Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

LSM-KinD: An Attention-Guided Low-LightEnhancement Model for Railway Fasteners

Yijie Cai
Hubei University of Technology
Jiannan Kang
Hubei University of Technology
Xing Dai
Hubei University of Technology
Xuehai Liu
Gemac Engineering Machinery Co., Ltd.
Feng Guo
Gemac Engineering Machinery Co., Ltd.
Changhao Chen
Institute of Scientific and Technological Research, China Railway Wuhan Group Co., Ltd.
Wupeng Deng
Hubei University of Technology
Xiaojun Yang
19910055@hbut.edu.cn

Hubei University of Technology

Research Article

Keywords: Railway Fastener, KinD, SENet, Low-Light Enhancement, Object Detection

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8393482/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

LSM-KinD: An Attention-Guided Low-Light Enhancement Model for Railway Fasteners

Yijie Cai[1,2], Jiannan Kang[1,2], Xing Dai[1,2], Xuehai Liu[3], Feng Guo[3], Changhao Chen[4], Wupeng Deng[1,2], Xiaojun Yang[1,2]

[1]School of Mechanical Engineering, Hubei University of Technology, Wuhan, 430068, Hubei, China.
[2]Hubei Provincial Key Laboratory of Modern Manufacturing Quality Engineering, Hubei University of Technology, Wuhan, 430068, Hubei, China.
[3]Gemac Engineering Machinery Co., Ltd., Xiangyang, 441000, Hubei, China.
[4]Institute of Scientific and Technological Research, China Railway Wuhan Group Co., Ltd., Wuhan, 430071, Hubei, China.

Contributing authors: yijie.cai@hbut.edu.cn; 102410099@hbut.edu.cn; 102310087@hbut.edu.cn; lxhonlyone@163.com; 18986396626@163.com; cch890907@whut.edu.cn; dengwupeng@hbut.edu.cn; 19910055@hbut.edu.cn (Corresponding author);

Abstract

In railway fastener monitoring, poor lighting leads to issues like shadow distortion, overexposure, and loss of texture details, affecting system reliability and analysis accuracy. To address these, we propose the LSM-KinD model, which integrates the KinD architecture with SENet attention and multi-scale residual connections. The model uses an image decomposition network to separate reflection and illumination components, enhancing accuracy with multi-scale feature extraction and attention mechanisms. The illumination adjustment network applies residual blocks and channel attention to precisely control illumination, improving image quality. The image restoration network, based on U-Net and channel attention, optimizes the enhancement results. Experimental results show that LSM-KinD outperforms other algorithms, achieving PSNR and SSIM scores of 23.921 and 0.901 on the RAILF dataset. The model also improves detection

accuracy, reaching a mAP@0.5 of 84.9% in low-light environments, demonstrating its broad application potential in railway safety and system maintenance.

Keywords: Railway Fastener, KinD, SENet, Low-Light Enhancement,Object Detection

1 Introduction

In modern railway systems, fasteners are essential components of the track structure, playing a crucial role in ensuring the safety and stability of railway transportation. During nighttime operations of automatic track inspection vehicles, the images of railway fasteners are often compromised by factors such as insufficient lighting and shadow artifacts, making precise target recognition a challenging task. This significantly impacts the accuracy of detection systems, thereby increasing the potential risk of accidents [1]. Consequently, enhancing the quality of fastener images in low-light conditions to improve detection accuracy is of great practical importance.

Object detection, a fundamental task in computer vision, focuses on identifying and locating objects within images or videos. Early detection methods relied on traditional image processing techniques and handcrafted feature extraction, including edge detection, color analysis, and shape matching. However, these approaches struggled with challenges like scale and rotation variations. The introduction of feature extraction techniques such as SIFT[2] and HOG[3] enhanced the robustness of object detection. With the advent of deep learning, advanced algorithms like Faster R-CNN[4], YOLO[5], SSD[6], and RetinaNet[7] were developed, along with architectural innovations such as FPN[8], Mask R-CNN[9], and DETR[10], significantly improving detection performance.

Despite these advancements, low-light environments, especially during nighttime railway fastener inspections, present a significant challenge. Factors such as light attenuation, dynamic motion blur, and noise interference degrade image quality, resulting in reduced contrast of surface textures and substantial loss of key component edge features. These issues severely limit the accuracy of fastener defect detection and undermine the effectiveness of machine vision systems.

To address this challenge, Huang et al.[11] developed an efficient digital image contrast enhancement method that combines adaptive gamma correction with weighted histogram to adjust the histogram, thereby enhancing the image contrast. He et al.[12] introduced a deep dual-domain semi-blind network (D3SN) that detects the image compression quality factor (QF) using a quality factor detection module (QFD), and subsequently applies a dual-domain compressed image enhancement network to improve the quality of compressed images. Zhuang et al.[13] proposed a variational model based on the super-Laplacian reflectance prior, employing the l1/2 norm to enhance underwater image details and colors, and the l2 norm to accurately estimate illumination, thus improving image quality. While traditional algorithms still retain application value in simpler scenarios, they often produce artifacts such as halos, local contrast distortion, and other issues in complex environments. Furthermore, as they rely on fixed rules, they are limited in their

ability to adapt to changing image conditions and to process local details with precision.

Compared with traditional approaches, deep learning has brought about a paradigm shift in image enhancement. Ko et al.[14] proposed a lightweight deep convolutional neural network (CNN) that employs similarity-preserving knowledge distillation to guide the student network in replicating the spatial relationships among intermediate features of a teacher network. This method achieves efficient low-light image enhancement while maintaining low computational complexity. Building upon Retinex theory, Wei et al.[15] introduced the LOL dataset and developed Deep Retinex-Net, a supervised, end-to-end framework that leverages dual-stream decomposition and enhancement for low-light image restoration. Similarly, Wang et al.[16], also grounded in Retinex theory, designed a three-component architecture comprising a decomposition network, a lightweight illumination adjustment module, and a residual reflectance reconstruction network. Their approach effectively optimizes the illumination and reflectance components of low-light images captured in coal mine environments, thereby enhancing contrast and recovering texture details, which addresses common shortcomings of conventional methods such as detail loss and blurred edges.

