

Article

Target Detection in Underground Mines Based on Low-Light Image Enhancement

Haodong Guo , Kaibo Lu, Shanning Zhan, Jiangtao Li and Zhifei Wu *

College of Mechanical Engineering, Taiyuan University of Technology, Taiyuan 030024, China; 2023510068@link.tyut.edu.cn (H.G.); lvkaibo@tyut.edu.cn (K.L.); 2022520148@link.tyut.edu.cn (S.Z.); 2023520199@link.tyut.edu.cn (J.L.)

* Correspondence: wuzhifei@tyut.edu.cn

Abstract

Underground mines' complex environments with dim lighting and high dust and humidity hamper feature extraction and reduce detection accuracy. To address this, we propose a low-light image enhancement-based target detection algorithm. Firstly, LIENet enhances low-light image quality and brightness via a dual-gamma curve and non-reference loss function-guided iterations. Secondly, the hierarchical feature extraction (HFE) method with a dual-branch structure captures long-term and local correlations, focusing on critical corner regions. Finally, HFE is combined with a feature pyramid structure for comprehensive feature representation through a top-down global adjustment. Our method, validated on a self-built dataset, outperforms other algorithms with an mAP@0.5 of 96.96% and mAP@0.5:0.95 of 71.1%, proving excellent low-light detection performance in mines.

Keywords: low-light enhancement; target detection; long-term correlation; unsupervised learning; feature pyramid structure; hierarchical feature extraction

1. Introduction

Target detection technology in coal mines is an important technology for safety monitoring in coal mine production, contributing to the supervision and early warning of mine safety accidents. Due to the complex underground environment, narrow passages, and limited light sources, images captured in some scenarios often exhibit characteristics such as low lighting, high noise, and blurry edges. In such cases, detection systems find it challenging to accurately capture the information of hidden target objects in each image, especially small target objects, leading to a significant number of false detections and missed detections. Therefore, the development of image enhancement and target detection technology for adverse lighting conditions can effectively improve the accuracy of target detection in coal mines, which is of great importance to promote the safety and sustainable development of the coal mining industry.

In recent years, there has been extensive research at home and abroad on low-light image enhancement. Currently, low-light image enhancement is mainly divided into model optimization methods and deep learning-based methods. The optimization methods of the model are mainly based on Retinex theory [1–3], which suggests that the original image S can be decomposed into the product of the illumination image L and the reflectance image R , that is, $S = LR$. By decomposing the original image, the influence of uneven illumination can be eliminated, thereby improving the visual effect of the image. However, existing methods based on Retinex theory often rely on multiple assumptions about the



Academic Editor: Phivos Mylonas

Received: 26 December 2025

Revised: 7 February 2026

Accepted: 20 February 2026

Published: 25 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

real environment, making the handling of image brightness less flexible and often resulting in color distortion and detail loss. In comparison, deep learning-based image enhancement algorithms [4,5] demonstrate superior and more stable performance in low-light image enhancement, mainly due to the powerful learning ability of Convolutional Neural Networks (CNNs), which can quickly establish the mapping relationship from low-light images to high-light images.

It is noteworthy that for extreme visual environments such as underground mines, infrared imaging technology offers another important sensing approach. This technique utilizes the thermal radiation characteristics of objects themselves to form images, providing all-weather operational capability. It can still generate stable images in scenarios where visible light is severely restricted, such as in complete darkness, dense fog, or smoke-filled conditions, significantly expanding human capabilities for detection and perception in complex environments. It has been widely applied in fields such as reconnaissance and security surveillance [6]. However, infrared imaging systems typically output grayscale images, which commonly suffer from a lack of color information, blurred detail texture, and low overall contrast. The human visual system, evolved in a color-rich natural environment, is more sensitive to chromatic information. Prolonged observation of monochromatic grayscale images can easily lead to visual fatigue and reduced interpretation efficiency. Furthermore, the nature of thermal radiation can cause objects of different materials to exhibit similar temperature signatures in infrared images, posing challenges for target differentiation and fine-grained recognition. Consequently, although infrared imaging holds clear advantages in terms of penetration capability, its inherent imaging characteristics limit its direct application potential for tasks requiring high-precision color and texture discrimination, such as the identification of safety helmets of specific colors or visual inspection of equipment status.

Some supervised learning algorithms [7,8] can be trained using prematched and annotated datasets to fit the mapping relationship between high- and low-light images as closely as possible. However, obtaining paired data in real-world scenarios is quite challenging. When the training dataset is too small or the model complexity is too high, supervised learning algorithms are prone to overfitting issues, resulting in good performance in training data but poor performance in practical applications. In contrast, unsupervised learning networks have advantages in generality and data processing. EnlightenGAN proposed by Jiang et al. [9] is the first network model to enhance low-light images using unsupervised training methods, including global and local adversarial losses and normalized perceptual losses, to make generated images closer to the illumination distribution of normal-light images. However, this GAN-based model is large, requiring larger-scale datasets and more powerful computing resources for training, which is not conducive to network deployment.

Therefore, Guo et al. [10] proposed a lightweight zero-reference enhancement network, Zero-DCE, which uses deep network learning of high-order curve estimation models and applies them to pixel-level dynamic range adjustment of input images. The introduction of this zero-reference learning method provides a new approach to image enhancement, learning only the illumination enhancement process from training data without the need for paired or unpaired data during application.

With the rapid development of deep learning technology, deep learning-based object detection techniques have been widely applied in various low-light scenarios. Du et al. [11] proposed a hybrid zero-reference and dehazing network (Z-DCE-DNet) specifically for enhancing low-light and hazy images in underground mines. This method combines low-light enhancement and dehazing techniques to address issues of uneven lighting and fog distortion, aiming to produce clearer images that improve the performance of downstream object detection tasks. Li et al. [12] proposed RW-DM, a diffusion model for

coal mine low-light enhancement that combines Retinex and Wavelet transforms to improve efficiency. It demonstrates superior image quality and a 3.1% increase in detection mAP@50. Han et al. [13] proposed UM-GAN, a generative adversarial network-based method for enhancing low-light images in underground mines. The model employs an encoder–decoder structure, fuses information from inverted grayscale and low-light images, and incorporates a noise reduction module to effectively restore details and improve the overall image quality. Experiments on diverse datasets validate its effectiveness. These algorithms enhance the overall object detection performance to a certain extent by enhancing low-light images in advance. However, there are still the following issues: (1) pre-adopted image enhancement algorithms rely on reference images for enhancement, with complex network structures and poor adaptability to specific scenarios; (2) the detection methods used focus too much on inter-layer feature interactions, neglecting intra-layer features, resulting in a large number of effective features not being extracted.

In summary, considering image enhancement quality and model lightweight, we propose a zero-reference low-light image enhancement algorithm, LIENet. In this work, considering the fact that the illumination enhancement function adjustment range in the Zero-DCE algorithm is insufficient, we have redesigned a set of illumination enhancement functions to increase local pixel values. Compared to the original enhancement functions, the designed enhancement functions require only a small number of iterations to adaptively adjust a wide range of pixels, avoiding the additional parameter and computational cost losses generated by multiple high-order iterative functions. Additionally, we propose a hierarchical feature extraction (HFE) method to simultaneously capture long-range correlation and local correlation information. Relevant studies have shown that shallow features mainly contain non-global information such as color and texture [14–16]. The introduction of HFE can enhance the expression information of shallow features to some extent, and, coupled with the feature pyramid’s top-down feature extraction structure, it can adjust feature information effectively.

Our contributions are summarized as follows:

1. We propose a lightweight zero-reference image enhancement algorithm, LIENet, which can quickly adjust low-light images at the pixel level;
2. We propose a full-layer feature extraction method, which contains two branch structures that can extract long-distance feature information and local area feature information respectively;
3. We validate the superiority of the proposed methods in our self-built low-light dataset mines, achieving performance improvements compared to other detection algorithms.

2. Materials and Methods

In this section, we will introduce the LIENet image enhancement method and the implementation details of the HFE. First, we will discuss the specific implementation of the LIENet algorithm in Section 2.1, followed by introducing the loss functions used in the enhancement algorithm in Section 2.2; then, we will demonstrate the implementation details of the HFE method in Section 2.3; finally, we will introduce the loss functions used in the object detection model in Section 2.4.

