

Multi-Feature Fusion ECANet-Based Adaptive Dehazing Algorithm for Aerial Images

Yunyun Li *, Juan Duan, Cong Zhai

Department of Data Science and Engineering, Yan'an University Xi'an Innovation College, Xi'an, Shaanxi, China

* Corresponding author: Yunyun Li (Email: 2674663785@qq.com)

Abstract: Aiming at the problems of blurred details, insufficient utilization of multi-scale features, brightness shift and color distortion in aerial hazy images, this paper proposes a multi-feature fusion attention-based adaptive dehazing network (MFFAMA-Net), which is composed of three cascaded modules, feature extraction, multi-feature fusion attention and adaptive brightness correction. The feature extraction module preserves shallow edge and texture details, and adopts three-branch parallel multi-scale convolutions to enlarge the receptive field level by level while keeping the number of channels constant, with skip connections linking the shallow and deep information. In the multi-feature fusion attention module, global max pooling (GMP) is introduced in parallel into the channel attention of ECA-Net without dimensionality reduction, and is combined with global average pooling (GAP) to form a dual-path encoding, so as to realize channel recalibration and adaptive fusion of shallow and deep features. The adaptive brightness correction module predicts pixel-wise and channel-wise Gamma parameters through the network to adaptively correct the brightness and color of the restored result. The network is optimized in an end-to-end manner with a multi-constraint joint loss function composed of a pixel-level reconstruction loss, a perceptual loss and a color consistency loss. Experimental results show that on the RESIDE dataset the proposed method outperforms the mainstream dehazing methods in objective metrics such as PSNR and SSIM, and the information entropy also indicates that the restored images contain richer details. Validation on the real aerial hazy images selected from HazyDet shows that the proposed method still maintains a good visual restoration effect in real scenes and has favorable practical applicability. In addition, the inference time of the proposed method for a single image is 0.216 s, which can meet the requirement of fast image dehazing.

Keywords: Image Dehazing; Feature Extraction; Multi-feature Fusion; ECANet; Gamma.

1. Introduction

Owing to the advantages of flexibility and strong timeliness, unmanned aerial vehicle (UAV) aerial photography has been widely used in remote sensing monitoring, disaster assessment, intelligent transportation, military reconnaissance and other fields. However, in hazy weather, the scattering and absorption of atmospheric particles degrade the imaging quality, resulting in decreased image contrast, color distortion and blurred details. Such degradation not only reduces the visual readability of images, but also restricts the accuracy and reliability of high-level vision tasks such as object detection and semantic segmentation. Therefore, it is of great theoretical significance and application value to study the dehazing methods for aerial images.

Traditional dehazing methods are divided into two categories, physical-model-based restoration and image-enhancement-based methods[1]. The former is based on the atmospheric scattering model and restores the scene by estimating the transmission and the atmospheric light component. He[2-3] proposed the dark channel prior (DCP), which estimates the transmission according to the dark channel statistics of local regions in haze-free images, refines it with guided filtering, and then substitutes it into the atmospheric scattering model to solve the scene radiance. Berman[4] estimated the transmission by using a non-local prior and optimized it with a regularization method. Such methods have strong physical interpretability and achieve good restoration results when the prior assumptions hold. However, their priors tend to fail under complex conditions such as sky regions, dense haze areas and drastic depth-of-field changes, leading to color distortion and halo artifacts.

Moreover, the iterative optimization process is computationally expensive and is difficult to meet the requirement of real-time processing. The latter does not rely on a physical model, but improves the visual quality by enhancing contrast and edge details. Soni[5] applied contrast limited adaptive histogram equalization (CLAHE)[6-7] to image dehazing, and enhanced the contrast by redistributing the local brightness. Liu et al.[8] estimated the scene reflectance by separating the illumination component and the reflection component based on the Retinex theory, thereby realizing image enhancement. These methods are simple to implement and computationally efficient. However, since they do not model the physical process of haze degradation and the spatial heterogeneity of the haze concentration distribution, and only improve the visibility through global contrast and brightness adjustment, they are insufficiently adaptable to complex scenes and are difficult to effectively handle images with non-uniform haze distribution.

In recent years, with its powerful feature learning and nonlinear fitting capabilities[9], deep learning has made remarkable progress in the field of image dehazing. Early studies mostly took the atmospheric scattering model as the framework and used convolutional neural networks (CNN) to estimate the key parameters in the model for dehazing. Ren[10] proposed MSCNN, which adopts multi-scale convolution kernels to extract features and establishes the mapping relationship between hazy images and transmission maps. Cai[11] proposed DehazeNet, which extracts haze-related features through multi-scale mapping and max pooling operations, and then estimates the transmission map. However, such methods are not end-to-end mappings from hazy images to haze-free images; instead, they first estimate

