An Image Dehazing Algorithm Based on the Improved CGAN

Xuelian Li, Zhibin Chen*, Wenqian Zheng and Ying Chang
University of Science and Technology Liaoning, Anshan, China

*Corresponding author e-mail: 272523125@qq.com

Abstract. In order to improve the haze removal effect of image, a modified Conditional Generative Adversarial Nets (CGAN) based algorithm is proposed. In the new algorithm, the pre-trained visual geometry group (VGG) model is adopted, the DenseNet instead of the traditional U-net as the network structure of the generator, the Patch-GAN as the network structure of the discriminator, and the loss function is modified by the total variation regularization gradient. The de-fogged image can be obtained without estimating the projection map and the related defogging feature. The experiments indicate that our new algorithm effectively reduces the halo phenomenon and haze residue problem caused by the traditional dehazing method, and can preserve more details of the image, the structural similarity is improved from 75.9% to 92.6%.

1. Introduction

Under foggy conditions, some media such as dust particles and water vapour are present in the air. These media are scattered into the atmosphere, which weakens the direct transmission of the scene radiation, resulting in a decrease in the sharpness of the image captured by the image acquisition device and a decrease in contrast. He [1] et al. presented a dark channel prior method, which estimated the global atmospheric light and atmospheric transmission rate, and by further optimized it to recover the image through soft matting or guided filter [2]. Meng [3] et al. added boundary constraints on the transfer function to make the transmission map prediction more accurate. Berman [4] et al. proposed a non-local method to perform dehazing. However, above mentioned algorithms have some drawbacks. He’s method may overestimate the concentration of haze, resulting in excessive defogging. Especially, when the sky area or large areas of white buildings appear, the estimated global atmospheric light will no longer be accurate, beside too much calculation when using soft matting to optimize transmission. Meng’s method, for some images, there is excessive defogging and halo effects. Berman’s method has the problem of large computational complexity.

Recently, some researchers have begun to use deep learning frameworks perform dehazing. Ren [5] et al., considering the multiscale features of fog images, proposed a multiscale Convolution Neural Network (CNN) to obtain the feature estimation transmittance of fog images. Cai [6] et al. proposed the DehazeNet network structure, using a specially designed CNN to learn the characteristics of foggy images and solve the difficulties of manual feature design. In this paper, we present an image dehazing algorithm by modifying the Conditional Generative Adversarial Networks (CGAN) [7]. The original idea of our algorithm came from the work in [8]. We mainly modify the Generator and Discriminator architectures, use the DenseNet [9] as the structure of generator, increase the number of network layers of the discriminator and change the size of the convolution kernel in the convolution layer.
2. Image Dehazing Based on CGAN

In this section, a CGAN based image dehazing method is proposed. Firstly, the transmission map is estimated by using the atmospheric scattering model [10]. Secondly, the network of CGAN model is designed to learn the mapping relationships between dehazing images and the corresponding estimated transmission maps. Finally, the new loss function for training is constructed, and by minimizing the loss function leads to our dehazing method.

2.1. Atmospheric Scattering Model

Atmospheric scattering model is first introduced by McCartney in [10], and then further developed by Nayer and Narasimhan [11]. McCartney's atmospheric scattering model is defined in (1).

\[ I(x) = J(x) \cdot t(x) + A(1 - t(x)) \]  

In (1), \(I(x)\) and \(J(x)\) represent the dehazing image and the haze-free image respectively, \(A\) is the global atmospheric light, \(t(x)\) is the transmission map, which represents the ability that the scene reaches to the sensor, \(x\) is the position of the pixel in the image.

2.2. Generative Adversarial Networks

The Generative Adversarial Network (GAN) was proposed by Goodfellow et al. in 2014 [12]. The GAN mainly includes a generator and a discriminator two parts. The generator is able to capture the potential distribution of real data samples and generate the new data samples. The discriminator essentially is a binary classification, and distinguishes whether the input sample is real or fake. Its structure is shown in Figure 1:

The CGAN is based on the condition \(y\) imposed on the original GAN to generate an image with certain requirements. The confrontation process is defined as follows:

\[ \min G \max D \mathbb{E}_{x \sim p_{data}(x)} \log(D(x|y)) + \mathbb{E}_{z \sim p_{data}(z)} \log(D(G(z|y))) \]  

In (2), \(p_{data}(x)\) and \(p_{data}(z)\) are distributions of the real data and the generated data, respectively. In the generator \(G\), the prior input noise \(z\) and the condition \(y\) are combined in the hidden layer, and a conditional generated data \(G(z|y)\) can be output. The discriminator \(D\) combines the real data \(x\) and \(y\) as real data inputs for the discriminator function.

2.2.1. Generator

The model consists of 8 dense blocks and 20 layers of bottleneck layers and two single-layer convolutions of input and output. The number of layers of dense blocks (DB) is 6, composed of Batchnorm-Relu-Convolution(1*1)-Batchnorm-Relu-Convolution(3*3) layers. Each DB is followed by a Transition Down (TD) Layer. The TD Layer comprises of Batchnorm-Relu-Convolution-Dropout-pooling operations. Similarly, on the decoder side, each DB is connected with a
Transition Up (TU) Layer. The TU Layer has only a Deconvolution operation. The growth rate is fixed at 12. The structure of the generator is shown in Figure 2.

