Original Research Article

An Aerial Image Dehazing Algorithm Using a Prior-Based Dense Attentive Network

ZHAO Hang

Civil Aviation for Flight Technology and Safety Research Base，Civil Aviation Flight University of China，Guanghan 618307，China

Abstract: To address the problem that the acquired images tend to degrade in clarity and fidelity in aerial hazy conditions to the extent that the target is difficult to detect, this paper proposes an aerial image dehazing algorithm using a prior-based dense attentive network. The network is based on dense blocks and attention blocks with an encoder-decoder structure, which can directly learn the mapping between the input image and the corresponding haze-free image without relying on the traditional atmospheric scattering model. In addition, to better handle inhomogeneous hazy images, the initial fuzzy density map is first extracted from the original hazy images and then used as a common input to the network together with the original hazy images. Finally, this paper synthesizes a large -scale aerial image dehazing dataset containing two subsets of uniform and non -uniform images. The experimental results and data analysis show that the proposed method exhibits better performance of dehazing with other algorithms on both synthetic and real images.

Keywords: Aerial image; Image dehazing; Deep learning; Dense network; Attention mechanism.

1. Introduction

Aviation images are easily affected by haze, leading to a decrease in image clarity and fidelity, resulting in low downstream analysis efficiency for application tasks such as object detection. Therefore, haze removal from aerial images is a crucial and indispensable preprocessing task[1].

According to the main idea of image dehazing, they can mostly be divided into two categories: models based on image enhancement and models based on image restoration[2]. Some early image dehazing methods utilized image enhancement techniques, including histogram based and contrast based methods. However, these enhancement methods did not take into account image degradation, leading to common problems such as overestimation, underestimation, and color shift.

The model based on image restoration aims to restore image distortion and correct blurry pixels. Most of these methods are based on atmospheric scattering models, which define the relationship between blurred images and true mappings as "transmission maps" to describe haze distribution[3]. Dark channel prior (DCP)[4] is a well-known prior knowledge applied to image restoration based models, which indicates that most pixels have very low intensity in at least one color channel. This information is used to calculate the dark channel prior, which is used to estimate the transition matrix from the original blurred image and restore a clear image. However, traditional methods have certain limitations when applied to complex aerial image scenes, and cannot achieve the expected results.

Therefore, in recent years, researchers have explored other methods to complete image dehazing tasks. Convolutional neural networks have been proven to be an effective supervised learning model. Cai et al.[4] proposed an end-to-end dehazing convolutional neural network (DehazeNet), which inputs a blurred image and then restores the medium transmission map output to a haze free image through an atmospheric scattering model; Ren et al[6]. proposed a Multi scale Convolutional Neural Network (MSCNN), which consists of a coarse scale sub network for overall transmission of images and a fine scale sub network for local refinement; Li et al[7].

Received: 11 February 2023 Accepted: 22 February 2023 Available online: 2 April 2023

proposed a lightweight defogging network (AOD Net) that directly generates clear images, unlike most previous models that estimate transmission maps and atmospheric light separately; Zhang et al[8]. proposed a new edge preserving dense connection encoder and decoder dehazing network (DCPDN) with multi-layer pyramid pooling modules for estimating transmission maps, and optimized the network using a newly introduced edge preserving loss function.

Although deep learning based dehazing methods have shown significant performance improvements compared to traditional prior methods, due to the significant differences between aerial images and conventional natural images in many aspects, these commonly used dehazing methods are prone to failure in aerial images. In addition, previous algorithms did not have good generalization ability for scenes with uneven haze applied in aerial images, so there is room for further improvement.

