

# Image Defogging Algorithm Based on Deblurgan Network

Yuanyuan Cheng\*, Xiaorong Cheng

North China Electric Power University (Baoding), Hebei, 071003, China

\* Corresponding author: Yuanyuan Cheng (Email: fighting\_cheng@163.com)

**Abstract:** The reduced visibility in foggy days impairs the quality of captured images and videos to varying degrees, leading to limited applications of these images in the field of computer vision. To solve this problem, direct recovery of fog-free images based on an improved Deblurgan network is proposed. We add the expanded convolution (Dilated Conv) module in the generator to expand the perceptual field to extract richer semantic information, and add the spatial attention mechanism module at specified locations to facilitate the elimination of residual fog; the discriminator uses the traditional PatchGAN for chunk determination to improve the discriminative accuracy; the loss function adds BCE loss to improve the discriminative accuracy of the discriminator and the pixel-level detail retention of the generator. In the synthetic fogged dataset RESIDE, the subjective visual comparison with dark channel, DehazeNet, AOD-Net, and defogging effect shows that the defogging effect of this network model and the detail information and color contrast of defogged images are improved; meanwhile, the objective evaluation indexes Peak Signal to Noise Ratio, PSNR and Structure Similarity (SSIM) were also improved respectively.

**Keywords:** Single image defogging; Deblurgan; dilation convolution; Spatial attention mechanism; BCE loss.

## 1. Introduction

The role of image defogging algorithms in computer vision applications such as target detection and traffic control is becoming more and more important. In foggy conditions, the quality of images and videos is drastically degraded by the scattering of suspended particles in the air and the ambient light caused by suspended particles, which causes great difficulties in fields such as traffic detection. Therefore, image defogging algorithms are of practical importance and have received a lot of attention from researchers in recent years.

With the proliferation of image defogging algorithms and the rapid development of deep learning, defogging algorithms are mainly divided into two categories: traditional defogging methods, and neural network-based defogging methods. Traditional defogging methods are divided into image enhancement-based defogging algorithms and physical model-based defogging algorithms. Image enhancement-based image defogging algorithms use contrast enhancement to highlight the information of foggy images, such as the sub-block histogram equalization algorithm and the color recovery multi-scale Retinex method to enhance the contrast of local information proposed by KIM [1] and Jobson [2], respectively. However, these types of algorithms do not consider the reason of foggy image degradation, which will result in reduced or over-enhanced detail information retention. The image defogging algorithms based on physical models estimate the transmittance values and atmospheric illumination values of foggy images by a priori information, and then recover clear images based on atmospheric scattering models, such as the dark channel a priori defogging algorithm proposed by He et al [3]. However, there is often a large computational effort in estimating the transmittance value and atmospheric illumination value, and the sky area effect is not satisfactory.

The neural network-based defogging algorithm is divided into two-stage defogging algorithm and single-stage defogging algorithm. The basic principle of the two-stage

defogging algorithm is still to use the atmospheric scattering model backward derivation to recover clear images, but the difference from the traditional defogging algorithm is to use the neural network to learn to estimate the transmittance and atmospheric light values. However, due to the cumulative effect of the two stages, the overall deviation will be enlarged and the clear image recovered by backward derivation is prone to color distortion. Cai et al [4] proposed a DehazeNet defogging network to estimate the transmittance in the atmospheric model, pioneering the use of neural networks for image defogging; Zhang [5] et al proposed a CNN-based DCPDN defogging network, using densely connected pyramidal modules to estimate the transmittance and sub-networks to estimate the atmospheric light values. A single-stage defogging algorithm, which uses a neural network to directly restore fogged images to clear images, i.e., end-to-end defogging, without a priori estimation of atmospheric light values and transmittance, achieves detachment from the atmospheric scattering model. Li et al [6] proposed an AOD-Net defogging network, which directly achieves end-to-end defogging by a lightweight CNN and can be connected with a faster R-CNN to improve performance. Mei et al [7] proposed a deep network based on feature fusion with a U-Net-like encoder-decoder to recover fog-free images by directly learning the mapping relationship between fogged and clear images. With the introduction of GAN networks into the image domain by Goodfellow et al [8], a series of algorithms for defogging using GAN networks have emerged [9-11].

The Deblurgan-based defogging network proposed in this paper adds the Dilated Convolution (DCC) module in the generator to expand the perceptual field to extract richer semantic information and adds the Spatial Attention Mechanism module at the specified location, which is beneficial to eliminate fog; the discriminator uses the traditional  $34 \times 34$  PatchGAN for chunking determination to improve the resolution of the image. The loss function adds BCE loss on top of the adversarial loss and content loss

function to better solve the problem of blurred generated images.

