Journal of Artificial Intelligence Practice (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2025.080317
ISSN 2371-8412 Vol. 8 Num. 3

Research on Underground Non-uniform Fog Removal Method Based on Enhanced Parallel Attention Mechanism

Huan Zhang, Lingfei Cheng*, Kui Tang

School of Physics and Electronic Information, Henan Polytechnic University, Jiaozuo, Henan, 454003, China

Keywords: Downhole Image, Image Dehazing, U-Net, Deep Learning

Abstract: The image quality in the underground environment is limited by insufficient lighting and the interference of non-uniform dust and mist generated by work activities. This non-uniform fog results in low image visibility, blurry details, and color distortion, which hinders underground safety monitoring. For this purpose, a model was designed for the removal of non-uniform fog underground. Firstly, the module includes multi-scale convolution and parallel attention mechanism. Multi scale convolution can obtain more feature information from images in order to restore texture information. Parallel attention can better capture multi-dimensional global information, improve the comprehensiveness of feature extraction, and perform well in removing non-uniform fog. In addition, the SE attention module is introduced to automatically learn the sensitivity of different channels to fog concentration, with high weights for dense fog areas, enhancing the dehazing effect. Finally, the PSNR and SSIM of the Haze4K dataset were verified to be 32.18 and 0.963, respectively. The validation indicators for the self-made non-uniform fog dataset are PSNR of 32.37dB and SSIM of 0.981. This provides a certain reference value for obtaining high-quality images for underground monitoring.

1. Introduction

The clarity of images and videos collected underground depends on the shooting equipment and collection environment. However, the underground environment is complex, and underground mining operations can generate non-uniform dust and mist. This non-uniform dust and mist cause severe absorption and scattering effects on light, resulting in loss of image details and contrast, as well as severe color distortion in the collected images. This problem leads to a sharp decline in the accuracy and robustness of subsequent advanced visual tasks, such as object detection, device state recognition, obstacle avoidance, and other algorithms. Therefore, researching non-uniform fog and dust image enhancement and restoration techniques for special underground environments has crucial theoretical value and practical significance for improving underground safety monitoring levels and ensuring production efficiency.

Most methods based on image restoration use the Atmospheric Scattering Model (ASM) [1] [2], which projects images and atmospheric light to achieve dehazing effects on the images. The Dark

Channel Prior (DCP) dehazing algorithm proposed by He et al. [3] uses an atmospheric scattering model to quickly estimate atmospheric light and transmittance. When the brightness is sufficient, it can achieve good restoration results. However, due to the use of the minimum value in the image, this method may cause certain color distortion. The underground image dehazing algorithm proposed by Wang et al. [4] is based on an adaptive dual channel prior algorithm for dehazing underground coal mine images, which can effectively remove dust and haze from the images, improve the recognizability of detailed information in the images, and greatly reduce computation time. Cao et al. [5] introduced a boundary constrained dehazing algorithm for underground coal mine images, which combines the DCP algorithm with boundary constraints and context regularization methods to achieve the goal of dehazing underground coal mine images. Overall, traditional dehazing methods can still achieve good dehazing effects, but due to their relative simplicity, there are certain shortcomings in accuracy.

Based on deep learning methods, the difference between foggy and non foggy images can be directly generated through ASM or learning. Li et al. [6] modified the formula based on ASM, combined the transmission map with atmospheric light to form a single parameter, and constructed an end-to-end integrated dehazing network (AOD-Net) to achieve fast parameter estimation. However, there may be phenomena such as excessive smoothness or insufficient removal in object edges or complex texture areas. The underground image dehazing algorithm proposed by Wang et al. [7] is based on a haze prior residual perception learning dehazing framework, which has significant advantages in ensuring the naturalness of edge structure and color, but the computational complexity is relatively high. Cai et al. [8] designed a network from start to finish to estimate the haze map transmission map, and then restored the haze free map based on ASM. However, the removal effect on outdoor defogging environments is not ideal. The improved downhole fog algorithm proposed by Huang et al. [9] can greatly improve processing speed, but its color retention ability is poor in strong fog areas. Li et al. [10] proposed a dual branch fusion network for enhancing edge feature extraction, which can be effectively applied to real underground dehazing tasks. However, its performance under extreme lighting conditions still needs improvement. Qin et al. [11] proposed a Feature Fusion Attention Network (FFA-Net) for underground fog, which utilizes feature attention mechanism to enhance the expression of Convolutional Neural Network (CNN). However, the resulting image often lacks clear details and boundaries. Chen et al. [12] proposed the Global Context Aggregation Network (GCA-Net), which introduces smooth expansion implementation in the network structure to avoid grid artifacts, and uses gate controlled sub networks to fuse multi-scale features together, improving the dehazing effect of underground images. Li et al. [13] designed an algorithm that combines global residual attention and gating features, which can effectively solve the problem of restoring dense fog areas underground, but has poor robustness to motion blur and noise.

