Singh *et al*. [18], we adopted iterated UNet block and Back Projected Pyramid architecture for the generator in our model, thereby making the generator able to learn diverse and intricate features of haze without losing global and local structural information. Our key contributions are summarized as follows:

- We incorporate the novelties of two GAN-based prior works so that the network not only can learn multiple levels of complexities at different scales but also generate realistic and natural haze-free images.
- We propose a novel end-to-end single image dehazing algorithm BF-GAN, which is able to output clear images without the estimation of intermediary parameters, such as I, J and A in the classical atmospheric scattering model.
- We adopt the data augmentation method to let our model be trained only with a limited amount of real-world hazy images.
- Comprehensive experiment has been done on two contemporary challenging datasets, namely O-Haze of NTIRE 2018 challenge and NH-Haze2 of NTIRE 2021 challenge. Through this experiment, our model is shown to have comparable performance with the state-of-the-art model and improve color reconstruction issues at the same time.

2. Related Work

There have been various methods tackling the single image dehazing task. Most of these early works utilized the physical scattering model that is highly ill-posed. As briefly mentioned in the introduction, earlier works are largely categorized into prior-based and learning-based methods. In recent years, GAN-based methods also have been actively studied and have achieved impressive performance.

Prior-based methods: Based on the physical scattering model, a number of conventional dehazing algorithms aim to estimate $t(x)$ and A from the given input hazy image $I(x)$. The estimation of $t(x)$ and A, however, suffers from the ill-posed nature of the problem. Early physical prior-based methods attempt to estimate the transmission map $t(x)$ by utilizing the statistical aspects of clear images (e.g., dark channel prior [9] and color-line prior [7]). However, these priors are hardly applicable in general due to airlight-albedo ambiguity [13] and unreliability of these priors.

Learning-based methods: To handle these issues, convolutional neural networks (CNNs) have been adopted to estimate transmissions [4, 7, 9] or predict clear images in an end-to-end fashion [5, 12, 15]. These methods show superior performance compared to the prior-based approaches. However, they are likely to suffer from color distortion, artifacts and insufficient removal of haze due to the absence of priors. In addition, deep learning on a small dataset to dehaze a single RGB image has remained as a notoriously difficult task.

GAN-based methods: GAN [8] is an architecture that comprises two models: a generator and a discriminator and they are trained with minimax two-player game. With the promising performance of GAN, multiple dehazing approaches have been proposed with GAN architecture and continuously achieved a great performance. In particular, Singh *et al*. [18] and Dong *et al*. [6] were proposed and achieved state-of-the-art at the time. Those GAN-based methods propose a special architecture for the generator and the discriminator so that the models can effectively learn the dehazing task. For instance, Singh *et al*. [18] addressed the issue of deficiency in the training dataset by using patch-discriminator structure. However, GAN-based methods still suffer from lingering issues such as undesirable color distortion and color casting.

3. Data

For training and testing purposes, we adopted real-world hazy image datasets that are annually offered by New Trends in Image Restoration and Enhancement (NTIRE) workshop. We used the NTIRE 2018 outdoor dataset (O-Haze) [2] for the training purpose and NTIRE 2021 non-homogeneous dataset (NH-Haze2) [3] for the testing purpose.

3.1. Training Datasets

In order to measure our model's relative performance against the baseline model [18], we have trained our model on the O-Haze dataset. This dataset includes 25 hazy training images of size 2833×4567 pixels. This dataset contains 5 hazy images for the validation purpose, together with their corresponding ground truth data. Since we used this dataset only for training purposes, we utilized all the available images in the dataset as training images.