## 2. Network model in this paper

The DeblurGAN network proposed by Orest Kupyn et al [12] has a simple structure and fast processing speed. However, the traditional DeblurGAN network has poor color recovery and fog residue when performing defogging. The improved DeblurGAN network proposed in this paper solves the above problems better and achieves direct mapping from foggy images to fog-free images. In this paper, the network consists of generator G and discriminator D. The generator G generates sample data that approximates the real data by learning the input data, and the input has fog images and directly outputs fog-free images; the role of discriminator D is to distinguish the real data from the samples generated by the generator G. The input is a combination of clear image - fog image or defog image - fog image. Inspired by the idea of game theory, generator G and discriminator D are trained adversarially and progress together until discriminator D cannot distinguish the samples generated by generator G from the real data samples, at which time the discriminator D output is stabilized at about 0.5, indicating that the model reaches the optimal state - Nash equilibrium.

### 2.1. Generators

In order to obtain deeper features as well as multi-scale information and enhance the ability of the generator to extract

features, the expanded convolution module [13] and the spatial attention mechanism module [14] are added to the generator. The network structure of the generator is shown in Fig. (1) and contains two  $7 \times 7$  normal convolution blocks, two stepwise convolution blocks with 1/2 step, nine ResBlock residual blocks, four Dilated Conv expansion convolution blocks, a spatial attention mechanism module and two transposed convolution blocks. Each ResBlock consists of a  $3 \times 3$  layer of convolution, an InstanceNorm (instance normalization layer), and a RELU activation function, and a loss regularization with probability 0.5 is added after the convolution to combat overfitting. The ResNet network linearly superimposes the input and output to achieve residual connectivity and solve the problem of gradient dispersion.

In this paper, we add four successive dilation convolutions to the generator, and the dilation factors of each two successive dilation convolutions are not multiplicative between them, and the dilation factors are 2, 3, 4, and 2 respectively, i.e., zeroes are introduced between the convolution kernel parameters according to the given dilation factors to expand the perceptual field of the convolution process and enable the network to extract multi-scale information and learn more features without reducing the resolution. However, the null convolution is computed in a tessellation-like format, resulting in no correlation between the pixels obtained from the convolution calculation and loss of local information. The problem of grid artifacts occurs when superimposing two convolutions with a dilation factor of 2, as shown in the figure 2 and 3.