Deep learning models offer superior flexibility and representational capacity, making them particularly well-suited for complex image enhancement tasks involving large-scale datasets and high-dimensional feature spaces. Among these, Retinex-based methods have gained considerable attention in deep learning applications by decomposing images into illumination and reflectance components to improve brightness and preserve structural details. However, applying fully convolutional architectures to enhance the illumination map may lead to the loss of deep semantic features and introduce local unevenness in illumination. Moreover, independently processing the reflectance and illumination components can weaken their intrinsic correlation, potentially degrading the perceptual quality of the enhanced image. Despite recent advances in low-light target detection, significant challenges remain, particularly in recovering fine texture details in severely dark regions. In the specific context of railway fastener inspection, enhancing image quality under low-light conditions, especially with regard to revealing critical texture details of fasteners, has emerged as a key research focus.

To address the issue of unclear texture details in dark regions of images, Zhang et al.[17] proposed KinD and its improved version, KinD++[18], a Retinex-based image enhancement algorithm for low-light image quality. This method is formulated as an image enhancement model grounded in the imaging process[19] and adheres to the color-perception consistency hypothesis[20], thereby improving the recognizability of low-light images. The algorithm maps low- and high-light images into multi-scale feature spaces through a decomposition network, employs an illumination adjustment network to dynamically allocate regional illumination intensity, and finally reconstructs clear images via a reflectance recovery network. Although deep learning methods for low-light image enhancement

can significantly improve image clarity, issues such as overexposure and distortion during image processing remain unavoidable.

This paper proposes a novel low-light image enhancement model, LSM-KinD. Through systematic architectural innovations, a holistic enhancement paradigm, from physical decomposition to semantic restoration, is constructed, providing new insights for improving image quality under extreme lighting conditions.

First, the image decomposition network precisely decomposes the reflectance and illumination components, and, through cross-level feature interaction and dynamic weight allocation mechanisms, significantly improves the decoupling accuracy of these components.Second, the illumination adjustment network introduces a collaborative architecture combining residual learning and channel attention, leveraging deep nonlinear mapping to model complex illumination distributions and achieving fine-grained, pixel-level brightness control that effectively balances local overexposure and underexposure.Finally, a dual-path U-Net-based detail restoration framework is constructed, which enhances high-frequency information reconstruction in extremely dark regions through dense feature reuse and a mixed attention gating strategy, thereby overcoming the performance bottleneck of traditional methods in weak-light detail recovery.Experimental results show that, through a comparative study of low-light images enhanced by LSM-KinD and other state-of-the-art low-light enhancement algorithms using object detection models, the proposed method effectively avoids the overexposure and distortion commonly encountered in image processing.

2 Method

2.1 Enhancement Model

This paper proposes an improved KinD-based low-light image enhancement model, named LSM-KinD, which consists of an image decomposition network, an illumination adjustment network, and an image restoration network. The overall network architecture of the LSM-KinD model is illustrated in Figure 1. By leveraging multi-scale feature extraction, feature fusion, and attention mechanisms, the model addresses common issues in low-light images, such as detail loss, blurring, and increased noise.

2.2 Image Decomposition Network

2.2.1 Theoretical Framework

The network is based on a image enhancement architecture using Retinex theory, employing a dual decomposition mechanism to separate the input low-light image into the reflectance component R and the illumination component L, as shown in Formula (1).

$$I = R \times L \tag{1}$$

The core technology comprises the following three modules:

(1) Feature extraction based on residual learning and multi-scale attention mechanisms[21], which captures detail features at different scales through multi-layer convolutions.

(2) Fusion of feature information across multiple scales.

(3) A channel attention mechanism to dynamically adjust feature weights.

This fully convolutional structure effectively enhances the visual quality of low-light images while suppressing noise interference.

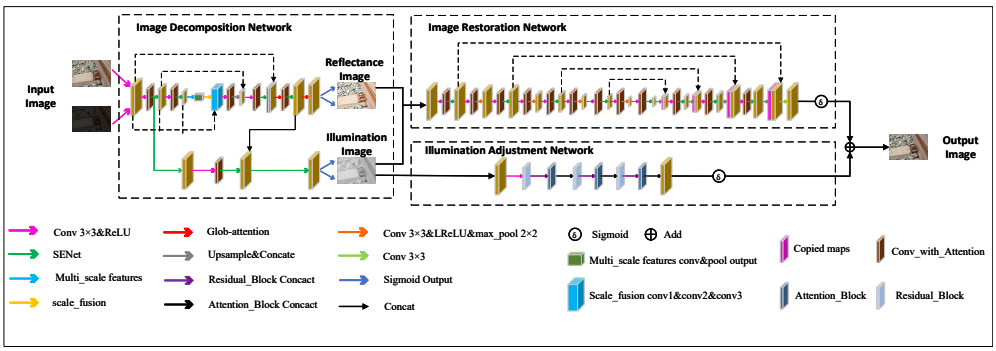


Fig. 1 Overall Architecture of the Network Model

2.2.2 Optimization Process

To improve the decomposition quality, the decomposition network undergoes significant architectural improvements in three key aspects:

(1) Activation function optimization: Leaky ReLU (hereafter referred to as "LR") is used to replace the standard ReLU, as it enables faster convergence and effectively preserves gradient flow for negative inputs. Its mathematical expression is given in Equation (2).

