processes

MDPI

Article

A Transformer-Based Low-Light Enhancement Algorithm for Rock Bolt Detection in Low-Light Underground Mine Environments

Wenzhen Yan [1,2], Fuming Qu [1,*], Yingzhen Wang [1,2], Jiajun Xu [1,2], Jiapan Li [1,2] and Lingyu Zhao [1,3]

[1] Institute of Minerals Research, University of Science and Technology Beijing, Beijing 100083, China
[2] School of Resources and Safety Engineering, University of Science and Technology Beijing, Beijing 100083, China
[3] School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China
* Correspondence: qufuming@ustb.edu.cn

Abstract

Underground roadway support is a critical component for ensuring safety in mining operations. In recent years, with the rapid advancement of intelligent technologies, computer vision-based automatic rock bolt detection methods have emerged as a promising alternative to traditional manual inspection. However, the underground mining environment inherently suffers from severely insufficient lighting. Images captured on-site often exhibit problems such as low overall brightness, blurred local details, and severe color distortion. To address the problem, this study proposed a novel low-light image enhancement algorithm, PromptHDR. Based on Transformer architecture, the algorithm effectively suppresses color distortion caused by non-uniform illumination through a Lighting Extraction Module, while simultaneously introducing a Prompt block incorporating a Mamba mechanism to enhance the model's contextual understanding of the roadway scene and its ability to preserve rock bolt details. Quantitative results demonstrate that the PromptHDR algorithm achieves Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) index scores of 24.19 dB and 0.839, respectively. Furthermore, the enhanced images exhibit more natural visual appearance, adequate brightness recovery, and well-preserved detailed information, establishing a reliable visual foundation for the accurate identification of rock bolts.

Keywords: low-light image enhancement; rock bolt detection; PromptHDR; transformer

check for updates

Academic Editor: Wei Sun

Received: 3 November 2025
Revised: 27 November 2025
Accepted: 2 December 2025
Published: 3 December 2025

Citation: Yan, W.; Qu, F.; Wang, Y.; Xu, J.; Li, J.; Zhao, L. A Transformer-Based Low-Light Enhancement Algorithm for Rock Bolt Detection in Low-Light Underground Mine Environments. *Processes* **2025**, *13*, 3914. https://doi.org/10.3390/pr13123914

Copyright: © 2025 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) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Mineral resources play a very important role in the world economy [1]. Meanwhile, large-scale mining operations are prone to safety issues [2]. As the depth and scale of metal mining continue to expand [3], the surrounding rock of roadways is subjected to increasingly severe conditions including high in situ stress [4], elevated temperatures, high water pressure, and mining-induced disturbances [5]. These factors make roadway support operations paramount for ensuring underground safety production [6]. The acceptance inspection of support quality is directly related to the stability of mining production and personnel safety, constituting an indispensable critical link in safety management [7]. However, traditional support acceptance primarily relies on manual methods conducted through visual inspection, manual measurement, and sounding. This approach is not only inefficient but also highly susceptible to the complex underground environment and the subjective state of inspectors, leading to difficulties in guaranteeing the accuracy and

consistency of acceptance results. As the scale of mining expands, the limitations of manual inspection become increasingly apparent, creating an urgent industry demand for the development of intelligent and automated support quality acceptance technologies.

Computer vision-based intelligent detection technology offers a viable solution to the aforementioned problems. However, the underground mine environment suffers from severely insufficient illumination [8,9]. Captured rock bolt images commonly exhibit issues such as low brightness, blurred details, and color distortion, which significantly constrain the performance of subsequent object detection algorithms. To overcome the bottleneck of low-light image quality, researchers have proposed various enhancement algorithms. Nevertheless, these methods still demonstrate notable limitations when dealing with the complex scenarios encountered underground [10,11]. That can be broadly categorized into three types:

(1) Physics-based and supervised learning methods, such as RetinexNet [12–14] and its variants, which employ CNNs to decompose images into illumination and reflectance components based on Retinex theory [15]. However, they are prone to color cast and detail blur under complex lighting conditions. RetinexFormer integrates Retinex priors with Transformers [16–18], achieving excellent enhancement effects, but its high computational cost makes real-time deployment on underground edge devices challenging. Methods like LightenNet [19] and KinD [20], while reducing computational demands, heavily rely on synthetic data, leading to poor generalization in real underground scenarios and susceptibility to artifacts.

(2) Unsupervised learning and Generative Adversarial Network (GAN)-based methods, such as EnlightenGAN [21,22], circumvent the need for paired data. However, their enhancement quality depends heavily on carefully designed prior loss functions, often resulting in under-enhancement, over-enhancement, or unstable outcomes in complex underground environments, thus exhibiting relatively poor reliability.

(3) Curve estimation and lightweight optimization-based methods, such as SCI [23] and Zero-DCE [24,25], achieve rapid enhancement by estimating mapping curves but tend to lose details and introduce color distortion under extreme low-light or high-noise conditions. RUAS [26] relies on handcrafted priors, limiting its generalization capability. Sparse-based methods are sensitive to noise and prone to artifacts. Approaches like NeRCO [27] and DRBN [28] lack stability under complex lighting, often causing local over-exposure or under-enhancement. URetinex-net [29] faces real-time performance bottlenecks. General restoration models like Restormer [30] and MIRNet [31] are not well-adapted to low-light characteristics and involve substantial computational overhead. Although PromptIR [32] shows progress in generalization and detail preservation, it still lacks specific optimization for underground low-light scenes.

In summary, existing methods struggle to achieve an effective balance between generalization capability and detail preservation, failing to fully meet the practical requirements for intelligent rock bolt detection in underground mines. Specifically, models relying on synthetic data or generic priors exhibit weak generalization in real underground environments, while those with strong detail reconstruction capabilities are often computationally complex or inadequately model low-light characteristics, making it difficult to stably retain fine rock rod features under complex illumination.

Given the practical needs of the mining industry, this study aims to design a low-light image enhancement algorithm specifically tailored for underground environments, capable of improving overall image quality while accurately preserving rock bolt details.