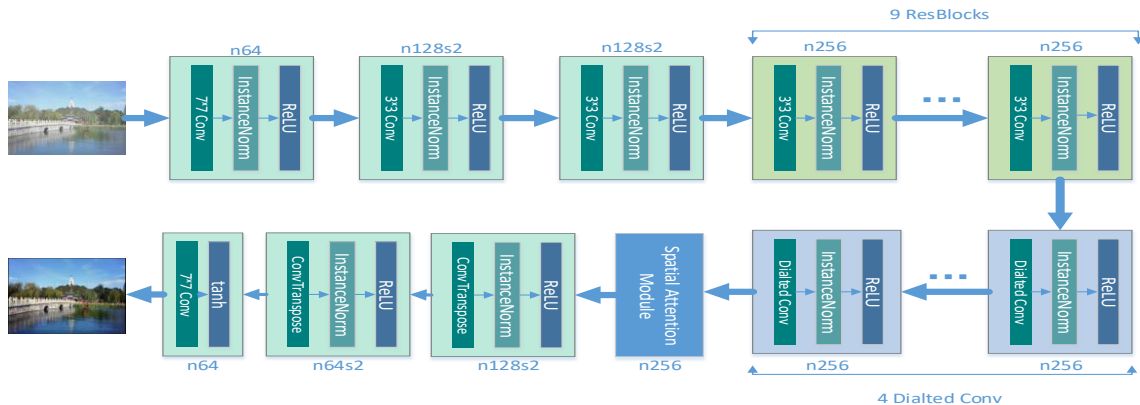


Figure 1. Generator network structure

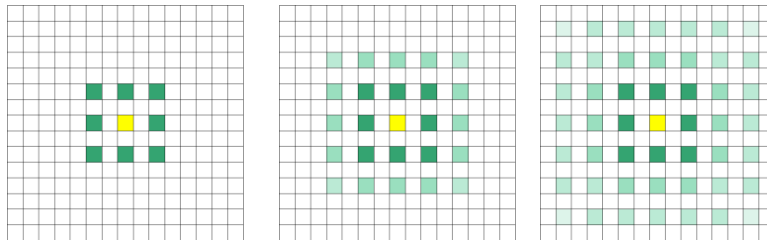


Figure 2. The expansion factor is multiplied

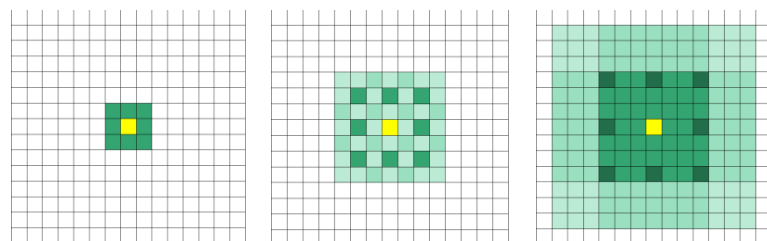


Figure 3. Dilation factor is not multiplied

Therefore, the above problem is solved by setting the expansion convolution to a continuous expansion factor that is not multiplicative, as shown in the figure 4.

In this paper, a spatial attention mechanism is added to the generator, in which two feature maps representing different information are generated based on channel-based global average pooling and global maximum pooling operations, and then combined and downscaled to 1 channel number by a convolution operation, and finally a weight map is generated by a Sigmoid operation, i.e., a weight mask is generated for each location and weighted to output. Then superimposed back to the original input feature map, thus making the target region enhanced, the module is shown in Figure (4) below. For the generator's task of removing fog from the image, the spatial attention mechanism enables the generator to spatially focus its attention on the foggy locations and enhance the recognition of fog with different concentrations in the image, thus achieving better defogging results.

## 2.2. Discriminator D

The discriminator is a fully convolutional network with the network structure shown in Fig. 5, which uses a PatchGAN with block determination, consisting of six convolutional blocks containing six  $4 \times 4$  convolutions, four instance normalization (BatchNorm) layers, and the Leaky ReLU activation function. The output of the discriminator is a matrix, and the value of each pixel point in the matrix corresponds to a  $34 \times 34$  image block (patches) of the input image, which is used to determine the true/false of each patch, and finally the mean value of the output matrix is taken as the output of True/False. The difference judgment by each patch changes the traditional discriminator's "one-word" discriminative result, realizes the extraction and characterization of local image features, and helps to maintain the high resolution and detail of the image.

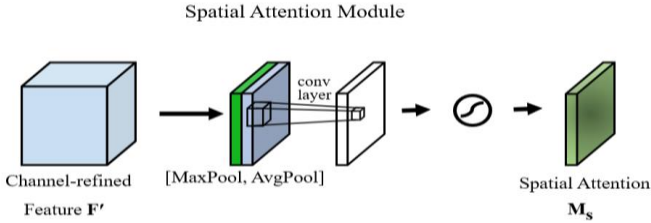


Figure 4. Network structure of spatial attention mechanism

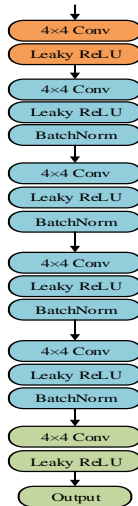


Figure 5. Discriminator network structure

## 3. Loss function

### 3.1. Adversarial loss (Adversarial loss)

In order to solve the problem of slow convergence and possible training instability during training, the adversarial loss function uses WGAN-GP with gradient penalty [15], which is calculated as follows.

$$L_{adv} = E_{x \sim P_r}[D(x)] - E_{x \sim P_g}[D(x)] + \lambda E_{x \sim \chi} [\|\nabla_x D(x)\|_p - 1] \quad (1)$$

where  $P_r$  denotes the real fog-free image data,  $P_g$  denotes the generated fog-free image data,  $\chi = \alpha x_r + (1-\alpha)x_g$ , where  $\alpha \in (0,1)$  denotes the randomly selected data samples in  $P_r$  and  $P_g$ ,  $D$  denotes the probability that the discriminator discriminates the input image as the real fog-free image, and  $E$  denotes the average statistics of the discriminator on the discriminated results.

### 3.2. Content loss

In this paper, we adopt the content loss based on L1 loss, which can identify low-frequency information more effectively, reduce the artifact phenomenon when the generator generates images, and make the generated image content more realistic, and its calculation formula is as follows.

$$L_s = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\Phi_{i,j}(I^s)_{x,y} - \Phi_{i,j}(G_{\theta_G}(I^B))_{x,y})^2 \quad (2)$$

where  $P_r$  denotes the real fog-free image data,  $P_g$  denotes the generated fog-free image data,  $\chi = \alpha x_r + (1-\alpha)x_g$ , where  $\alpha \in (0,1)$  denotes the randomly selected data samples in  $P_r$  and  $P_g$ ,  $D$  denotes the probability that the discriminator discriminates the input image as the real fog-free image, and  $E$  denotes the average statistics of the discriminator on the discriminated results.