the atmospheric light value and the transmission map separately, and then substitute them into the atmospheric scattering model to obtain the restored result analytically. This stage-wise paradigm tends to cause the estimation bias of intermediate parameters to be transferred, accumulated and amplified during the inversion process, thus degrading the quality of the restored image. To this end, Li [12] equivalently reformulated the atmospheric scattering model and proposed a lightweight end-to-end network, AOD-Net, which directly reconstructs the haze-free image and avoids the performance degradation caused by the bias of transmission estimation. Thereafter, the attention mechanism was introduced into the dehazing task to improve the adaptability of the model to the spatial difference of haze concentration[13]. Qin[14] proposed FFA-Net, which uses channel attention and pixel attention to characterize the importance difference of different feature channels and different pixel positions respectively, and realizes cross-layer information aggregation with the help of a feature fusion structure. Nevertheless, while alleviating color distortion, the existing end-to-end networks still face problems such as the attenuation of shallow spatial details with the deepening of the network and insufficient multi-scale feature fusion, and the ability of detail restoration remains to be further improved.

In summary, although the above methods have made some progress in general dehazing tasks, problems such as insufficient multi-scale haze modeling, illumination distortion after restoration and single pooling representation still widely exist when they are directly transferred to aerial images. To this end, this paper proposes a multi-feature fusion attention mechanism based adaptive dehazing network for aerial images (MFFAMA-Net). The network adopts a cascaded structure and consists of three modules, which are respectively used for multi-scale feature extraction, channel-level adaptive fusion of shallow and deep features, and joint correction of the brightness and color of the restored result, and is optimized end-to-end under the supervision of a multi-constraint joint loss. Experimental results show that the proposed method achieves a good visual restoration effect on UAV aerial images, and significantly improves the image

clarity and color fidelity, with the restored scenes being natural and realistic.

2. Overall Framework Design of the Algorithm

To address the shortcomings of the existing aerial image dehazing methods in multi-scale feature utilization, channel representation and illumination consistency, this paper proposes MFFAMA-Net, whose architecture is shown in Fig. 1. It is composed of a feature extraction module, a multi-feature fusion attention module and an adaptive brightness correction module in a cascaded manner. Given a hazy aerial image, the feature extraction module first extracts shallow features with a single 3×3 convolutional layer to preserve edge and texture details, and then captures multi-scale deep features through multi-level stacked 3×3 , 5×5 and 7×7 three-branch parallel convolutions which progressively enlarge the receptive field while keeping the number of channels constant, with skip connections linking the shallow and deep information. Next, the multi-feature fusion attention module concatenates the shallow and deep features along the channel dimension, and introduces global max pooling (GMP) in parallel within the channel attention framework of ECA-Net[15] without dimensionality reduction, forming a dual-path pooling encoding together with global average pooling (GAP), so as to complete channel recalibration and weighted fusion, enhance haze-related responses and suppress redundant information, and then obtains the preliminary restored result through 1×1 convolution reconstruction. Finally, the adaptive brightness correction module predicts a pixel-wise and channel-wise Gamma map in a residual form, which is zero-initialized to ensure that the training starts from an identity mapping, and adjusts the brightness and color of the preliminary restored result to output the dehazed image. The whole network is optimized end-to-end under the supervision of a multi-constraint joint loss composed of a pixel-level reconstruction loss, a perceptual loss and a color consistency loss.

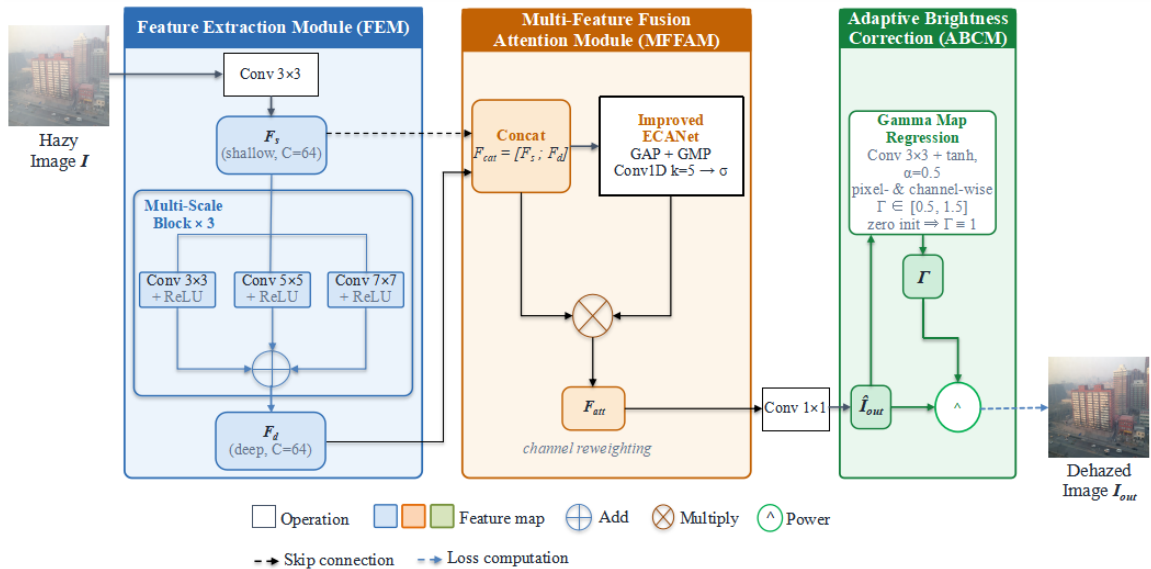


Fig 1. The network model of MFFAMA-Net

2.1. Feature Extraction Module

Haze in aerial images is manifested both as the subtle

texture degradation in local thin haze regions and as the global structural occlusion in dense haze regions. To simultaneously capture fine-grained structural details and multi-scale haze-

related features, this paper divides the feature extraction module into two parts, shallow feature extraction and deep multi-scale feature extraction.