2. Basic Principles

2.1 Atmospheric Scattering Model

According to the light scattering phenomenon that occurs under outdoor foggy conditions, scattered light (air light) mixes with image signals, causing changes in the visual brightness and color of the scene. In addition, the signal received by the imaging device from the scenic spot decays along the line of sight, and the degradation amount increases with depth. The degradation of foggy images can be modeled as a combination of the above factors. Based on this principle, the atmospheric scattering model[9] describes the formation mechanism of degraded images in foggy weather. This model is widely used in related research on video and image dehazing, and can be expressed as:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad \text{(Eq.1)}$$

Among them, $I(x)$ and $J(x)$ represent the observed foggy blurred image and clear restored image, A represents the global atmospheric light, $t(x)$ represents the medium transmission map, and x represents the pixel position. When the atmospheric light value is uniform, the medium transmission map can be further defined as:

$$t(x) = e^{-\beta d(x)} \quad \text{(Eq.2)}$$

Where β is the atmospheric scattering coefficient, and $d(x)$ is the scene depth (the distance between the object and the imaging device). From the model equation, it can be seen that the essence of the dehazing problem is to estimate two key parameters from the foggy image $I(x)$, namely atmospheric light intensity A and transmission map $t(x)$, and then recover and solve for the clear image $J(x)$.

2.2 Densely Connected Convolutional Networks

In recent years, with the deepening of neural network research, DenseNet[10] has been widely used. Inspired by deep residual learning networks, it extends the traditional layer by layer connection method of neural networks to cross layer dense connections.

There is a direct connection between any two layers in DenseNet, and the input of each layer in the network can be regarded as the union of the outputs of all previous layers. The feature maps learned by that layer will also be directly transmitted to all subsequent layers as new inputs. It is worth noting that in order to avoid excessive reuse of features caused by network deepening, the Dense type network module only learns very limited features at each layer, in order to better achieve the goal of reducing redundancy. In addition, DenseNet also proposes the concepts of bottleneck layer and transition layer, which compress the computational load of each layer of the network locally through the bottleneck layer, improve the efficiency of feature reuse through the global nature of the transition layer, and further expand and explore the advantages of this model.

3. Network Models

3.1 Network Structure

The network structure proposed in this article combines prior knowledge with deep learning networks to better describe the distribution of fog. To address the irregular distribution of haze in aerial images, a fuzzy density map is extracted from the original input foggy image, and then used as a common input to the network along with the original foggy image. Inspired by the use of dense networks and attention mechanisms in existing computer vision tasks, this paper designs a network with an encoder decoder structure consisting of dense blocks and attention blocks. The overall network structure is shown in Figure 1.

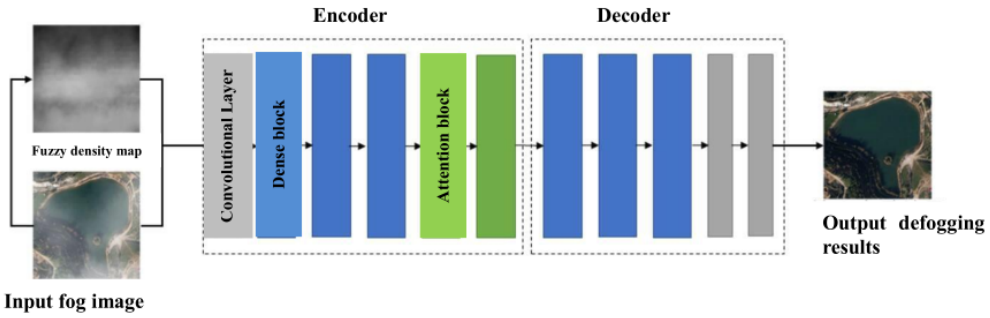


Figure 1 Network structure diagram.

3.2 Fuzzy Density Map

According to the assumption proposed in reference [11], for a blurred area in a given image, the minimum intensity value is higher than the intensity value in the fogless area. Therefore, the minimum intensity in channels R, G, and B is extracted to roughly describe the fuzzy distribution in the original blurred image. Therefore, the original definition was:

$$H_{raw}(x) = \min_{c\in(r,g,b)} I_n^c(x) \tag{Eq.3}$$

In order to further improve the accuracy of extracting fuzzy density maps, we found that in fogless areas, saturation will be higher than in fuzzy areas. Therefore, the modified fuzzy density map can be represented as:

$$H_{\text{mod}}(x) = \max\left(H_{raw}(x) - \alpha S(x), 0\right) \tag{Eq.4}$$

$$S(x) = 1 - \frac{3\min\left(R(x), G(x), B(x)\right)}{R(x) + G(x) + B(x)} \tag{Eq.5}$$

Among them, $S(x)$ represents saturation, and α serves as a regulating factor to control the darkness of the fogless area, and is set to 2 based on experience to ensure that the fogless area is sufficiently dark.