### 3.3. Content loss

BCE loss [16] is a loss function based on pixel points, which is predicted for each pixel and is calculated as follows.

$$L_{BCE} = -1/n \sum_i (t(i) * \log(o(i)) + (1-t(i)) * \log(1-o(i))) \quad (3)$$

Where,  $t$  denotes the target value,  $o$  denotes the predicted value, and  $i$  is the pixel point. For the task of defogging in this thesis,  $t(i)$  denotes the pixel point value of the clear and fog-free image, and  $o(i)$  denotes the pixel point value of the generated image of the corresponding generator, which can distinguish the real clear and fog-free image from the generated image more precisely based on pixel points, making the generated image of the generator more detailed and clearer, and the discriminator more precise based on pixel discrimination.

In this paper, the content loss and BCE loss are used in the generator and the BCE loss is used in the discriminator, and the total loss function expression is as follows.

$$L_{total} = L_{adv} + L_{BCE} + \lambda L_s \quad (4)$$

## 4. Experiments and Analysis of Results

### 4.1. Dataset

The dataset used in this paper is from the outdoor dataset OTS (outdoor training set) added to the RESIDE dataset. This dataset uses 2061 real outdoor maps of Beijing real-time weather, and uses the algorithm proposed by Liu et al [17] to reduce the error of depth and the visual artifacts that may be generated. The parameter settings of beta are [0.04, 0.06, 0.08, 0.1, 0.12, 0.16, 0.2] and the parameter settings of atmospheric light value A are [0.8, 0.85, 0.9, 0.95, 1], so that 35 fog maps of different degrees are generated for each clear image, and a

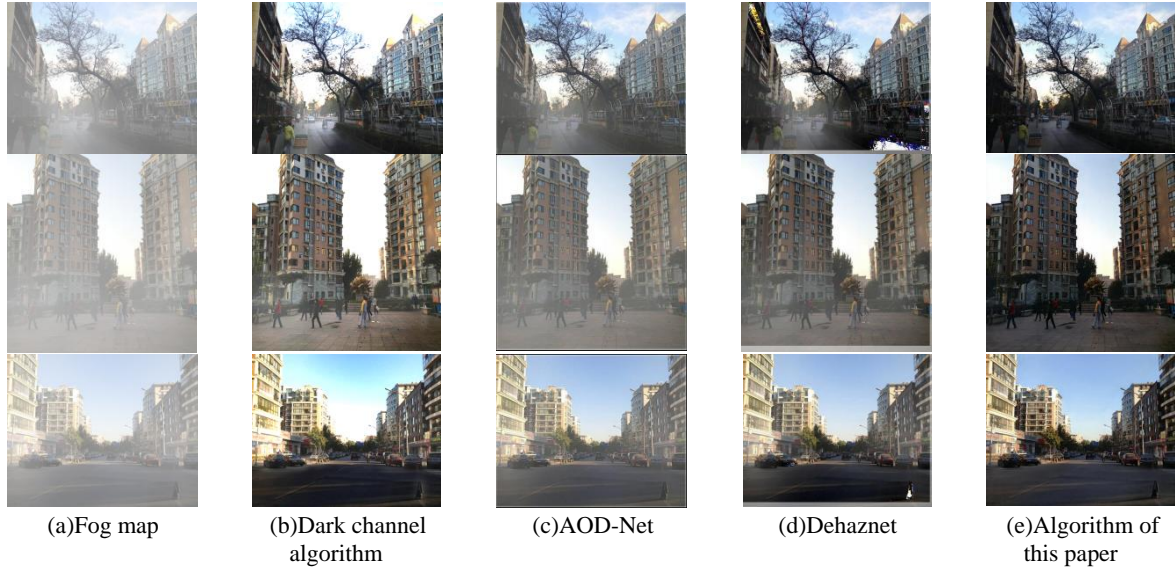


Figure 6. Fog removal effect of different algorithms

Through the subjective comparison figure Fig. 6 can be seen, the dark channel defogging algorithm for the bright region of the transmittance and atmospheric light value estimation inaccurate make the sky region overexposure phenomenon, and the defogging image color darkened, there are fog residue; AOD-Net algorithm can obviously see the defogging effect is not ideal, the fog residue most; Dehaznet algorithm compared to the AOD-Net algorithm fog residue less, but will The fog residue of Dehaznet algorithm is less than that of AOD-Net algorithm, but there are small areas of distortion, such as the lower right part of the defogged image in the first row of Fig. 6.