![Generator Diagram](image)

**Figure 2.** The Structure of Generator.

2.2.2. **Discriminator.** The discriminator is mainly to make true and false judgments on the image generated by the generator. The discriminator used in this paper adopts PatchGAN, which maps the input to N*N patch. When N=1, it is equivalent to performing pixel by pixel. When N=256, it is equivalent to directly operating on one image. It is proved by experiments that when N=70, the effect is the best. The basic operating units include the Batchnorm-LeakyRelu-Convolution layer. In the network structure, the number of the first layer filters of the discriminator is 64, and the total number of defined network structure layers is 5, and the number of filters per layer is 64*[1, 1, 2, 4, 8]. The convolution "padding" uses the "VALID" approach. The convolution kernel has a size of 3 and a step size of 2. The convolution of the last layer, the convolution kernel is 4, and the step size is defined as 1. The structure of the discriminator is shown in Figure 3.

![Discriminator Diagram](image)

**Figure 3.** The Structure of Discriminator.

2.2.3. **Loss Functions.** The loss function in this paper consists of three parts: the loss of conditional CGAN, the loss of VGG network, and the loss of \( L_p \). The CGAN loss is defined as below:

\[
L_{Adv} = E_{(x,y)}[\log(D(x,y))] + E_{(x,z)}[\log(1 - D(x, G(x,z)))] \quad (3)
\]

As mentioned in the earlier section, \( D \) refers to the discriminator and \( G \) refers to the generator. The input hazy image is defined as \( x \), and its corresponding haze-free counterpart is defined as \( y \).

The network in this paper is trained by VGG19, so we consider introducing a pre-trained VGG feature-aware loss to constrain the generator when calculating the loss. The loss function is defined as:

\[
L_{VGG} = \frac{1}{cwh} \sum_{c=1}^{C} \sum_{w=1}^{W} \sum_{h=1}^{H} \| V(G(x,z)^{c,w,h} - V(y^{c,w,h}) \|^2_2 \quad (4)
\]

In (4), \( C, W, \) and \( H \) represent the output channels, width and height respectively. \( V \) represents the non-linear CNN transformation, which is performed by the VGG network.

Define the loss between the target image and the generated image as \( L_p \), \( \lambda \) is the regular term coefficient, \( \| \nabla G(x,z) \|_1 \) is the total variation regular term, which is defined as:

\[
L_p = E_{xy,z}[\| y - G(x,z) \|_2^2] + \lambda \| \nabla G(x,z) \|_1 \quad (5)
\]

Therefore, the total loss of the entire network model is:
In (6), $\alpha$, $\beta$, and $\gamma$ are the weighted values of adversarial loss, $L_p$ loss, and VGG perception loss, respectively. The algorithm is achieved through minimizing $Loss_{total}$ by continuously training the network.

3. Experiments

3.1. Experiments Setting

All the data trainings are performed on the computer configuration for Intel (R) (TM) i5 / 9400F CPU@2.90GHZ, and 8G DDR3 RAM. Set up the TensorFlow framework and program in the Python language. The network training was performed on the synthesized indoor dataset NYU depth image dataset, and the picture size is (256, 256, 3). The adaptive optimization algorithm was used, the learning rate defaults set to 0.001, GAN_wt=100, VGG_wt=10.

Our new algorithm is evaluated by Peak Signal to Noise Ratio (PSNR) and structural similarity index (SSIM). PSNR is one of the commonly and widely used metrics in objective image evaluation indicators. It is the ratio of the maximum signal amount to the noise intensity. The PSNR refers to the ability of the algorithm to remove haze from the noise image. Two identical images will have infinite PSNR values. The higher PSNR value means the better the performance when determining the dehazing ability. Structural similarity index (SSIM) measures the similarity of two images. It is based on the brightness, contrast, and structure of the local pattern. Structural similarity index refers to the degree of similarity between the two images, and two identical images have a SSIM value of one. Score is the weighted sum of PSNR and SSIM, and the $W_{PSNR}$ and $W_{SSIM}$ are the weight coefficients. Similarly, the higher score value means the better the network performance. It is defined as:

$$Score = W_{PSNR} \cdot PSNR + W_{SSIM} \cdot SSIM$$

In (7), the weight coefficient of $W_{PSNR}$ is set to 0.05, and $W_{SSIM}$ is set to 0.95.

Table 1 shows the quantitative indexes of different numbers of training pictures under condition of 200 iterations. From the quantitative indexes in Table 1, it can be seen that the number of pictures affect the performance of the network.

| Factors | Images 521 | Images 850 | Images 1449 |
|---------|------------|------------|-------------|
| PSNR    | 20.31      | 24.231     | 24.472      |
| SSIM    | 0.899      | 0.907      | 0.926       |
| Score   | 1.870      | 2.073      | 2.103       |

3.2. Experiments on synthetic images and Natural Images

In order to verify the validity of the proposed method, the defogging results of the algorithms of He [1], Meng [3], Berman [4], Raj N [8] on different images are compared.