2.1. Dataset Construction

2.1.1. Data Sources and Composition

The dataset used in this paper for the underground mine was established through the method of independent collection. Part of the data comes from the CUMT-HelmeT and CUMT-BelT datasets released by China University of Mining and Technology [17]; the dataset was captured by the KBA12(B) mining intrinsically safe alarm camera in multiple

underground coal mines. This camera features a maximum resolution of 2560×1920 (5 megapixels), a fill-light distance of 30 m, and characteristics such as compact structure, small size, and explosion-proof and moisture-proof design, making it suitable for use in underground coal mine environments, while another part was collected through web crawlers. In this work, we collected 6500 images of different working scenarios in the mine. Figure 1 shows that the dataset mainly includes three types of annotations: worker, helmet, and anchor bolt. Table 1 displays detailed information for each category in the dataset. It is worth noting that collecting data in real mines under extremely low-light conditions is very challenging, and the quality and quantity of corresponding images are limited. Therefore, following the approach in reference [18], we adjusted normal images and artificially synthesized a low-light dataset. In addition, in order to demonstrate the effect of the algorithm proposed in this paper on low-light enhancement, relevant verification was carried out on the basis of the publicly available low-light datasets LOL [19], Exdark [20] and LIME [21].



Figure 1. Self-built dataset partial display.

Table 1. Information statistics of self-built dataset.

Category	Number of Labels	Number of Pictures
Worker	9758	1185
Helmet	14,541	3261
Anchor bolt	6531	2054
Sum	30,830	6500

The LOL dataset consists of 500 pairs of low-light and normal-light images, which are divided into 485 pairs of training images and 15 pairs of validation images. The resolution of the images is 400×600 . The Exdark dataset contains 7363 low-light images with annotations for 12 target categories. The LIME dataset is reference-free low-light image data, composed of 10 low-light images of different scenes.

2.1.2. Low-Light Synthesis Method

A uniform distribution of $\gamma \sim \text{Uniform}(2,5)$ with random variable γ is selected to result in training patches that are darkened to a varying degree. To simulate low-quality cameras used to capture images, these original training patches are corrupted by Gaussian noise via the MATLAB R2023b function `imnoise` with standard deviation of

$\sigma = \sqrt{B(25/255)^2}$, where $B \sim \text{Uniform}(0, 1)$. Hence, the final corrupted image and the original image exhibit the following relationship:

$$I_{train} = n\left(g\left(I_{original}\right)\right) \quad (1)$$

where function $g(\cdot)$ represents the gamma adjustment function and $n(\cdot)$ represents the noise function.

Random gamma darkening with random noise levels result in a variety of training images that can increase the robustness of the model. In reality, natural low-light images may also include quantization and Poisson noise (e.g., images captured with imaging sensors such as CCD and CMOS) in addition to Gaussian noise. We chose to focus on the Gaussian-only model for the ease of analysis and as a preliminary feasibility study of the framework trained on synthetic images and applied to natural images. Furthermore, since Gaussian noise is a very familiar yet popular noise model for many image denoising tasks, we can acquire a sense of how well LIENet performs with respect to other image enhancement algorithms. While training the model, the network attempts to remove the noise and simultaneously enhance the contrast of these darkened patches. The reconstructed image is compared against the clean version (i.e., bright, noiseless image) by computing the mean-squared error.

In this study, the paired synthetic low-light images and their corresponding original normal-light images generated by the above method will serve as the ground truth for subsequent objective image quality assessment (PSNR, SSIM).

2.2. Network Architecture

2.2.1. LIENet

Due to the limited lighting conditions in underground mines, most of the collected images exhibit characteristics of low light and low contrast. To enhance the brightness and contrast of images captured in underground mines while minimizing the burden on model parameters, we designed the lightweight image enhancement network LIENet (as shown in Figure 2) as an image preprocessing method, referencing the Zero-DCE algorithm. It is worth noting that the convolutional kernels used in the model have a size of 3, padding of 1, and 32 channels.

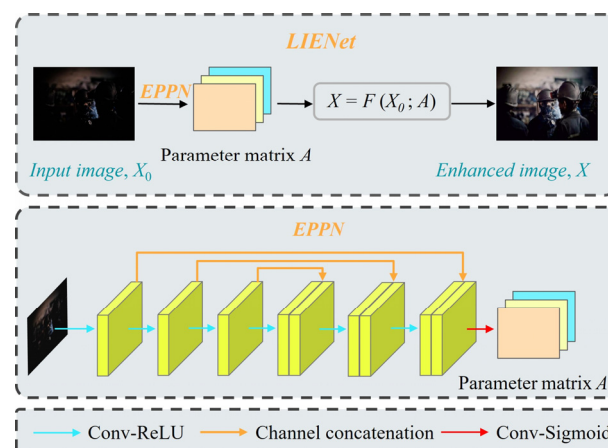


Figure 2. Low-light enhancement network model structure (LIENet).

2.2.2. Low-Light Enhancement Algorithm

We analyzed the pixel distribution in the dataset and found that the majority of image pixels are distributed within the range of $[0, 0.32]$ (as shown in Figure 3). Therefore, the

image enhancement network model in this paper mainly focuses on the characteristics of low pixel regions.

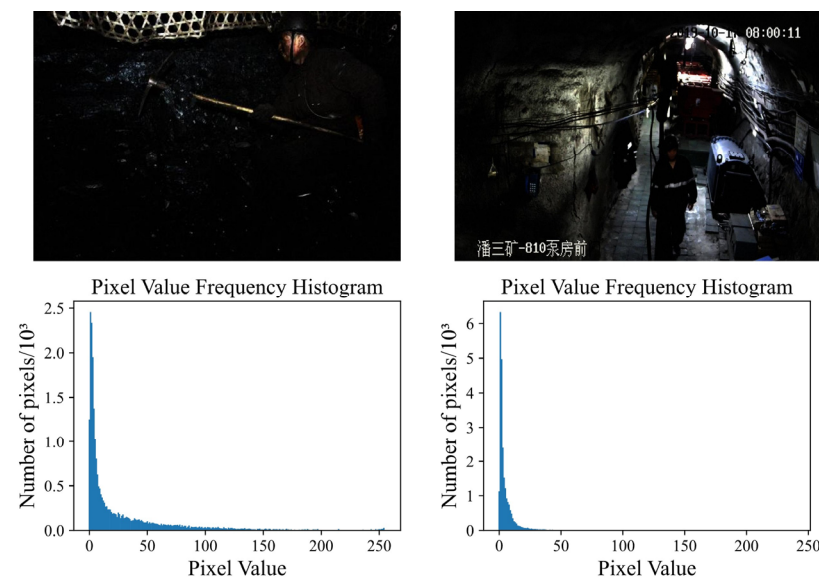


Figure 3. Low-light images and grayscale histogram.

To this end, this paper redesigns a set of illumination enhancement curves based on double gamma function correction for low-light images collected in underground mines, as shown below:

$$F(x; \alpha) = \alpha G_a(x) + (1 - \alpha) G_b(x), 0 \leq \alpha \leq 1 \quad (2)$$

where $G_a(x) = x^{1/\gamma}$, $G_b(x) = 1 - (1 - x)^{1/\gamma}$ and γ is an adjustable variable used to adjust the degree of image enhancement. After related testing, we set it to 4; this value is chosen based on the alignment between the specific distribution of low-light mine images (as shown in Figure 3, where pixels are highly concentrated in the $[0, 0.32]$ interval) and the characteristics of the dual-gamma function. When $\gamma = 4$, the function $x^{(1/\gamma)}$ exhibits a pronounced stretching slope in the low-value region, effectively enhancing the dominant dark pixels. Simultaneously, the overall morphology of the function remains smooth, avoiding excessive enhancement in mid-tone and high-light areas. This achieves an optimal balance between dark detail recovery and overall visual naturalness. Fixing this key parameter aligns with our design philosophy of building a lightweight, stable, and easily deployable enhancement network (LIENet). x is the input pixel value, $F(x; \alpha)$ is the corresponding output pixel value, and α is the illumination enhancement parameter, which can be learned by the enhancement parameter prediction network in the proposed model. In addition, we compared the quadratic iterative enhancement function used in the original ZeroDCE algorithm, with the following function expression:

$$E(x; \alpha) = x + \alpha x(1 - x) \quad (3)$$

The corresponding illumination mapping curves under different α values are shown in Figure 4:

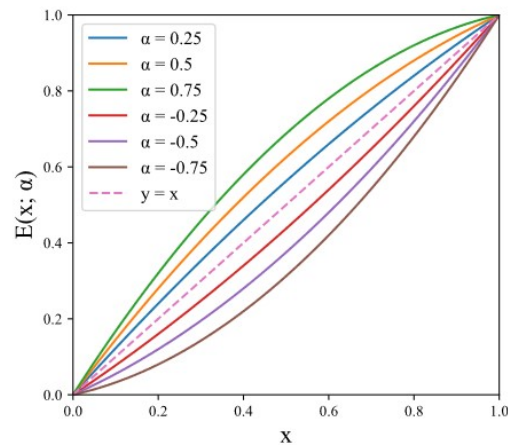


Figure 4. Mapping curve under quadratic iterative function.

