Single image dehazing algorithm using generative adversarial network based on feature pyramid network

Xuejie Cao
School of Internet of Things, Nanjing University of Posts and Telecommunications
No. 9 Wenyuan Road, Qixia District
Nanjing City, Jiangsu Province, China
B17070729@njupt.edu.cn

Jiangwei Dong
School of Internet of Things, Nanjing University of Posts and Telecommunications
No. 9 Wenyuan Road, Qixia District
Nanjing City, Jiangsu Province, China
B17070730@njupt.edu.cn

Shiqi Zhou
School of Internet of Things, Nanjing University of Posts and Telecommunications,
No. 9 Wenyuan Road, Qixia District
Nanjing City, Jiangsu Province, China
B17070721@njupt.edu.cn

Shasha Zhao
School of Internet of Things, Nanjing University of Posts and Telecommunications
No. 9 Wenyuan Road, Qixia District
Nanjing City, Jiangsu Province, China
zhaoss@njupt.edu.cn

Dengyin Zhang*
School of Internet of Things, Nanjing University of Posts and Telecommunications
No. 9 Wenyuan Road, Qixia District
Nanjing City, Jiangsu Province, China
+86 18951896076
zhangdy@njupt.edu.cn

***ABSTRACT*-This paper proposes a single image dehazing algorithm using generative adversarial network (GAN) based on feature pyramid network (FPN). This method is an end-to-end image dehazing method, avoiding the physical model dependence. The generator uses MobileNet-V2 as the backbone network, and uses the FPN structure to improve the feature utilization rate of the image, combined with the discriminator formed by the convolutional neural network to form GAN that can improve the training stability and convergence of the generator. The model uses a lightweight MobileNet-V2 network, and the FPN structure also enables multiple-scale feature maps to be obtained while avoiding the use of direct scaling, thus reducing the computational power and memory requirements and allowing the model to operate with limited computational resources. We used the RESIDE training set to train our proposed model and conducted extensive experiments on the test set. The experimental results show that the algorithm has satisfactory results in terms of quality and speed.**

Keywords-Dehaze; Generative Adversarial Network; Feature Pyramid Network; Convolutional Neural Network; MobileNet-V2;

I. INTRODUCTION

In hazy weather, there are many suspended particles in the air, which absorb and scatter light, resulting in color distortion and reduced contrast in the pictures taken, reducing the value of the pictures for use in computer vision applications such as target identification and security monitoring. Therefore, the study of image dehazing is of great practical importance. At present, the image dehazing algorithm can be divided into three types: The first type is the dehazing algorithm based on image enhancement, It enhances the contrast, saturation, and sharpness features of an image by means of image enhancement. However, it does not fundamentally eliminate the cause of fog, because image degradation is not considered. The second type is the dehazing algorithm based on physical model. Based on physical models such as the atmospheric scattering model, uses various methods to estimate the parameters in the model to solve the original image, but this method requires manual summary of the a priori knowledge of the image and lacks generalizability for complex scenes. The third type is the dehazing algorithm based on deep learning. It learns the features of haze through the feature extraction capability of neural networks to achieve the dehazing effect, it has achieved good results, but there is also the problem that the processing results of real hazy images are not stable enough.

McCartney first proposed the classical atmospheric scattering model [1], and then Narasimhan studied and derived it further and systematically described the degradation mechanism of hazy images [2]. Tan proposed a dehazing algorithm based on the contrast prior [3]. It has