3.2. Test Datasets

After training our model on images from the O-Haze dataset, we used NH-Haze2 images for testing the performance of each model. This dataset contains 35 hazy images

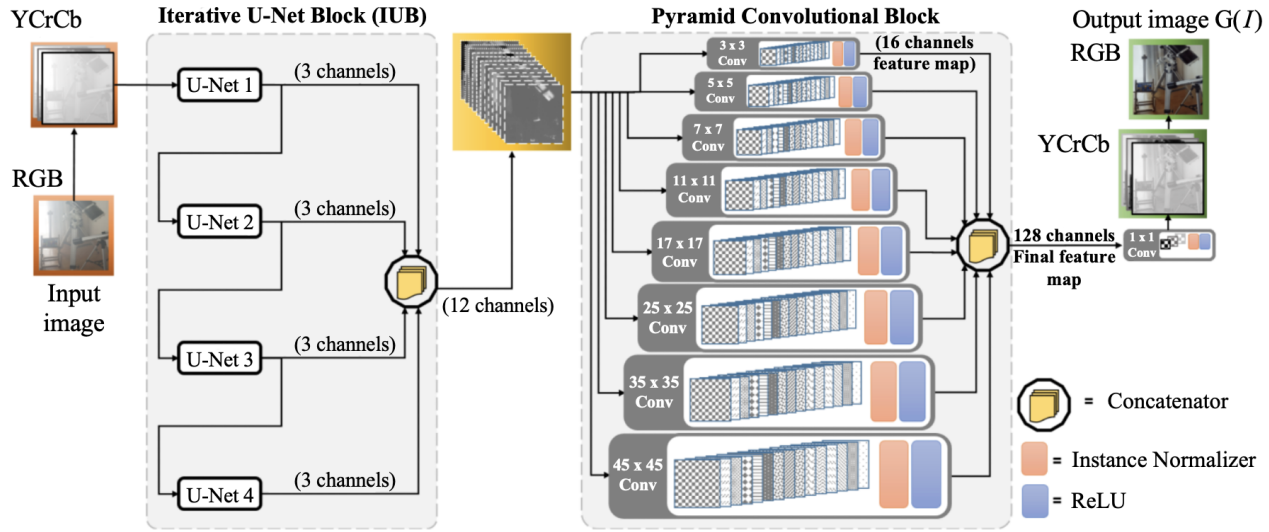

Figure 2: The architecture of the generator [18]

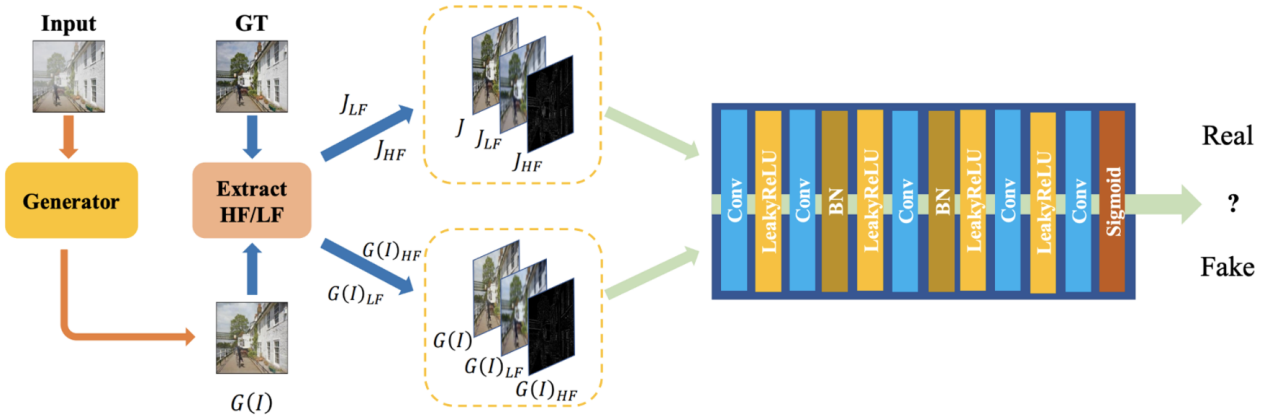

Figure 3: The architecture of the discriminator [6]

and their corresponding hazy-free (ground truth) images. With this dataset, we measured the test performance of our model and baseline model by evaluating PSNR, SSIM metrics which are widely adopted evaluation metrics in various computer vision tasks.

4. Methods

In this section, we introduce the design of our new Single Image Dehazing framework. Our framework collaboratively leverages the architectural designs of the back-projected pyramid network (BPPNet) [18] and the GAN with Fusion-Discriminator (FD-GAN) [6]. Both works propose end-to-end trainable GAN architectures with their novel designs of generator and discriminator. To effectively improve the image dehazing performance, we adopt the architectures of the generator and discriminator from [18] and [6], respectively.