Correspondingly, under the condition of setting γ to 4, the results of the redesigned dual-gamma illumination mapping curve under different α values are shown in Figure 5:

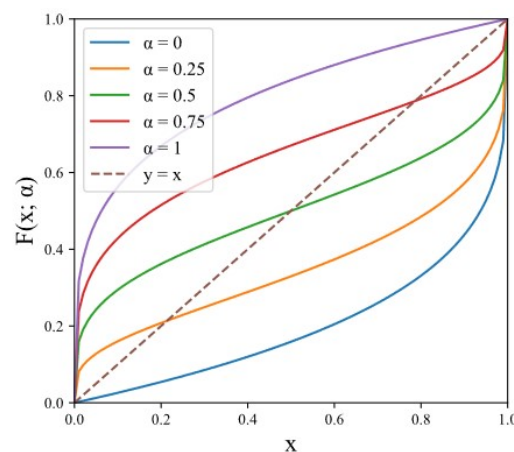


Figure 5. Mapping curve under double gamma function.

Through comparison, it was found that the quadratic iterative enhancement function in the original algorithm only has an enhancing or weakening effect for each α value, and the adjusted illumination range is relatively small. Therefore, in the original algorithm, dynamic adjustment of light intensity was achieved through eight iterations to control the exposure level and increase the range of light adjustment. The illumination mapping curve proposed in this paper based on dual-gamma correction will enhance the low pixel area within a certain range based on different illumination enhancement parameters, while suppressing the high pixel area, thereby avoiding excessive exposure in the enhanced image. In addition, the proposed illumination mapping curve can adjust the illumination over a large range without the need for function iteration, requiring fewer learning parameters and a lighter network.

In the actual model, to ensure that each pixel in the image can achieve dynamic adjustment, the relationship between the enhanced image and the low-light image is established as follows:

$$F(X; A) = AG_a(X) + (1 - A)G_b(X) \quad (4)$$

where A is the enhancement parameter matrix composed of the illumination enhancement parameter a , corresponding in scale to the input image. X is the input image composed

of a large number of pixels. By substituting the obtained output enhancement parameter matrix A into Equation (3), the output image with a bright effect can be obtained.

2.3. Illumination Model Loss Function Design

Many illumination enhancement models require a large amount of manually labeled or paired datasets for supervised learning, which is time consuming, inefficient, and less applicable in different scenarios. In order to achieve lightweight network model design and improve the generalization of the network, a set of non-reference loss functions is designed to guide the network to learn the mapping relationship between low illumination and normal illumination. By introducing this loss function, the model can achieve zero-reference learning without relying on any data labels. This loss function consists of four parts specifically:

1. Spatial Consistency Loss (L_{spa}):

$$L_{spa} = \frac{1}{S} \sum_{i=1}^S \sum_{j \in \Omega(i)} (|Y_i - Y_j| - |I_i - I_j|)^2 \quad (5)$$

where S represents the number of local regions partitioned from the image, $\Omega(i)$ denotes the four neighboring regions (up, down, left, and right) with region i as the center, Y and I respectively represent the average intensity values of local regions in the low-light image and the enhanced image.

2. Exposure Control Loss (L_{exp}):

$$L_{exp} = \frac{1}{M} \sum_{k=1}^M |Y_k - E| \quad (6)$$

In the equation, M represents the number of non-overlapping local regions with a size of 16×16 pixels, and the average intensity value of local regions in the enhanced image is denoted as Y . The exposure control loss measures the distance between the average intensity value of a local region to the well-exposedness level E . We follow existing practices [10] to set E as the gray level in the RGB color space. We set E to 0.6 in our experiments, although we do not find much performance difference by setting E within [0.4, 0.7].

3. Color Constancy Loss (L_{col}):

$$L_{col} = \sum_{\forall (p,q) \in \varepsilon} (J^p - J^q)^2 \quad (7)$$

where J^p and J^q represent the average intensity values of channels p and q in the enhanced image, and $\varepsilon = \{(R, G), (R, B), (G, B)\}$ represents any two combinations of channels in the RGB color space.

4. Illuminance Smoothness Loss (L_{tvA}):

$$L_{tvA} = \frac{1}{N} \sum_{n=1}^N \sum_{c \in \delta} (|\nabla x A_n^c| + |\nabla y A_n^c|)^2, \delta = \{R, G, B\} \quad (8)$$

In the equation, N is the number of iterations and ∇x and ∇y represent the horizontal/vertical gradient values of corresponding channel pixel values.

5. Total Loss (L_{total}):

$$L_{total} = L_{spa} + L_{exp} + W_{col} L_{col} + W_{tvA} L_{tvA} \quad (9)$$

In the equation, W_{col} and W_{tvA} act as weight parameters to balance the weight scales between different losses. Based on multiple comparative experiments, W_{col} is set to 0.5 and W_{tvA} is set to 20 for better performance.

2.4. Improvement of YOLOv8 Object Detection Algorithm

The YOLOv8 [22] network model structure (as shown in Figure 6) mainly consists of three parts: Backbone, Neck, and Head. The input image is processed by the Backbone network and output three different scale feature maps at stage 3, stage 4, and stage 5, with spatial sizes of $1/8$, $1/16$, and $1/32$ of the input image, respectively. The three feature maps generated by the Backbone network are then fused into multi-scale feature maps by the Neck section, which adopts the structure of feature pyramid and path aggregation networks to further enhance the feature representation ability, thereby improving the accuracy and efficiency of object detection. After processing by the Neck section, three feature maps of different scales are still output, and the final processing by the Head section produces object detection results at three different scales: large, medium, and small. In the original YOLOv8 algorithm, although the FPN structure used in the Neck section can effectively integrate multi-scale inter-layer features at different levels, it ignores intra-layer features at the same level, which are essential for object detection and recognition. The application of visual transformers and attention mechanisms can to some extent learn compact intra-layer feature representations, but there are still disadvantages such as insufficient attention to corner regions and high computational complexity. Inspired by references [23–25], we replaced the attention module in the transformer model with a spatial pooling layer, which can simplify the network structure while extracting the long-term correlation of input features. In addition, in order to obtain locally correlated features of input features, we designed a lightweight encoder structure as an additional parallel branch structure to extract these features from input feature maps, and the final network model structure is shown in Figure 7.

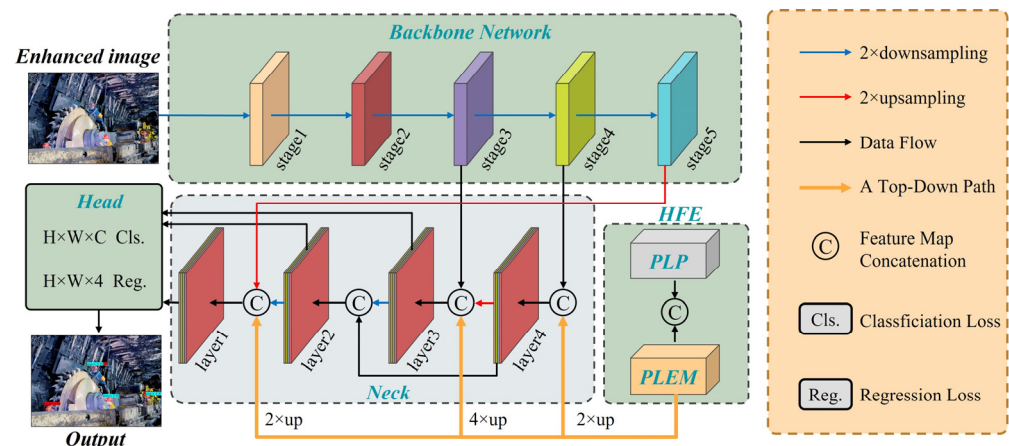


Figure 6. Improved structure of YOLOv8 object detection algorithm.