978-1-7281-9481-3/20/$31.00 ©2020 IEEE

DOI 10.1109/CVIDL51233.2020.00017

satisfactory results in details. However, the recovered image is prone to oversaturation and localized halo. He proposed a dehazing algorithm based on dark channel prior (DCP), which has obvious dehazing effect and simple implementation, but has the problems of slow processing, overexposure of sky, and reduced brightness [4]. Zhu proposed a dehazing method based on Color Attenuation Prior (CAP) to estimate the haze concentration by the difference in image saturation and brightness on fog days [5]. Cai applied the convolutional neural network (CNN) to the image dehazing task for the first time, and obtained better dehazing results [6]. Ren proposed a multi-level transmission map estimation network (MSCNN) to estimate image transmission maps from multiple scales and then combine it with an atmospheric scattering model to restore the images [7]. Li proposed an end-to-end dehazing network model, called all-in-one dehazing network (AOD-Net), which unified the transmittance and atmospheric light value in the atmospheric scattering model into one variable, and achieved a better dehazing effect [8]. The current image dehazing algorithm have a certain degree of deficiencies. Image enhancement-based algorithm are prone to lose image detail information. Physical model-based algorithm require manual design to extract the feature, not universal. Existing deep learning -based dehazing methods have achieved good dehazing results, but most of them only estimate the atmospheric light value and transmittance in the atmospheric scattering model through CNN, which is essentially a physical model-based method. In addition, those methods usually calculate the mean square loss only, and the loss tends to be the average of pixel points, easily causing the phenomenon of image smoothing, losing image detail and reducing image clarity.

This paper proposes a single image dehazing algorithm using generative adversarial network (GAN) [9] based on feature pyramid network (FPN) [10]. The network model generator fuses the feature map of multiple scales extracted by the backbone network MobileNet-V2 [11] through the structure of FPN, and then reconstructs and outputs the clear image through concatenate, convolution, and deconvolution, avoiding the image scaling during the feature fusion process and improving the model efficiency. At the same time, we use a discriminator constructed from convolutional neural networks to solve the problem of generator instability and slow convergence. Experiments proved that this method has a certain dehazing effect, and has some advantages in terms of quality and efficiency compared to several other dehazing algorithms.

II. RELATED WORK

Generating adversarial network model is a framework for generating models through adversarial process estimation, which has a simple structural design, does not require a pre-designed complex function model, and can train functions by backpropagation, but it is difficult to guarantee stability and convergence due to its high degree of freedom. Then, Conditional Generative Adversarial Nets (CGAN) [12], Deep Convolutional Generative Adversarial Networks (DCGAN) [13], Wasserstein GAN (WGAN) [14], PatchGAN [15] and other models have been gradually developed in response to these deficiencies. These models are widely used with image super-resolution, image-to-image translation, image restoration, and achieved good results

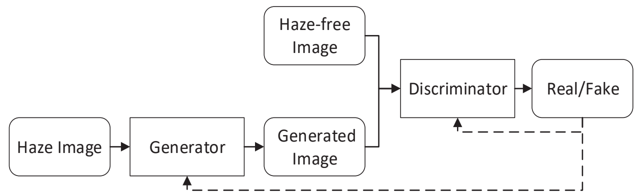

Figure 1: The architecture of dehazing algorithm using GAN

The process of the image dehazing task using this framework is shown in Figure 1, which contains generator and discriminator models. Generator maps the hazy images to the haze-free images and minimizes its error with the objective function to deceive the discriminator. And the input of discriminator includes the clear image and the image generated by generator, and try to discriminate them. These two networks play against each other and finally reach Nash equilibrium.

However, since the build of the generator often extracts the features of the hazy image through the multi-layer encoder, then uses the feature converter to realize the conversion between the hazy image and the clear image from the feature level, and finally uses the decoder to reconstruct the haze-free image, there is often a problem of underutilization of feature information. Inspired from [10, 11, 16] in this paper, we proposes a single image dehazing algorithm using generative adversarial network based on feature pyramid network, which improves the feature utilization of the generator, and the quality and efficiency of image dehazing.

III. THE PROPOSE METHOD

A. Generator Network

The purpose of the generator network is to complete the mapping of hazy images to clear images, and the generator in this paper uses the FPN structure to fuse multiple scales of feature images generated by the MobileNet-V2 model. There are multiple bottom-up feature maps within MobileNet-V2, and the FPN can enhance the expression of those feature maps by fusing these different scales of feature maps and using information from both low-level and high-level features, as shown in Figure 2.

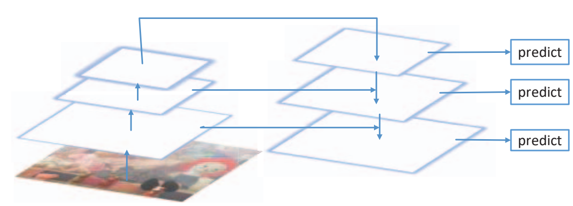

Figure 2: The architecture of FPN

FPN takes advantage of the bottom-up, multi-scale feature representation within the regular CNN model to generate multi-scale feature representation of images in a single image view. It generates more expressive feature maps by corresponding to the different feature maps of the element-sum-fusion feature extraction network. Compared to image pyramids, FPN avoids direct image scaling operations, greatly reducing the computational power and memory requirements of the model.