4.1. Generator

For the generator of our framework, we adopt the generator architecture proposed by Singh *et al*. [18]. It is proposed to handle challenging haze conditions such as dense or non-homogeneous haze. As mentioned earlier, BPPNet adopts GAN architecture and introduces a novel generator architecture. The generator comprises Iterative UNet Block (IUB) and pyramid convolution (PyCon) block as shown in Fig. 2.

Iterative UNet Block (IUB) is a serie of multiple UNet

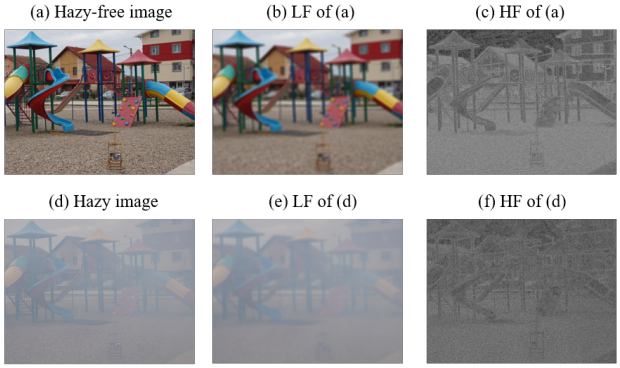

Figure 4: Compared to haze-free images, hazy images are often characterized by their low contrast, color saturation and shortage of edge information.

[16] units. The first UNet takes the hazy image (in YCrCb space) as input and subsequent UNet units take the 3-channel output of the previous UNet unit. All 3-channel outputs from each UNet unit are then concatenated to create an intermediate feature map. The following equations describe how IUB works:

$$I_1 = \text{UNet}_1(I_{haze}); \;\; I_i = \text{UNet}_i(I_{i-1}) \quad \text{for } \; i > 1, \quad (2)$$

$$\hat{I}_{\text{IUB}} = I_1 \oplus I_2 \oplus ... \oplus I_M \quad (3)$$

where $\oplus$ is a concatenation operator and M is the number of UNet units in IUB. We used $M = 4$ as suggested by the authors. Since each UNet is an encoder-decoder pair, IUB can be interpreted as a sequence of encoder-decoder pairs. This allows the generator of our framework to learn the complex features of haze without losing the structural information and the spatial features.

Pyramid convolution (PyCon) block mitigate the problem that the output of IUB lacks the global and local structural information for objects with different scales. As shown in Fig. 3, PyCon block employs 8 different kernel sizes in parallel on the 12-channel intermediate output $\hat{I}_{\text{IUB}}$ computed from IUB. For each kernel size, there are 16 different kernels applied to the input. Since different kernel size results in different output dimensions, proper padding values are used to keep the output dimensions the same. 8 16-channel feature maps are then concatenated into 128-channel. The final feature map is then passed to a 1x1 convolution layer and the sigmoid activation function to reconstruct the final image (in YCrCb space). By utilizing the PyCon block, spatial features of different scales structural information can be obtained by the generator.

4.2. Discriminator

For the discriminator, we decided to adopt the architecture of the fusion-discriminator proposed by Dong *et al*. [6]. Therefore, our discriminator utilizes the frequency information as additional priors. An image can be decomposed into high-frequency (HF) and low-frequency (LF) components that respectively contain different information of the image. We used the Laplace operator with a window size 3 to generate the HF image. Also, a Gaussian filter with window size 15 and σ=5 is used to generate the LF image. As shown in Fig. 4, HF and LF images are generated for both ground-truth and fake images, and then concatenated together before being passed to the discriminator. Accordingly, the discriminator can better distinguish the differences between the hazy and ground truth images. The discriminator consists of 4 convolutional layers followed by LeakyReLU and another convolutional layer followed by Sigmoid activation. For the second and third convolutional layers, a batch normalization layer is attached between the convolutional layer and the LeakyReLU activation layer.