3.3 Dense Blocks

In order to ensure the maximum information flow between layers of the network, we chose to use Dense type modules to connect all matched feature maps in the network, reusing low-level features to high-level ones to improve the accuracy of dehazing, as shown in Figure 2. This preserves the characteristics of forward propagation, and the results of each layer are obtained based on the results of the previous layer. This can effectively enhance the feature propagation transmitted by the hierarchical network, while promoting feature reuse. The relationship between the input and output of any layer L can be expressed as follows:

$$X_L = H_L\left(\left[x_0, x_0, \cdots, x_{L-1}\right]\right) \tag{Eq.6}$$

In the formula, [...] represents concatenating the outputs of all previous layers together, and HL represents

the mapping function of the L-th layer. This dense connection method allows each layer of the network to receive gradient signals from all subsequent layers during backpropagation, so as the network depth increases, the gradient near the input convolutional layer does not become smaller and smaller, reducing the problem of gradient dissipation during training. Meanwhile, due to the reuse of a large number of features, using a small number of convolution kernels can generate a large number of features, ultimately resulting in a smaller model size[12].

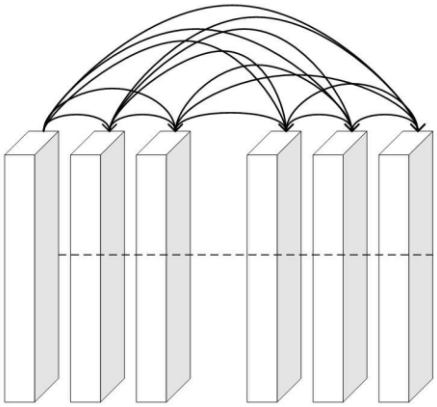


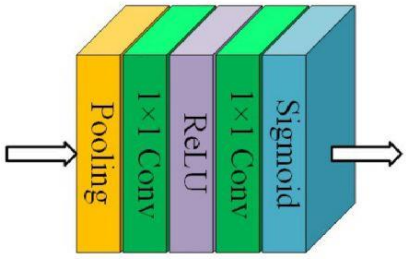
Figure 2 Schematic diagram of densely connected convolutional block structure.

3.4 Attention Block

Attention mechanisms play an important role in human perception, utilizing a series of local observations and selectively focusing on salient parts to better capture visual structures[13]. The attention module in this article is mainly composed of two parts in series: one is the channel attention module, and the other is the spatial attention module.

Each channel feature map of the depth features extracted by the channel attention module can be viewed as a response to a specific category, and different semantic responses are interrelated. Its structure is shown in Figure 3. By constructing a channel attention mechanism to express the interaction between channels, the input feature maps are compressed in spatial dimensions through average pooling and maximum pooling, respectively. The channel attention map is calculated by sharing multi-layer perceptrons.

Unlike channel attention, spatial attention mainly focuses on the position information of the target on the image. The network structure of the spatial attention module is shown in Figure 4. Input a feature map and perform global pooling on elements at the same position for each channel to obtain a spatial feature description map. Extract edge information from the image, and then process it using a Sigmoid activation function to obtain a spatial attention weight map.



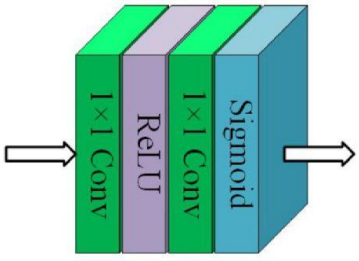


Figure 3 Schematic diagram of channel attention module structure. **Figure 4** Schematic diagram of spatial attention module structure.

By utilizing channel attention and spatial attention modules separately, important features can be focused on from both channel and spatial dimensions, and unimportant features can be filtered out for recalibration of the extracted feature map, making the entire network more focused on important features and better describing the uneven haze distribution[14].