Low-light image enhancement algorithm has been applied to rock bolt detection in underground mines. To comprehensively evaluate the algorithm's performance, we conducted not only traditional image quality assessment (using metrics such as PSNR

and SSIM) and visual comparisons, but also designed a task-driven validation experiment based on the YOLOv8 object detection model. By comparing the mAP@0.5, Precision, and Recall metrics for bolt detection tasks on images before and after enhancement, we comprehensively demonstrate the effectiveness and practical value of our algorithm from an application perspective. The main contributions of this detection model can be summarized as follows:

(1) A Lighting Extraction Module based on the Retinex physical model is introduced, which computes illumination priors from the image and utilizes parallel convolutional layers to extract spatial features, explicitly decomposing an illumination map that characterizes the distribution of light. This mechanism provides a robust and physically meaningful illumination representation for subsequent processing.

(2) A Prompt Illumination Block is designed, which leverages illumination features as dynamic prompts to guide the model in adaptively focusing on global scene structures and local rock rod details, thereby achieving scene-aware image enhancement.

(3) A powerful sampling module is introduced, enhancing the feature extraction performance of the Transformer module. Down-sampling extracts multi-scale information, preserves global structure, and reduces noise; up-sampling balances global and local features, improving overall detail recovery.

2. The PromptHDR Algorithm

2.1. Algorithm Introduction: PromptHDR

PromptHDR primarily consists of two processing stages. First, to mitigate the common issue of color distortion in traditional methods, an Lighting Extraction Module is constructed to extract illumination features from the low-light image and reconstruct them into an illumination map. Second, aiming to tackle problems such as computational redundancy in long-range dependency modeling and weak local detail representation in the Transformer architecture, a U-shaped encoder–decoder network is designed. This network incorporates a prompt learning mechanism to enhance feature representation capability and embeds a multi-scale sampling module to improve the recovery quality of local details. Finally, by fusing the illumination map output from the illumination module with the residual features generated by the decoder, an enhanced image with superior visual quality is produced.

PromptHDR comprises the Lighting Extraction Module (LEM), the Feature Fusion Module (FFM), and the Image Reconstruction Module (IRM). The overall workflow is illustrated in Figure 1. The workflow begins by feeding the input image into the Lighting Extraction Module (LEM). This module processes the input through parallel 1×1 convolutional and 5×5 depthwise separable convolutional layers to extract spatial features, generating an illumination map (illu_map) that represents the illumination distribution and the illumination features (illu_fea) as a crucial intermediate result.

Subsequently, both the original image and the illumination features are input into the Feature Fusion Module (FFM). This module employs a U-shaped network structure containing multiple Transformer Blocks. In the encoding path, the input features are processed by the first three Transformer Blocks interleaved with HWD downsampling modules, which extract multi-scale local features while preserving global semantics. In the decoding path, the features are processed by the subsequent three Transformer Blocks and DySample upsampling modules. Throughout this stage, U-shaped skip connections concatenate shallow details with deep semantic features to enhance contextual information processing. Integrated across the FFM, the Prompt module—composed of GCP and HFP components—incorporates the illumination features from the LEM to guide, fuse, and refine the feature flow, thereby optimizing image quality and naturalness.