The algorithm in this article utilizes multi-scale convolution to extract image feature information, and utilizes parallel global grouping coordinate attention module, Triplet Attention module, and enhanced channel pixel attention module to capture the interaction of cross dimensions through mixed attention and calculate weights. It has a huge effect on dehazing processing in special underground environments. The cross dimensional interaction between the spatial dimension and channel dimension of the Triplet Attention module enhances the performance of the dehazing network. The global grouping coordinate attention module aims to capture the spatial global information of the feature map and generate an attention map to weight the features, thereby enhancing their feature expression ability. The introduction of SE module focuses on improving the network representation ability, prompting the model to focus more on the dehazing task itself.

2. Research design

2.1 Dehazing network model structure

The complexity of the underground environment is high, the lighting conditions are weak, and the dust and fog in the underground working environment make the obtained images blurry. Therefore, a dehazing model was designed. The network structure is based on UNet as the backbone, and the multi-scale convolution module can effectively expand the receptive field and obtain rich information from underground images. Large convolutional kernels focus on areas with significant dense fog, while small convolutional kernels focus on detailed features and restore edge texture information. Designing parallel attention can enable the network to focus on detailed feature maps.In addition, the SE attention module is introduced to enhance the model's attention to dense fog areas, making the model more effective in removing non-uniform fog. The network model is shown in Figure 1.

