ITC 1/54

Information Technology and Control
Vol. 54 / No. 1/ 2025
pp. 5-15
DOI 10.5755/j01.itc.54.1.38427

Image Enhancement Model for Open-Pit Mine Monitoring Based on Parallel Multi-Scale Feature Fusion

Received 2024/08/09 | Accepted after revision 2024/12/16

HOW TO CITE: Tao, X., Guo, S., Zhang, Z., Wu, X. (2025). Image Enhancement Model for Open-Pit Mine Monitoring Based on Parallel Multi-Scale Feature Fusion. *Information Technology and Control*, 54(1), 5-15. https://doi.org/10.5755/j01.itc.54.1.38427

Image Enhancement Model for Open-Pit Mine Monitoring Based on Parallel Multi-Scale Feature Fusion

Xuewei Tao

School of Safety Engineering, China University of Mining and Technology; Zhonglian Runshi Xinjiang Coal Industry Co., Ltd; China; e-mail: 125710054@qq.com

Shike Guo

Zhonglian Runshi Xinjiang Coal Industry Co., Ltd; China; e-mail: 284673993@qq.com

Ziheng Zhang

Zhonglian Runshi Xinjiang Coal Industry Co., Ltd; China; e-mail: 1021622377@qq.com

Xuchu Wu

School of Energy and Mining Engineering, China University of Mining & Technology-Beijing, China; e-mail: 1437699376@qq.com

Corresponding author: 1437699376@qq.com

The workspace in open-pit mining systems often suffers from insufficient or uneven illumination due to spatial constraints and obstructions caused by large equipment or geotechnical structures, leading to de-graded surveillance imagery and consequently impacting safety monitoring efforts. The quality of surveil-lance determines the real-time monitoring of personnel safety at night, such as during night-time opera-tions, remote maintenance of equipment, and various simple remote control tasks This study designed an open-pit mine surveillance image enhancement model based on a parallel multi-scale feature fusion Trans-former to address the degradation of surveillance video images and leverage the superior expressive power of Transformer networks in visual image processing compared to other networks. The network architecture mainly processes and integrates full-size feature maps and various levels of downsampled feature maps in parallel, preserving both the semantic relationships of image elements and their overall structure. The downsampling process of the network aims to maximize the extraction and restoration of the luminance features of small-sized objects from low-resolution images. By integrating features from downsampling, full-size image processing effectively restores illumination, thereby enhancing the accuracy of the images. To reduce the computational demands of the Transformer structure and facilitate its application in moni-toring imagery, we employed an orthogonal self-attention mechanism along both the rows and columns of the image to be processed. This mechanism shifts the network's computational demand from exponential to linear growth. During the training phase, the network model was trained using a dataset collected on-site to enhance the model's adaptability to field conditions. Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) test results confirm that this model performs exceptionally well in open-pit mining production systems.

KEYWORDS: Image Restoration; Open-Pit Mine Environmental Monitoring Image Restoration; Parallel Multi-Scale Attention Feature Fusion; Parallel Multi-Resolution Image Restoration.

1. Introduction

The surveillance videos of open-pit mines are significant for daily production work. For instance, they can provide a comprehensive grasp of the dynamics of the mining area, better coordinate various processes, make overall arrangements, and optimize production organization. They can also monitor the operation situation of the mining area in real-time and promptly discover and prevent potential safety hazards. However, due to the dust in the mining area and uneven illumination, the quality of surveillance video images often deteriorates. The main reasons include:

1. Explosion-proof standard restrictions: Due to the explosion risk in open-pit mines, the design of explosion-proof equipment limits the image acquisition performance, resulting in frequent occurrences of many noise points and low image resolution.
2. Blockage by equipment and buildings: Large equipment and buildings in open-pit mines (such as crushers and silos) block light during production, leading to insufficient and uneven illumination of the images captured by cameras.
3. Severe dust in the operation site: Whether in the process of mining, shovel loading, or external transportation and loading, the transfer and transportation of materials will generate a large amount of dust, seriously blocking the light and line of sight and affecting the clarity and visibility of the images.

The above factors jointly lead to severe degradation of surveillance images in open-pit mines, affecting dispatchers' production judgment and work safety.

Therefore, developing high-performance surveillance solutions and advanced image restoration technologies to improve image quality is crucial for enhancing production efficiency and ensuring safety.

In the current mining industry, most image enhancement methods are based on improvements in the Retinex approach. These enhancement techniques primarily consider the relationships between different lighting conditions, without addressing the enhancement of different image features. As a result, the enhanced images often exhibit a noticeable sense of artificiality, and artifacts or noise frequently appear in areas with significant illumination changes.