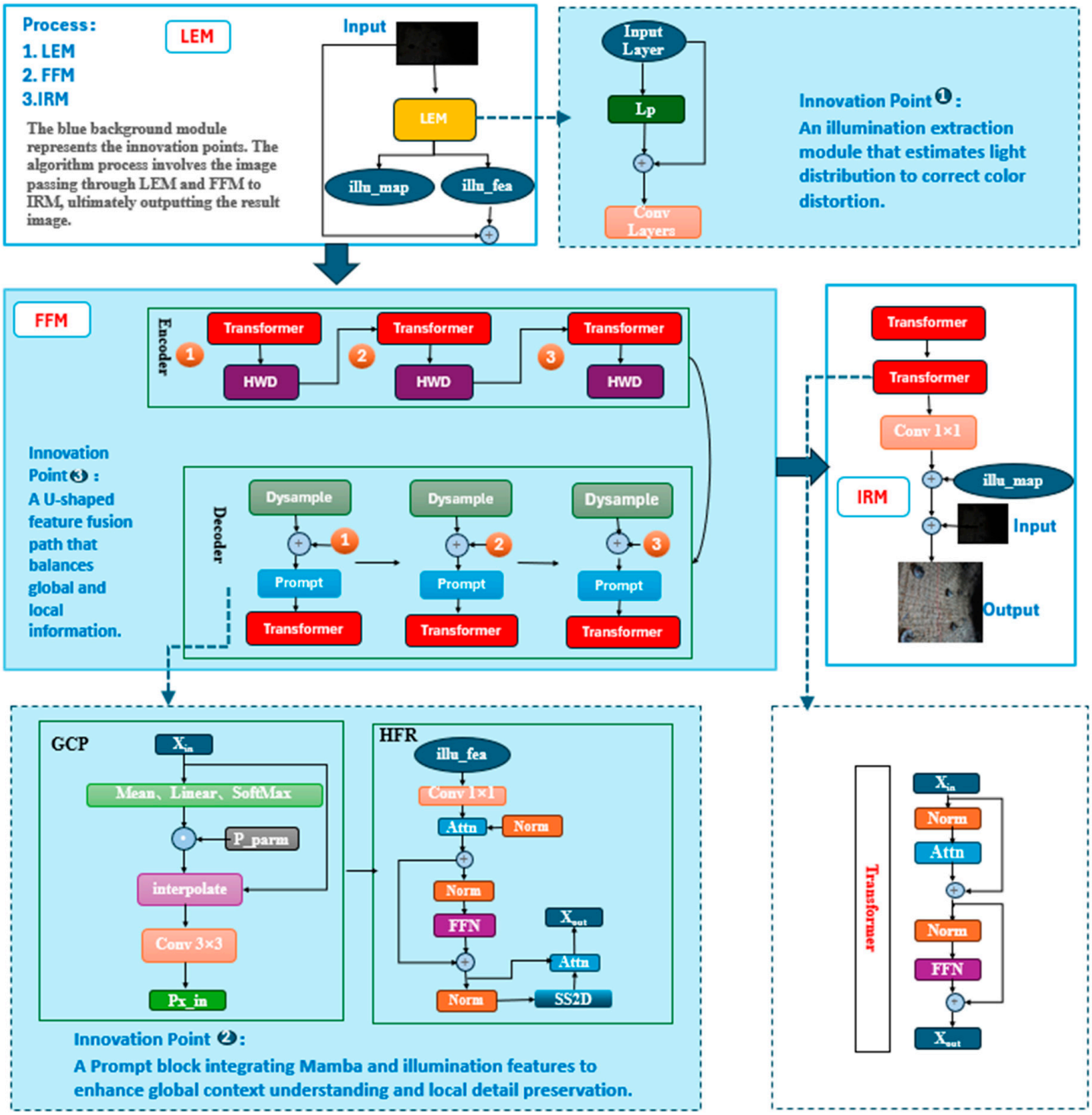


Figure 1. Flowchart of the PromptHDR Algorithm.

The processed features then enter the Image Reconstruction Module (IRM). Here, the deep features undergo further extraction via two additional Transformer Blocks. After processing by an output transformer, they are combined with the initial input image through a residual connection, ultimately producing a high-quality image with enhanced details and significantly improved illumination.

2.2. Core Component Design

2.2.1. Lighting Extraction Module

Rock bolt images captured in underground low-light environments suffer not only from insufficient overall brightness but also from complex illumination conditions characterized by intertwined local highlights and shadows, caused by point light sources and reflections from metal surfaces. This non-uniform illumination readily leads to color distortion and the submersion of local details, severely hindering the accurate judgment of rock

bolt morphology and position. To suppress these illumination disturbances at the source and provide a more robust visual representation for subsequent enhancement and detection tasks, a Lighting Extraction Module (LEM) is designed, aiming to explicitly separate and model the illumination distribution information within the image.

The detailed structure of the Lighting Extraction Module is illustrated in the LEM section of Figure 1. First, the network receives an RGB image as input. The initial step involves computing an illumination prior, achieved by averaging all color channels of the input image. Subsequently, the original RGB image and the newly computed illumination prior are concatenated. This concatenated result then passes through the first 1×1 convolutional layer for inter-channel feature reorganization. The resulting feature map is processed by a 5×5 depthwise separable convolutional layer, enabling the network to efficiently extract spatial features and capture local illumination patterns and variations. Finally, the feature map passes through a second 1×1 convolutional layer, which maps the extracted features back to the original image space, generating the final illumination map. The outputs are the illumination map (illu_map), used for feature extraction in subsequent networks, and the intermediate illumination features (illu_fea), utilized by the Prompt block for feature optimization.

Given an input RGB image I $\in$ R^(b $\times$ 3 $\times$ h $\times$ w), the final output of the Lighting Extraction Module is formulated as

$$X = f(I) = \text{Conv2}(\text{DepthConv}(\text{Conv1}(I \oplus Lp))) \tag{1}$$

where the illumination prior Lp is defined as

$$Lp = \left(\frac{1}{3}\right) * \sum(R, G, B) \tag{2}$$

where $\oplus$ denotes the concatenation operation; Conv1, DepthConv, and Conv2 represent the three convolutional operations, respectively.

This module analyzes the input image to provide the model with crucial prior knowledge of the underground illumination. This prior information directly guides the image enhancement process towards adaptive adjustment, effectively overcoming disturbances from low light and local highlights. Consequently, it presents the true details and color characteristics of the rock bolts clearly to subsequent detection models, ultimately significantly improving the accuracy of rock bolt identification and acceptance.

2.2.2. Prompt Block Based on Mamba and Illumination Features

In the low-light environment of underground roadways, rock bolt targets are often intermixed with complex backgrounds (e.g., rock walls, support nets), and their metallic properties readily cause local reflections or shadows. This poses significant challenges for traditional enhancement models in balancing global illumination and local details. To overcome this difficulty, a Prompt block is introduced. The core idea of this module is to leverage the structured prior information obtained from the Lighting Extraction Module, using "prompts" to dynamically guide the enhancement process. This enables the model to adaptively focus on both the structural context of the entire roadway and the fine-grained features of individual rock bolts (such as nuts and washers), thereby achieving more perceptually intelligent image enhancement.

The structure of the Prompt block is shown in the lower part of Figure 1. It consists of two components: Global Context Prompting (GCP) and Hierarchical Feature Refinement (HFR). The first stage focuses on capturing global information and generating prompts. By producing guided prompt information, it enhances the model's understanding of the overall

scene structure. The second stage performs multi-level feature fusion and self-attention-based refinement, deeply integrating the prompt information with the illumination features. This achieves a balance between detail and global information, ultimately generating enhanced features for image enhancement.

Integration of Mamba Mechanism: The Mamba mechanism is a novel state-space model (SSM) based architecture that offers efficient long-range sequence modeling with linear time complexity. Unlike traditional self-attention, which has quadratic complexity, Mamba uses a selective SSM that adaptively focuses on relevant information based on the input. In the Prompt block, Mamba is integrated into the HFR stage to replace the standard self-attention in some steps. This allows the model to capture global context more efficiently while reducing computational cost. Specifically, the Mamba layer processes the feature sequences by maintaining a hidden state that evolves linearly, enabling it to model long-range dependencies without the need for expensive attention computations.

The mathematical formulation of the Mamba mechanism for a feature sequence $x \in R^{L \times D}$ is as follows:

$$h_t = \overline{A}_t h_{t-1} + \overline{B}_t x_t \tag{3}$$

$$y_t = C_t h_t \tag{4}$$

where $\overline{A}_t$ and $\overline{B}_t$ are discretized parameters derived from the continuous parameters A and B using a step size Δ, and Ct is the output matrix. This formulation allows Mamba to handle sequences of length L with O(L) complexity, making it suitable for high-resolution images.

Mamba replaces the self-attention mechanism with a linear state space model (SSM) for modeling long-range dependencies. While the computational complexity of self-attention grows quadratically with sequence length ($O(L^2)$), Mamba achieves linear complexity (O(L)). In terms of parameter count, the core parameters of Mamba remain constant and are independent of sequence length, offering an advantage over the variable parameter count of self-attention.