2.1.1. Shallow Feature Extraction

A single 3×3 convolution is adopted to extract the shallow features $F_s \in \mathbb{R}^{H \times W \times C}$, from the input hazy image I . It expands the channel dimension while preserving spatial details such as edges and textures, providing a basis for the detail supplement of the subsequent deep features.

2.1.2. Deep Feature Extraction

Since the haze concentration in aerial scenes varies significantly with the spatial position and the depth of field, the receptive field of a single-scale convolution kernel is limited, and it is difficult to take into account both the local details and the modeling of the wide-area haze distribution. To this end, this paper adopts three convolution branches with different scales to extract deep feature information in parallel. Let the input feature be $F_{in} \in \mathbb{R}^{H \times W \times C}$. The three branches adopt 3×3 , 5×5 and 7×7 convolutions respectively, so that the receptive fields are enlarged step by step, which are used to capture local details, mid-range structures and wide-area context in turn. The number of output channels of the three branches is kept consistent with that of the input. After element-wise addition and fusion, the result is skip-connected with the input feature:

$$F_d = \delta(\text{Conv}_{3 \times 3}(F_{in})) + \delta(\text{Conv}_{5 \times 5}(F_{in})) + \delta(\text{Conv}_{7 \times 7}(F_{in})) \quad (1)$$

$$F_{out} = F_{in} + F_d \quad (2)$$

where $\delta(\cdot)$ is the ReLU activation function. Equations (1) and (2) describe the forward process of a single multi-scale feature extraction block. When three multi-scale feature extraction blocks are stacked in series, the input of the first block is F_s and the output of the last block is the deep feature F_d . The skip connection provides a direct path for the shallow details to be transferred to the deep layers, and alleviates the gradient attenuation caused by the serial stacking at the same time.

2.2. Multi-Feature Fusion Attention Module

After being processed by the feature extraction module, the obtained features already have multi-scale context information. However, there is still redundancy among the channel responses, and the targeted modeling of the non-uniform spatial distribution of haze concentration is still insufficient. To this end, this paper introduces the lightweight channel attention mechanism ECA-Net, which replaces the fully connected layers in SE-Net[16] with one-dimensional convolution, and realizes channel attention with only a few parameters while avoiding channel dimensionality reduction and maintaining cross-channel interaction. However, ECA-Net only adopts global average pooling to encode the channel responses, which only reflects the overall mean statistics of the features and ignores the maximum response information of the salient regions, so that the representation of channel weight allocation under non-uniform haze distribution is relatively single. To this end, on the basis of retaining its non-dimensionality-reduction property, this paper introduces global max pooling in parallel to form a dual-path pooling encoding, and extends this structure to the channel recalibration and adaptive fusion of the concatenated shallow and deep features. Its structure is shown in Fig. 2. Different from CBAM[17], which aggregates dual-path pooling with a dimensionality-reducing shared multi-layer perceptron, this

paper fuses the dual-path statistics within the one-dimensional convolution framework of ECA-Net without dimensionality reduction throughout, thus retaining rich channel representation while having the property of being lightweight.

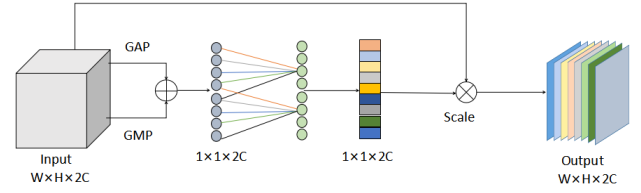


Fig 2. Structure of the improved ECANet model

2.2.1. Dual-Path Pooling Encoding

Taking the concatenated shallow and deep features $F_{cat} \in \mathbb{R}^{H \times W \times 2C}$ as the input, global average pooling (GAP) and global max pooling (GMP) are used to encode the channel responses in a dual-path parallel manner, reflecting the overall response intensity and the strongest response intensity of the channels respectively:

$$z_{avg} = \text{GAP}(F_{cat}), \quad z_{max} = \text{GMP}(F_{cat}) \quad (3)$$

where $z_{avg}, z_{max} \in \mathbb{R}^{1 \times 1 \times 2C}$ are the mean descriptor and the peak descriptor respectively. They compress the features in the spatial dimension, reducing the dimension from $H \times W \times 2C$ to $1 \times 1 \times 2C$, which correspond to the two branches GAP and GMP in Fig. 2.