Figure 4 shows the comparison of the haze removal effects of different methods for the synthetic haze image. Figure 5 shows the comparison of the dehazing effects of different algorithms for natural scenes.

From these figures, we can see that He's method has a better haze removal effect, but after haze removal, the halo phenomenon is more serious and the colour of the sky area image deviates too much from the original image. The image processed by the Meng method can better retain the true colour of the entire image, but its haze removal effect is not ideal, and the halo phenomenon of the image after defogging is the most serious. The Berman method has a clearer image after haze removal, and the halo phenomenon is smaller, but the colour distortion is more serious. The haze-remove image is not obvious by the Raj method. Compared with the other four defogging methods, the defogging effect of this paper is more obvious, the defogging is more thorough, and the image is clearer. Obviously, the presented algorithm performs well, and can retain more image details from the input blurred image.
Relatively speaking, the halo phenomenon is not obvious in our defogging images. However, there are also colour distortion problems in the images processed by our algorithm. This reminds us to do further research and improvement on our algorithm.

**Figure 4.** Comparison of dehazing effects of different algorithms for synthetic images.

**Figure 5.** Comparison of dehazing effects of different algorithms for natural images.

### 3.3. Quantitative Assessment

This subsection gives a quantitative comparison of other classical dehazing algorithms (Dark channel prior haze removal, Boundary constraint and context regularization, Non local image dehazing, GAN dehazing) with the proposed algorithm under the same parameter settings.

Table 2 shows the quantitative indexes of different defogging methods. Experiments show that the new algorithm obtains the higher PSNR and SSIM values and Score. The structure similarity index increased from 75.9% to 92.6%, the Score increased from 1.321 to 2.103 as well. The improvement effect is obvious.

**Table 2.** Quantitative comparison of dehazing effect of different algorithms.

| Model     | He [1] | Meng [3] | Berman [4] | Raj N [8] | Proposed Model |
|-----------|--------|----------|------------|-----------|----------------|
| PSNR      | 13.89  | 14.48    | 12.48      | 20.32     | 24.47          |
| SSIM      | 0.659  | 0.651    | 0.649      | 0.759     | 0.926          |
| Score     | 1.321  | 1.342    | 1.241      | 1.737     | 2.103          |
4. Conclusion

An image dehazing method based on the improved CGAN is proposed. The algorithm adopts different constructs of Discriminator and Generator. The loss function is modified by adding the total variation regular term. The experiments reveal that the presented algorithm has a very good defogging effect. This algorithm largely solves the halo phenomenon and haze residue problems caused by traditional methods, and can better retain the images details after defogging. Experimental results indicate that the structure similarity index increased from 75.9% to 92.6%. However, the size of the experimental images is restricted (256×256), and the number of real smog pictures trained is small, so the real defogging effect is not as good as the dehazing effect of the synthetic indoor smog image.

5. Acknowledgments

This work was financially supported by Liaoning Education Department fund (2017FWDF06).

References

[1] He K, Jian S, Tang X. Single image haze removal using dark channel prior. Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, San Francisco, CA, USA, 2010, pp 2341-2353.

[2] Li Z, Li J Z, Hu Y J, Zhang Y. Mixed prior and weighted guided filter image dehazing algorithm. Journal of Image and Graphics. 24 (2019) 2 0170-0179.

[3] Meng G, Ying W, Duan J, et al. Efficient image dehazing with boundary constraint and contextual regularization. Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 2013, pp617-624.

[4] Berman D, Treibitz T, Avidan S. Non-local image dehazing. Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 2016, pp1674-682.

[5] Ren W, Si L, Hua Z, et al. Single image dehazing via multi-scale convolutional neural networks. Proceedings of the European Conference on Computer Vision, Amsterdam, Netherland, 2016, pp154-69.

[6] Cai B, Xu X, Jia K, et al. DehazeNet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing. 25 (2016) 11 5187-5198.

[7] Mirza M, Osindero S. Conditional Generative Adversarial Nets. IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, USA, 2014. arXiv:1411.1784.

[8] Bharath Raj N, Venkateswaran N. Single Image Haze Removal using a Generative Adversarial Network. IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA,2018. arXiv:1810.09479.

[9] Huang G, Liu Z, Maaten L V D, et al. Densely connected convolutional networks. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017. arXiv:1608.06993.

[10] McCartney E.J. Optics of the atmosphere: Scattering by molecules and particles. USA: John Wiley and Sons Inc., 1976.

[11] Narasimhan S G, Nayar S K. Contrast restoration of weather degraded images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 25 (2003)6 713-724.

[12] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. International Conference on Neural Information Processing Systems, Kuching, Malaysia, 2014, pp2672-2680.