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ ax, & \text{otherwise} \end{cases} \tag{2}$$

In the equation, a is a slope parameter less than 1. By retaining this parameter in the negative region, it addresses the issue where the ReLU function outputs zero for negative inputs, thereby avoiding zero gradients that freeze weight updates[22].

(2) Multi-scale feature fusion: A transposed convolution cascade structure is constructed, achieving hierarchical fusion of multi-resolution features through feature upsampling and cross-layer concatenation. Let the network contain S convolution blocks for S scales, and the output of the feature map F_S at the $S-th$ scale is given by equation (3):

$$F_S = LR\left(W_s * X_{S-1} + b_S\right), \quad X_S = F_S \tag{3}$$

Where $*$ denotes 2D convolution,W_s and, b_S are the convolution weights and biases. LR is used after each layer in equation (2), which preserves gradient flow for negative activations and reduces the likelihood of neuron deactivation. The multi-scale fused features are given by equation (4):

$$F_{\text{fusion}} = \emptyset(F_1, F_2, \ldots, F_S) \tag{4}$$

(3) Channel attention module: Global average pooling is used to compress and obtain the descriptor:$z = GAP(F_fusion)$,A LR dimensionality reduction fully connected layer:$h = W_2LR(W_1GAP(F_fusion)+b_1)$ is then applied, followed by a Sigmoid function to output the channel weights: $a = \sigma(W_2h+b_2)$,generating a normalized attention vector. This vector is multiplied with the original feature map, allowing the network to adaptively focus on key channel features, ultimately enhancing the overall performance of the model. The process is shown in equation (5):

$$F'_{\text{fusion}} = F_{\text{fusion}} \times \sigma\left(W_2 h + b_2\right) \quad (5)$$

2.2.3 Loss Function

The loss function for training the decomposition network includes mutual information loss, total variation loss, gradient consistency loss, color consistency loss, L1 loss, and pyramid loss. The total loss function for the decomposition network is shown in equation (6).

$$L_{\text{DEC}} = 0.2L_{\text{MUT}} + 0.15L_{\text{i_mut}} + 0.01L_{\text{tol}} + 0.05L_{\text{col}} + 0.1L_{\text{py}} \quad (6)$$

Where L_{MUT} represents the mutual information loss,$L_{\text{i_mut}}$ represents the input mutual information loss, L_{tol} represents the total variation loss, L_{col} represents the color consistency loss, L_{py} represents the pyramid loss.

Mutual Information Loss is used to measure the degree of shared information between the reflectance image and the input image. By maximizing mutual information, details and features obtained from the decomposition process are preserved, thereby improving restoration accuracy. Input Mutual Information Loss maintains the consistency of input image information to ensure that the reflectance and illumination components retain important information after decomposition. Total Variation Loss smooths the image to reduce noise while preserving edges and structural details, resulting in a clearer image. Color Consistency Loss ensures that the colors of the decomposed image match those of the input image, preventing color distortion and maintaining visual realism. Pyramid Loss effectively captures multi-scale features through multi-scale processing, ensuring consistent fidelity at both global and local levels, thereby enhancing visual quality.

2.3 Illumination Adjustment Network

2.3.1 Theoretical Framework

This network adopts a modular architecture to achieve precise brightness control. The core structure consists of two parts: the Residual Block and the Attention Block (as shown in Figures 2 and 3). The model employs a dual-stream input mechanism to receive the initial illumination image and brightness scaling parameters, and combines these two information sources into a feature input through channel concatenation.

2.3.2 Optimization Process

The network structure incorporates the following key improvements:

(1) Encoder Design: A three-level progressive feature enhancement strategy is employed. The first-level features undergo convolution for basic feature mapping. Subsequently, the first residual block processes features through two parallel 3×3 convolutional paths, followed by normalization and LR activation. A cross-layer concatenation mechanism maintains feature transfer stability across scales. The second-level processing integrates a channel attention module (inspired by Squeeze-and-Excitation blocks) for feature refinement. Global average pooling extracts channel-wise statistics, and a two-layer fully connected network generates channel attention weights to dynamically calibrate feature channels. Finally, the deep feature refinement layer, composed of the third residual block and third attention block, progressively

enhances the network's capacity to represent higher-level semantic features of the illumination distribution. The adjustment of the illumination component is formalized in Equation (7):

$$L_{\text{ad}} = LR\left(W_3 \cdot L_i + b_3\right) \cdot A_c + \Delta L \quad (7)$$

Where L_i is the input illumination image; W_3 and b_3 are the learnable parameters of the residual block; A_c is the attention weight obtained through the channel attention module; ΔL is the brightness scaling parameter, representing the adjustment of the illumination intensity. This process can be represented by equation (8):

$$L_f = \sigma\left(W_4 \cdot L_{\text{ad}} + b_4\right) \quad (8)$$

Where W_4 and b_4 are the learnable parameters of the convolution layer in the decoder, and σ is the Sigmoid activation function.

(2) Decoder Design: Optimized high-dimensional features are mapped to the illumination adjustment space via the $conv_finalconvolutional$ layer. This layer applies a Sigmoid activation to constrain output values within $[0, 1]$, ensuring the enhanced image remains physically plausible.

2.3.3 Loss Funtion

The training of the illumination adjustment network employs multiple loss functions for comprehensive optimization, including gradient loss, mean square error(MSE) loss, structural similarity (SSIM) loss, and KL divergence loss. The total loss function for the illumination adjustment network is formulated in Equation (9):

$$L_{\text{ADJ}} = L_{\text{grad}} + L_{\text{ssim}} + L_{\text{squ}} + L_{\text{kl}} \quad (9)$$

Where L_{grad} represents the gradient loss function,L_{squ} represents the MSE loss function,L_{ssim} represents the structural similarity index(SSIM) loss function, and L_{kl} represents the KL divergence loss function. Gradient Loss is used to preserve the edge and structural information of the image, ensuring that the adjusted image maintains preserves edge and structural details under illumination changes. MSE Loss minimizes the difference between the predicted illumination and the target illumination, enabling the model to adjust the image brightness more accurately. SSIM Loss measures the similarity in structural information between the output image and the reference image, ensuring structural consistency during the illumination adjustment process. KL Divergence Loss aligns the statistical distributions, improving the distribution alignment of the illumination image and ensuring that the predicted distribution is as close as possible to the target distribution.