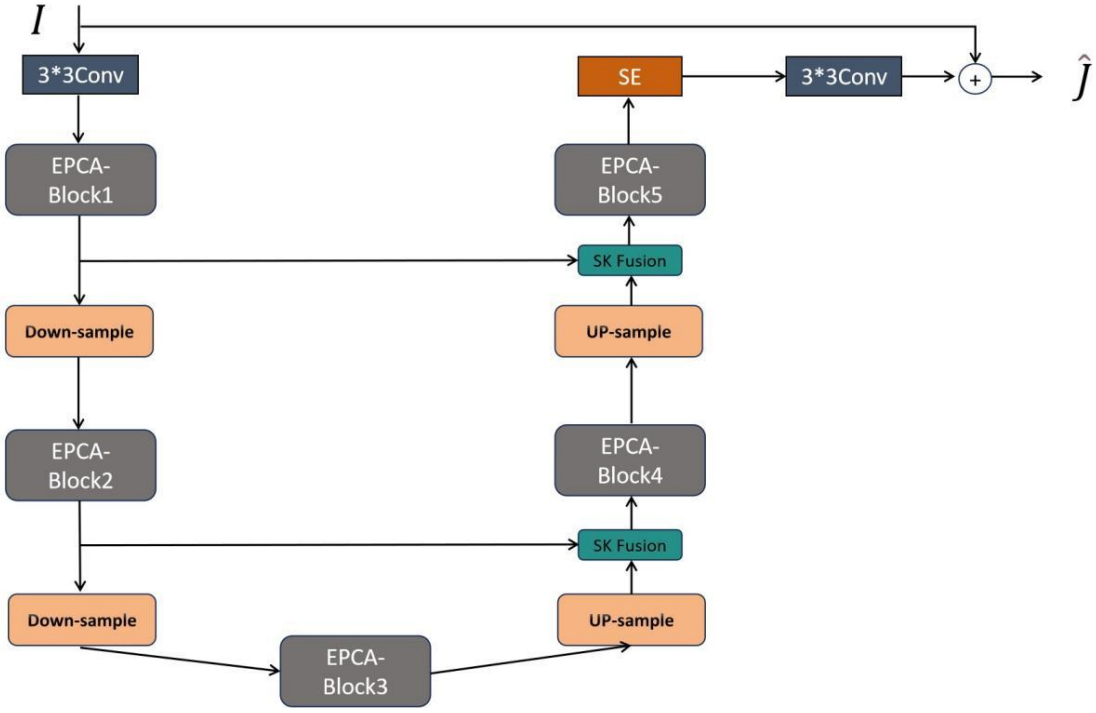


Figure 1 Dehazing network flowchart

2.2 EPCA module

This module consists of a multi-scale convolution module and a parallel attention module. Firstly, let x be the original feature map and normalize it using BatchNorm, let $\hat{x} = \mathrm{BatchNorm}(x)$. BatchNorm can accelerate network convergence, improve generalization ability, and prevent overfitting. As shown in formulas 1-5 below.

$$x1 = \mathrm{PWConv}(\hat{x}) \quad (1)$$

$$x2 = \mathrm{Conv}(x1) \quad (2)$$

$$x3 = \mathrm{Concat}(\mathrm{DWDConv13}(x2)) \quad (3)$$

$$\mathrm{DWDCov9}(x2) \quad (4)$$

$$\mathrm{DWDCov5}(x2)) \quad (5)$$

Here, PWConv refers to point wise convolution. Conv refers to a convolution with a kernel size of 5. DWDConv13 refers to a 7 × 7 depth dilated convolution with an dilated kernel size of 13 and an dilation rate of 2. DWDConv9 refers to a 5 × 5 depth dilated convolution with a kernel size of 9 and an dilation rate of 2. DWDConv5 represents a deep dilated convolution with a kernel size of 5 and a dilation rate of 2, which is 3 × 3. Finally, Concat represents the concatenated features in the channel dimension. The large convolution kernel focuses on the dense fog areas in the image, while the small convolution kernel restores the detailed texture features of the image. As shown in formula (6).

$$y = x + PW\ Conv(GELU(PW\ Conv(x_3))) \quad (6)$$

The enhanced parallel convolutional attention module combines different types of attention mechanisms. It includes a global grouping coordinate attention, an enhanced pixel attention, and triplet attention. Triplet attention does not use any dimensionality reduction operations, but captures cross dimensional interactions through three parallel branches, utilizing C-W, C-H, and H-W to capture the dependencies between channel spatial dimensions, thereby more accurately estimating fog concentration, restoring obscured details, and enhancing image contrast. The core idea of enhanced pixel attention is to regress to the pixels themselves and perform extremely refined analysis and enhancement on a pixel by pixel basis. Capable of finely processing non-uniform haze, complex textures, and lighting changes in underground images. The Global Grouping Coordinate Attention Module (GGCA) generates attention maps by extracting global information of feature maps in spatial dimensions (height and width). This module utilizes the generated attention map to weight the input features, effectively enhancing the model's representational ability. The EPCA module structure diagram is shown in Figure 2.