2.2.2. Channel Weight Generation

After the two descriptors are added and fused, one-dimensional convolution Conv1D_k is used to realize local cross-channel interaction, and the channel weights are generated by the Sigmoid activation:

$$s_c = \sigma(\text{Conv1D}_k(z_{avg} + z_{max})) \quad (4)$$

where $\sigma(\cdot)$ is the Sigmoid function and s_c is the channel weight with a tensor shape of $1 \times 1 \times 2C$. The one-dimensional convolution keeps the channel dimension without dimensionality reduction, avoiding the information loss caused by the fully connected dimensionality reduction in the SE module.

2.2.3. Adaptive Kernel Size

The kernel size k of the one-dimensional convolution is adaptively determined by the number of channels after concatenation:

$$k = \left\lfloor \frac{\log_2(2C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}, \quad \gamma=2, b=1 \quad (5)$$

where $\lfloor \cdot \rfloor_{\text{odd}}$ means rounding down the result in the brackets first, and then taking the smallest odd number not less than the integer. For $C=64$, the number of channels after concatenation is $2C=128$, and $\lceil \log_2 128 / 2 + 1/2 \rceil = 4$, which is an even number, so $k=5$.

2.2.4. Per-Channel Recalibration

The channel weights act on the input concatenated features along the skip path, and the per-channel weighting is completed at the Scale node and then output:

$$F_{att} = s_c \odot F_{cat} \quad (6)$$

where $\odot$ denotes element-wise multiplication and $F_{att} \in \mathbb{R}^{H \times W \times 2C}$. Since s_c acts on the concatenated $2C$ -dimensional channels, its first C -dimensional weights correspond to the shallow feature channels and the last C -

dimensional weights correspond to the deep feature channels. The network adaptively assigns channel-level weights to the two types of features through global pooling encoding, and adjusts their overall contribution ratio in the channel dimension, so as to strike a balance between detail preservation and concentration estimation, and realize the channel-level weighted fusion of shallow and deep features.

2.3. Adaptive Brightness Correction Module

Gamma correction[18] is a classical power-law transformation that performs nonlinear gray mapping through a single parameter. However, traditional methods need to set the Gamma exponent manually and perform a globally consistent transformation on the whole image, which is difficult to adapt to the spatially non-uniform distribution of illumination and haze concentration in aerial scenes. To this end, this paper designs ABCM, which generalizes the fixed Gamma exponent into a pixel-wise and channel-wise Gamma parameter map predicted by the network.

ABCM takes the preliminary restored result $\hat{I}_{out}$ as the input, and predicts the Gamma map Γ through a 3×3 convolution layer and the tanh activation:

$$\Gamma = 1 + \alpha \cdot \tanh(\text{Conv}_{3 \times 3}(\hat{I}_{out})) \quad (7)$$

where $\alpha=0.5$ and $\Gamma \in \mathbb{R}^{H \times W \times 3}$. The limiting effect of tanh makes $\Gamma \in [0.5, 1.5]$, which guarantees the boundedness of the correction amplitude. The Gamma map is applied to the preliminary restored result, and $\hat{I}_{out}$ is numerically truncated to avoid gradient explosion, so as to obtain the final restored image:

$$I_{out} = \text{clamp}(\hat{I}_{out}, \varepsilon, 1)^\Gamma \quad (8)$$

where $\varepsilon=10^{-4}$. The regions with $\Gamma < 1$ are brightened to restore the details in dark areas, while the regions with $\Gamma > 1$ are darkened to suppress over-exposure, and the Gamma exponents of the color channels are independent of each other, which can alleviate the color shift while correcting the brightness. During training, the weights and bias of this convolution layer are both initialized to zero, so that $\Gamma=1$ and $I_{out}=\hat{I}_{out}$ in the initial state. The network starts from an identity mapping, which guarantees the stability of training. Although ABCM can be removed as a whole for ablation, its parameters are optimized end-to-end with the main network and it is not an independent post-processing. The Gamma transformation is completed by a single convolution layer at one time, and thus the overhead is smaller.

2.4. Multi-Constraint Joint Loss Function

After being processed by the adaptive brightness correction module, the network finally outputs the dehazed image I_{out} . In order to make the restored result optimal in pixel accuracy, perceptual quality and color fidelity at the same time, this paper designs a multi-constraint joint loss function to supervise the training process in an end-to-end manner. The loss function consists of three parts, a pixel-level reconstruction loss, a perceptual loss and a color consistency loss:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_{per} + \lambda_3 \mathcal{L}_{col} \quad (9)$$

where λ_1 , λ_2 and λ_3 are the hyperparameters that balance the weights of each loss term, which are set to 1.0, 0.01 and 0.1 respectively. The definitions of each loss term are as follows.

2.4.1. Pixel-Level Reconstruction Loss

In order to effectively train the network and make the restored image consistent with the clear label in pixel values, this paper adopts the L1 loss. Compared with the L2 loss, it can effectively alleviate the problem that the restored result is too smooth and preserve more texture details:

$$\mathcal{L}_{rec} = \frac{1}{HWC} \sum_{i,j,c} \|I_{out}(i,j,c) - I_{gt}(i,j,c)\|_1 \quad (10)$$

where I_{gt} is the corresponding clear haze-free label image.