2.4 Image Restoration Network

2.4.1 Theoretical Framework

This network achieves precise control over image restoration through a collaborative mechanism that integrates physical modeling and data-driven learning. The architecture employs a dual-stream input encoding module and a single-stream decoding pathway. The core design concatenates the reflectance map and illumination map along the channel dimension to form a multi-channel input feature map, explicitly separating the image's

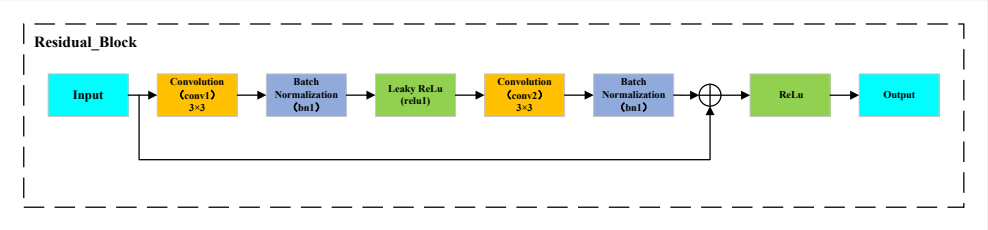


Fig. 2 Structure of the Residual Block

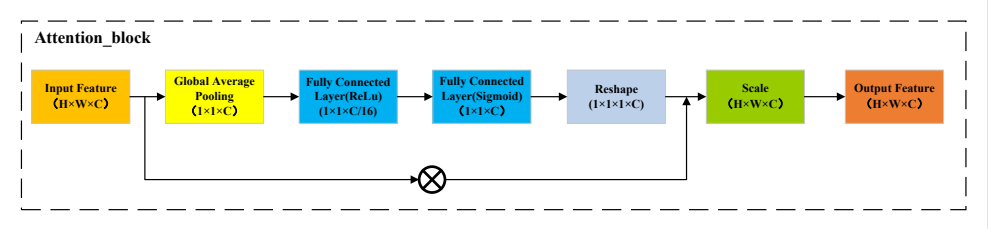


Fig. 3 Structure of the Attention Block

illumination attributes from its material features. This framework retains the physical interpretability of Retinex theory while leveraging deep learning to model complex illumination patterns.

2.4.2 Optimization Process

(1) Encoder Design: The encoder employs a progressive multi-stage feature extraction architecture. The reflectance map and illumination map are concatenated along the channel dimension, followed by two 3×3 convolutional layers (each with 32 channels) to extract initial features. Subsequently, three downsampling blocks are applied, each comprising a 3×3 convolution (with channel counts increasing from 64 to 128 to 256) and a 2×2 max-pooling layer, reducing spatial resolution to 1/8 of the original. Within each downsampling block, a channel attention module dynamically adjusts feature importance via global average pooling and two fully connected layers, enhancing critical details for restoration. The initial feature representation derived from the concatenated input is formalized in Equation (10):

$$R_{\text{ini}} = \text{Conv}(R_i, L_i, W_{\text{conv1}}, b_{\text{conv1}}) \quad (10)$$

Where R_i is the reflectance image; W_{conv1} and b_{conv1} are the weights and biases of the convolution layer; then, downsampling is performed and the feature map is further compressed through the convolution layer, as shown in equation (11):

$$R_{\text{down}} = \text{Conv}\left(\text{Pool}(R_{\text{ini}}), W_{\text{conv2}}, b_{\text{conv2}}\right) \quad (11)$$

(2) Decoder Design: The decoder implements progressive reconstruction through transposed convolution and skip connections. Starting from the top-level

encoder features, three upsampling stages are applied, each using a 2×2 transposed convolution to restore spatial resolution to the original scale. At each stage, corresponding encoder features are fused via channel concatenation and refined through 3×3 convolutions (with channel counts decreasing from 128 to 64 to 32). A final 1×1 convolution reduces channels to 3, followed by Sigmoid activation to constrain output to [0, 1], ensuring physically plausible illumination values. The progressively restored image is generated through sequential convolution operations, formalized in Equation (12):

$$R_{\text{up}} = \text{Conv}\left(\text{USp}(R_{\text{down}}), W_{\text{conv3}}, b_{\text{conv3}}\right) \tag{12}$$

Where $\text{USp}(R_{\text{down}})$ denotes the upsampling of R_{down} through a deconvolution operation to restore the resolution, and W_{conv3} and b_{conv3} are the weights and biases of the convolution layer.

2.4.3 Loss Function

This network uses several loss functions to optimize the model's training process and improve reflectance restoration. These include gradient loss, structural similarity loss, contrast loss, Total Variation(TV) loss, and MSE loss. The combined loss function for the reflectance restoration network is shown in equation (13):

$$L_{\text{RES}} = L_{\text{grad_r}} + L_{\text{ssim_r}} + L_{\text{squa_r}} + L_{\text{cont}} \tag{13}$$

Where $L_{\text{grad_r}}$ represents the gradient loss function, $L_{\text{ssim_r}}$ represents the SSIM loss function, L_{cont} represents the contrast loss function, and $L_{\text{squa_r}}$ represents the MSE loss function.

Gradient Loss ensures that the restored reflectance image retains edge and texture details, preventing detail loss during the restoration process.SSIM Loss ensures structural similarity between the output and the original image, making the enhanced image appear more natural.MSE Loss measures the pixel differences between the restored result and the target image, improving restoration accuracy.Contrast Loss enhances the contrast in local regions, improving the image's depth perception and visual layers.TV Loss suppresses noise during the restoration process, ensuring that the output image remains clear while retaining details.