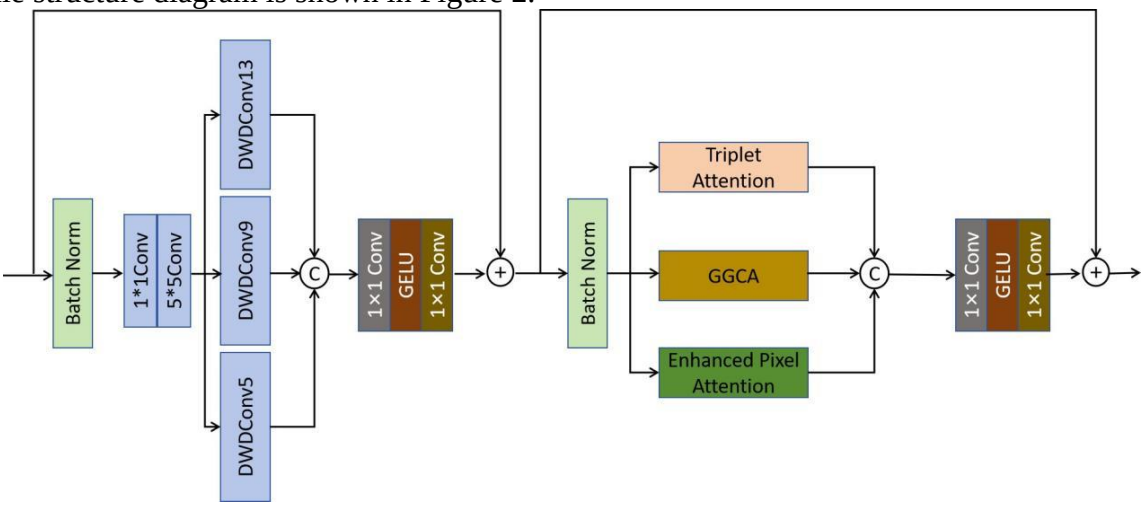


Figure 2 EPCA module structure diagram

2.3 GGCA module

The core idea of the global grouping coordinate attention module is to construct an attention map by capturing the long-range spatial dependencies of feature maps. As shown in formula (7).

$$X \in \mathbb{R}^{B\times C\times H\times W} \quad (7)$$

Firstly, based on the total number of channels C, we evenly divide all channels into G independent groups. Subsequently, allocate C/G channels to each group for further processing. The feature maps after grouping are represented by formula (8).

$$X \in \mathbb{R}^{B\times G\times \frac{C}{G}\times H\times W} \quad (8)$$

Subsequently, we perform pooling on the grouped feature maps along both the height and width directions. Specifically, we applied both global average pooling and global maximum pooling operations simultaneously. As shown in formula (9).

$$\begin{aligned} X_{h,avg} &= AvgPool(X) \in \mathbb{R}^{B\times G\times \frac{C}{G}\times H\times 1} \\ X_{h,max} &= MaxPool(X) \in \mathbb{R}^{B\times G\times \frac{C}{G}\times H\times 1} \\ X_{w,avg} &= AvgPool(X) \in \mathbb{R}^{B\times G\times \frac{C}{G}\times 1\times W} \\ X_{w,max} &= MaxPool(X) \in \mathbb{R}^{B\times G\times \frac{C}{G}\times 1\times W} \end{aligned} \quad (9)$$

For each grouped feature map, we apply shared convolutional layers for feature processing. This shared convolutional layer consists of two 1x1 convolutional layers, a batch normalization layer, and a ReLU activation function, used to reduce and restore channel dimensions. As shown in formulas (10) and (11).

$$Y_{h,avg} = Conv(X_{h,avg}), \quad Y_{h,max} = Conv(X_{h,max}) \quad (10)$$

$$Y_{w,avg} = Conv(X_{w,avg}), \quad Y_{w,max} = Conv(X_{w,max}) \quad (11)$$

By adding the outputs of the convolutional layers and applying the Sigmoid activation function, attention weights are generated for the height and width directions. As shown in formulas (12) and (13).

$$A_h = \sigma(Y_{h,avg} + Y_{h,max}) \in \mathbb{R}^{B\times G\times \frac{C}{G}\times H\times 1} \quad (12)$$

$$A_w = \sigma(Y_{w,avg} + Y_{w,max}) \in \mathbb{R}^{B\times G\times \frac{C}{G}\times 1\times W} \quad (13)$$

Among them, σ represents the Sigmoid activation function.

Finally, the input feature map is multiplied element by element with the obtained attention weights to achieve feature reweighting and generate the output feature map. As shown in formula (14).

$$O = X \times A_h \times A_w \in \mathbb{R}^{B\times C\times H\times W} \quad (14)$$