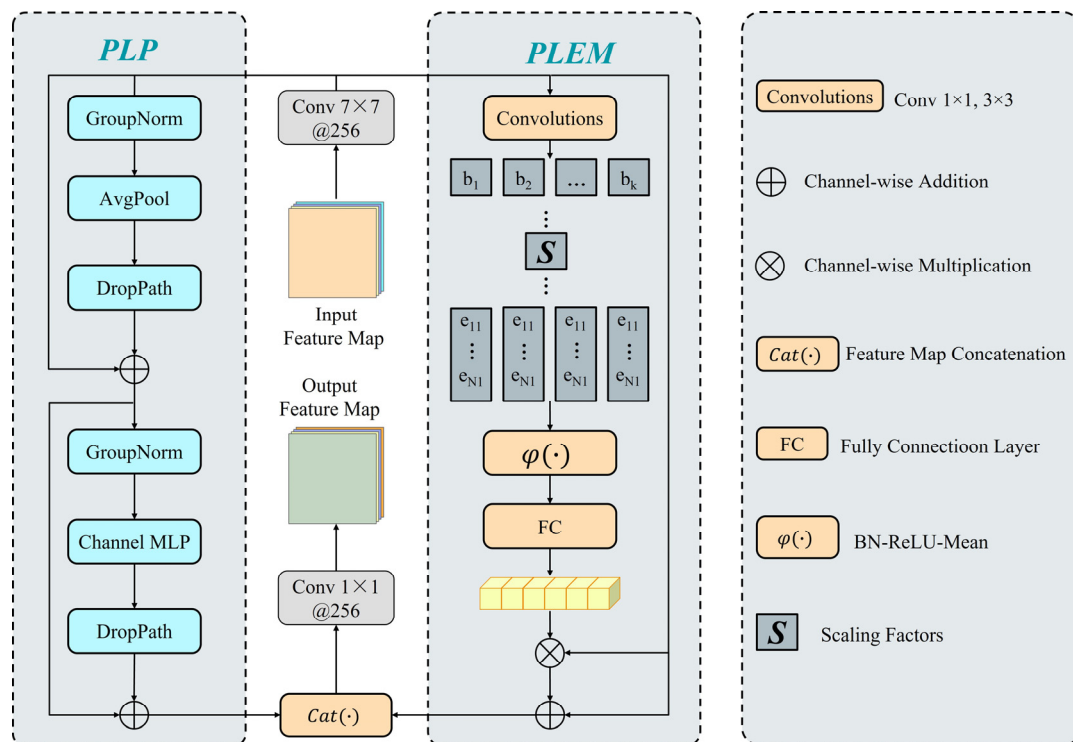


Figure 7. Network model structure diagram.

As shown in Figure 7, the proposed HFE network is composed of two parallel parts, with the PLP module obtained by concatenating two residual modules, which is a general structure also used in some transformer models. What sets it apart is that in the PLP module, we use spatial pooling layers instead of attention modules to mix information between tokens and employ DropPath operations to improve the model's generalization ability and robustness. This structure has been proven in subsequent experiments to effectively extract the long-term correlated characteristics of image features, thereby enhancing the model's recognition accuracy. In addition, another structure, PLEM, is parallelly connected to obtain the local correlation information of the input image. This module is mainly composed of a codebook and fully connected layers. The specific processing flow of the corresponding module is as follows:

PLP: Specifically, after the input features undergo group normalization, they are used as inputs to the spatial pooling layer. The computational complexity of pooling is linearly related to the sequence length and does not require learnable parameters, which significantly reduces the model's parameter quantity and complexity. The above process can be expressed using the following formula:

$$Y = \text{Pooling}(\text{GN}(X)) + X \quad (10)$$

In the context of a residual module, X represents the input to the first residual module, Y represents the output of the residual module, GN stands for group normalization, and "Pooling" represents the spatial average pooling operation, with pool size set to 3 and padding set to 1.

For the residual structure based on Channel MLP, where Y serves as the input to this structure, it undergoes group normalization operation first, and then the input features are processed using Channel MLP. Compared with space MLP, Channel MLP not only

effectively reduces computational complexity but also meets the requirements of general visual tasks. The above process can be specifically represented as follows:

$$Z = CM(GN(Y)) + Y \quad (11)$$

In the above equation, CM represents Channel MLP and Z represents the output of the second residual module.

PLEM: The composition of the codebook mainly includes two parts: visual codewords $B = b_1, b_2, \dots, b_K$ and a set of learnable scaling factors $S = s_1, s_2, \dots, s_K$. In the model construction process, we set K to 64.

The processing flow of the inherent dictionary can be expressed as follows:

Using a set of scaling factors S to map the positional information of the input image features and B correspondingly, the information e_{ik} of the i-th pixel x_i with respect to the k-th codeword can be obtained from the following equation:

$$e_{ik} = \frac{e^{-s_k \|x_i - b_k\|^2}}{\sum_{k=1}^K e^{-s_k \|x_i - b_k\|^2}} (x_i - b_k) \quad (12)$$

where b_k represents the k-th learnable visual codeword, s_k represents the k-th scaling factor, and $x_i - b_k$ represents the positional information of each pixel relative to the codeword. Similarly, the positional information e_k of the entire image pixels with respect to the k-th codeword can be expressed as follows:

$$e_k = \sum_{i=1}^N e_{ik} \quad (13)$$

where $N = H \times W$ represents the total number of features in the input features, and H and W represent the spatial size of the input features in height and width, respectively. After obtaining the output e_k of the codebook, the complete information of the entire image with respect to all codewords can be calculated using the following equation:

$$e = \sum_{k=1}^K \varphi(e_k) \quad (14)$$

where φ includes operations such as ReLU, normalization BN, and average layer. The e is sent into the fully connected layer and 1×1 convolutional layer to predict the features of prominent key categories. Then, the obtained scaling factor coefficient $\delta(\cdot)$ is multiplied with the input feature X_{in} of the B module in the channel dimension, and the corresponding process can be expressed as follows, where δ represents the sigmoid function operation and $\otimes$ represents channel multiplication. Finally, the obtained Out is added to the input feature X_{in} of the B module in the channel dimension on a per-channel basis.

$$LVC(X_{in}) = X_{in} \oplus Out \quad (15)$$

2.5. Loss Function of Object Detection Model

The detection model's loss function includes both classification loss and regression loss. YOLOv8 uses the same classification head structure as YOLOv5, using BCE (Binary Cross-Entropy) loss function as the classifier's evaluation criterion, as shown in Equation (15):

$$L = \frac{1}{N} \sum_{i=1}^N -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (16)$$

where y_i represents the true label of sample i and p_i represents the predicted label of sample i . For regression loss, the detection task commonly uses IoU to measure the overlap between the predicted box and the true box, which can be expressed as follows:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (17)$$

where $A \cap B$ and $A \cup B$ represent the intersection and union areas of the predicted box and the true region. In order to consider the distance factors between the predicted box and the true box and accelerate the convergence process, YOLOv8 uses CIoU Loss as part of the regression loss, which can be specifically expressed as follows:

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu \quad (18)$$

where b and b^{gt} respectively represent the center coordinates of the predicted box and the true box; $\rho^2(\cdot)$ represents the Euclidean distance between the two coordinates; c represents the diagonal length of the minimum bounding rectangle of the predicted box and the true box; in addition, α and ν represent

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{W^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \quad (19)$$

$$\alpha = \frac{\nu}{(1 - \text{IoU}) + \nu} \quad (20)$$

2.6. Overall Training and Inference Pipeline

Figures 8 and 9 clearly illustrate the collaborative workflow between the proposed LIENet enhancement module and the YOLOv8-HFE detection module during the training and inference stages.

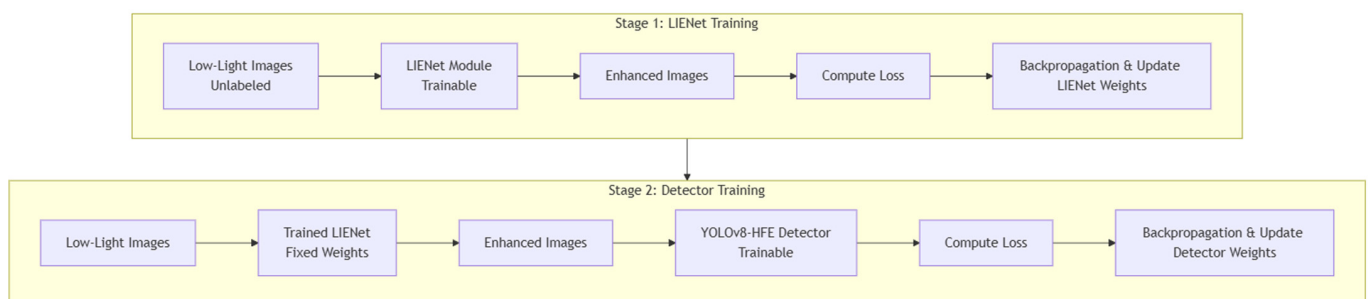


Figure 8. Training phase: two-stage independent training pipeline.