In our implementation, the Mamba layer is applied after the feature fusion in HFR. The input features are first flattened into a sequence, processed by Mamba, and then reshaped back. This integration enhances the model's ability to capture global context while maintaining efficiency.

The final output of this module is formulated as

$$O(x_{in}, I(x)) = HFR(Cat(x_{in}, P(x_{in})), I(x_{in})) \tag{5}$$

where x_{in} denotes the input features, $I(x_{in})$ represents the illumination features generated in the first stage, and $P(x_{in})$ signifies the prompt information produced by the Prompt module.

The GCP stage is responsible for extracting global information from the input features and generating guided prompt information. Its output, $P(x_{in})$, is formulated by

$$P(x_{in}) = Conv3x3\left(Interp\left(\sum_{i=0}^{L} \omega_i \cdot P_i\right)\right) \tag{6}$$

The HFR stage is responsible for the deep integration and refinement of the prompt information from GCP, the original input features, and the illumination features. This process comprises three sub-steps:

(1) Integration of the prompt information output by the Prompt module with the illumination features:

$$I = Conv1x1(I(x)) \tag{7}$$

$$P'(x_{in}) = \text{Norm}(P(x_{in})) + \text{Attention}(\text{Norm}(P(x_{in})), I) \quad (8)$$

where I(x)I(x) denotes the illumination features and P(x)P(x) represents the input features.

(2) Enhancing the fused features via a Feed-Forward Network (FFN):

$$P''(x_{in}) = P'(x_{in}) + \text{FFN}\big(\text{Norm}\big(P'(x_{in})\big)\big) \quad (9)$$

(3) Refining the features and upsampling to restore spatial information:

$$x_{out} = P''(x_{in}) + \text{DropPath}(\text{SS2D}(\text{Norm}(P''(x_{in})))) \quad (10)$$

Here, the Mamba layer replaces the previously used SS2D operation, providing more efficient sequence modeling.

Among peer algorithms with prompt mechanisms similar to PromptHDR, Retinexformer also adopts a Transformer + CNN architecture and illumination-guided logic; however, it primarily relies on illumination-guided self-attention for single-stage training and lacks the scene adaptability enabled by dynamic prompts. PromptIR, while achieving multi-decoder-level prompt injection via a plug-and-play prompt module, lacks prior information support in its illumination processing—this results in lower accuracy in complex illumination analysis compared to PromptHDR, which leverages Retinex decomposition and explicit illumination map generation. NeRCo, on the other hand, additionally integrates MLP and CLIP modules, focusing on optimization guided by neural normalization and CLIP. This stands in distinct contrast to PromptHDR, whose design centers on the dual enhancement pathway of efficiency and accuracy driven by "dynamic illumination prompts + Mamba mechanism."

Furthermore, nine categories of algorithms without prompt mechanisms (e.g., Zero-DCE, MIRNet, and KinD) exhibit advantages in specific scenarios, such as lightweight deployment and reference-free enhancement. For instance, Zero-DCE achieves fast inference based on zero data dependency; MIRNet enhances detail representation through dual attention; and KinD supports customizable illumination adjustment. However, these algorithms generally take a single CNN (or GAN + CNN) as their core network, lacking the dynamic adaptability to complex scenes that prompt mechanisms provide. Moreover, they differ significantly from PromptHDR in terms of the explicitness of illumination processing and the dynamism of feature fusion—this further underscores PromptHDR's comprehensive advantages in complex illumination scenarios.

2.2.3. Sampling Module

Rock bolt detection in low-light underground environments requires enhancement algorithms capable of restoring both the overall image brightness and contrast while clearly preserving the fine features of the rock bolts. However, in simple encoder–decoder architectures, local detail information tends to be lost as the network depth increases during feature extraction. To address this, a "U-shaped" structure incorporating DySample upsampling and HWD downsampling modules is constructed. This aims to establish a multi-scale feature fusion pathway that connects the deep and shallow layers of the network. It ensures that during high-quality image reconstruction, uniform illumination can be restored based on global contextual information, while the local details critical for rock bolt recognition can be accurately reconstructed.

The model utilizes the DySample and HWD modules for upsampling and downsampling, respectively, to realize this "U" structure. The coordinated use of upsampling and downsampling achieves a balance between global and local features, effectively con-

trols computational costs, and avoids the efficiency issues associated with processing high-resolution data.

The DySample module implements a dynamic sampling mechanism, the core idea of which is to achieve flexible spatial transformation by learning offsets from the input features. First, the input feature map *x* undergoes two parallel convolutional operations: an offset convolution and an optional scope convolution. The offset convolution (raw_offset) employs a 1×1 convolution to generate raw offset values, while the scope convolution (raw_scope), which can be optionally enabled, is used to modulate these offsets, producing init_pos. These raw values are then processed to yield the final offsets: processed_offset. Subsequently, the input features are resampled using the grid-sample function, resulting in the upsampled features.

The overall output of the DySample module is formulated as

$$\mathrm{y} = \mathrm{DySample}(\mathrm{x}) = \mathrm{F.grid}_{\mathrm{sample(x,coords)}} \tag{11}$$

where y is the output feature map after upsampling, x is the input low-resolution feature map, $\mathrm{F.grid}_{\mathrm{sample(x,coords)}}$ is the grid sampling function in PyTorch 1.12, which extracts values from x using bilinear interpolation based on the provided sampling coordinates coords to construct the output feature map y; and coords is the final, dynamically modulated sampling coordinate grid.

The calculation formula for coords is

$$\mathrm{coords} = \frac{2 * (\mathrm{grid} + \mathrm{processed}_{\mathrm{offset}})}{[\mathrm{W}, \mathrm{H}]} - 1 \tag{12}$$

where grid is the standard pixel grid, normalized by dividing by [W, H], with its range subsequently adjusted to $[-1, 1]$.

The core concept of the HWD module is to leverage the Haar Wavelet Transform (HWT) to reduce the spatial resolution of the feature map while retaining all information, encoding partial spatial information into the channel dimension. Subsequent convolutional layers are then used to extract discriminative features and filter out redundant information.