Tian et al. [20] introduced an improved Retinex fusion algorithm to address the issues of halo enlargement and overexposure in traditional image enhancement methods. By combining homomorphic filtering, bilateral filtering, and Multi-scale Retinex with Color Restoration (MSRCR), the method enhanced the luminance while preserving the hue and applied adaptive nonlinear stretching to the saturation. The method proposed by Tian et al. [21] enhances coal mine low-light images through a Transformer-based model and adaptive feature fusion. The innovation includes a generative adversarial framework for adaptive image enhancement, decoupling of illuminance and reflection components to prevent color distortion, and the use of a CEM-Transformer Encoder for improved brightness and reduced local unevenness. Additionally, the method integrates a skip connection with CEM-Cross-Transformer Encoder to preserve fine details and employs ECA-Net for more efficient feature extraction. Experimental results show that this approach significantly improves image quality in both objective metrics and subjective visual evaluation, outperforming existing low-light enhancement methods.

The study by Kong et al. [4] focused on improving the image quality and detection accuracy of safety monitoring for personnel in coal mines. The research team found that due to insufficient underground illumination, traditional surveillance images often suffer from recognition difficulties and low detection accuracy. To address this, they proposed an innovative low-light image enhancement method. This method employs an adaptive enhancement strategy, including a local enhancement module and a Transformer-based global adjustment module. The local enhancement module maps low-light areas to normal illumination at the pixel level while preserving as much detail as possible. The global adjustment module is used to prevent over-enhancement of bright areas and insufficient illumination in dark areas, while also avoiding color bias. Furthermore, to ensure that image enhancement does not negatively impact personnel detection, the research team introduced a feature similarity loss to maintain the consistency of target features.

2. Image Enhancement Algorithms

Image enhancement is an important image processing technique aimed at improving the visual effect of an image and making it more suitable for specific applications or more in line with human visual preferences. This technique can be applied to reduce noise and enhance multiple aspects, such as an image's contrast, brightness, details, and colors.

With the development of technology, related processing methods have evolved from finding mapping models for image illumination restoration to establishing self-learning neural networks.

1 Histogram Equalization

Histogram equalization adjusts the image brightness distribution and automatically enhances the contrast, suitable for extremely dark or bright backgrounds and foregrounds. Contrast Limited Adaptive Histogram Equalization (CLAHE) is its improved version [14] which optimizes local detail preservation and reduces over-enhancement problems by independently applying equalization to image blocks. Although CLAHE has a significant effect in adjusting local brightness, it may still lead to color distortion and increased noise.

2 Retinex Algorithm

The Retinex theory is based on the human visual system and simulates the adaptability of the human eye to light to improve image brightness [6], contrast, and color. Among them, Single-Scale Retinex (SSR) and Multi-Scale Retinex (MSR) algorithms are particularly suitable for low-light enhancement and color restoration [20]. Although Retinex has progressed in color fidelity, its detail recovery and application breadth under complex illumination are still limited, especially in complex illumination environments such as open-pit mines.

3 CNN**s** and GAN

Convolutional Neural Networks (CNNs) have made significant progress in image enhancement. By learning a large amount of data, CNNs can optimize the visual effect of images [12]. Generative Adversarial Networks (GANs) [2, 19] can generate high-quality images for tasks such as low-light image enhancement, High Dynamic Range (HDR) image generation, and color enhancement. These methods, from basic image processing techniques to advanced machine learning algorithms, are constantly evolving to meet the growing demand for image quality and visual effects. In the application process of CNNs, the performance of image enhancement has been greatly improved, and there have also been obvious improvements in generalization and other aspects. However, because the structure is still relatively simple, its expressive ability is still limited even if a very deep network is established. It will encounter problems such as losing many details due to the too-deep structure, and the enhancement effect is significantly reduced.

4 Transformer Model

Thanks to the attention structure in the network, the network of the Transformer model can obtain the relationship between all features and solve the problem of feature forgetting caused by the too-deep network structure, thereby making its expressive ability unprecedentedly improved and its generalization ability qualitatively transformed [22, 9, 23, 11, 1]. The emergence of the Transformer structure has made large models the basis of intelligence in various industries today. Still, it poses an unprecedented challenge to the storage and computing power of the GPU of the computer.

3. Analysis of Current Problems

When open-pit mine monitoring images degrade, multiple problems coexist, such as the simultaneous presence of low resolution, uneven illumination, insufficient illumination, and many noise points. Existing methods often focus only on a single problem and have obvious limitations in the open-pit mine environment. Although the algorithms mentioned above have been developed systematically and have a broad application space, there are still limitations in the processing of open-pit mine monitoring images.

For example, histogram equalization or adaptive histogram equalization relies on the initial quality of the image to improve contrast and brightness. If the monitoring equipment has performance limitations due to explosion-proof requirements, it may lead to poor original image quality, such as many noise points and low resolution, which will affect the effect of these methods.

When encountering physical occlusion and uneven light, methods such as the Retinex Algorithm, which relies on the internal illumination consistency of the image, face a challenge because they need to evaluate the illumination conditions of the entire image to adjust the brightness and contrast.

Although deep learning methods have potential in low-light image enhancement, they require a large amount of high-quality, well-labeled training data to effectively learn how to process images under extremely low-lighting conditions. In practical applications, such data may be difficult to obtain.