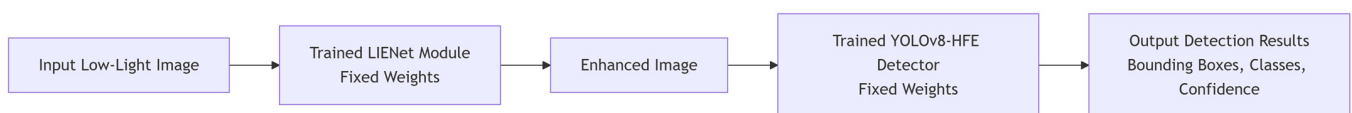


Figure 9. Inference phase: sequential processing pipeline.

(1) The training stage adopts a two-stage independent training strategy (Figure 8). In the first stage, the LIENet network is independently trained in an unsupervised manner on low-light images using the non-reference loss function L_{total} . In the second stage, the weights of the trained LIENet from the first stage are fixed and it serves as a preprocessing module to enhance the low-light images in the training set. Subsequently, the YOLOv8-HFE detection network is trained under supervision using these enhanced images along with their corresponding annotations. The two networks are not trained jointly in an end-to-

end manner. This design ensures the stable optimization of each module and maintains methodological flexibility.

(2) The inference stage follows a sequential pipeline mode (Figure 9). For an input unknown low-light image, it is first enhanced by the fixed-weight LIENet module. The enhanced image is then fed directly into the fixed-weight YOLOv8-HFE detector to finally output the target detection results.

3. Results

3.1. Experiment Preparation

This paper's experiments consist of two parts: low-light enhancement and object detection. The low-light enhancement model adopts the LIENet proposed in this paper and uses 1200 low-light images as the training set for the model. Since the training process is driven by unsupervised learning based on the non-reference loss function, there is no need to pair or annotate the low-light dataset in advance. The experimental operating system is Ubuntu 20.04, the graphics card is NVIDIA GeForce RTX 3090, and the network framework is PyTorch 2.0.1. During the training process, we set the training epochs to 90 and batch size to 8, using the Adam optimizer, and set the learning rate to 0.0001. The object detection network uses the improved YOLOv8 model for detection. Before training, the images need to be resized to 640×640 , and the self-built mine dataset is divided into a training set and a validation set at a ratio of 4:1. The batch size is set to 32, epochs to 300, and SGD is used for optimization training, with momentum set to 0.9. Figure 10 shows the loss situation of the training set and the validation set during the training process. It can be seen that after 280 epochs, the loss of the training set tends to stabilize.

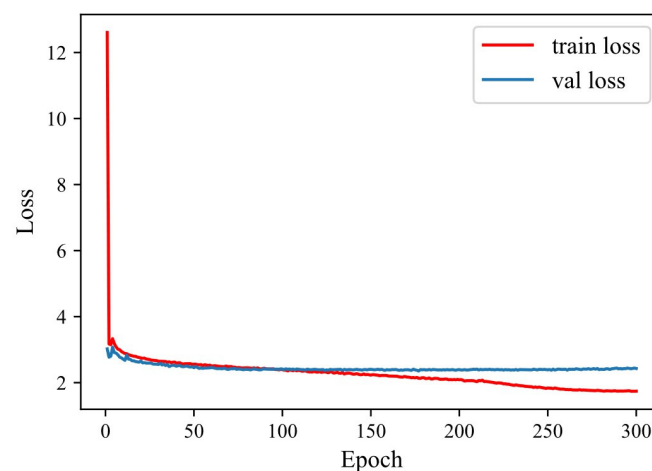


Figure 10. Improved YOLOv8 training loss convergence curve.

3.2. Image Enhancement Evaluation Index

The ultimate goal of this work is to improve target detection performance in low-light conditions. Therefore, we consider the mean average precision (mAP) of object detection as the core and final metric for evaluating the utility of image enhancement algorithms. Meanwhile, to analyze enhancement quality from different perspectives, we also report full-reference image quality metrics (PSNR, SSIM) and no-reference metrics (NIQE, BRISQUE). Among them, PSNR and SSIM are calculated only on our self-built synthetic dataset which has paired ground truth, providing a reference for pixel-level restoration fidelity.

To verify the enhancement effect of the low-light image enhancement algorithm proposed in this paper, five mainstream algorithms were selected for comparative experiments, including Retinex-Net [26], MIRNetv2 [7], MBLLEN [26], Zero-DCE [10], and SCI [27]. In

addition, three low-light enhancement algorithms based on traditional algorithms were also selected for comparison, including LIME, Dong [28] and BIMEF [29].

Among them, Retinex-Net, MIRNetv2, and MBLLEN are supervised learning algorithms, SCI is an unsupervised learning algorithm, and Zero-DCE is a zero-reference learning algorithm. This article will evaluate the effects of image enhancement from both subjective visual aspects and objective metrics. In terms of objective evaluation, to simultaneously assess the quality of image processing and its impact on target detection performance, we will use peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), Natural Image Quality Evaluator (NIQE), and mean average precision (mAP) in detection tasks as evaluation metrics.

(1) Peak signal-to-noise ratio (PSNR) is used to evaluate the quality of image processing. It is calculated by comparing the grayscale values of image pixels before and after enhancement, and is measured in decibels. A higher value indicates less distortion and better image quality. The calculation process is as follows:

$$P_{SNR} = 20 \times \log_{10} \left(\frac{I_{Max}}{\sqrt{M_{se}}} \right) \quad (21)$$

In the formula, I_{Max} represents the maximum value of the input data, and M_{se} represents the mean squared error between the images before and after enhancement.

(2) Structural similarity is a metric for measuring the similarity between two images, ranging from 0 to 1, where closer to 1 indicates better similarity. The formula is shown below:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (22)$$

(3) Natural Image Quality Evaluator (NIQE) is a no-reference, distortion-free image quality assessment metric, where a lower value indicates better image quality. The expression is shown below:

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{\left((v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (v_1 - v_2) \right)} \quad (23)$$

Here, v_1 , v_2 , Σ_1 , and Σ_2 represent the mean and covariance matrices of the natural MVG model and the distorted image MVG model, respectively.

(4) mAP is the average value of AP (average precision) for all categories, used to measure the accuracy of the model in detecting multiple classes of objects. AP is calculated by computing the area enclosed by the precision–recall (PR) curve, where precision and recall are calculated by the following formulas.

$$Precision = \frac{TP}{TP + FP} \quad (24)$$

$$Recall = \frac{TP}{TP + FN} \quad (25)$$

TP represents the number of true positive predictions, FP represents the number of false positive predictions, and FN represents the number of false negative predictions. Different confidence threshold values can be selected to plot different PR curves, and the corresponding AP value can be obtained by calculating the area under the curve.

3.3. Image Enhancement Effect Analysis

Comparative experiments were conducted on the image enhancement algorithm proposed in this paper and the other eight low-light enhancement algorithms on the self-built dataset, LOL, Exdark and LIME. The comparison results in six scenes were selected for subjective analysis. Among them, Scene 4 was taken from the Exdark dataset, Scene 5

was sourced from the LIME dataset, and Scene 6 was sourced from the LOL dataset. The comparison results are shown in Figure 11.

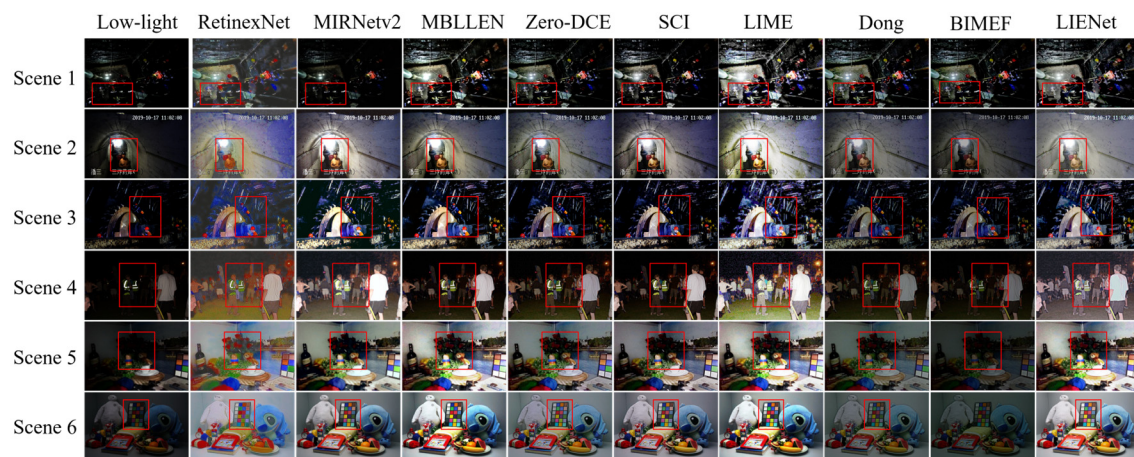


Figure 11. Comparison results of different image enhancement algorithms on self-built dataset.