First, the feature map is decomposed using HWT, yielding one low-frequency component (A) and three high-frequency components (H, V, D). During this process, the spatial resolution of each component is reduced to half, while the number of channels is doubled. A 1×1 convolutional layer is then applied to adjust the channel dimensionality of the feature map and filter redundant information. Finally, by incorporating Batch Normalization and the ReLU activation function, discriminative features are extracted, producing the final downsampled output.

The overall output of the HWD module is formulated as

$$\mathrm{y} = \mathrm{HWT}(\mathrm{x}) = \mathrm{ReLU}(\mathrm{BatchNorm}(\mathrm{Conv1x1}(\mathrm{x_{cat}}))) \tag{13}$$

where $\mathrm{x_{cat}}$ represents the concatenated output derived from the property of the Haar Wavelet Transform (HWT) when processing a 2D signal (such as an image or feature map): a single application of the transform yields one low-frequency component and three high-frequency components corresponding to different orientations.

$$\mathrm{x_{cat}} = \mathrm{Concat}([\mathrm{yL}, \mathrm{y_{HL}}, \mathrm{y_{LH}}, \mathrm{y_{HH}}]) \tag{14}$$

where yL represents the low-frequency coefficients, while y_{HL}, y_{LH}, and y_{HH} denote the high-frequency coefficients in the horizontal, vertical, and diagonal directions, respectively.

Through the proper use of the upsampling and downsampling modules and the implementation of the "U" structure, the model achieves simultaneous understanding of the global image context and fine-grained local prediction. Consequently, this provides the Prompt module with comprehensive information, ultimately leading to superior image enhancement results.

3. Experiment

3.1. Dataset

This study employs a self-constructed low-light dataset of underground tunnel support for training and validation of the proposed PromptHDR algorithm. The data were collected from typical tunnels in an iron mine, featuring straight-wall semi-circular arch sections with dimensions of 4.2 m (width) × 3.5 m (height). The tunnels adopt a combined support system of 'rock bolts + steel mesh + shotcrete.' The ore body exhibits thick layered distribution with an average thickness of 115 m. As mining depth increases, the rock mass conditions become increasingly complex, with background interference from cables, air ducts, and other obstructions, making the environment representative and suitable for experimental data requirements.

Data acquisition was conducted using a mine-intrinsically safe explosion-proof camera ZHS2640 (received mine product safety certification), with key parameters including an effective pixel count of 20 million, maximum resolution of 5120 × 3840, and aperture range of F2.0–F8.0, ensuring operational safety in complex underground environments. The collection process was supplemented with mine-intrinsically safe LED lighting. To ensure sufficient diversity and representativeness, the dataset covers various lighting conditions (normal, low-light, side-lighting) and background complexities (clean rock surfaces, pipeline interference), as shown in Figure 2. For each tunnel scenario, paired images were captured at a fixed camera position (without any movement)—one low-light image and one normal-light image—resulting in a total of 400 paired tunnel images. The dataset was divided into training/validation/test sets in an 8:1:1 ratio. All comparative experiments were conducted under the same partition to ensure fair comparisons.

During construction, the dataset strictly adhered to the FAIR (Findable, Accessible, Interoperable, Reusable) principles for organization and annotation. Given that the data contain sensitive geographic information from underground mine sites, they are not publicly available at this time. Dataset will be sharing upon reasonable academic requests under signed data usage agreements after publication.

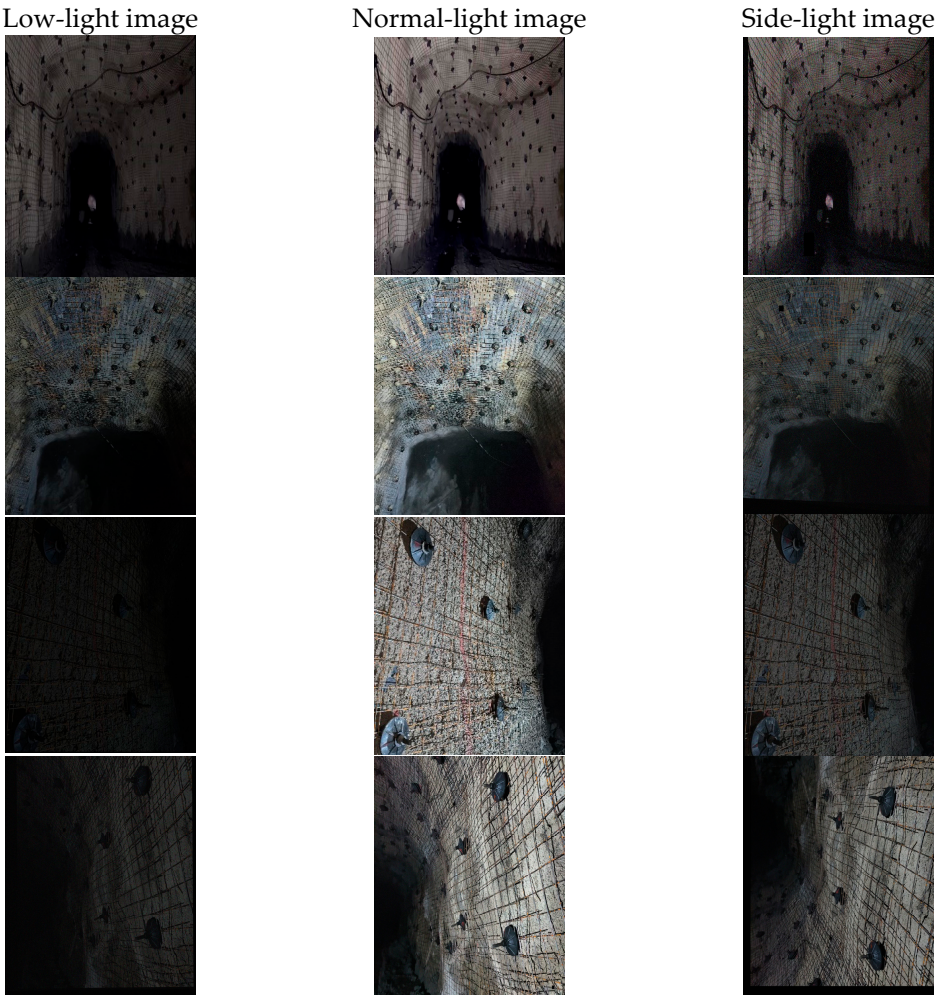


Figure 2. Selected images of the roadway low-light image dataset.