The LSM-KinD method proposed in this paper, through the synergistic effect of these three sub-networks, effectively enhances illumination and clarity of low-light images, addressing issues such as noise, detail loss, and uneven illumination. It significantly enhances the image processing performance in low-light environments and is suitable for object detection and image enhancement tasks.

3 Experiment

3.1 Dataset

To validate the LSM-KinD model, two datasets are employed: the Railway Fastener Dataset (RAILF) and the publicly available Low-Light Image Dataset (LOL). In low-light enhancement research, high-quality paired images under both low-light and normal-light conditions are essential for supervised learning. While existing benchmarks (e.g., LOL) provide general-purpose low-light samples, domain-specific datasets for railway fastener applications remain scarce. To address this gap, we construct the RAILF dataset with 4000 in-situ

captured images (2000 low-light + 2000 normal-light), forming 1800 paired low-light/normal-light samples for training and 200 pairs for testing. The dataset is split into training-validation sets with an 8:2 ratio to ensure robust model evaluation. All images are resized to 600×400 pixels with RGB channels. For cross-domain performance assessment, the LOL dataset is selected as a benchmark for comparative experiments. During LSM-KinD training, 1800 paired low-light/normal-light samples from RAILF are randomly sampled for supervised learning, while the remaining 200 pairs serve as the test set.

In the application of object detection, low-light images from real outdoor working environments were selected from the RAILF dataset. Data augmentation techniques such as flipping, scaling, translation, and random cropping[23] were applied to expand the dataset to 3000 images. These augmentation techniques not only enriched the dataset but also simulated different lighting, angles, and background conditions, significantly improving the model's generalization ability and robustness, thereby enhancing the recognition accuracy of railway fasteners in complex scenes.

3.2 Experiment Setup

For the LSM-KinD model, experiments were conducted in an environment with an Nvidia RTX 4090D GPU. During training and testing of the dataset, the input image size was set to 600×400 pixels to balance training efficiency and model performance. In the low-light decomposition network module and illumination adjustment module, the batch size was set to 10, and the image block size was set to 48. In the reflectance restoration network, the batch size was set to 16, and the image block size was set to 384. During training, the initial learning rate for the optimizers of the decomposition network, adjustment network, and restoration network was set to 0.0001, with a linear learning rate schedule applied. Specifically, after 800 iterations, the learning rate was adjusted to half of the initial learning rate; after 1250 iterations, it was adjusted to one-quarter of the initial learning rate; and finally, after 1500 iterations, it was adjusted to one-tenth of the initial learning rate. The total training iterations were 2000, with 1000 iterations for the reflectance restoration network. A larger learning rate was set in the early stages to ensure fast convergence, while the learning rate was reduced later to avoid oscillations during training. The object detection model used the YOLOv5s model, and experiments were conducted in the Nvidia RTX 4090D GPU environment. The batch size was set to 16, and the SGD (stochastic gradient descent) strategy was used for network optimization, with a total of 300 epochs for training.

3.3 Low-Light Enhancement Results

The LSM-KinD algorithm was experimentally compared with several other representative low-light image enhancement algorithms, such as LIME[24], Retinex, Zero-DCE[25], Enlighten-GAN[26], and MBLLEN[27]. The LSM-KinD algorithm was tested on both the RAILF railway fastener dataset and the LOL publicly available low-light image dataset, and the corresponding image evaluation metrics, including PSNR, SSIM, MSE, LPIPS, and NIQE, were collected. Among these, Peak Signal-to-Noise Ratio (PSNR) and SSIM Index are two widely accepted image

quality assessment metrics, and their values are higher when the image quality is better. On the other hand, Natural Image Quality Evaluator (NIQE) is used to measure the reconstruction quality of real images, and its unique feature is that it does not require any reference image. The lower the NIQE value, the better the image quality. Additionally, MSE measures the pixel difference between the enhanced image and the reference image, and the lower the MSE, the smaller the difference between the two images, thus higher image quality. As for the Perceptual Similarity Index (LPIPS), it focuses on assessing the perceptual differences in images that are noticeable to the human eye, and a lower value indicates higher perceptual similarity between the two images. The quantitative experimental results of different low-light image enhancement algorithms on the RAILF railway fastener dataset for PSNR, SSIM, MSE, LPIPS, and NIQE are shown in Table 1. The comparison results of different low-light image enhancement algorithms on the LOL low-light image public dataset are shown in Table 2.

From the experimental results in Table 1, it can be seen that on the RAILF dataset fastener dataset, the proposed method outperforms other enhancement algorithms in all four evaluation metrics. Specifically, the PSNR of the proposed method is 3.141 higher than the second-best method, EnlightenGAN, SSIM is 0.084 higher, MSE is 274.613 lower, and LPIPS is 0.102 lower. Additionally, the NIQE is 0.224 higher than the second-best method, LIME. To further verify the generalization ability and feasibility of the LSM-KinD algorithm, a quantitative comparative experiment was conducted on the LOL dataset. From the experimental results in Table 2, it can be seen that the four evaluation metrics of the proposed method still outperform other enhancement algorithms on the LOL dataset, further validating the generalization ability and feasibility of the model.

To further verify the feasibility and generalization of the proposed LSM-KinD algorithm for low-light image enhancement in railway fastener scenarios, a visualization-based comparison with state-of-the-art methods was conducted. The experiments utilized the RAILF dataset (Railway Fastener Dataset) and the LOL dataset (Low-Light Image Dataset). Figures 5 and 6 present the visual comparison results of various low-light image enhancement methods on these datasets.