3.5 Loss Function

Mean squared error (MSE) is the most widely used loss function for image dehazing. Although minimizing the MSE loss directly can achieve good results, it is difficult to avoid ambiguity in details. Therefore, this paper adopts a multi task loss function, which consists of MSE and perceptual loss, to achieve anomaly control and performance improvement. The formula for calculating MSE loss is as follows:

$$L_{MSE} = \frac{1}{N}\sum_{x=1}^{N}\sum_{i=1}^{3}\left\| G\left(I(x)\right) - I(x)\right\|_2 \quad \text{(Eq.7)}$$

Where $G\left(I(x)\right)$ and $I(x)$ represent the images before and after dehazing, respectively, and N represents the total number of images included in the training set.

Perceived loss utilizes multi-scale features extracted from pre trained deep neural networks to quantify the visual differences between estimated images and ground truth images. This article uses VGG16 pre trained on ImageNet as the loss function, defined as follows:

$$L_{vgg} = \frac{1}{CWH}\sum_{c=1}^{C}\sum_{w=1}^{W}\sum_{h=1}^{H}\left\| V\left(G(x,z)^{c,w,h}\right) - V\left(y^{c,w,h}\right)\right\|_2^2 \quad \text{(Eq.8)}$$

C, W, and H represent the output channel, width, and height, respectively. The total loss is defined by combining MSE loss and perceptual loss, as follows:

$$L_{Total} = L_{MSE} + \lambda L_{vgg} \quad \text{(Eq.9)}$$

Among them, pose is the relative weight coefficient used to adjust the two loss components, which is set to 0.04 in this article.

4. Experimental Process

4.1 Datasets

Usually, it is impractical to collect a large number of blurry real-world images and their corresponding haze free images from the same perspective and scene. Therefore, data-driven dehazing methods typically rely on synthesizing blurred images. Here, we established aerial image pairs of uniform and non-uniform fog as training datasets, where clear background images were selected from the aerial image dataset (AID)[15], which was originally developed for aerial scene classification tasks. In order to synthesize a uniform foggy blurry image, atmospheric light is uniformly set between [0.5, 1], and $t \in \{0.4, 0.6\}$ is selected. Then, using clear images from the AID dataset, a uniform foggy image is generated using formula (1). For non-uniform foggy blurred images, this paper adopts the synthesis method proposed in reference [16], first extracting 40 different transmission images from real aerial images, and then adding them to the clear images of the AID dataset. Finally, a total of 12000 pairs of uniform fog images and 8000 pairs of non-uniform fog images were synthesized. During the experimental process of this article, 80% of it was selected for training, and the remaining 20% was used for validation.

4.2 Training Details

Training refers to the process of continuously updating network information and minimizing the loss function through error backpropagation. The network training in this article is conducted on a Nvidia RTX 2080Ti GPU (16GB) computing graphics card. Using Gaussian random variables to initialize weights, momentum and decay parameters are set to 0.9 and 0.0001, learning rate value is set to 0.001, Batchsize value is set to 16, and the network performs gradient descent training under the Adam optimizer, iteratively training for a total of 120 cycles until convergence.

5. Experimental Results and Analysis

To demonstrate the effectiveness and robustness of the proposed network, we conducted qualitative and quantitative comparative experiments on synthetic and real data, and compared and analyzed it comprehensively with the four existing mainstream image dehazing algorithms mentioned earlier, namely DCP, MSCNN, AOD Net, and DCPDN.

5.1 Synthetic Image Comparison Experiment

5.1.1 Subjective Evaluation

Firstly, we tested and evaluated our network on a synthetic dataset, and Figure 5 selected two sets of comparison results.