From Figure 11, it can be seen that after enhancement by the RetinexNet network, the brightness of the image is significantly improved, but the corresponding colors and textures are not harmonious, and there is insufficient clarity in the highlighted area, leading to limited information content. MIRNetv2 performs well in color restoration but has limited enhancement effects on darker areas. MBLLEN and ZeroDCE show good visual restoration effects and no texture distortion, but the overall brightness improvement is limited. The image enhanced by SCI also exhibits low brightness and contrast and the other three enhancement algorithms based on traditional algorithms also have problems such as noise, color deviation and relatively low brightness to a greater or lesser extent. In contrast, the algorithm proposed in this paper can achieve better color saturation after enhancing the images. It retains the detailed information of the images while enhancing the contrast, significantly improves the brightness, and makes the colors more natural.

To intuitively demonstrate the differences in detail recovery among various enhancement algorithms, Figure 12 provides a zoomed-in comparison of the critical region in Scene 1 from Figure 11. In the original low-light image, this region is almost devoid of recognizable target features due to severe underexposure. After enhancement by LIENet, details such as the texture of the safety helmet and the contours of tools are clearly restored. This directly explains why, in the corresponding detection results shown in Figure 12, the detection model using LIENet-enhanced input can identify targets in this region more accurately. In contrast, other enhancement algorithms exhibit limited detail recovery capabilities in this area, which contributes to missed or false detections in the subsequent detection task.

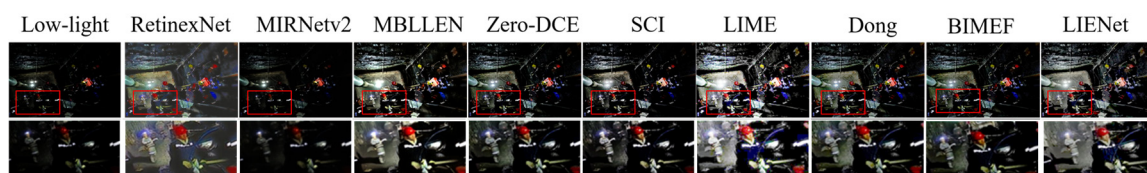


Figure 12. Zoomed-in comparison of different enhancement algorithms on the critical region (red box in Scene 1 of Figure 11).

Objective Quantitative Analysis: In order to better verify the enhancement effect and lightweight model structure of the LIENet proposed in this paper, the eight low-light enhancement algorithms mentioned above were selected to be compared with the LIENet

proposed herein. In terms of performance metrics, PSNR, mAP, the number of model parameters and running time were chosen. The experimental results corresponding to each enhancement algorithm on the self-built dataset are shown in Table 2, where the black bold font represents the optimal results and the underlined font represents the second-best results.. It should be noted that the original size of the input images was all 640×640 , and the improved YOLOv8 detection model was used for the detection network. For the objective quality assessment on the self-built dataset, we used the original normal-light images corresponding to the synthetic low-light images as the ground truth for calculating PSNR and SSIM metrics. These reference images are the original ones before darkening and noise addition in the synthesis process, thus providing a reasonable basis for evaluating the restoration capability of enhancement algorithms.

Table 2. Performance comparison of various low-light image enhancement algorithms.

Enhanced Network	PSNR↑	SSIM↑	mAP@0.75	Model Parameter/ 10^3	Inference Time/ms	Platform
RetinexNet	18.01	0.545	80.86%	555.21	128	Pytorch (GPU)
MIRNetv2	19.73	0.613	81.74%	5858.56	802	Pytorch (GPU)
MBLEN	<u>21.49</u>	<u>0.662</u>	<u>82.54%</u>	450.17	7890	Tensorflow (GPU)
Zero-DCE	18.44	0.582	82.13%	79.42	8.89	Pytorch (GPU)
SCI	18.14	0.513	82.11%	0.26	0.574	Pytorch (GPU)
BIMEF	15.62	0.452	76.54%	-	-	Python 3.9
Dong	16.21	0.482	77.13%	-	-	Python 3.9
LIME	17.58	0.515	79.65%	-	-	Pytorch (GPU)
LIENet	22.59	0.721	83.44%	<u>67.3</u>	<u>7.4</u>	Pytorch (GPU)

The tabulated data show that the images processed by the proposed algorithm achieve the highest mean average precision (mAP) after passing through the detection model, with an mAP@0.75 of 83.44%. Simultaneously, the algorithm also outperforms other comparative methods in image fidelity metrics, PSNR and SSIM, indicating that its enhancement process effectively improves image quality while avoiding over-enhancement artifacts, thereby reducing distortion. These results collectively demonstrate that LIENet is superior to other illumination enhancement models in terms of overall color restoration and downstream detection effectiveness.

Regarding model efficiency, LIENet exhibits a notable lightweight advantage. Its parameter count is only 67.3K, which is significantly lower than other deep learning-based models (e.g., RetinexNet: 555.21K, MIRNetv2: 5858.56K). Although the SCI algorithm has an even lower parameter count, LIENet performs better in both color restoration and detection accuracy, achieving a more favorable trade-off between performance and efficiency. It should be noted that while our method achieves relatively excellent results in inference time, the absolute fairness of cross-platform comparisons may be influenced by architectural differences, as different algorithms rely on disparate software frameworks, implementation optimizations, and hardware adaptations. Nonetheless, the extremely low parameter count of LIENet provides a fundamental guarantee for its efficient inference in practical deployment.

In addition, we also verified different enhancement algorithms in the LOL, ExDark and LIME datasets respectively. In the LOL dataset, the commonly used image quality assessment metrics, namely peak signal-to-noise ratio (PSNR) and Structural similarity index measure (SSIM), were adopted for evaluation. Since the LIME and ExDark datasets do not have corresponding low-light reference images, and low-light images in real life do not have corresponding ground truth either, we used the no-reference evaluation

metrics, Naturalness Image Quality Evaluator (NIQE) and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), instead. The corresponding experimental results are shown in Table 3, where the black bold font represents the optimal results and the underlined font represents the second-best results.

Table 3. Performance comparison of various low-light image enhancement algorithms.

Dataset	LOL		ExDark		LIME	
Evaluation	PSNR↑	SSIM↑	NIQE↓	BRISQUE↓	NIQE↓	BRISQUE↓
RetinexNet	17.64	0.47	4.42	31.82	5.26	29.47
MIRNetv2	23.54	<u>0.84</u>	<u>3.15</u>	<u>24.53</u>	<u>3.82</u>	21.24
Zero-DCE	14.86	0.56	3.22	25.95	3.96	23.73
SCI	15.12	0.51	3.95	27.56	4.28	24.41
LIME	16.75	0.56	3.38	26.43	4.35	22.31
Dong	16.72	0.48	3.85	29.51	4.24	26.22
BIMEF	13.86	0.60	3.18	25.46	3.92	24.25
LIENet	<u>21.54</u>	0.86	3.03	23.47	3.78	<u>22.22</u>

In order to more intuitively reflect the recognition effect of different enhancement algorithms on the detection model, the improved YOLOv8 is used as the target detection model, and the outputs of different image enhancement algorithms are used as inputs to the detection model. The corresponding visual results are shown in Figure 13. From the figure, it can be seen that the recognition effect of the detection model under low-light conditions is not ideal, with a large number of missed detections and false alarms. RetinexNet and MIRNetv2 algorithms exhibit some missed detections for small targets such as safety helmets. MBLLEN, Zero-DCE, and SCI algorithms show false alarms in the dark areas; for example, in scenario 1, all three algorithms mistakenly identify the tool bag carried by the worker as a helmet. The traditional Dong algorithm and the BIMEF algorithm have limited detection effects on small target objects. Moreover, due to insufficient brightness and a large amount of noise information, the detection effect in dark areas is rather poor. The LIME algorithm also has a large number of missed detections in the areas where targets are concentrated. For example, in Scene 4, some people were not detected due to the interference of noise information. In contrast, the images enhanced by LIENet are more conducive to the recognition of the detection model. It can identify the targets that are easily overlooked in other enhanced images, and the situations of missed detections and false detections have been significantly improved. In addition, there are remarkable improvements in indicators such as the recognition rate and confidence level.