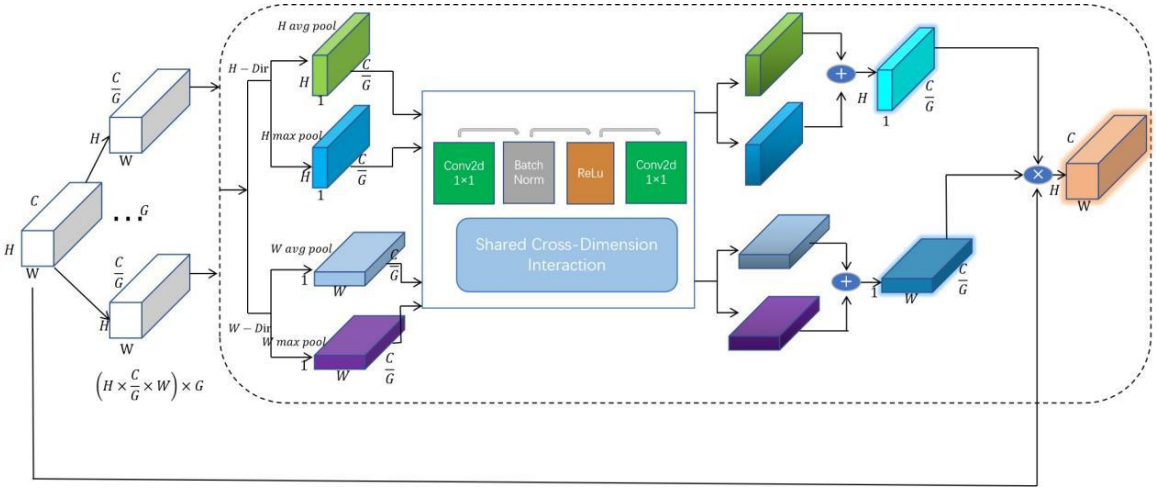


Figure 3 GGCA module structure diagram

GGCA achieves efficient fusion of local details and global contextual information at the channel grouping level by grouping channels and applying global pooling and coordinate attention separately. It can accurately perceive the distribution of non-uniform haze underground, restore cross regional object structures, and enhance the representativeness of key channel features, thereby achieving natural dehazing effects. The structure diagram of GGCA module is shown in Figure 3.

2.4 Triplet Attention Module

The first branch is responsible for calculating channel attention: the input features first pass through the Z-Pool layer, then through a 7x7 convolutional layer, and finally generate channel attention weights using the Sigmoid activation function.

The other two branches: Channel C interacts with the space W and H dimensions to capture the input features. The input features are first permed, then Z-Pool is performed on the W or H dimensions, followed by 7 × 7 convolution and batch Norm layers. Finally, add the output features of the three branches and calculate Avg. The structural diagram is shown in Figure 4.

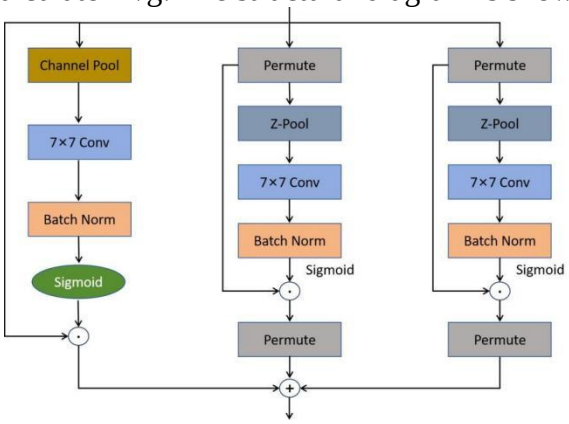


Figure 4 Triplet Attention Module Structure Diagram

2.5 SE module

The SE module divides the image dehazing process into three steps: compression, excitation, and reweighting. Firstly, the spatial information (i.e. height and width) of each channel is compressed into a single value to obtain a feature vector that contains global information for all channels. The dimension of this vector is 1 × 1 × C, where C is the number of channels. The input-output definition is shown in formula (15).

$$F_{tr}: X \to U, X \in R^{W' * H' * C'}, U \in R^{W * H * W} \quad (15)$$

The calculation formula is a conventional convolution operation, as shown in (16).

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * x^s \quad (16)$$

Among them, v_c represents the c-th convolution kernel, x^s represents the s-th input covered by the current convolution kernel, and C' represents the number of convolution kernels. This operation results in the second matrix in the figure, with dimensions of [H, W, C].

Next, perform the global average pooling operation using the formula shown in (17).

$$z_c = F_{sq}(u_c) = \frac{1}{W * H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i, j) \quad (17)$$

Then perform Excitation operation, and the calculation formula is shown in (18).

$$s = \text{sigmoid}(W_2 * \text{Relu}(W_1 z)) \quad (18)$$

Among them, z represents the previous step's z, W_1, and W_2 represent linear layers. The s calculated here is the core of the module, used to represent the weights of each channel, and this weight is learned through the fully connected and nonlinear layers mentioned earlier.