4. Model Structure Design

The minimum processing unit of the model includes convolution filtering enhancement, a transformer with feature orthogonal enhancement, and a SKFF unit. The network structure consists of full-resolution feature enhancement and down-resolution feature enhancement.

The convolution filtering enhancement and the Transformer with feature orthogonal enhancement are mainly responsible for enhancing the feature brightness at this resolution. The SKFF unit is responsible for paying attention to the channel level and selectively fusing the full-resolution structure and the features processed by variable resolution. The stacking of the minimum processing units forms a residual context block, as shown in Figure 2. The stacking of the basic structure forms a Multi-scale Residual Block (MRB). The stacking of MRB and residuals constitutes a recursive residual group. The image is enhanced step by step through the recursive residual group to form a Transformer network with parallel multi-scale feature fusion, as shown in Figure 1.

MRB: The residual block mainly includes the following elements: (a) Parallel multi-resolution convolutional streams for extracting semantic richer fine-to-coarse and coarse-to-fine spatial accurate feature representations; (b) Information exchange between multi-resolution streams; (c) Aggregation of different stream features based on attention; and (d) Residual Contextual Block (RCB) for extracting attention-based features.

The overall processing process mainly includes inputting a three-channel image I, then applying a convolutional layer to extract the low-level feature map F0 and passing the feature map F0 through N Recursive Residual Group (RRG) to generate deep features Fn. Each RRG contains multiple MRB. Next, apply a convolutional layer to the deep feature Fn and obtain a residual image R. Finally, the restored image can be obtained by I + R. Use the Charbonnier loss to optimize the proposed network, and its formula is:

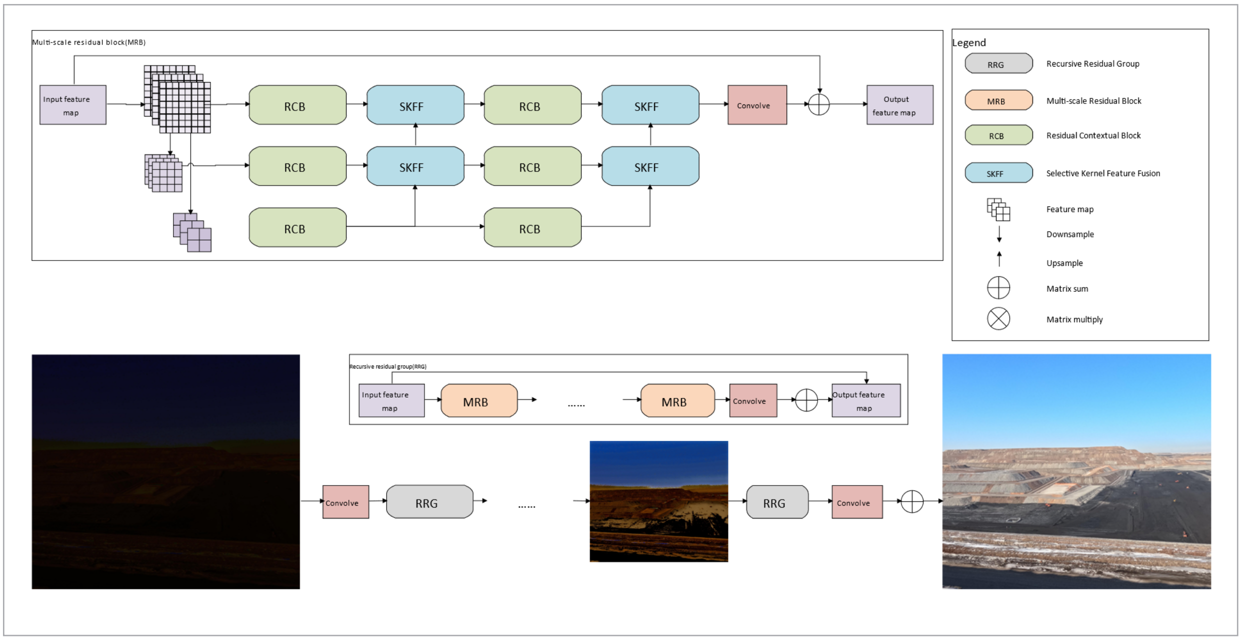
$$L(\hat{I}, I^*) = \sqrt{\left\| \hat{I} - I^* \right\|^2 + \varepsilon^2}, \tag{1}$$

where $\hat{\mathbf{I}}$ is the processed image, $\mathbf{I}^*$ is the supervised image, and ε is an empirical constant, which is set to 10^{-3} in all experiments.

Figure 1 shows the model structure diagram, whose backbone network is based on the recursive residual structure. The backbone network is used to preserve the full resolution of the processed pictures and enhance the detail retention ability of the network, while the branch network is used to decode the feature maps in the entire network through a kind of codec so that the full-resolution image is decomposed into low-resolution and high-feature image representations, enhancing the spatial expression ability of the processed image, so that the position information represented by the image will not deviate.