2.4.2. Perceptual Loss

Relying only on the pixel-level loss tends to make the restored image lack high-frequency details. To this end, this paper introduces a perceptual loss calculated by extracting high-level semantic features with the pre-trained VGG-19 network[19], so as to improve the visual perceptual quality:

$$\mathcal{L}_{per} = \frac{1}{N\phi} \sum_{l=1}^L \|\phi_l(I_{out}) - \phi_l(I_{gt})\|_2^2 \quad (11)$$

where $\phi_l(\cdot)$ denotes the feature map extracted by the l -th layer of the VGG-19 network, and $N\phi$ is the dimension of the corresponding feature map. This loss constrains the distance between the restored image and the real image in the semantic feature space, so that it has more natural texture and global structure.

2.4.3. Color Consistency Loss

Considering the color shift that may occur in the process of dehazing and Gamma adaptive brightness correction, this paper introduces a color consistency loss to constrain the mean difference between the restored image and the real image on the three RGB channels:

$$\mathcal{L}_{col} = \sum_{c \in \{R,G,B\}} \|\mu_c(I_{out}) - \mu_c(I_{gt})\|_1 \quad (12)$$

where $\mu_c(\cdot)$ denotes the pixel mean of the image on the c -th channel. This loss term can effectively suppress the color distortion in the restoration process. Through the optimization of the above multi-constraint joint loss, the network can take into account the pixel-level accuracy, high-level perceptual quality and color fidelity at the same time, thus generating dehazed images with better visual effects.

3. Experimental Results and Analysis

3.1. Experimental Environment and Datasets

The experiments in this paper are carried out on the following hardware and software platforms. The hardware configuration is an Intel(R) Core(TM) i7-7700K CPU @ 4.20 GHz processor, 32 GB memory and an NVIDIA GeForce RTX 2080 Ti GPU (11 GB memory). The software environment is built based on Python 3.8 and the PyTorch 1.12 deep learning framework. In the training stage, the input images are uniformly resized to 384×384 resolution, and the network weights are randomly initialized with a Gaussian distribution. In terms of the optimization strategy, the Adam optimizer combined with the back-propagation algorithm is used to update the parameters, and the specific hyperparameter configuration is as follows: the initial learning rate is 1×10^{-3} , the batch size is 64, and the number of epochs is 100.

In order to fully train and evaluate the proposed algorithm, this paper uses the RESIDE[20] and HazyDet[21] datasets to construct the experimental data. There are 30,000 hazy images in total, which are composed of the ITS subset of RESIDE (10,000 indoor images), the OTS subset (15,000

outdoor images) and the synthetic subset of HazyDet (5,000 UAV aerial images). Each subset is randomly divided according to the ratio of 8:2, of which 24,000 images are used for training and 6,000 images for validation. The introduction of HazyDet aerial samples aims to enable the model to learn the haze distribution characteristics and the scale characteristics of ground objects from the top-view perspective in the training stage. In the test stage, the SOTS subset of RESIDE (1,000 images, including 500 indoor images and 500 outdoor images) is used for quantitative evaluation. In addition, 200 UAV aerial images are selected from the real hazy subset of HazyDet to verify the applicability of the model in real complex scenes. Different haze concentration levels, flight altitudes and ground object types are taken into account in the selection to ensure the representativeness of the samples. There is no overlap between the above test data and the training data.

3.2. Evaluation Metrics

In order to overcome the limitations of subjective evaluation and objectively and comprehensively verify the effectiveness of the proposed method, this paper selects three objective metrics, information entropy (IE), structural similarity (SSIM) and peak signal-to-noise ratio (PSNR), for quantitative evaluation. IE measures the information richness of the restored image, and a larger value indicates that the texture and details are more completely preserved. SSIM and PSNR measure the closeness between the restored image and the clear reference image from the aspects of structural similarity and pixel fidelity respectively, and a higher value indicates better restoration quality. The definitions of the three metrics are as follows:

$$IE = -\sum_{l=0}^{L-1} p_l \log_2 p_l \quad (13)$$

where p_l denotes the probability of the occurrence of the gray level l , L is the total number of gray levels, and IE reflects the information amount and detail richness of the image.

$$SSIM(I_{out}, I_{gt}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

where μ_x and μ_y are the means of the two images, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance, and C_1 and C_2 are stability constants. The closer SSIM is to 1, the better the structure is preserved.

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{\frac{1}{HW} \sum_{i,j} |I_{out}(i,j) - I_{gt}(i,j)|^2} \right) \quad (15)$$

where MAX is the maximum pixel value. PSNR reflects the restoration quality from the perspective of pixel fidelity, and the higher the value, the better.

3.3. Ablation Study

To verify the effectiveness of each module, this paper sets up four variants for comparison on SOTS outdoor hazy images: (a) removing the multi-scale feature extraction module; (b) retaining the attention structure, but changing the dual-path pooling to single-path GAP; (c) removing the adaptive brightness correction module; (d) the complete network of this paper. The four variants adopt the same training data and hyperparameter configuration, and the experimental results are shown in Fig. 3.