The final Fscale operation is calculated using the formula shown in (19).

$$\tilde{x} = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (19)$$

u_c represents a channel in u, and s_c represents the channel weight. Therefore, it is equivalent to multiplying the value of each channel by its weight.

As shown in Figure 5, the SE module enhances the critical channels for dehazing tasks and suppresses irrelevant information by adaptively recalibrating channel feature responses, thereby significantly improving the overall performance of the underground image dehazing network.

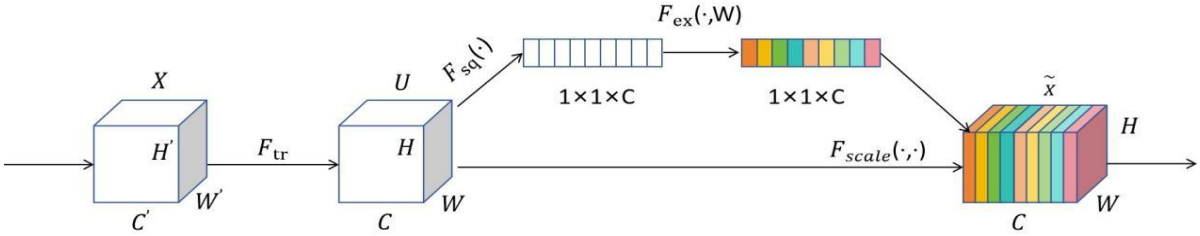


Figure 5 SE module structure diagram

3. Experimental configuration and result analysis

3.1 Experimental dataset

This article uses the publicly available dataset Haze-4K and a self-made non-uniform fog underground dataset for experimental verification.

Haze-4K is a high-resolution dataset specifically designed for image deblurring research. It consists of 4000 pairs of blurred images with real blurring effects and corresponding clear images after deblurring, and is divided into training and testing sets in a 3:1 ratio. Including various types of images of cities, nature, and indoors, as well as varying degrees of haze effects.

Given the current lack of publicly available underground coal mine image datasets, a self-made underground image non-uniform fog dataset is needed. This dataset is constructed by extensively collecting images and monitoring videos of various typical scenarios such as underground work faces, conveyor belts, and tunnels, ensuring its diversity and representativeness. The dataset consists of 5000 images and is divided into training and testing sets in a 4:1 ratio. Figure 6 shows some of the dataset images.

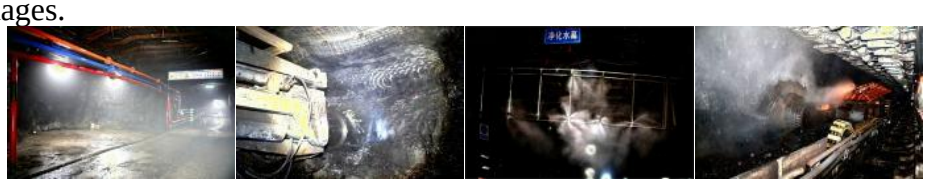


Figure 6 Partial non-uniform fog images in the dataset

3.2 Experimental Details

The system used in this experiment is the Windows 11 system. CPU: Gen Intel(R) Core(TM) i7-13620H 2.40 GHz. GPU: NVIDIA GeForce RTX4060. This model is developed using the Python programming language and the deep learning framework PyTorch.

This experiment uses the AdamW optimizer to train the dehazing model, with momentum parameters β 1 and β 2 set to 0.9 and 0.999, respectively, and an initial learning rate of 2×10^{-4}. The learning rate decays from the initial value to 10^{-6} through cosine annealing strategy, which helps guide the model to explore the local optimal solution of the loss function, thereby improving convergence speed and generalization performance. Set the hyperparameter β to 0.1.

3.3 Evaluation indicators

To objectively and fairly evaluate the dehazing performance of various algorithms, this study jointly uses two objective indicators, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM), supplemented by subjective visual analysis for comprehensive evaluation. PSNR quantifies the reconstruction quality by calculating the mean square error between the original image and the dehazed image, with higher values indicating less distortion. By combining subjective evaluation, the removal effect of the model on underground non-uniform haze can be comprehensively evaluated.