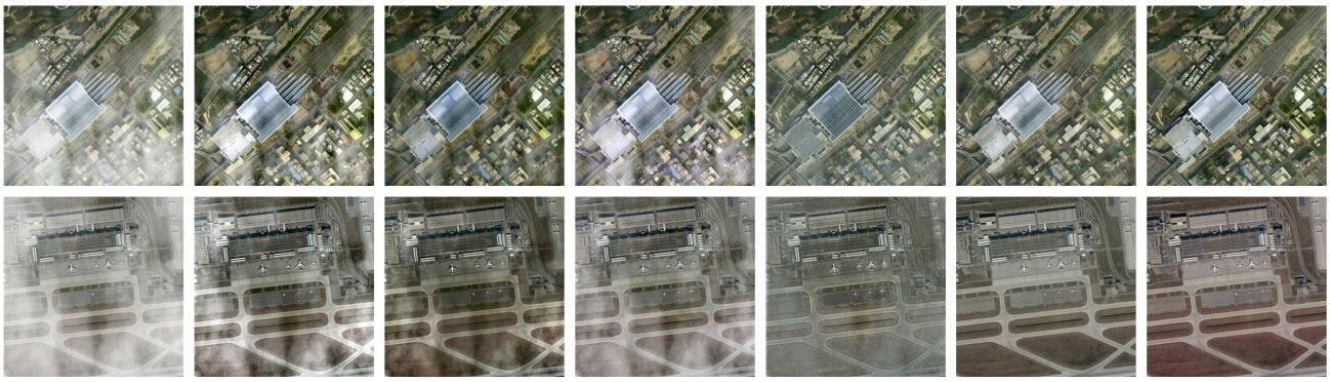


Figure 5 Comparison results of composite images.

We can intuitively see that most algorithms do not perform well in dealing with uneven fog. Due to inaccurate estimation of fog thickness, the results of DCP tend to cause severe color distortion, thereby affecting the quality of its output. For MSCNN, there is still residual haze that cannot be eliminated, and the output suffers from color distortion. AOD Net has insufficient defogging processing intensity and is ineffective for complex aerial image scenes. Although DCPDN exhibits a certain defogging effect, unfortunately, it loses the detailed information in the image, which is unacceptable for subsequent advanced aviation missions. Therefore, compared with existing technologies, the method proposed in this article has the best performance in defogging, suppressing artifacts, distortion, and detail restoration.

5.1.2 Objective Evaluation

Subjective evaluation alone cannot demonstrate the effectiveness of our method. Due to the presence of paired labels in synthesized images, we use two image quality evaluation metrics, PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index), to evaluate the image quality after dehazing. PSNR represents the peak signal-to-noise ratio, which evaluates the quality of an image based on pixel to pixel error. The larger the value, the closer the image is to the ground without fog. SSIM represents structural similarity, which mainly evaluates the similarity of images from the aspects of brightness, contrast, and structure. The larger the value, the smaller the degree of image distortion. The detailed experimental results are shown in Table 1.

Table 1 Quantitative comparison of different algorithms on synthetic data.

Methods	PSNR/dB	SSIM/%
DCP	16.3782	0.7725
MSCNN	15.4825	0.7582
AOD-Net	14.9105	0.7541
DCPDN	17.2170	0.791
This paper	**24.2996**	**0.9010**

From the perspective of defogging performance indicators, the algorithm in this article has the highest PSNR and SSIM values on the synthesized image, indicating that the defogging results are closer to the real ground.

5.2 Real Image Comparison Experiment

5.2.1 Subjective Evaluation

In order to further verify the effectiveness of the method proposed in this article under real-world conditions, real foggy aerial images were selected for comparative experiments. Figure 6 shows one of the comparison scenarios.

Input **DCP** **MSCNN** **AOD-Net** **DCPDN** **This article's algorithm**

Figure 6 Comparison results of real images.

We have noticed that DCP successfully removes most of the haze, but often excessively enhances the image, especially for large white areas, because when the color of the object approaches atmospheric light, DCP priors become invalid. For the three deep learning methods of DehazeNet, AOD-Net, and DCPDN, dehazing is obviously not thorough. After defogging on the internet in this article, the observation of buildings, image edge features, and detailed information on the ground is more complete, and the overall color tone is more in line with human visual perception.

5.2.2 Objective Evaluation

This article also uses two indicators, IE (Information Entropy) and AG (Average Gradient), to compare and analyze the experimental results. IE represents information entropy, and the larger the value, the greater the amount of information contained in the image, resulting in a clearer image. AG represents the average gradient, which reflects the rate of detail contrast change in the multi-dimensional direction of the image. The larger the value, the more layers the image has and the clearer it becomes. The results are shown in Table 2.