The RRG is an advanced structure used in deep learning, especially in the field of image processing. It uses the idea of recursion and residual learning to construct a deep network, aiming to improve the restoration and enhancement quality of images effectively.

Figure 1
Model structure



Its main structure includes the residual block and the recursive structure, with the residual block as the basic unit, which helps information across multiple layers in the network by introducing a skip connection. This design allows the input to be directly added to the output, thereby alleviating the problem of vanishing or exploding gradients in deep networks, enabling the network to go deeper without losing its learning ability. The recursive structure means that within the recursive residuals, the residual block or a group of residual blocks will be reused multiple times. This design not only reduces the number of model parameters (because the same block is used multiple times) but also increases the model depth and complexity without the need to increase the network width.

The design inspiration for the MRB in the main network of the model comes from the research on the human visual cortex in neuroscience. Because the local receptive field size of neurons in the same area of the visual cortex differs, a similar mechanism-collecting multi-scale spatial information in the same layer and integrating it into the CNNs has been proven more effective. Based on the above principles, the MRB is proposed to solve the problem of receiving rich contextual information from the low-resolution while maintaining the full-resolution representation, generating that is, while maintaining the full-resolution, it also allows the transmission of contextual information from the network structure in the low-resolution. The MRB is composed of multiple full convolution flows connected in parallel, and these flows operate on feature maps of different resolutions, ranging from low to high. This allows the transmission of contextual information from the low-resolution flow to consolidate the full-resolution features.

The RCB in the basic structure of the model is shown as following Figure 2.

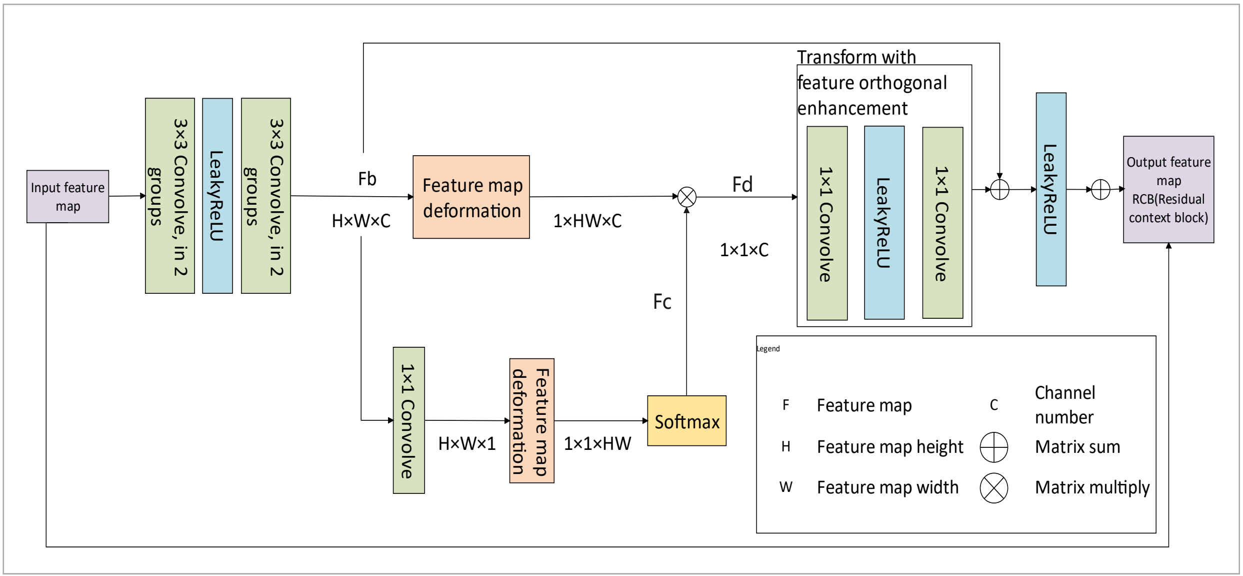
The main process of the RCB is represented by the following formula:

$$F_{RCB} = F_a + \mathrm{W}(\mathrm{CM}(F_b)). \quad (2)$$

Here, F_b is the input feature map. Firstly, a 1x1 convolution is applied, then the matrix is reshaped into a matrix of single pixels and single channels, and a softmax operation is performed to generate a new feature F_c. Next, the feature map F_b is reshaped as H×W×C and multiplied with F_c to obtain the global feature description feature F_d. To capture the interdependence between channels, the description feature F_d passes

Figure 2

Residual Contextual Block structure



through two 1x1 convolutions to generate a new attention feature F_e. Using an element-wise addition operation, the context feature is aggregated to each position of the original feature F_b.

The RCB enhances the network's processing ability for important information in the image by filtering and strengthening the features that are more helpful for the final task, making the feature representation more abundant and effective. This mechanism helps to improve the performance of image restoration and enhancement tasks, especially when dealing with complex scenes and details.