3.4 Ablation Experiment

In order to verify the effectiveness of the model in removing non-uniform fog underground, an ablation study was designed. In this study, Model 1 represents the basic model; Model 2 represents a model that uses only the designed GGCA module on top of the base model; Model 3 represents a model that uses only the Triplet module on top of the base model; Model 4 represents the use of only the SE module on the base model; Model 5 represents a model that uses the GGCA+Triplet module on top of the base model; Model 6 represents a model that uses the GGCA+SE module on top of the base model; Model 7 represents a model that uses the Triplet+SE module on top of the base model. The objective indicators of non-uniform fog removal by the model are shown in the figure.

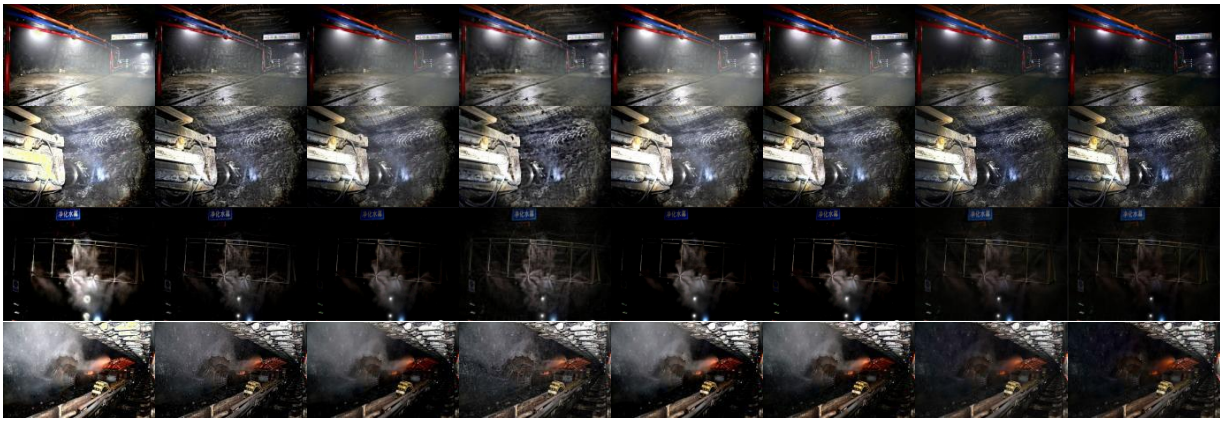
This study used two indicators, PSNR and SSIM, to quantitatively evaluate the proposed algorithm. The ablation experiment data in Table 1 clearly shows that the comprehensive performance of the model has been substantially improved.

Table 1 Results of ablation experiment

Model Number	foundation model	GGCA	Triplet	SE	PSNR/dB	SSIM
1	√	×	×	×	30.56	0.949
2	√	√	×	×	30.93	0.963
3	√	×	√	×	30.70	0.956
4	√	×	×	√	31.16	0.971
5	√	√	√	×	30.99	0.964
6	√	√	×	√	30.48	0.943
7	√	×	√	√	31.68	0.978
8	√	√	√	√	32.37	0.981

3.5 Comparison with existing dehazing methods

Comparing this model with existing dehazing methods, experimental results were compared and analyzed with DCP, AOD-Net, DehazeNet, GCA-Net, and FFA-Net algorithms under the same configuration. This study extracted 4 dehazing images from the underground dataset for subjective visual evaluation, as shown in Figure 6.



Haze DCP AOD-Net Dehazenet GCA-Net FFA-Net Our GT

Figure 7 Comparison of dehazing effects of various models

From Figure 7, it can be seen that the use of DCP for image dehazing is effective but accompanied by color distortion. After being processed by AOD-Net network, there is still haze in the image and the details and texture information are lost. Dehazenet has a certain effect on removing fog, but the edges of the fog are blurred and the details are distorted. The effectiveness of GCA-Net and FFA-Net networks in removing non-uniform fog underground needs to be improved. In contrast, the model proposed in this article outperforms other algorithms in removing non-uniform fog images underground. While ensuring effective restoration of clear images, it preserves the color and texture information of objects, making it visually closer to clear images.