From the visual experimental results in Figure 4, it can be observed that methods (b) and (c) suffer from overexposure and color distortion, especially in the enhancement of texture details and brightness in extremely dark areas, where performance is poor. Other methods, such as (d), (f), and (g), fail to clearly display the details and texture structure of the fastener in the red-boxed area, and cannot clearly show the details of the cushion near the rail. In contrast, the method (e) demonstrates some cushion information but still has blurred areas in the extremely dark regions, with noticeable color differences and distortions. In comparison, the visual results of the proposed LSM-KinD method do not show color distortion, and the enhancement results are closer to natural light, especially showing significant improvement in recovering the illumination and texture details in the extremely dark areas. Additionally, no overexposure or noise issues appear in the visual results, and the color and texture details are

Table 1 Quantitative Comparison Results on the RAILF Railway Fastener Dataset

Dataset	Method	PSNR↑	SSIM↑	MSE↓	LPIPS↓	NIQE↓
RAILF	LIME	19.568	0.793	718.173	0.309	**3.677**
	MBLLEN	17.453	0.778	1168.915	0.373	4.769
	Retinex	17.297	0.660	1211.43	0.705	4.596
	Zero-DCE	16.911	0.730	1324.420	0.635	6.765
	EnlightenGAN	20.780	0.817	545.951	0.218	4.021
	KinD	17.731	0.778	1124.316	0.236	4.720
	Ours	**23.921**	**0.901**	**271.338**	**0.116**	3.901

Table 2 Quantitative Comparison Results on the LOL Dataset

Dataset	Method	PSNR↑	SSIM↑	MSE↓	LPIPS↓	NIQE↓
LOL	LIME	18.844	0.732	1344.584	0.344	**3.620**
	MBLLEN	17.658	0.787	1114.896	0.247	5.358
	Retinex	17.428	0.623	1753.209	0.432	4.494
	Zero-DCE	15.706	0.715	1747.864	0.442	3.689
	EnlightenGAN	19.660	0.817	905.238	0.256	4.458
	KinD	17.472	0.657	1823.450	0.389	4.985
	Ours	**23.739**	**0.929**	**274.895**	**0.082**	3.698

well preserved, with good illumination recovery, no overexposure, noise suppression, and excellent retention of color and details. This demonstrates that the proposed image enhancement model has significant advantages and feasibility for low-light image enhancement tasks on the RAILF railway fastener dataset.

From the visual experimental results in Figure 5, it can be observed that for the LOL dataset, methods (b) and (g) also suffer from overexposure and color distortion in the enhanced images. Compared to the proposed method, method (g) exhibits a darker region in the rail, where the clarity and contrast of the restored edge details and texture details are not as good as those of the proposed method. The enhancement results of methods (c), (d), and (f) show darker colors in the hole in the rail, with unclear contour details inside the door. Additionally, method (e) shows block effect blur in the enhanced results. The visual experimental results on the LOL dataset further demonstrate that the proposed method not only excels in low-light image enhancement of railway fasteners but also proves the superiority and generalization ability of the method in the field of low-light image enhancement.

3.4 Object Detection Performance Evaluation

To comprehensively evaluate the effectiveness of the proposed LSM-KinD model in enhancing low-light railway fastener images, and its improvement in object detection accuracy, comparative

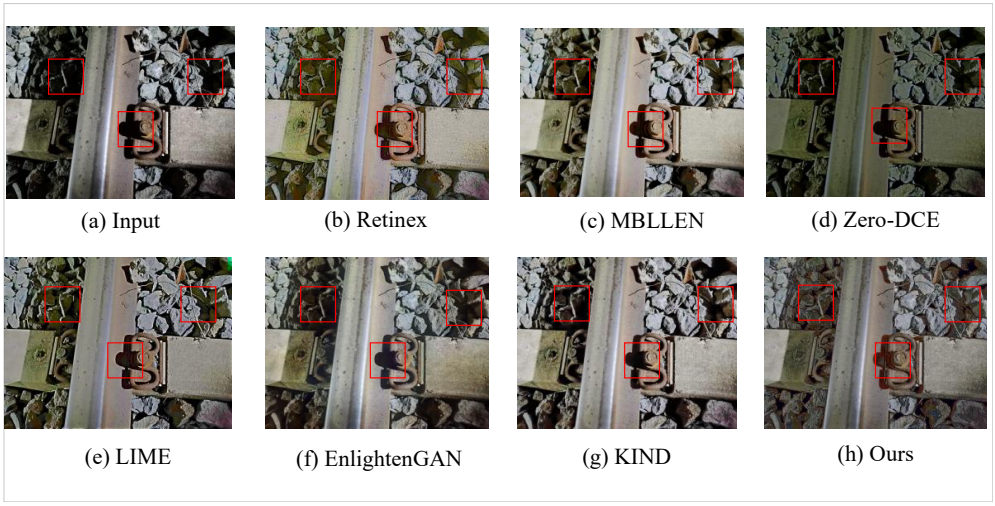


Fig. 4 Visualization Comparison of Different Low-Light Enhancement Methods on the RAILF Railway Fastener Dataset



Fig. 5 Visualization Comparison of Different Low-Light Enhancement Methods on the LOL Low-Light Image Dataset

experiments were conducted by contrasting it with other enhancement algorithms. These experiments aim to demonstrate the model's performance in object detection under low-light conditions for railway fasteners, through image preprocessing and performance evaluation of downstream object detection tasks.