3.2. Experimental Setup and Parameters

The configuration of the experimental system is summarized in the table below (Table 1).

Table 1. System configuration.

Parameter	Value
Operating System	Windows11
CPU	i9-13900K
GPU	RTX NVIDIA A6000
CUDA	11.8
Python	3.10
Pytorch	1.12
Training Image Resolution	128 × 128
Optimizer	Adam

3.3. Evaluation Metrics

In the field of low-light image enhancement, two metrics are commonly employed to evaluate algorithm performance: SSIM and PSNR. The descriptions of these metrics are as follows:

(1) SSIM (Structural Similarity Index):

The SSIM index is designed to measure the perceptual similarity between two images. As it considers the structural information of the image, it generally aligns better with

human visual perception in quality assessment. The SSIM value ranges from −1 to 1. A value closer to 1 indicates higher similarity between the two images, a value of 0 suggests no similarity, and a value of −1 signifies that the images are completely different. The SSIM is calculated as follows:

$$\mathrm{SSIM}(X,Y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{15}$$

where μ_x and μ_γ are the mean values of image x and image y, respectively; ${\sigma_x}^2$ and ${\sigma_\gamma}^2$ represent the variances of image x and image y; $\sigma_{x\gamma}$ denotes the covariance between image x and image y; C_1 and C_2 are small constants typically set to avoid division by zero.

(2) PSNR (Peak Signal-to-Noise Ratio):

PSNR is a widely used metric for evaluating image quality, particularly in the contexts of image compression or reconstruction. It quantifies the restoration quality by measuring the difference between the original and processed images. A higher PSNR value indicates better image quality with lower distortion. Generally, values between 30 and 40 dB represent good to excellent quality, often indistinguishable to the human eye, while values between 20 and 30 dB suggest moderate quality with potential perceptible distortion. The metric is defined as follows:

$$\mathrm{MSE} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j) - K(i,j)]^2 \tag{16}$$

where I (i, j) and K (i, j) denote the pixel values at position (i, j) of the original image and the processed image, respectively.

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{R^2}{\mathrm{MSE}}\right) \tag{17}$$

where R represents the dynamic range of the image pixel values, typically set to 255.

(3) NIQE (Natural Image Quality Evaluator):

NIQE is a no-reference image quality assessment metric capable of evaluating image quality without the need for an original high-quality reference image. This metric is based on spatial natural scene statistics (NSS) features extracted from natural scenes. It quantifies the naturalness of an image by measuring the deviation between the test image and a pre-trained multivariate Gaussian model (MVG). There is no absolute normal range for the NIQE score, with a lower value indicating better image quality and a closer resemblance to a natural image. This metric is particularly suitable for image quality assessment in real-world scenarios. The specific formula is as follows:

$$NIQE = \sqrt{(v_1 - v_2)^T\left(\frac{(\sum_1 + \sum_2)}{2}\right)^{-1}(v_1 - v_2)} \tag{18}$$

where v_1 and v_2 are the mean vectors of the multivariate Gaussian models for natural and distorted images, respectively, and $\frac{(\sum_1 + \sum_2)}{2}$ represents the mean of their covariance matrices.

(4) LPIPS (Learned Perceptual Image Patch Similarity):

LPIPS is a deep learning-based perceptual similarity metric that aligns more closely with human visual judgment. Unlike traditional pixel-level difference measures, LPIPS extracts features from a pre-trained deep neural network and computes the distance in this feature space to assess the perceptual similarity between images. The LPIPS value ranges from 0 to 1, where a lower value indicates higher similarity between the two images, and a higher value signifies a greater perceptual difference. This metric effectively captures the

impact of structural distortions and textural changes on human perception. Let V represent the normal-light image and I denote the low-light enhanced image. The specific formula for LPIPS is as follows:

$$\mathrm{LPIPS}(\mathrm{V},\mathrm{I}) = \sum_i w_i \cdot |\varphi_i(\mathrm{V}) - \varphi_i(\mathrm{I})|_2^2 \tag{19}$$

where φ_i represents the feature layers of the network, and w_i denotes the weight associated with feature layer i.

(5) BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator):

BRISQUE is a no-reference image quality assessment algorithm that evaluates image quality based on spatial domain statistical features. The algorithm analyzes the locally normalized luminance coefficients of an image to extract its statistical features, and subsequently employs a support vector regressor (SVR) model to predict the image quality score. The BRISQUE score typically ranges from 0 to 100, where a lower value indicates superior image quality, and a higher value suggests the presence of more distortions and noise. This metric demonstrates good sensitivity to common image degradation types, such as blur, noise, and compression artifacts. Detailed algorithmic steps can be found in reference [4].

4. Results

To rigorously evaluate the performance of the proposed PromptHDR algorithm, extensive experiments are conducted on the self-constructed Roadway Support Low-light Dataset. This section is organized into three parts: first, a qualitative visual comparison with state-of-the-art methods is provided to intuitively demonstrate the enhancement effects; second, quantitative results using PSNR and SSIM metrics are presented for objective performance comparison; finally, ablation studies are detailed to validate the contribution of each core component within the PromptHDR architecture.

4.1. Qualitative Results

A qualitative comparison between the proposed PromptHDR algorithm and other low-light enhancement methods is presented in Figure 3. When visually compared against high-quality reference images acquired with artificial lighting, it is observed that on the roadway low-light image dataset: the enhanced results from EnlightenGAN and KinD exhibit issues with dark color tones; Zero-DCE suffers from significant noise amplification; URetinex-Net demonstrates severe color distortion; and although PromptIR shows some improvement in details, it still falls short of the reference image in terms of overall brightness and color naturalness. In contrast, the proposed PromptHDR algorithm most closely approximates the reference image in color fidelity, brightness uniformity, and edge sharpness, achieving the best visual enhancement performance.