The generator network model combines the lightweight MobileNet-V2 network model and the efficient feature extraction capability of FPN to achieve the image dehazing task in a low hardware environment, and its structure is shown in Figure 3.

The model can be divided into three parts. The first part is the backbone network consisting of MobileNet-V2, which is mainly responsible for the extraction of fog features. It extracts and outputs four feature maps of the network at four different scales: [112×112, 56×56, 28×28, 14×14]. In addition, these feature maps are also passed through a 1×1 convolutional layer before output to reduce the amount of model calculations by reducing the dimensionality. The second part is multi-scale feature fusion, which is achieved by convolution, activation and summation of the feature map. In this paper, we use deconvolution instead of upsampling, and we find that this operation can significantly improve the problem of image reimaging. The third part is the reconstruction of the image, which is achieved by restoring the image through convolution, activation and upsampling. Considering that the FPN is prone to losing some of the underlying detail feature information, an addition layer with the input image was added before the output to enhance the underlying detail.

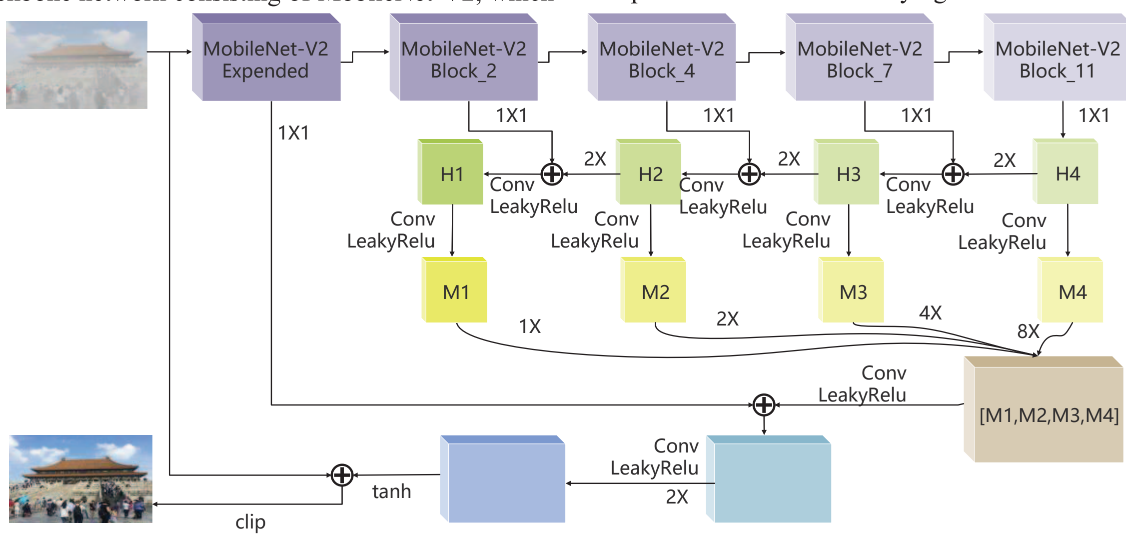

Figure 3: The architecture of generator network

B. Discriminator Network

The function of the discriminator network is to distinguish as much as possible whether the input image is a real image or not, and the goal is to make the generated image through the discriminator network output of the discriminator results close to 0, the real image through the discriminator network output of the discriminator results close to 1. The network discriminator in this paper consists of seven convolutional layers. The network structure is shown in Figure 4. In this model, every layers except the input and output layers is extracted in a Conv-BatchNorm-Relu way to extract the feature information of the image, with the addition of a fully connected layer at the end of the network and a Sigmoid function to normalize the discriminant results to between [0, 1]. The discriminator implements the ability to judge the authenticity of the input image, and enhances the stability and convergence of the generator.