In this paper, two image quality evaluation indexes, PSNR and SSIM, are used to demonstrate the superiority of this method from the objective aspect. PSNR peak signal-to-noise ratio, which is the most common and widely used image objective evaluation index, is based on the error between corresponding pixel points in dB, and the larger value means the smaller distortion. SSIM structural similarity, which is also a full-reference image quality evaluation index, measures image similarity in terms of brightness, contrast, and structure, respectively, and SSIM takes values in the range [0,1], with larger values indicating higher similarity to the original image.

The following comparison table 1 shows that the PSNR and SSIM values of this paper are not only the highest but also have been improved.

total of 72135 maps are generated. Before training, the dataset is preprocessed to get 10,000 pairs of datasets with consistent names, and 1000 sheets are tested. This experiment is trained under Ubuntu system with GPU of Nvidia 1080Ti. optimizer is Adam, learning rate is 0.0001, batchsize is 2, epoch is 100.

### 4.2. Experimental results

In this experiment, 1000 images are selected and compared with three algorithms with better defogging effect, which are dark channel a priori algorithm, AOD-Net algorithm and Dehaznet algorithm, for defogging effect, as shown in the figure 6.

Table 1. Comparison of PSNR and SSIM of different algorithms

De-fogging algorithm	SSIM/%	PSNR/dB
dark channel	0.8833	15.5543
AOD-Net	0.8894	18.7977
DehazeNet	0.8831	18.9424
My Method	<b>0.9505</b>	<b>26.8199</b>

## 5. Conclusion

The image defogging algorithm based on DeblurGAN network proposed in this paper adds dilated convolution to the generator to expand the perceptual field without changing the resolution, and also adds a spatial attention mechanism at specified locations to optimize the detail information of the generated fog-free images and reduce the fog residue. The loss function replaces the traditional adversarial loss with a pixel-based BCE loss to improve detail retention. By comparing the subjective defogging effect graph and the objective experimental data, we can see that the defogging effect of this algorithm is obvious.

## References

- [1] KIM T K, PAIK J K, KANG B S. Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering [J]. IEEE Transactions on Consumer Electronics, 1998, 44(1): 82-87.

- [2] JOBSON D J, RAHMAN Z, WOODDELL G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes[J]. IEEE Transactions on Image Processing, 2002, 6(7): 965-976.
- [3] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2341-2353.
- [4] CAI B L, XU X M, JIA K, et al. Dehaze Net: an end-to-end system for single image haze removal[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5187-5198.
- [5] He Zhang, Vishal M. Patel. Densely Connected Pyramid Dehazing Network [J]. arXiv: 1803.0839 [cs. CV] <https://doi.org/10.48550/arXiv.1803.08396>
- [6] LI B Y, PENG X L, WANG Z Y, et al. AOD-net: All-in-one dehazing network [C]. IEEE International Conference on Computer Vision, 2017: 4770-4778.
- [7] MEI K F, JIANG A W, LI J C, et al. Progressive feature fusion network for realistic image dehazing[C]. Asian Conference on Computer Vision, 2018: 203-215.
- [8] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Proceedings of the 2014 Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 2672-2680.
- [9] ZHU H Y, PENG X, CHANDRASEKHAR V, et al. Dehaze GAN: When image dehazing meets differential programming [C]. The Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018: 1234-1240.
- [10] Zhong W F, Zhao J. Image defogging algorithm based on generative adversarial network[J]. Advances in Laser and Optoelectronics 2022,59(4):337-345.
- [11] Tu, H Y, Wang W L, Chen J C, et al. A generative adversarial network defogging algorithm combining atmospheric scattering model [J]. Journal of Zhejiang University (Engineering Edition), 2022,56(02),225-235.
- [12] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, et al. Deblurgan: Blind motion deblurring using conditional adversarial networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8183-8192.
- [13] C. Liang-Chieh, G. Papandreou, I. Kokkinos, k. murphy, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. In ICLR, 2015.
- [14] YING X Y, WANG Y Q, WANG L G, et al. A stereo attention module for stereo image super-resolution[J]. IEEE Signal Processing Letters, 2020, 27: 496-500.
- [15] Gulrajani I, Ahmed, F, Arjovsky, M, et al. Improved Training of Wasserstein GANs[J]. arXiv:1704.00028v3,2017.
- [16] De Boer P T, Kroese D P, Mannor S, et al. A tutorial on the cross-entropy method [J]. Annals of operations research, 2005, 134 (1): 16-67.
- [17] Liu F Y, Shen C H, Lin G S, and Reid L. Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields[J]. IEEE transactions on pattern analysis and machine intelligence, VOL. 38, NO. 10, 2016.