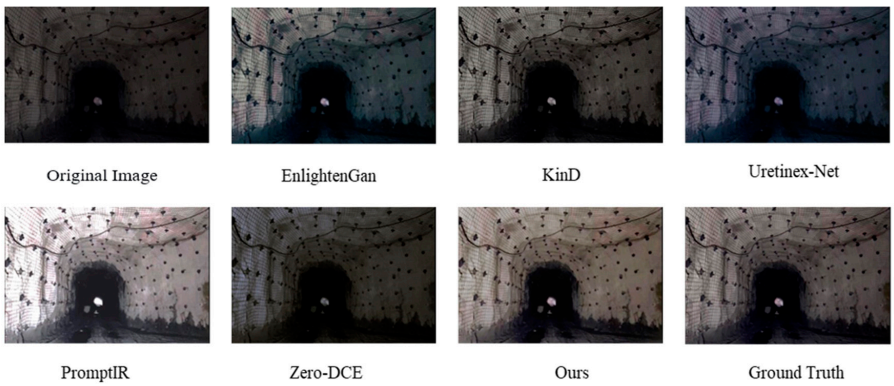


Figure 3. Comparison of the effect of each algorithm in the homemade roadway dataset.

Figure 4 further displays the enhancement results of the PromptHDR algorithm on the roadway low-light dataset. Through direct comparison with the original low-light images and their corresponding reference images, it can be intuitively observed that the enhanced results exhibit excellent performance in color fidelity, noise suppression, and control of color distortion. This further confirms the effectiveness of the proposed algorithm in restoring the authentic visual attributes of the images.

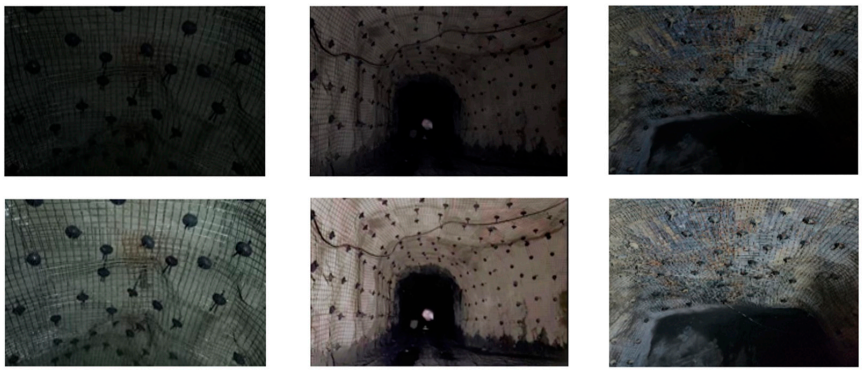


Figure 4. Original low-light image at top, enhanced image at bottom.

A comparative analysis in both 2D and 3D HSV color space was conducted on normal-light, low-light, and enhanced images of the same scene, with results shown in Figures 5 and 6. The enhanced image distribution exhibits minimal differences from the normal-light image distribution, with a significant improvement in the V-channel while maintaining structural consistency in the H and S distributions. This validates the effectiveness of the enhancement method in brightness recovery and color restoration.

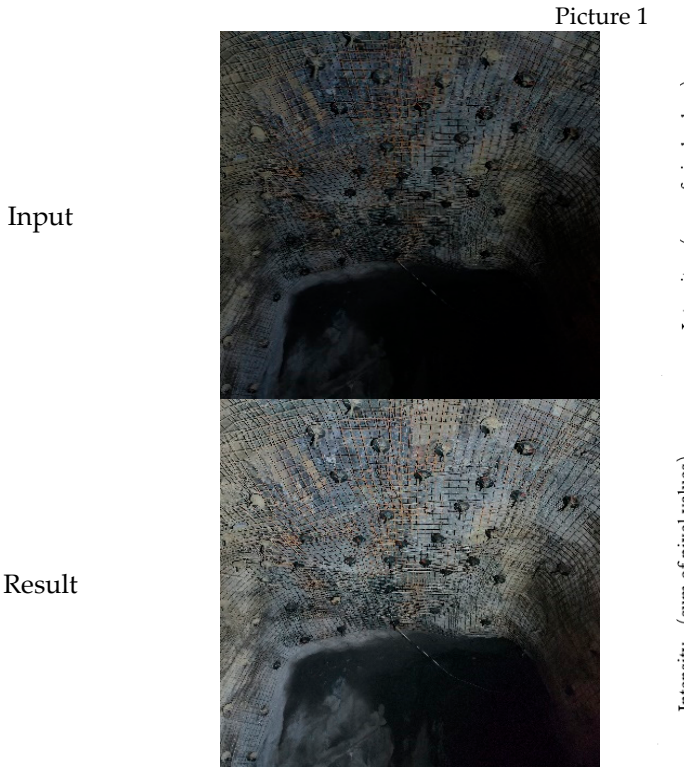
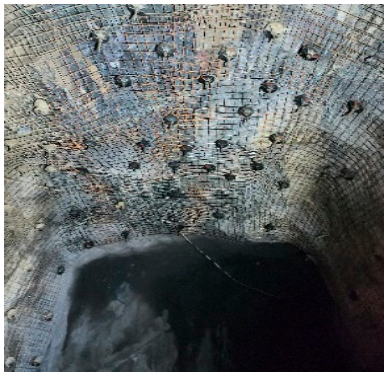
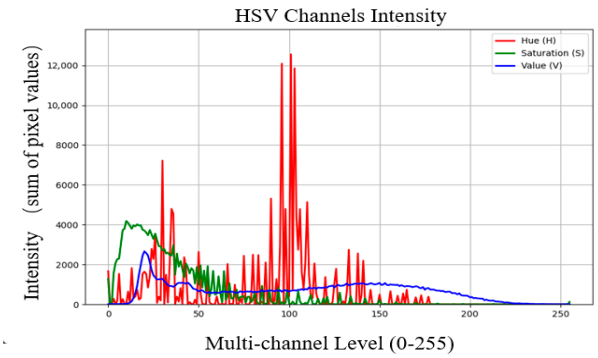