The experimental results were meticulously analyzed with both subjective qualitative evaluation and objective quantitative analysis, aimed at thoroughly verifying and assessing the proposed model's object detection performance in low-light conditions for railway fasteners. Figure 6 shows the comparison of performance in object detection tasks between the proposed model, current mainstream image enhancement algorithms, and four recognition algorithms (YOLOv8s, SSD300, RetinaNet, and YOLOv5s). Specifically, this paper not only tests the four recognition algorithms mentioned above but also provides a detailed performance comparison analysis between the proposed model and four detection algorithms. From Figure 6, it can be seen that the proposed enhancement algorithm performs excellently in terms of detection accuracy, especially when combined with YOLOv8s, where the mAP50 reaches 84.9%, significantly outperforming all other comparison algorithms. This result strongly demonstrates the effectiveness and superiority of the proposed model in railway fastener object detection under low-light conditions.

To further validate the generality and transferability of the proposed enhancement model, three different object detection models (SSD300, RetinaNet, and YOLOv5s) were selected, and the proposed model was applied as preprocessing for object detection experiments on enhanced illumination images. The experimental results show that, even in different object detection algorithms, the detection accuracy of the proposed model still outperforms other enhancement algorithms, further verifying the effectiveness of the enhancement model in improving the quality of low-light railway fastener images, as well as its consistent performance improvement across various object detection algorithms. To visually present the performance of different algorithms in improving image quality and object detection accuracy, the YOLOv8s object detection model was selected, and some visual experimental results of each image enhancement model in the object detection task are shown in Figure 7.

3.5 Ablation Study

To verify the effectiveness of the three modules:decomposition network, illumination enhancement network, and reflectance restoration network in low-light image enhancement, ablation experiments were conducted on the RAILF and LOL datasets, with the results shown in Tables 3 and 4. The experiments separately analyzed the impact of multi-scale feature fusion, channel attention mechanism, and the combined loss function in the decomposition network; the influence of attention blocks, residual blocks, and combined loss in the illumination enhancement network; and the effect of the channel attention mechanism and combined loss in the reflectance restoration network. Common image evaluation metrics such as PSNR and SSIM were selected for analysis.

As summarized above, it is visually evident that the absence of any network component has varying degrees of impact on the final enhanced image's quality metrics. This further demonstrates the necessity and effectiveness of each part of the network structure.

4 Conclusion

(1) This paper proposes the LSM-KinD model for enhancing railway fastener

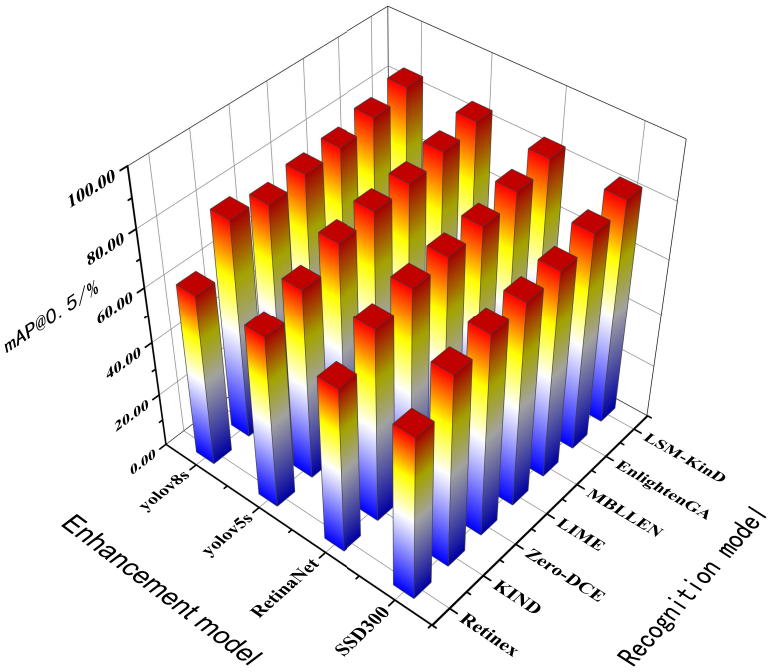


Fig. 6 Performance Comparison of Different Recognition Models and Mainstream Image Enhancement Algorithms in Object Detection Tasks

Table 3 Quantitative Ablation Comparison Results on the RAILF Dataset

Network module	Method	Structural improvement	Comprehensive loss	PSNR↑	SSIM↑
Decomposition network	W/O			13.320	0.697
	W/O	✓(Channel Attention + Multi-Scale)		13.829	0.789
	W/O		✓	14.012	0.791
	Ours	✓(Channel Attention + Multi-Scale)	✓	14.456	0.805
Illumination adjustment network	W/O			14.456	0.805
	W/O	✓(Residual Block + Attention Block)		19.837	0.865
	W/O		✓	19.477	0.816
	Ours	✓(Residual Block + Attention Block)	✓	20.102	0.843
Image restoration network	W/O			20.102	0.843
	W/O	✓(Channel Attention)		22.847	0.850
	W/O		✓	23.024	0.856
	Ours	✓(Channel Attention)	✓	23.821	0.901

images under low-light conditions. The model integrates the SENet attention mechanism and multi-scale residual connections in KinD, effectively separating the reflectance and illumination components while improving the accuracy of the decomposition process. The illumination adjustment network precisely controls the illumination intensity through residual blocks and channel attention mechanisms, while the image restoration network combines the channel attention mechanism with the U-Net architecture to optimize the enhancement effects.



Fig. 7 Visualization of Object Detection Results

Table 4 Quantitative Ablation Comparison Results on the LOL Dataset

Network module	Method	Structural improvement	Comprehensive loss	PSNR↑	SSIM↑
Decomposition network	W/O			13.472	0.657
	W/O	✓(Channel Attention + Multi-Scale)		13.531	0.660
	W/O		✓	13.722	0.665
	Ours	✓(Channel Attention + Multi-Scale)	✓	13.619	0.672
Illumination adjustment network	W/O			13.619	0.672
	W/O	✓(Residual Block + Attention Block)		16.993	0.775
	W/O		✓	18.024	0.798
	Ours	✓(Residual Block + Attention Block)	✓	19.793	0.822
Image restoration network	W/O			19.793	0.822
	W/O	✓(Channel Attention)		21.847	0.853
	W/O		✓	22.524	0.898
	Ours	✓(Channel Attention)	✓	23.739	0.929

(2) Experimental results show that the model effectively avoids overexposure and color distortion in the enhanced images, with particularly excellent detail performance in extremely dark areas. When combined with YOLOv8s for object detection, the mAP@0.5 in low-light environments reached 84.9%, demonstrating strong generalization ability.