In order to objectively evaluate the algorithm performance of the model, tests were conducted on the public dataset Haze4K and the underground non-uniform fog dataset under the same configuration. Use PSNR and SSIM as evaluation criteria. Both indicators of the algorithm in this article are slightly better than other algorithms, indicating that the model has good performance and practical significance in removing non-uniform fog underground. The results are shown in Table 2.

Table 2 Model comparison evaluation indicators

Algorithm model	Haze4K dataset		Underground non-uniform fog dataset	
	PSNR/dB	SSIM	PSNR/dB	SSIM
DCP	19.52	0.858	19.30	0.782
DehazeNet	19.60	0.870	19.95	0.777
AOD-Net	17.97	0.733	19.08	0.784
GCANet	24.47	0.919	27.82	0.881
FFA-Net	24.26	0.928	28.26	0.926
Our	32.18	0.963	32.37	0.981

4. Conclusion

The introduction of GGCA module greatly enhances the network's understanding and modeling ability for complex non-uniform fog distributions, achieving precise preservation of image detail information. The Triplet Attention module captures cross latitude interaction dependencies in a lightweight and efficient manner through three parallel branches, accurately modeling the distribution of fog, dust, and light in complex underground environments. In addition, the SE attention module enhances the dehazing effect by giving high weight to dense fog areas through sensing them. The experimental results show that this model has a certain effect on removing non-uniform fog underground, providing a reference for obtaining high-quality underground images.

References

[1] McCartney, E.J. Optics of the Atmosphere: Scattering by Molecules and Particles[M]. New York: Wiley, 1976.

[2] Nayar, S.K.; Narasimhan, S.G. Vision in bad weather[C]. In Proceedings of the Seventh IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999: 820–827.

[3] He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior[C]. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 1956–1963.

[4] Wang, Y.; Wei, S.; Duan, Y.; Wu, H. Defogging algorithm of underground coal mine image based on adaptive dual-channel prior[J]. Journal of Mine Automation, 2022, 48: 46–51+84.

[5] Cao, H.; Yao, S.; Wang, Z. Defogging algorithm of underground coal mine dust and fog image based on boundary constraint[J]. Journal of Mine Automation, 2022, 48, 139–146.

[6] Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. AOD-Net: All-in-One Dehazing Network[C]. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy: IEEE, 2017: 4780–4788.

[7] Wang, K., Liu, Y., Yang, Y., Zhang, G., & Qian, W. Single Image Dehazing Based on Haze Prior Residual Perception Learning[J]. Circuits, Systems, and Signal Processing, 2025. doi: 10.1007/s00034-025-03058-0.

[8] Cai, B.; Xu, X.; Jia, K.; Qing, C. Tao, D. Dehazenet: An end-to-end system for single image haze removal[J]. IEEE Transactions on Image Processing, 2016, 25(11), 5187–5198.

[9] Huang, H., Ouyang, H., Dong, Y., He, X., & Zhao, X. Fast dehazing for large format oblique images based on improved dark channel prior[J].Optics and Precision Engineering, 2025, 33(3): 476–485. doi: 10.37188/OPE. 20253303.0476.

[10] Li, X., Xia, F., Zhang, K., Wang, H., & Xie, T. Enhanced edge feature extraction dual branch fusion network for real image dehazing[J]. Optics and Precision Engineering, 2025, 33(2), 247–261. doi: 10.37188/OPE.20253302.0247.

[11] Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34: 11908–11915.

[12] Chen, D.; He, M.; Fan, Q.; Liao, J.; Zhang, L.; Hou, D.; Yuan, L.; Hua, G. Gated Context Aggregation Network for Image Dehazing and Deraining[C]. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2019: 1375–1383.

[13] Li, H.-Y., Qiao, R.-C., Li, H.-J., & Chen, Q. CNN-Transformer Dehazing Algorithm Based on Global Residual Attention and Gated Feature Fusion[J]. Journal of Northeastern University, 2025, 46(1): 26–34. doi: 10.12068/j.issn. 1005-3026.2025.20239041.