4.3. Loss Function

Pixel-wise loss. Given an input hazy image I_h, the output of the generator is $G(I_h)$ and the ground truth is J_h. Then l_2 loss, also referred to as MSE loss, is defined to be the Euclidean distance between I_h and J_h and can be written as:

$$l_2 = \sum_{h=1}^{N} ||G(I_h) - J_h||_2 \quad (4)$$

SSIM loss. The structural similarity over reconstructed image $G(I_h)$ and ground truth J_h is proposed to measure the similarity between two images. It can be written as:

$$SSIM(G(I_h), J_h) = \frac{2\mu_{G(I_h)}\mu_{J_h} + \epsilon_1}{\mu^2_{G(I_h)} + \mu^2_{G(J_h)} + \epsilon_1} \cdot \frac{2\sigma_{G(I_h)J_h} + \epsilon_2}{\sigma^2_{G(I_h)} + \sigma^2_{G(J_h)} + \epsilon_2} \quad (5)$$

where μ_x, σ^2_x are the mean and variance of x, respectively. In addition, σ_{xy} is the covariance of x and y and ϵ_1 and ϵ_2 are constants for ensuring numerical stability. The SSIM range is between 0 and 1 and the loss l_{SSIM} is defined as follows:

$$l_{\text{SSIM}} = 1 - SSIM(G(I), J) \quad (6)$$

Content loss. In addition to pixel-wise loss, we introduce VGG-based perceptual loss l_C [10] as our content loss to evaluate the perceptual similarity in feature space:

$$l_C = \sum_{k=1}^{L} ||\phi_k(G(I)) - \phi_k(J)||_1 \quad (7)$$

where ϕ_k denotes the feature maps obtained by the ith activation layers of the VGG-19 network.

Figure 5: Qualitative comparison of the baseline [18] with our model on NH-Haze2 [3] dataset

Adversarial loss. The adversarial loss for generator l_G and discriminator l_{D} is defined to be:

$$l_G = \log(D(G(I) \oplus G(I)_{\text{LF}} \oplus G(I)_{\text{HF}}) \tag{8}$$

$$l_D = \log(D(J \oplus J_{\text{LF}} \oplus J_{\text{HF}}) + \log(1 - D(G(I) \oplus G(I)_{\text{LF}} \oplus G(I)_{\text{HF}}) \tag{9}$$

where $D(.)$ denotes the predicted probability that input is indeed real, which is estimated by the discriminator. In addition, $\oplus$ denotes the concatenation operation. This loss term lets our network favor solutions that have a similar manifold with the real clear images.

Overall loss function. Ultimately, we combine the aforementioned loss terms together to have generator loss l_{GEN} and discriminator loss l_{DISC}:

$$l_{\text{GEN}} = \alpha_1 l_2 + \alpha_2 l_{\text{SSIM}} + \alpha_3 l_C + \alpha_4 l_G \tag{10}$$

$$l_{\text{DISC}} = \alpha_5 l_D \tag{11}$$

where α_1, α_2, α_3, α_4 and α_5 are positive weights. For our model, we have heuristically chosen the values of the above constants as $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.7, 0.5, 1.0, 1.0, 1.0)$

4.4. Our Architecture

In this project, we propose to adopt the Fusion-discriminator from FD-GAN [6] to the BPPNet [18] architecture. BPPNet has achieved state-of-the-art performance on the NTIRE datasets by using a simple discriminator. We believe that the incorporation of the frequency information will further enhance the performance. Therefore, our

proposed architecture replaces the discriminator of BPPNet with the Fusion-discriminator.

5. Experiments

5.1. Implementation Details

We used the Adam [11] optimizer for the training with the initial learning rate of $\lambda_{gen} = 0.001$ and $\lambda_{disc} = 0.0005$ for the generator and discriminator respectively. In order to address the problem with the shortage of high-quality real-world hazy images, we adopted the random crop technique which is one of the most popular data augmentation methods. We have randomly cropped square patches of size 1024 $\times$ 1024 from the training images. After applying random crop, we resize each square patch into 512 $\times$ 512 using bicubic interpolation. This technique has created a more extensive dataset from the small-sized real hazy image dataset. Furthermore, we transformed the color space of each image patch from RGB space to YCbCr space.