In the basic structure of the network, the structure of Schematic for Selective Kernel Feature Fusion (SKFF), which fuses the full-resolution features and the low-resolution features, is shown in the following Figure 3.

In the MRB, a nonlinear procedure is introduced to fuse the features from different resolution streams using the self-attention mechanism. This procedure is called SKFF. The SKFF module dynamically adjusts the receptive field through two operations, Fuse and Select, enabling the selective fusion of two feature maps of different sizes.

Figure 3

Selective Kernel Feature Fusion structure

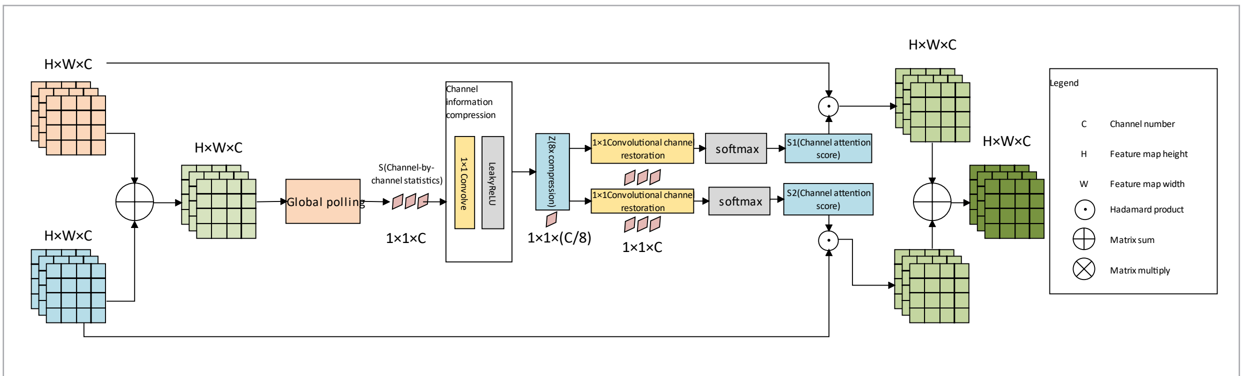
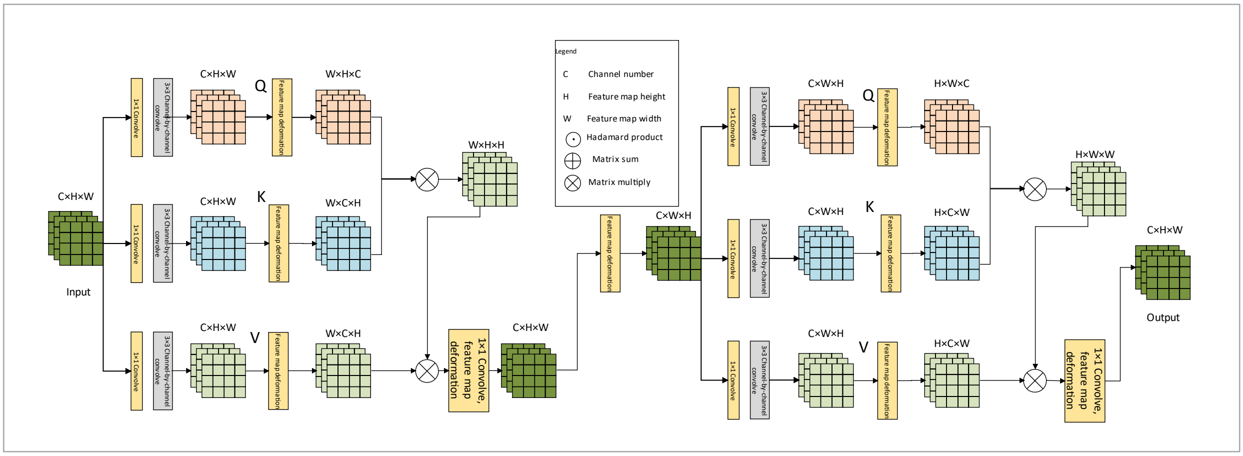


Figure 4

Transformer with feature orthogonal enhancement



Fuse: SKFF receives the input from two parallel convolution streams, which carry information of different scales. Firstly, the multi-scale features are merged using an element-wise summation: L = L1 + L2. Then, Global Average Pooling (GAP) is applied to the spatial dimension of L to calculate the channel-level statistics, which serve as important information for different channels. Next, a channel reduction convolutional layer is applied to generate a compact feature representation z. Finally, the feature vector z passes through two parallel channel enhancement convolutional layers (one for each resolution stream) and provides two feature descriptors, v1 and v2.

Select: This operation applies the softmax function to v1 and v2, generating attention activations s1 and s2. These activations are used to recalibrate the multi-scale feature maps L1 and L2 adaptively. The entire process of feature recalibration and aggregation is defined as U = s1 * L1 + s2 * L2.