Fig 3. Experimental results of network structure verification. From left to right: hazy input, (a), (b), (c) and (d).

It can be seen from Fig. 3 that variant (a), due to the lack of multi-scale feature extraction, has insufficient modeling ability for the wide-area haze distribution, and there is still obvious residual haze after dehazing. Variant (b) only retains global average pooling, and the channel weights are driven only by mean statistics, which lacks the response characterization of salient activation regions, resulting in insufficient discrimination of channel importance under non-uniform haze distribution, and both the detail preservation and contrast restoration of the restored image are inferior to those of the complete model. Variant (c) does not introduce adaptive brightness correction, so the output image is dark as a whole and the brightness effect is not good. In contrast, variant (d), the complete network of this paper, shows clearer edge details and better dehazing effect, which benefits from the hierarchical capture of haze features under different receptive fields by multi-scale convolution and the channel-level weighted fusion of shallow and deep features by the attention mechanism. At the same time, the Gamma adaptive brightness correction further optimizes the image brightness, making the restored result more realistic and natural.

To further verify the effectiveness of the network structure, IE, SSIM and PSNR are used as the evaluation metrics for outdoor dehazing images. The experimental results are shown in Table 1.

Table 1. Ablation study results

Evaluation metric	(a)	(b)	(c)	(d) ours
IE	7.1941	7.0823	7.1255	7.3356
SSIM	0.8136	0.8357	0.8776	0.9289
PSNR/dB	25.8236	28.7836	29.8968	31.5817

It can be seen from Table 1 that the three metrics of the complete network model proposed in this paper are all better than those of the ablation variants, which verifies the effectiveness of the design of each module in this paper. Among them, after removing the multi-scale feature extraction module, PSNR drops the most (from 31.5817 dB to 25.8236 dB), indicating that multi-scale modeling contributes most significantly to the dehazing performance.

3.4. Comparative Experiments

To further prove the effectiveness of the model proposed in this paper, DCP[2-3], CLAHE[5], DehazeNet[11], AOD-Net[12], FFA-Net[14] and other algorithms are selected for comparison. To ensure a fair comparison, all the deep learning comparison methods are retrained on the same mixed training set as this paper, while DCP and CLAHE directly adopt the original implementations. All the algorithms are run in the

same hardware and software environment, and are tested on the SOTS dataset and the real aerial hazy images of the HazyDet dataset respectively, and both subjective and objective aspects are used as the evaluation criteria of the image dehazing effect.

3.4.1. Experimental Analysis on Synthetic Hazy Images

First, two indoor hazy images and two outdoor hazy images are randomly selected from the synthetic hazy image dataset SOTS, and are compared with other dehazing algorithms for subjective and objective evaluation. The experimental results are shown in Fig. 4.

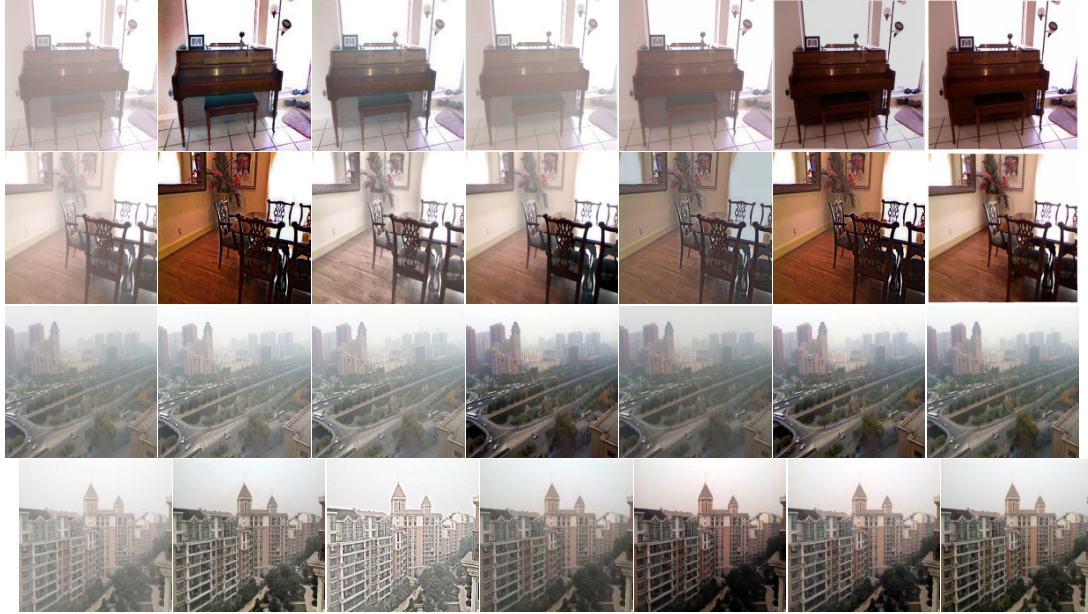


Fig 4. Performance of different methods on the SOTS test set. From left to right: hazy input, (a) DCP, (b) CLAHE, (c) Dehaze-Net, (d) AOD-Net, (e) FFA-Net, (f) Ours.