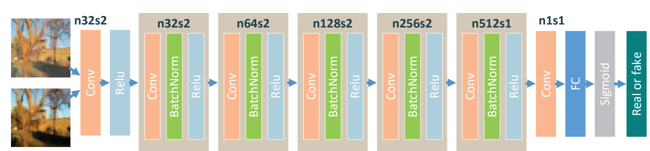

Figure 4: The architecture of discriminator network

C. Loss Function

The loss function used for the training network model is the weighted sum of mean square error loss (MSE) and adversarial loss. The loss of mean square error based on pixels can restore the low-frequency information of the image, while the adversarial loss can restore the high-frequency information of the image. The loss function is defined as:

$$L = L_{MSE} + \lambda L_{GAN} \quad (1)$$

Where L is the overall loss of the generative network, L_{MSE} is the mean square error loss, and L_{GAN} is the adversarial loss. λ is the weighting factor, which is set to 0.01 in this paper.

L_{MSE} is used to evaluate the gap between the generated image and the clear image. The image generated by the network model that uses it as the objective function tends to be smooth. Although a part of the high-frequency information of the image will be lost, and it is not conducive to capturing information on image perception, the low-frequency information such as the color of the image can be well restored.

$$L_{MSE} = \frac{1}{\mathrm{C} \times \mathrm{W} \times \mathrm{H}} \sum_{i=1}^{\mathrm{C}} \left\| \mathrm{I}_i^* - \mathrm{I}_i^{\mathrm{G}} \right\|^2 \quad (2)$$

Where $\mathrm{I}_i^{\mathrm{G}}$ and I_i^* represent the label image and the generated image of the generator respectively, C is the image channel, and $\mathrm{W} \times \mathrm{H}$ is the image size.

Because the original GAN is easy to encounter problems, such as vanishing gradient and mode collapse. WGAN reduces the difficulty of network training by using the "Wassertein-1" distance and improves the stability of training. Therefore, we use it as the loss function for improving the stability of network training. L_{GAN} represents the adversarial loss, and its calculation formula is:

$$L_{GAN} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left[D\left(\mathrm{I}_i^{\mathrm{G}}\right) \right] - \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left[D\left(\mathrm{I}_i^*\right) \right] \quad (3)$$

Where $D\left(\mathrm{I}^{\mathrm{G}}\right)$ and $D\left(\mathrm{I}^*\right)$ represent the discriminator's discrimination results for the label image and the generated image of the generator respectively, and N represents the number of images.

IV. EXPERIMENTS AND ANALYSIS

A. Training Data and Experimental Settings

Under normal circumstances, it is impossible to obtain the hazy image and its corresponding clear image at the same time, so the dataset in this paper is composed of clear images and their corresponding synthetic hazy images. The data set is outdoor training set (OTS), which contains 8970 pairs of clear images and the corresponding synthetic hazy images of different concentrations. In this paper, 3000 clear images and their corresponding synthetic hazy images were selected from the OTS dataset to form the training set, and 300 images were extracted from the remaining images for the testing set. In addition, adjust the size of the image to 224*224, so that the picture can adapt to the input of the network model.

The training of the network uses alternating iteration optimization. First fix the generator network parameters, input the hazy image into the generator network for generator optimization. Then fix the discriminator network parameters, input the generated image and the haze-free image into the discriminator network simultaneously for discriminator optimization. The discriminator and generator are iteratively optimized until the discriminator network is unable to discriminate the authenticity of the generated image and the loss of the generator is stable. The optimizer of the generator and discriminator is Adam, the learning rate is set to 0.0005, the update ratio of the discriminator and generator is set to 5, the batch size is adjusted to 2, and 10 epochs were run on the Tensorflow 2.0 platform of the computer with GeForce GTX 1080 GPU.

B. Objective Evaluation

In order to objectively evaluate the algorithm, this paper adopts peak signal to noise ratio (PSNR) and structural similarity (SSIM) as the evaluation indicators of the dehazing effect. We compare the proposed algorithm with several state-of-the-art dehazing algorithm [4-7, 17]. We randomly selected 100 images from synthetic objective testing set (SOTS) as a sample, and used these algorithms for dehazing operations, and calculate the SSIM and PSNR from the corresponding original clear image.