The application of the feature orthogonal Transformer structure in the RCB is mainly divided into two parts: the high-axis self-attention structure and the wide-axis self-attention result. The result is to orthogonally decompose the attention of the full feature map into the attention of the two axes. The specific structure is as following Figure 4: The input feature map (C×H×W) is convolved into three matrices of Q, K, and V, respectively, and after deformation, the high-axis attention score is multiplied by the value, and in the same way, the wide-axis attention score and the final result are obtained.

5. Experimental Analysis and Discussion

When designing the deep learning network structure, considering the observed advantages in the experiments, an innovative method combining progressive learning and MRB was adopted to handle image restoration and enhancement tasks. The design of this structure aims to solve the challenges faced by traditional convolutional neural networks when dealing with image details and context information while optimizing training efficiency and model performance. To compare the algorithm effects, four indicators were selected for evaluation in the final results:

1 PSNR is a common indicator for measuring the quality of image reconstruction, especially in the field of image and video compression [7, 18]. It is calculated based on the Mean Square Error (MSE) between the original image and the processed image. The higher the PSNR value, the better the image quality. PSNR is calculated through the following formula:

$$PSNR = 20 * \log\left(\frac{MAX}{\sqrt{MSE}}\right). \quad (3)$$

Here, MAX is the maximum pixel value of the image, and MSE is the mean square error between the original image and the processed image. The formula is as follows:

$$MSE = \frac{\sum(I_a - I_b)^2}{N}. \quad (4)$$

2 Mean Absolute Erron (MAE) is another simple method for measuring image differences. It calculates the average absolute pixel differences between the original image and the processed image. The smaller the MAE value, the more similar the images are; that is, the better the image processing effect. Here, I_a and I_b bear the original image and the processed image, respectively, and N is the number of pixels. The formula is as follows:

$$MAE = \frac{\sum|I_a - I_b|}{N}. \quad (5)$$

3 SSIM [16,17] is a more advanced indicator for measuring image quality, which considers the image's luminance, contrast, and structural information. Unlike PSNR, SSIM attempts to evaluate image quality from the perspective of visual perception, making it closer to the evaluation of the human visual system. The value of SSIM is between 0 and 1, and 1 indicates that the two images are the same.

4 Learned Perceptual Image Patch Similarity (LPIPS) [3, 25] is a recently proposed indicator for evaluating image quality. It uses deep learning to measure the perceptual similarity between images. Unlike PSNR and SSIM, which are based on traditional mathematical formulas, LPIPS predicts the visual differences between images by training a deep network, which can more accurately reflect the differences in human visual perception. The lower the LPIPS value, the more similar the images are in perception and the closer the quality is.

The experiments selected two different scenarios, indoor and outdoor, namely the images obtained from the outdoor monitoring of the open-pit mine and the monitoring images obtained from the silo within the production system. The images obtained consist of photos taken at different times by a GoPro and a mobile phone. These photos were later adjusted by researchers using Adobe Lightroom to modify the brightness curves according to the required different illumination levels. Then, 20 volunteers evaluated and filtered the adjusted images. Finally, a dataset comprising both indoor and outdoor images was formed. The dataset was split into a training set and a validation set in a ratio of 0.8 to 0.2. The comparison diagram of the experimental results is shown in Table 1, and the calculation results of the evaluation

Table 1

Comparison of results

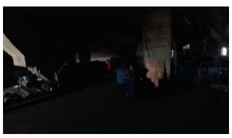
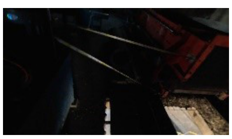
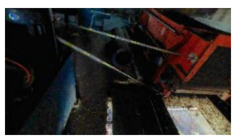
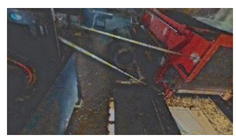
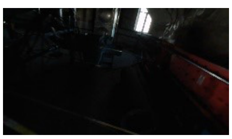
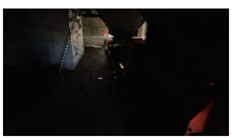
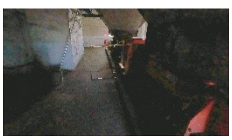
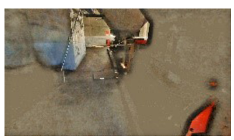
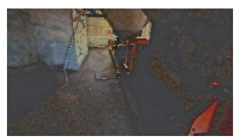
Original image	Algorithm proposed in this paper	DRBN [8]	DSLR [13]	EnlightenGAN [10]
				
				
				
				

Table 2
Index comparison table

Method	Indoor scene of open-pit mine				Outdoor scene of open-pit mine			
	PSNR	SSIM	MAE	LPIPS	PSNR	SSIM	MAE	LPIPS
Method in this paper	22.88	0.88	0.013	0.1234	28.58	0.91	0.010	0.0989
DRBN	16.65	0.82	0.244	0.2104	12.65	0.41	0.051	0.6154
DSLR	13.56	0.73	0.428	0.3512	24.71	0.83	0.018	0.4312
EnlightenGAN	21.12	0.84	0.084	0.1111	25.12	0.81	0.124	0.2121
EBDB [24]	17.23	0.66	0.224	0.2931	25.23	0.82	0.023	0.1591
PDPD [15]	18.34	0.72	0.135	0.2145	22.34	0.76	0.025	0.3243
JNB [5]	14.53	0.66	0.623	0.3416	13.53	0.56	0.422	0.8892

indicators are shown in Table 2. The results in the table are the average values of the calculation indicators of multiple images within the scenarios.