It can be seen from the results in Fig. 4 that although the DCP algorithm can remove haze, it will cause color distortion and the overall image is dark. The CLAHE algorithm has over-enhancement accompanied by local color distortion. The DehazeNet algorithm does not remove haze thoroughly and the residual haze is obvious. The AOD-Net algorithm tends to cause the loss of scene details while improving the visibility, and the output image is dark. The FFA-Net algorithm has a

good dehazing effect, but there is still color deviation in some regions. Compared with the above algorithms, the algorithm in this paper can better restore the color information, making the dehazed image closer to the haze-free image, with a realistic and natural visual effect.

At the same time, objective metrics are used to evaluate different dehazing algorithms on the SOTS test set. The experimental results are shown in Table 2.

Table 2. Objective evaluation metrics

Model	Indoor IE	Indoor SSIM	Indoor PSNR/dB	Outdoor IE	Outdoor SSIM	Outdoor PSNR/dB
DCP	6.9831	0.7972	22.5217	6.9482	0.8879	21.1342
CLAHE	7.0342	0.7423	19.8235	7.0711	0.8287	19.0325
Dehaze-Net	7.0963	0.8165	25.2723	7.2353	0.8354	24.8536
AOD-Net	7.1823	0.8527	28.1527	7.2842	0.8992	27.9294
FFA-Net	7.2082	0.8602	30.9236	7.3026	0.9045	30.5235
Ours	7.2895	0.8795	31.3921	7.3356	0.9289	31.5817

It can be seen from Table 2 that the IE, SSIM and PSNR of the proposed method are all better than those of other comparison methods. Among them, the IE of outdoor images reaches 7.3356, SSIM reaches 0.9289 and PSNR reaches 31.5817 dB, which proves that the dehazing effect of the proposed method is better. Combined with the subjective and objective comparative experiments, it can be seen that compared with other dehazing comparison algorithms, the algorithm proposed in this paper has relatively optimal performance.

3.4.2. Experimental Analysis on Real Aerial Hazy Images

In order to verify the dehazing effect of the proposed method in real aerial image scenes, each dehazing method is used to conduct experiments on the real aerial hazy images selected from the HazyDet dataset. The dehazing effects of each method on two real aerial hazy images are shown in Fig. 5.

As can be seen from Fig. 5, the difference of the effects of various dehazing algorithms on real aerial hazy images is more obvious than that on synthetic hazy images. The DCP algorithm has a poor dehazing effect, with problems of detail loss and color distortion. The CLAHE algorithm is over-enhanced, resulting in color distortion. The dehazing performance of DehazeNet is not ideal, which may be due to its uneven processing of channels and pixels leading to errors in the estimation of the transmission map. The AOD-Net and FFA-Net algorithms have good dehazing effects, but the output images are dark and the details are not clear enough. The dehazing algorithm proposed in this paper has a good dehazing effect when processing aerial hazy images, but its performance in removing dense haze in large depth-of-field regions is still not ideal. It is speculated that the reason is that the training data come from synthetic hazy images rather than naturally collected real hazy images, which leads to a certain

deviation in the modeling of the real haze concentration

distribution by the model.

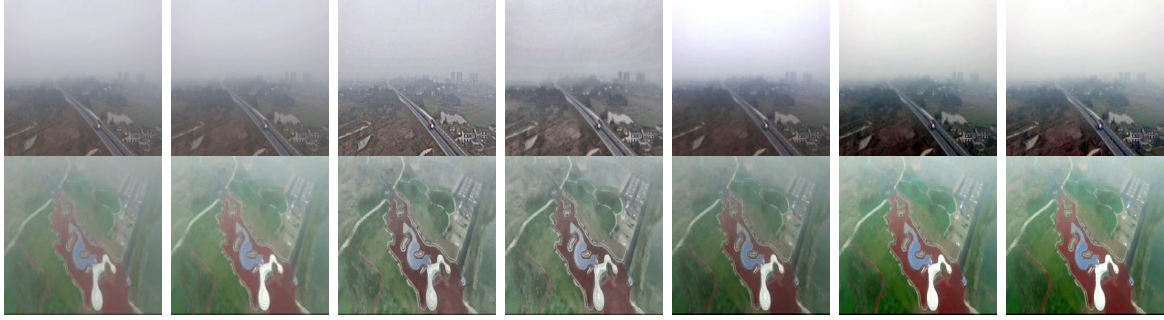


Fig 5. Visual comparison on the real aerial hazy images of HazyDet. From left to right: hazy input, (a) DCP, (b) CLAHE, (c) Dehaze-Net, (d) AOD-Net, (e) FFA-Net, (f) Ours

3.4.3. Runtime Comparison Results and Analysis

In order to further verify the processing efficiency of the proposed algorithm, the average dehazing time of a single image of each algorithm is counted on the SOTS test set, and the results are shown in Table 3.