(3) The model outperforms existing enhancement algorithms across multiple evaluation metrics, including PSNR, SSIM, MSE, LPIPS, and NIQE. Ablation experiments confirm the necessity of each sub-network structure, and the absence of any component negatively impacts the final image quality. This indicates that the LSM-KinD model has significant practical application value and development prospects in low-light image enhancement and object detection.

Author Contributions

Y.C contributed to the overall framework design, conducted model training

and performance analysis, and drafted the manuscript. J.K participated in model design and experimental setup, and assisted in data preprocessing and result evaluation. X.D assisted in model implementation and visual result generation. X.L contributed to data collection and field validation of enhancement performance. F.G participated in experimental environment setup and dataset labeling. C.C supported evaluation metric design and helped improve manuscript quality. W.D provided support in algorithm optimization and literature review. X.Y supervised the entire research process, guided the algorithm design, and finalized the manuscript as the corresponding author.

Funding

This work was supported by the Science and Technology Research and Development Plan of China Railway Wuhan Bureau Group Co., Ltd. (21G11), and the High-Level Talent Fund of Hubei University of Technology (XJ2025001402).

Declarations

The authors declare no competing interests.

References

[1] Bai, T., Yang, J., Xu, G., et al.: Research on railway fastener positioning based on Faster R-CNN. *J. Rail Sci. Eng.* **18**(2), 502–508 (2021)

[2] Tsai, T.H., Wang, R.Z., Tung, N.C.: Hardware architecture design for real-time SIFT extraction with reduced memory usage. *Multimed. Tools Appl.* **83**(2), 6297–6317 (2024)

[3] He, Y., Huang, J., Pan, Y.: A novel low-resource consumption and high-speed hardware implementation of HOG feature extraction on FPGA for human detection. *Integration* **97**, 102208 (2024)

[4] Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587 (2014)

[5] Redmon, J.: You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016)

[6] Liu, W., Anguelov, D., Erhan, D., et al.: SSD: Single Shot MultiBox Detector. In: *Computer Vision – ECCV 2016*. Lecture Notes in Computer Science, vol. 9905, pp. 21–37. Springer, Cham (2016)

[7] Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988 (2017)

[8] Lin, T.Y., Dollár, P., Girshick, R., et al.: Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125 (2017)

[9] He, K., Gkioxari, G., Dollár, P., et al.: Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969 (2017)

[10] Carion, N., Massa, F., Synnaeve, G., et al.: End-to-end object detection with transformers. In: *European Conference on Computer Vision*. Lecture Notes in Computer Science, vol. 12346, pp. 213–229. Springer, Cham (2020)

[11] Huang, S.C., Cheng, F.C., Chiu, Y.S.: Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Trans. Image Process.* **22**(3), 1032–1041 (2012)

[12] He, J., He, X., Zhang, M., et al.: Deep dual-domain semi-blind network for compressed image quality enhancement. *Knowl.-Based Syst.* **238**, 107870 (2022)

[13] Zhuang, P., Wu, J., Porikli, F., et al.: Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Trans. Image Process.* **31**, 5442–5455 (2022)

[14] Ko, S., Park, J., Chae, B., et al.: Learning lightweight low-light enhancement network using pseudo well-exposed images. *IEEE Signal Process. Lett.* **29**, 289–293 (2021)

[15] Wei, C., Wang, W., Yang, W., et al.: Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560* (2018)

[16] Wang, M., Zhang, H., Li, J., et al.: A deep neural network-based enhancement algorithm for low-light images in underground coal mines. *Coal Sci. Technol.* **51**(9), 231–241 (2023)

[17] Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: a practical low-light image enhancer. In: *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1632–1640 (2019)

[18] Zhang, Y., Guo, X., Ma, J., et al.: Beyond brightening low-light images. *Int. J. Comput. Vis.* **129**, 1013–1037 (2021)

[19] Land, E.H., Hubel, D.H., Livingstone, M.S., et al.: Colour-generating interactions across the corpus callosum. *Nature* **303**(5918), 616–618 (1983)

[20] Zhang, H., Wang, X., Wang, H., et al.: Improved Retinex-Net image enhancement method based on luminance component. *Acta Phys. Sin.* **71**(11), 107–115 (2022)

[21] Hu, Z., Ding, J., Tian, B.: Retinex low-light image enhancement algorithm based on residual and multiscale attention mechanism. *Sci. Technol. Eng.* **24**(35), 15118–15125 (2024)

[22] Xu, J., Wang, F.: Image classification method based on improved S-ReLU activation function. *Sci. Technol. Eng.* **22**(29), 12963–12968 (2022)

[23] Liu, F., Xing, C., Wei, G., et al.: An improved YOLOv8 method for object detection in underwater sonar images. *Appl. Sci. Technol.* **52**(1), 34–40 (2025)

[24] Guo, X., Li, Y., Ling, H.: LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **26**(2), 982–993 (2016)

[25] Guo, C., Li, C., Guo, J., et al.: Zero-reference deep curve estimation for low-light image enhancement. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1780–1789 (2020)

[26] Jiang, Y., Gong, X., Liu, D., et al.: EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **30**, 2340–2349 (2021)

[27] Lv, F., Lu, F., Wu, J., Lim, C.S.: MBLLEN: Low-light image/video enhancement using CNNs. In: *Proceedings of the British Machine Vision Conference (BMVC)* (2018)

References