For the training method, we chose to further train different methods based on the original network. The loss function used for training is Smooth L1, and the training was conducted for 300k epochs. The learning rate scheduler chosen was cosine annealing, with a minimum learning rate of 1e-6 and a maximum learning rate of 1e-3.

For image evaluation, a corresponding third-party library was selected, and the default parameters were used to evaluate four metrics: PSNR, SSIM, MAE, and LPIPS. The images evaluated were from the test set portion of the captured dataset.

By comparing the performance of different methods in the indoor and outdoor scenarios of the open-pit mine (using the four evaluation indicators of PSNR, SSIM, MAE, and LPIPS), it is concluded that the method proposed in the paper consistently outperforms other methods in both indoor and outdoor scenarios of the open-pit mine. Particularly, it shows significant advantages in the two indicators of PSNR and MAE, indicating that the method proposed in the paper has high accuracy and high-quality output in image restoration and enhancement tasks. The excellent performance of SSIM and LPIPS also demonstrates that the method proposed in the paper can not only accurately reconstruct images but is also very outstanding in maintaining the image structure and perceptual quality. These results prove the effectiveness and superiority of the network structure proposed in this paper in handling image restoration and enhancement tasks.

In terms of image metric evaluation, our method outperforms the compared methods. PSNR primarily represents image restoration at the pixel level. Our method achieves better results than other methods in this metric. SSIM evaluates the structural similarity of images, and our results surpass those of other methods, indicating that our images are better at maintaining semantic coherence. MAE reflects the model's fitting capability, and compared to other image processing methods, our method shows superior fitting ability. LPIPS evaluates the realism of images, and our method performs slightly worse than EnlightenGAN in indoor environments. The main reason is that in indoor settings where illumination changes are not significant, adversarial networks tend to produce images with a stronger sense of realism to the human eye than our network.

6. Conclusion

This study significantly improves the image brightness and clarity in low-illumination environments by integrating the improved PSNR loss, SSIM loss, and self-attention loss. More importantly, it can precisely retain the high-level features in the image, such as the details of mechanical equipment and the identity information of the operating personnel, ensuring that the restored image is both of high quality and practical. This is particularly important for applications that require precise image restoration, such as open-pit mine monitoring systems, especially when dealing with challenges such as low resolution and high noise.

When dealing with images under extremely low-illumination conditions, the method of this study shows excellent anti-noise ability, effectively reducing noise points and image distortion. Compared with methods such as DRBN, DSLR, and EnlightenGAN, the method in this paper performs outstandingly in maintaining the clarity and stability of the image enhancement effect, reflecting its strong performance when dealing with low-quality original images.

Through tests on open-pit mine images under different scenarios and conditions, the method in this paper has wide applicability and flexibility. It can achieve effective image enhancement effects, whether under extremely low-illumination conditions or facing complex environmental backgrounds, such as the challenges of many noise points and low resolution. At the same time, it has made a breakthrough in computing efficiency and can thus be quickly deployed in practical applications to meet the needs of real-time processing.

In the study of image enhancement for open-pit mines, we found that noise tends to appear in large areas of the same color during enhancement. Such areas are frequently seen in open-pit mine images. Therefore, future research will focus on handling image enhancement for large monochromatic areas and exploring how to avoid the appearance of noise artifacts.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Acknowledgement

This study was supported by the National Key Research and Development Program of China (No. 2022YFB4703701) and the Fundamental Research Funds for the Central Universities (No. 2023JCCX-NY01).