Table 1. Quantitative results on SOTS in terms of PSNR and SSIM

Metrics	BCCR[17]	DCP[4]	CAP[5]	DehazeNet[6]	MSCNN[7]	Ours
PSNR	15.9414	17.906	18.4533	19.6963	22.0332	**23.2252**
SSIM	0.7557	0.7909	0.7584	0.8281	0.8164	**0.8326**

The results are shown in Table 1. Compared with other algorithms, the SSIM index value of this algorithm is the highest, which indicates that the structure of the image before and after dehazing is more similar, and the recovery effect is better at the image edge details. In addition, the PSNR index value of ours algorithm also obtained the highest value, indicating that the image generated by ours algorithm is less different from the clear image in pixel level. Our proposed algorithm eliminates more noise effects, leading to a significant improvement in image quality.

C. Subjective Evaluation

In order to more visually demonstrate the defogging effect of this algorithm, we selected four images of different scenes for dehazing, and the effect of those algorithms is shown in Figure 5.

It can be seen from the figure that the traditional methods (CAP, DCP, BCCR) show color distortion, decrease in brightness, and scene recovery is not natural enough. DCP and other algorithms fail in the white area such as the sky due to the dark channel a priori, resulting in severe color distortion in the sky area of the second image. The learning-based approaches (DehazeNet, MSCNN) have incomplete dehazing and continue to be flawed in their detail handling. This algorithm is more natural for the more distant scenery processing effect, the object surface details of the restoration is more realistic. In addition, in the restoration of image color, especially the restoration of sky color has certain advantages. The proposed algorithm can produce a clearer image while protecting the image color and contrast information. Experimental results show that traditional dehazing algorithms tend to exhibit color distortion. Compared with the deep learning method

(DehazeNet, MSCNN), the image detail information after dehazing is preserved intact, color recovery is more natural, the degree of dehazing is moderate and the defogging effect is better.

In summary, the proposed algorithm is trained on synthetic hazy image dataset, but still has good results on real hazy images, and the dehazing effect is better than other comparison algorithms, verifying that this algorithm has good generalization ability and strong applicability.

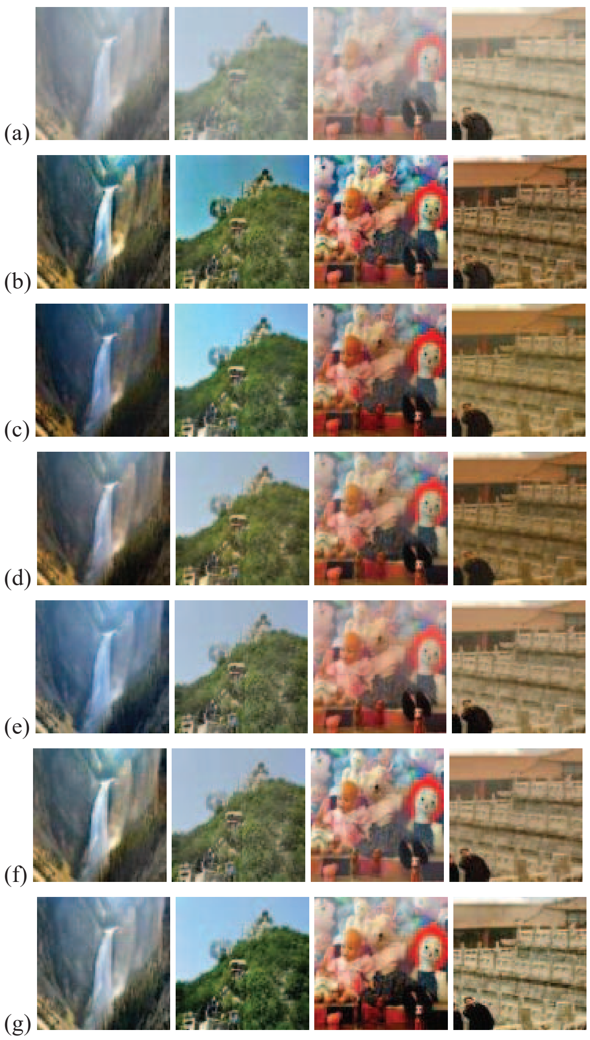

Figure 5. Dehazing results evaluated on the real-world images. (a) Sample hazy images. (b) BCCR [17]. (c) DCP [4]. (d) CAP [5]. (e) DehazeNet [6]. (f) MSCNN [7]. (g) Proposed method.