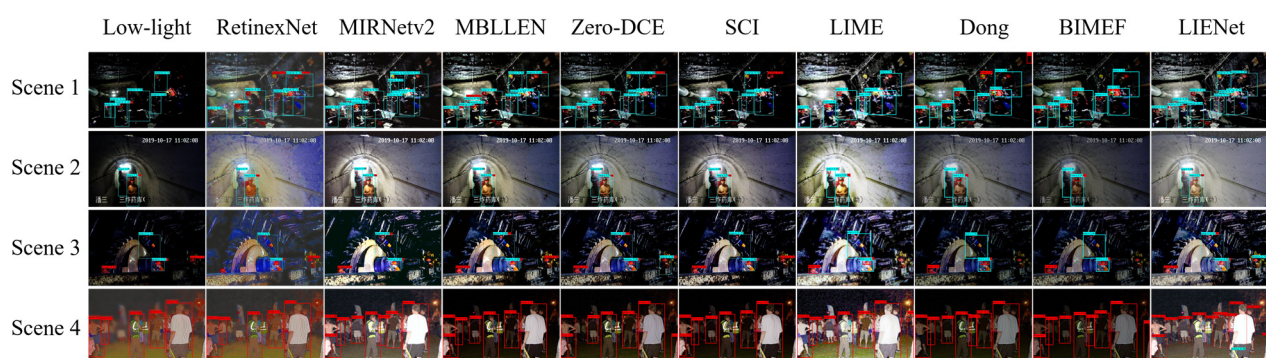


Figure 13. Detection results corresponding to different image enhancement algorithms on the self-built dataset.

3.4. Target Detection Performance Evaluation

In order to verify the accuracy of the algorithm proposed in this paper for the identification and positioning of different targets in underground mines, we selected some classic target detection algorithms for comparative experiments, including Faster R-CNN [30], SSD [31], RetinaNet [32], YOLOv5 [33], CenterNet [34], and FCOS [35]. Meanwhile, we also compared with some target detection algorithms based on low-light scenes in the past two years, including those proposed by Wang et al. [36], Peng et al. [37] and Zhou et al. [38]. Among them, Faster R-CNN is a classic two-stage target detection algorithm that requires generating candidate regions in advance. SSD, RetinaNet, and YOLOv5 are famous one-stage target detection networks that directly generate object category probabilities and position coordinate values without the need for proposing boxes. CenterNet and FCOS are anchor-free single-stage target detection algorithms based on the anchor-free concept, which can save memory consumption during training.

Table 4 shows the mAP and FPS of various comparative algorithms on a self-built mine dataset. It can be seen from the results that the detection accuracy evaluation indicator mAP@0.5 of the algorithm proposed in this paper reaches 96.96%, and mAP@0.5:0.95 can reach 71.1%, significantly higher than other comparative algorithms. Compared with the YOLOv8s baseline model, there is a 2.42% increase in accuracy, making the target detection performance in low-light mine environments more ideal. Additionally, while ensuring high accuracy, the model's inference FPS can still reach 35.2 frames per second, giving it a competitive advantage over other algorithms and meeting the real-time detection needs in mine environments.

Table 4. Evaluation of object detection performance metrics.

Enhanced Network	Detection Network	Backbone	mAP@0.5	mAP@0.75	mAP@0.5:0.95	FPS
LIENet	Faster R-CNN	ResNet50	94.25%	71.62%	63.42%	18.22
	SSD	VGG16	92.93%	64.78%	59.40%	45.63
	RetinaNet	ResNet50	94.85%	72.57%	63.23%	23.66
	YOLOv5s	CSP-DarkNet53	95.44%	76.24%	67.72%	31.94
	CenterNet	ResNet50	93.54%	66.26%	61.17%	35.43
	FCOS	ResNet50	94.27%	68.50%	60.54%	26.79
	YOLOv8s	CSP-DarkNet53	95.61%	78.51%	68.68%	39.37
	Wang et al. [36]	-	93.45%	75.56%	65.92%	27.54
	Peng et al. [37]	-	92.32%	72.45%	65.54%	25.4
	Zhou et al. [38]	-	95.94%	79.56%	69.53%	28.54
	proposed method	-	96.96%	83.44%	72.50%	35.2

In addition, we visualized the training results. Figure 14 compares the precision–recall (PR) curves of the participating models on the self-built dataset in this paper, and it shows the mAP situation corresponding to a threshold value of 0.5. It can be seen from it that the area formed by the PR curve of our proposed model is the largest, indicating that our method has the best detection performance.

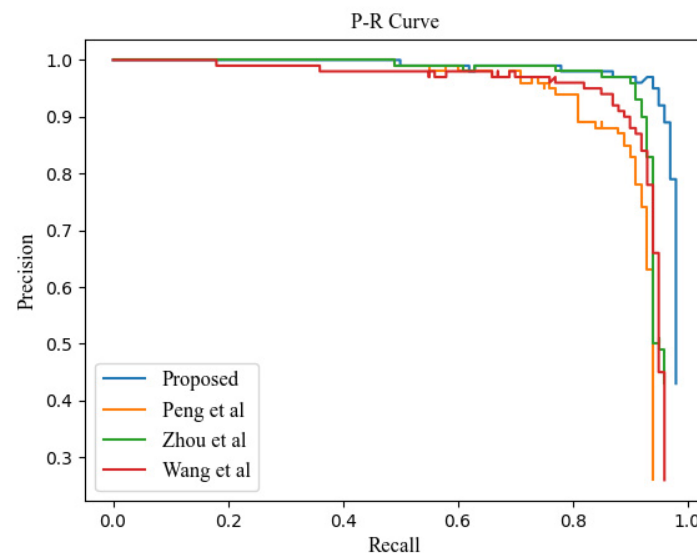


Figure 14. PR curves under different models (Wang et al. [36], Peng et al. [37] and Zhou et al. [38]).

3.5. Ablation Experiment

To verify the effectiveness of each improvement module in the target detection performance of the algorithm in this paper, ablation experiments were conducted on the baseline model, gradually introducing the LIENet, PLP, and PLEM modules, as shown in Table 5.

Table 5. Results of ablation experiment.

YOLOv8s	LIENet	PLP	PLEM	mAP@0.5	mAP@0.75	mAP@0.5:0.95	Runtime/ms
✓				94.70%	75.34%	64.52%	18
✓	✓			95.61%	78.51%	69.68%	25.4
✓		✓		96.27%	82.63%	71.86%	27.5
✓	✓		✓	96.8%	83.10%	72.17%	27.8
✓	✓	✓	✓	96.96%	83.44%	72.50%	28.4

From the data in the table, it is evident that when we only use the baseline model to detect images without undergoing image enhancement, the corresponding average precision is the lowest, with an mAP@0.5:0.95(%) value of 64.52%. When the LIENet algorithm is introduced for image enhancement, the model's detection accuracy immediately improves significantly. Compared to before the introduction, the corresponding mAP@0.5:0.95(%) increases by 5.16%, while the corresponding runtime only increases by 7.4ms. Furthermore, when the PLP module is introduced to capture the long-term correlations between input features, the model's accuracy increases by 2.18%. Similarly, with the introduction of the PLEM for extracting features of locally related regions, the model's accuracy increases by 2.49%. Finally, after introducing the LIENet, PLP, and PLEM modules simultaneously on the baseline model, the model achieves the highest detection accuracy, with an mAP@0.5 reaching 96.96%. Although the runtime increases by 10.4ms compared to not introducing any modules, there is a 7.98% improvement in the mAP@0.5:0.95(%), fully demonstrating the excellent detection performance of the proposed algorithm in low-light mine environments.

To evaluate the robustness of the final model, we conducted training with five different random seeds under identical configurations (with γ fixed at 4 and E at 0.6). The mean (μ) and standard deviation (σ) of the detection performance (mAP@0.5:0.95) are reported. The results show a mean performance of 72.3% with a standard deviation of $\pm 0.4\%$, demonstrating that our method maintains highly stable performance across different random initializations.