As explained in the previous section, the discriminator of our model receives 7 channel inputs which consist of 3 channels of real or fake images and 3 channels of Low Frequency (LF) images and 1 channel of High Frequency (HF) images. To extract LF data, we applied a Gaussian filter of window size 15 and standard deviation $\sigma = 3$ to the real or fake images. For HF data, we first transform the YCbCr images into a grayscale image and then apply the Laplacian operator to the transformed images. The Laplacian operator that we utilized had a kernel size of 3.

We reduced the learning rate of the generator by a factor of 10 when the loss becomes stagnated. We terminated the training as the learning rate λ_{gen} reached 0.00001. The network was trained for 300 epochs by Pytorch with two Nvidia RTX 3090 GPUs.

5.2. Comparison with state-of-the-art methods

We compare our proposed method with BPPNet [18], which showed state-of-the-art performance on O-Haze [2] and I-Haze [1] datasets. The evaluation is conducted on real-world images, namely, NH-Haze2 datasets [3]. We used the weights that were pre-trained and released by authors and adopted the evaluation metrics offered from NTIRE Challenge.

Model	SSIM ↑	PSNR ↑
Baseline	0.7309	14.9504
Ours	0.6595	13.4771

Table 1: Quantitative comparisons with baseline model [18]

5.3. Results

We have used Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) for quantitative analysis of our model. The results are presented in Table 1. The average SSIM and PSNR values are 0.6595, 13.4771 respectively. These SSIM and PSNR values show our model is as robust as the baseline models. Furthermore, our model improves the baseline model in terms of the following qualitative properties: (1) color preservation and (2) color cast. As shown in Figure 5, the BPPNet model loses color information of the original hazy image. For example, most of the green parts, such as plants and grass, in the original images are recognized as haze by the baseline model so that these elements are removed completely. These undesirable color destruction problems can be solved by our model. Our model is shown to be able to generate more naturally dehazed results with sharper textures and better color fidelity. These features make our results more visually akin to the ground-truth images.

6. Conclusion

In this project report, we propose an end-to-end GAN-based single image dehazing framework. The generator architecture with IUB and Pycon block enables learning of multiple levels of complexities and structural and spatial information at multiple scales. The incorporation of a fusion discriminator successfully complements the drawbacks of the baseline model by improving the color reconstruction. Despite the limited amount of time, we successfully showed that the adoption of frequency information can further improve performance. In the future, we plan not only to perform more hyperparameter tuning but also to utilize the synthetic dataset with a domain adaptation module to observe a better result than the current version of our model.

References

[1] C. O. Ancuti, C. Ancuti, R. Timofte, and C. D. Vleeschouwer. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 620–631, 2018.

[2] C. O. Ancuti, C. Ancuti, R. Timofte, and C. D. Vleeschouwer. O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 867–8678, 2018.

[3] C. O. Ancuti, C. Ancuti, F.-A. Vasluianu, and R. Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 627–646, 2021.

[4] D. Berman, T. Treibitz, and S. Avidan. Non-local image dehazing. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1674–1682, 2016.

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

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

[7] R. Fattal. Dehazing using color-lines. *ACM Transactions on Graphics*, 34(1):1–14, 2014.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2014.

[9] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.

[10] J. Johnson, A. Alahi, and F. Li. Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision*, pages 694–711, 03 2016.

[11] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations*, 2015.

[12] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. AOD-Net: All-in-one dehazing network. *IEEE International Conference on Computer Vision*, pages 4770–4778, 2017.

[13] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Z. Zhou, and L. F. Cheong. Simultaneous video defogging and stereo reconstruction. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4988–4997, 2015.

[14] E. J. Mccartney. Scattering phenomena (book reviews: Optics of the atmosphere. scattering by molecules and particles). *Science*, 196:1084–1085, 1976.

[15] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. H. Yang. Single image dehazing via multiscale convolutional neural networks. *European Conference on Computer Vision*, pages 154–169, 2016.

[16] O. Ronneberger, P. Ronneberger, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[17] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang. Domain adaptation for image dehazing. *IEEE International Conference on Computer Vision*, pages 2808–2817, 2020.

[18] A. Singh, A. Bhave, and D. Prasad. Single image dehazing for a variety of haze scenarios using back projected pyramid network. *European Conference on Computer Vision*, pages 166–181, 01 2020.