References

1. Bi, C., Qian, W., Cao, J., Wang, X. LightingFormer: Transformer-CNN Hybrid Network for Low-Light Image Enhancement. Computers & Graphics, 2024, 124, 104089. https://doi.org/10.1016/j.cag.2024.104089
2. Chen, Y., Yang, X.-H., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., Chen, H., Hu, H., Zhou, Q., Guan, Q. Gen-erative Adversarial Networks in Medical Image Augmentation: A Review. Computers in Biology and Medicine, 2022, 144, 105382. https://doi.org/10.1016/j.compbiomed.2022.105382
3. Ding, K., Liu, Y., Zou, X., Wang, S., Ma, K. Locally Adaptive Structure and Texture Similarity for Image Quality Assessment. In Proceedings of the 29th ACM International Conference on Multimedia, 2021, 2483-2491. https://doi.org/10.1145/3474085.3475419
4. Erwei, K. O. N. G., Yabang, Z. H. A. N. G., Jiayue, L. I. An Enhancement Method for Low Light Images in Coal Mines. Journal of Mine Automation, 2023, 49(4), 62-69.
5. Ferzli, R., Karam, L. J. A No-Reference Objective Image Sharpness Metric Based on The Notion of Just Noticeable Blur (JNB). IEEE Transactions on Image Processing, 2009, 18(4), 717-728. https://doi.org/10.1109/TIP.2008.2011760
6. Hong, Z. H. A. N. G., Tingfeng, S. U. O., Wanying, S. O. N. G. A Mine Image Enhancement Method Based on Structural Texture Decomposition. Journal of Mine Automation, 2024, 50(3), 56-64. DOI: 10.13272/j.issn.1671-251x.2023100005
7. Hore, A., Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In 2010 20th International Conference on Pattern Recognition, 2010, 2366-2369. https://doi.org/10.1109/ICPR.2010.579
8. gnatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L. DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision, 2017, 3277-3285. https://doi.org/10.1109/ICCV.2017.355
9. Jiang, N., Lin, J., Zhang, T., Zheng, H., Zhao, T. Low-Light Image Enhancement via Stage-Transformer-Guided Network. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8), 3701-3712. https://doi.org/10.1109/TCSVT.2023.3239511
10. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z. Enlightengan: Deep light enhancement without paired supervision. IEEE transac-

tions on image processing, 2021, 30, 2340-2349. https://doi.org/10.1109/TIP.2021.3051462

11. Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., Shah, M. Transformers in Vision: A Survey. ACM Computing Surveys (CSUR), 2022, 54(10s), 1-41. https://doi.org/10.1145/3505244
12. Li, C., Guo, J., Porikli, F., Pang, Y. LightenNet: A Convolutional Neural Network for Weakly Illuminat-ed Image Enhancement. Pattern Recognition Letters, 2018, 104, 15-22. https://doi.org/10.1016/j.patrec.2018.01.010
13. Lim, S., Kim, W. DSLR: Deep Stacked Laplacian Re-storer for Low-Light Image Enhancement. IEEE Transactions on Multimedia, 2020, 23, 4272-4284. https://doi.org/10.1109/TMM.2020.3039361
14. Mustafa, W. A., Abdul Kader, M. M. M. A review of histogram equalization techniques in image en-hancement application. In Journal of Physics: Confer-ence Series, 2018, 1019, 012026. IOP Publishing. https://doi.org/10.1088/1742-6596/1019/1/012026
15. Osher, S., Rudin, L. I. Feature-Oriented Image En-hancement Using Shock Filters. SIAM Journal on Numerical Analysis, 1990, 27(4), 919-940. https://doi.org/10.1137/0727053
16. Palubinskas, G. Image Similarity/Distance Measures: What Is Really Behind MSE and SSIM? International Journal of Image and Data Fusion,2017, 8(1), 32-53. https://doi.org/10.1080/19479832.2016.1273259
17. Sara, U., Akter, M., Uddin, M. S. Image Quality As-sessment Through FSIM, SSIM, MSE and PSNR-A Comparative Study. Journal of Computer and Com-munications, 2019, 7(3), 8-18. https://doi.org/10.4236/jcc.2019.73002
18. Tanchenko, A. Visual-PSNR Measure of Image Quali-ty. Journal of Visual Communication and Image Rep-resentation, 2014, 25(5), 874-878. https://doi.org/10.1016/j.jvcir.2014.01.008
19. Tang, G., Ni, J., Chen, Y., Cao, W., Yang, S. X. An Improved CycleGAN Based Model for Low-Light Im-age Enhancement. IEEE Sensors Journal, 2023, 24(14), 21879-21892. https://doi.org/10.1109/JSEN.2023.3296167
20. Tian, F., Chen, T., Zhang, J. Research on Improved Retinex-Based Image Enhancement Method for Mine Monitoring. Applied Sciences, 2023, 13(4), 2672. https://doi.org/10.3390/app13042672
21. Tian, Z., Wu, J., Zhang, W., Chen, W., Zhou, T., Yang, W., Wang, S. An Illuminance Improvement and Details Enhancement Method on Coal Mine Low-Light Images Based on Transformer and Adap-tive Feature Fusion. Coal Science and Technology, 2024, 52(1), 297-310. DOI: 10.13199/j.cnki.cst.2023-0112
22. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T. Ultra-High-Definition Low-Light Image En-hancement: A Benchmark and Transformer-Based Method. In Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(3), 2654-2662. https://doi.org/10.1609/aaai.v37i3.25364
23. Wen, Y., Xu, P., Li, Z., Ato, W. X. An illumination-Guided Dual Attention Vision Transformer for Low-Light Image Enhancement. Pattern Recognition, 2025, 158, 111033. https://doi.org/10.1016/j.patcog.2024.111033
24. Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., Shao, L. Learning Enriched Fea-tures for Fast Image Restoration and Enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2), 1934-1948. https://doi.org/10.1109/TPAMI.2022.3167175
25. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 586-595. https://doi.org/10.1109/CVPR.2018.00068