4. Discussion

Under the influence of the objective environment with low visibility and the imaging equipment with limited performance, the collected images inevitably have a relatively low visual quality, which brings certain difficulties to the target detection task. In order to improve the image quality and the efficiency of target detection, in this work, we first proposed a lightweight zero-reference image enhancement algorithm, LIENet, which utilizes a dual-gamma enhancement curve to achieve wide-range brightness adaptive adjustment with a small number of iterations, addressing issues such as insufficient contrast enhancement in enhanced images, limited enhancement in dark areas, and exposure phenomena. In addition, to address the issue of inadequate full-level feature extraction in target detection models, we proposed a global feature extraction method, HFE, where PLP is used to capture long-term correlations and PLEM is used to extract local relevant information of input features. A large number of comparative experiments were conducted on a self-built mine personnel dataset, comparing with some excellent target detection algorithms. The proposed method achieved an average accuracy of up to 96.96% in low-light scenes, with a single-frame image inference time of only 28.4ms, overall superior to other comparative methods. This indicates that the algorithm in this paper has high detection accuracy, lightweight model size, high flexibility, and fast detection speed. We also conducted extensive ablation experiments to demonstrate the effectiveness of each improvement module in enhancing model performance. Although the above-mentioned methods can effectively enhance low-light images and conduct efficient detection, there are still some limitations, mainly in the following aspects:

1. **Limitations of application scenarios:** The algorithm proposed in this paper is developed primarily to address the low-light conditions prevalent in underground mine environments. It is important to clarify that this work focuses specifically on illumination enhancement and does not explicitly handle other common visual degradations in such settings, including but not limited to dust scattering, lens contamination, motion blur from equipment vibration, strong local glare, and dense occlusion. Consequently, when deployed in scenarios where these additional factors are prominent, the performance of the proposed method may be suboptimal, and further adaptation or integration with complementary techniques would be necessary to achieve robust performance.
2. **Integration with downstream tasks:** The algorithm proposed in this paper has achieved certain effects in target detection under low-light environments, but the application of the algorithm goes far beyond this. For example, it can be combined with detection in the research on underground positioning in mines. In this case, targeted improvements to the algorithm are required to optimize the performance of specific downstream tasks.

In the follow-up, we will also conduct research on the above limitations to promote the informatization and intelligent development of coal mines.

Author Contributions: Conceptualization, H.G. and Z.W.; methodology, H.G.; software, H.G.; validation, H.G., K.L., S.Z. and J.L.; formal analysis, H.G.; investigation, H.G. and S.Z.; resources, Z.W. and K.L.; data curation, H.G. and J.L.; writing—original draft preparation, H.G.; writing—review and editing, Z.W., K.L. and S.Z.; visualization, H.G.; supervision, Z.W.; project administration, H.G.; funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shanxi Province Key Research and Development Program under Grant 202102110401020, the Research Project Supported by Shanxi Scholarship Council of

China under Grant 2021-050, and the Fundamental Research Program of Shanxi Province under Grant 202103021224040.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and ethical considerations.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HFE hierarchical feature extraction

References

- Shen, Z.; Xu, H.; Jiang, G.; Yu, M.; Du, B.; Luo, T.; Zhu, Z. Pseudo-retinex decomposition-based unsupervised underwater image enhancement and beyond. *Digit. Signal Process.* **2023**, *137*, 103993. [\[CrossRef\]](#)
- Yang, W.; Wang, W.; Huang, H.; Wang, S.; Liu, J. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Trans. Image Process.* **2021**, *30*, 2072–2086. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, B.; Guo, Z.; Yao, W.; Ding, X.; Zhang, D. A novel low-light enhancement via fractional-order and low-rank regularized retinex model. *Comput. Appl. Math.* **2023**, *42*, 7. [\[CrossRef\]](#)
- Yang, S.; Zhou, D. Single image low-light enhancement via a dual-path generative adversarial network. *Circuits Syst. Signal Process.* **2023**, *42*, 4221–4237. [\[CrossRef\]](#)
- Chen, Y.; Zhu, G.; Wang, X.; Shen, Y. Fmr-net: A fast multi-scale residual network for low-light image enhancement. *Multimed. Syst.* **2024**, *30*, 73. [\[CrossRef\]](#)
- Yang, X.; Tian, L.; Cai, F. Thermal infrared imaging for conveyor roller fault detection in coal mines. *PLoS ONE* **2024**, *19*, e0307591.
- Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.; Shao, L. Learning enriched features for fast image restoration and enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1934–1948. [\[CrossRef\]](#)
- Xie, S.; Ma, Y.; Xu, W.; Qiu, S.; Sun, Y. Semi-supervised learning for low-light image enhancement by pseudo low-light image. In Proceedings of the 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Taizhou, China, 28–30 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
- 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 Trans. Image Process.* **2021**, *30*, 2340–2349. [\[CrossRef\]](#)
- Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. *arXiv* **2020**, arXiv:2001.06826.
- Du, Q.; Zhang, S.; Wang, Z.; Liang, J.; Yang, S. A hybrid zero-reference and dehazing network for joint low-light underground image enhancement. *Sci. Rep.* **2025**, *15*, 10135. [\[CrossRef\]](#)
- Li, Y.; Tian, J.; Chen, Y.; Wang, H.; Yan, H.; Peng, Y.; Wang, T. Rw-Dm: Retinex and wavelet-based diffusion model for low-light image enhancement in underground coal mines. *Complex Intell. Syst.* **2025**, *11*, 327. [\[CrossRef\]](#)
- Han, W.; Xiao, Y.; Yin, Y. UM-GAN: Underground mine GAN for underground mine low-light image enhancement. *IET Image Process.* **2024**, *18*, 2154–2160. [\[CrossRef\]](#)
- Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 27–30 June 2016; pp. 2921–2929.
- Ru, L.; Zhan, Y.; Yu, B.; Du, B. Learning affinity from attention: End-to-end weakly-supervised semantic segmentation with transformers. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 16846–16855.
- Li, R.; Mai, Z.; Zhang, Z.; Jang, J.; Sanner, S. Transcam: Transformer attention-based cam refinement for weakly supervised semantic segmentation. *J. Visual Commun. Image Represent.* **2023**, *92*, 103800. [\[CrossRef\]](#)
- Chen, L.; Guo, L.; Cheng, D.; Kou, Q. Structure-preserving and color-restoring up-sampling for single low-light image. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1889–1902. [\[CrossRef\]](#)
- Lore, K.G.; Akintayo, A.; Sarkar, S. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [\[CrossRef\]](#)

19. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560. [CrossRef]
20. Loh, Y.P.; Chan, C.S. Getting to know low-light images with the exclusively dark dataset. *Comput. Vis. Image Underst.* **2019**, *178*, 30–42. [CrossRef]
21. Guo, X.; Li, Y.; Ling, H. Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [CrossRef]
22. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLO, January 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 5 October 2025).
23. Quan, Y.; Zhang, D.; Zhang, L.; Tang, J. Centralized feature pyramid for object detection. *IEEE Trans. Image Process.* **2023**, *32*, 4341–4354. [CrossRef]
24. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. Mlp-mixer: An all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24261–24272.
25. Yu, W.; Luo, M.; Zhou, P.; Si, C.; Zhou, Y.; Wang, X.; Feng, J.; Yan, S. Metaformer is actually what you need for vision. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 10819–10829.
26. Lv, F.; Lu, F.; Wu, J.; Lim, C. Mblen: Low-light image/video enhancement using cnns. *BMVC* **2018**, *220*, 4.
27. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5637–5646.
28. Dong, X.; Pang, Y.; Wen, J. Fast efficient algorithm for enhancement of low lighting video. In *ACM SIGGRAPH 2010 Posters*; Association for Computing Machinery: New York, NY, USA, 2010; p. 1.
29. Ying, Z.; Li, G.; Gao, W. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv* **2017**, arXiv:1711.00591.
30. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
31. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2016; pp. 21–37.
32. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
33. Jocher, G. YOLOv5 by Ultralytics, May 2020. Available online: <https://github.com/ultralytics/yolov5> (accessed on 5 October 2025).
34. Zhou, X.; Wang, D.; Krhenbuhl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
35. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. *arxiv* **2019**, arXiv:1904.01355.
36. Wang, T.; Qu, H.; Liu, C.; Zheng, T.; Lyu, Z. Lle-std: Traffic sign detection method based on lowlight image enhancement and small target detection. *Mathematics* **2024**, *12*, 3125. [CrossRef]
37. Peng, D.; Ding, W.; Zhen, T. A novel low light object detection method based on the yolov5 fusion feature enhancement. *Sci. Rep.* **2024**, *14*, 4486. [CrossRef] [PubMed]
38. Zhou, Q.; Zhang, D.; Liu, H.; He, Y. Kcs-yolo: An improved algorithm for traffic light detection under low visibility conditions. *Machines* **2024**, *12*, 557. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.