Table 3. Runtime comparison of different methods

Model	Average dehazing time/s
DCP	1.346
CLAHE	1.028
Dehaze-Net	0.857
AOD-Net	0.356
FFA-Net	0.297
Ours	0.216

It can be seen from Table 3 that the average dehazing time of the proposed method is 0.216 s, which is the shortest among the above comparison methods. The processing speed is fast, and it has the potential of real-time processing on resource-constrained platforms. In summary, the algorithm proposed in this paper is superior to other comparison algorithms in terms of the thoroughness of dehazing, the integrity of the restored scene information and the processing speed.

4. Conclusion

This paper proposes MFFAMA-Net, an adaptive dehazing network based on a multi-feature fusion attention mechanism, to address the problems of blurred details, insufficient utilization of multi-scale features, brightness shift and color distortion in aerial hazy image dehazing. The network obtains feature representations that take into account both local details and wide-area context through shallow and deep multi-scale feature extraction, alleviating the detail degradation of dense small targets. It completes channel recalibration and shallow-deep feature fusion without dimensionality reduction through an improved ECA-Net encoded by global average pooling and global max pooling in a dual-path parallel manner, overcoming the limitation of single pooling in characterizing the spatial difference of haze concentration. It predicts the pixel-wise and channel-wise Gamma correction amount in a residual form through the adaptive brightness correction module, and degenerates into an identity transformation at the early stage of training through zero initialization, which suppresses the brightness shift and color distortion of the restored result. Since there is no need to explicitly estimate the transmission and the atmospheric light, the method avoids the accumulation of stage-wise parameter estimation errors in model inversion, and improves the color fidelity and residual

haze suppression compared with existing methods. Experimental results show that on the RESIDE dataset, the proposed method outperforms the mainstream dehazing methods in objective metrics such as SSIM and PSNR, and maintains a good restoration effect on the real aerial hazy images of the HazyDet dataset. However, the restoration of severely degraded regions under dense haze conditions is still insufficient. In the future, we will study the haze-concentration-adaptive hierarchical dehazing strategy, and investigate the gain of dehazing for downstream object detection tasks.

References

- [1] Pandey, P., Gupta, R., & Goel, N. (2025). Comprehensive review of single image defogging techniques: enhancement, prior, and learning based approaches. *Artificial Intelligence Review*, 58(4), 116. <https://doi.org/10.1007/s10462-024-11034-4>.
- [2] He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), 2341–2353.
- [3] He, K., Sun, J., & Tang, X. (2013). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), 1397–1409.
- [4] Berman, D., Treibitz, T., & Avidan, S. (2016). Non-local image dehazing. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- [5] Soni, B., & Mathur, P. (2020). An improved image dehazing technique using CLAHE and guided filter. In *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 902–907). IEEE.
- [6] Stark, J. A. (2000). Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing*, 9(5), 889–896.
- [7] Kim, J. Y., Kim, L. S., & Hwang, S. H. (2001). An advanced contrast enhancement using partially overlapped sub-block histogram equalization. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(4), 475–484.
- [8] Liu, X., Liu, C., Lan, H., et al. (2020). Dehaze enhancement algorithm based on retinex theory for aerial images combined with dark channel. *Open Access Library Journal*, 7(4), 1–12.
- [9] Shi, Z. (2026). Uav path planning and obstacle avoidance based on deep learning and reinforcement learning. *Frontiers in Computing and Intelligent Systems*, 16(1), 178–184. <https://doi.org/10.54097/7g9fac16>.
- [10] Ren, W., Liu, S., Zhang, H., et al. (2016). Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*.

- [11] Cai, B., Xu, X., Jia, K., et al. (2016). DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11), 5187–5198.
- [12] Li, B., Peng, X., Wang, Z., et al. (2017). Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4770–4778).
- [13] Jiang, J. S., Huang, J. M., Yu, Y., et al. (2026). Traffic image defogging method under hazy environments based on frequency domain fusion and attention mechanism. *Transportation Research Record*, 2680(7), 104–116. <https://doi.org/10.1177/03611981251387597>.
- [14] Qin, X., Wang, Z., Bai, Y., et al. (2020). FFA-Net: Feature fusion attention network for single image dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(7), 11908–11915. <https://doi.org/10.1609/aaai.v34i07.6865>.
- [15] Wang, Q., Wu, B., Zhu, P., et al. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11534–11542).
- [16] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7132–7141).
- [17] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3–19).
- [18] Xu, S., Chen, Q., Peng, L., et al. (2026). Towards fast and effective low-light image enhancement via adaptive Gamma correction and detail refinement. *Journal of Visual Communication and Image Representation*, 116, 104724. <https://doi.org/10.1016/j.jvcir.2026.104724>.
- [19] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.
- [20] Li, B., Ren, W., Fu, D., et al. (2018). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1), 492–505.
- [21] Feng, C., Chen, Z., Li, X., et al. (2024). HazyDet: Open-source benchmark for drone-view object detection with depth-cues in hazy scenes. <https://doi.org/10.1109/TGRS.2026.3701819>.