D. Run Time

To verify the efficiency of this paper's algorithm, we compared it with several state-of-the-art dehazing algorithm [4-7, 17] on dehazing efficiency. We randomly selected 300 images from SOTS, ran all the dehazing algorithms using the same machine (Intel(R) Core(TM) i5 8400, 16GB) with GPU off, counted the time consumed by each algorithm and calculated the average time consumed by dehazing each image, the experimental results are shown in Table 2. The results show that compared to other dehazing algorithms, the algorithm we proposed has a significant advantage in the efficiency of dehazing.

Table 2. Average time taken to process single image

Method	**BCCR[17]**	**DCP[4]**	**CAP[5]**
Time/s	1.4731	0.7243	0.6023
Method	**DehazeNet[6]**	**MSCNN[7]**	**Ours**
Time/s	0.3832	1.4696	**0.2225**

V. CONCLUSIONS

This paper proposes a single image dehazing algorithm using generative adversarial network based on feature pyramid network. The generator of the network uses MobileNet-V2 as the backbone, while incorporating the structure of FPN to improve the efficiency and quality of the fusion of multi-scale feature maps and to form a GAN with a discriminator composed of convolutional neural network to improve its convergence and stability. The model is based on synthetic hazy images, which also has a certain effect on the real hazy image, so it can provide a reference for hazy images to perform computer graphics tasks. At the same time, due to the lightweight of the network model, the algorithm has low computing power and memory requirements for the computing platform, it can be widely applied to mobile terminal, edge computing, autopilot and other scenarios in the future. The proposed algorithm has been experimentally proven to have significant advantages in terms of both quality and efficiency. In the future, we will further optimize the generator structure of the model and explore its application in more complex dehazing scenarios to improve the generalizability and visual effect of the model.

ACKNOWLEDGMENTS

This work was partially supported by the National Natural Science Foundation of China (No. 61872423), the Scientific Research Foundation of the Higher Education Institutions of Jiangsu Province (No. 19KJA180006), and the Science and Technology Innovation Training Program (No. SZDG2019029).

REFERENCES

[1] McCartney, E. J. J. N. Y., John Wiley and Sons, I., . 421 p. Optics of the atmosphere: scattering by molecules and particles (1976).

[2] Narasimhan, S. G., Nayar, S. K. J. I. t. o. p. a. and intelligence, m. Contrast restoration of weather degraded images, 25, 6 (2003), 713-724.

[3] Tan, R. T. *Visibility in bad weather from a single image*. IEEE, City, 2008.

[4] He, K., Sun, J., Tang, X. J. I. t. o. p. a. and intelligence, m. Single image haze removal using dark channel prior, 33, 12 (2010), 2341-2353.

[5] Zhu, Q., Mai, J. and Shao, L. J. I. t. o. i. p. A fast single image haze removal algorithm using color attenuation prior, 24, 11 (2015), 3522-3533.

[6] Cai, B., Xu, X., Jia, K., Qing, C. and Tao, D. J. I. T. o. I. P. Dehazenet: An end-to-end system for single image haze removal, 25, 11 (2016), 5187-5198.

[7] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X. and Yang, M.-H. *Single image dehazing via multi-scale convolutional neural networks*. Springer, City, 2016.

[8] Li, B., Peng, X., Wang, Z., Xu, J. and Feng, D. *Aod-net: All-in-one dehazing network*. City, 2017.

[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. *Generative adversarial nets*. City, 2014.

[10] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. *Feature pyramid networks for object detection*. City, 2017.

[11] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C. *Mobilenetv2: Inverted residuals and linear bottlenecks*. City, 2018.

[12] Mirza, M. and Osindero, S. J. a. p. a. Conditional generative adversarial nets (2014).

[13] Radford, A., Metz, L. and Chintala, S. J. a. p. a. Unsupervised representation learning with deep convolutional generative adversarial networks (2015).

[14] Arjovsky, M., Chintala, S. and Bottou, L. J. a. p. a. Wasserstein gan (2017).

[15] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A. *Image-to-image translation with conditional adversarial networks*. City, 2017.

[16] Kupyn, O., Martyniuk, T., Wu, J. and Wang, Z. *Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better*. City, 2019.

[17] Meng, G., Wang, Y., Duan, J., Xiang, S. and Pan, C. *Efficient image dehazing with boundary constraint and contextual regularization*. City, 2013.