Table 2 Quantitative comparison of different algorithms on real data.

Methods	IE	AG
Input	7.3279	4.8889
DCP	7.4176	8.6398
MSCNN	7.6550	7.1184
AOD-Net	7.3855	9.9490
DCPDN	6.9144	7.8127
This paper	**7.8148**	**10.2104**

Both IE and AG have higher values, indicating that our method achieves better preservation of edge information, with better contrast, clarity, and detail information compared to other methods.

6. Conclusion

This article proposes a dense attention network that combines prior knowledge for aerial image dehazing. This method effectively describes the fuzzy distribution in aerial images, thus solving the problem of removing uneven haze. Specifically, the initial blur density map is first extracted from the original foggy blurred image, and then used as a common input to the network along with the original foggy image. In addition, a clear image estimation network based on dense blocks and attention blocks can directly learn clear images from input images

without estimating any intermediate parameters. The experimental results on the established dataset show that our algorithm has better performance than other commonly used dehazing algorithms. In future work, we will attempt to further enhance the potential of the network by introducing more background prior knowledge.

References

1. Wang Jingdong, Zhang Wentao, Wang Zirui, Xu Lihong. A fast aerial image dehazing algorithm [J]. Journal of Aeronautics, 2013, 34 (03): 636-643.
2. Wu Di, Zhu Qingsong. Latest research progress on image dehazing [J]. Journal of Automation, 2015, 41 (2): 221-239.
3. Fattal R. Single image dehazing [J]. ACM Transactions on Graphics, 2008, 27(3): 72.
4. He K, Sun J, Tang X. Single image haze removal using dark channel prior[C]//IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2009. IEEE, 2009: 1956-1963.
5. Cai B, Xu X, Jia K, et al. DehazeNet: An End-to–End System for Single Image Haze Removal [J]. IEEE Transactions on Image Processing, 2016, 25 (11): 5187-5198.
6. Ren, X., Bo, L. and Fox, D. (2012) Rgb-(d) Scene Labeling: Features and Algorithms. IEEE Conference on Computer Vision and Pattern Recognition, Providence, 2759-2766.
7. Li Yong-Fu, Cui Heng-Qi, Zhu Hao, Zhang Kai-Bi. A defogging algorithm for aerial image with improved AOD-Net. Acta Automatica Sinica, 2022, 48(6): 1543−1559.
8. Zheng, S., Jayasumana, S., Romera-Paredes, B., et al. (2015) Condi-tional Random Fields as Recurrent Neural Networks. Proceedings of the IEEE International Conference on Computer Vision, Santiago, 1529-1537.
9. Shel-hamer, E., Long, J. and Darrell, T. (2017) Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39, 640.
10. Csurka, G. and Perronnin, F. (2011) An Efficient Approach to Semantic Segmentation. International Journal of Com-puter Vision, 95, 198-212.
11. Xiao, J.X. and Quan, L. (2009) Multi-ple View Semantic Segmentation for Street View Images. International Conference on Computer Vision, Kyoto, 29 Sep-tember-2 October 2009, 686-693.
12. Shotton, J., Winn, J. and Bother, C. (2006) Textonboost: Joint Appearance, Shape and Context Modeling for Multiclass Object Recognition and Segmentation. ECCV, 1, 1-15.
13. Noh, H., Hong, S. and Han, B. (2015) Learning Deconvolution Network for Semantic Segmentation. Proceedings of the IEEE International Conference on Computer Vision, Santiago, 18 February 2015, 1520-1528.
14. Zhao Quanhua, Zhao Xuemei, Li Yu Fuzzy ISODATA high-resolution remote sensing image segmentation combined with HMRF model [J] Signal Processing, 2016, 32 (2): 157-166.
15. Li Quanwu Research on Road Background Detection Algorithm Based on K-means [J] Information and Computer (Theoretical Edition), 2020, 32 (15): 48-50.
16. Xu Xinzheng, Ding Shifei, Shi Zhongzhi, etc New theories and methods for image segmentation [J] Journal of Electronics, 2010, 38 (2): 76-82.