Figure 5. *Cont.*

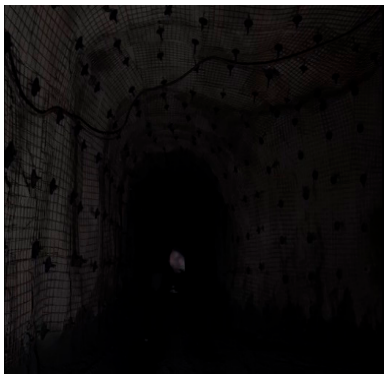
Ground Truth

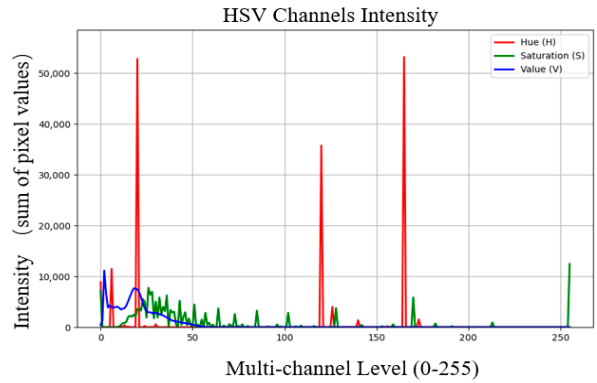




Picture 2

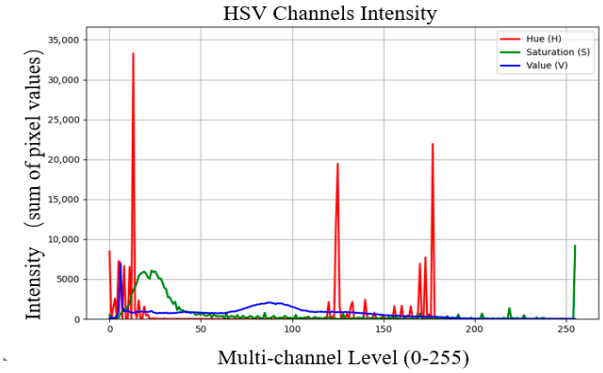
Input





Result





Ground Truth



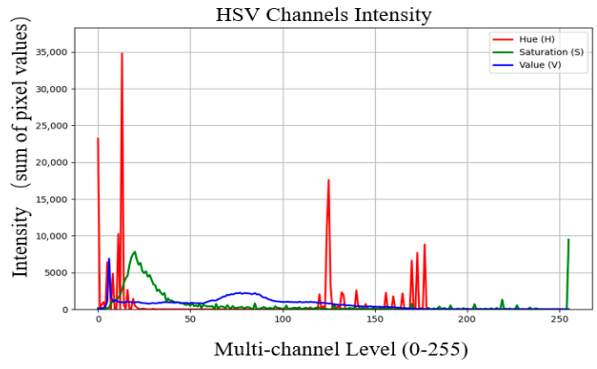
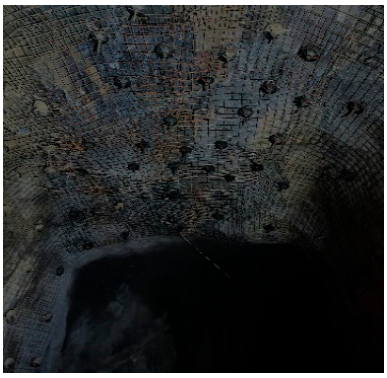


Figure 5. Comparison of 2D effect of HSV color space of enhanced image.

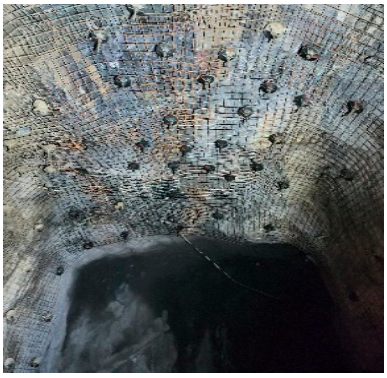
Input



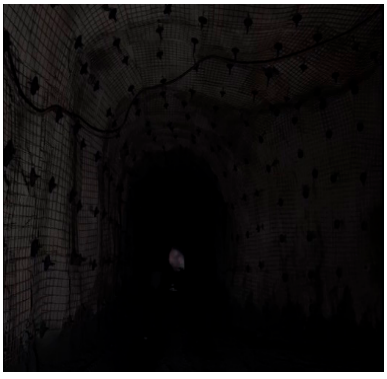
Result



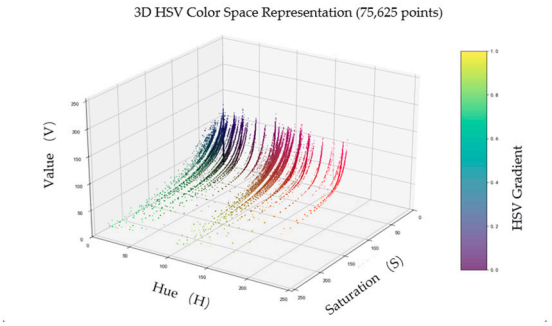
Ground Truth

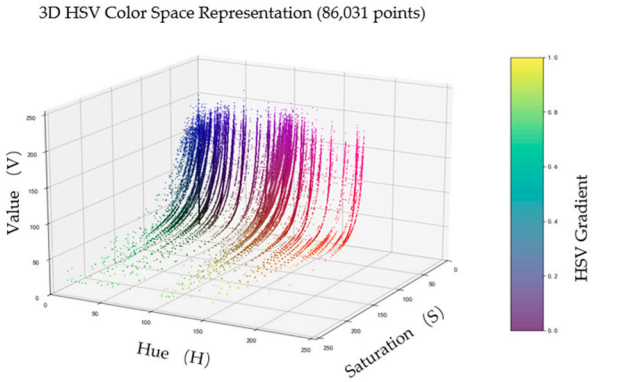


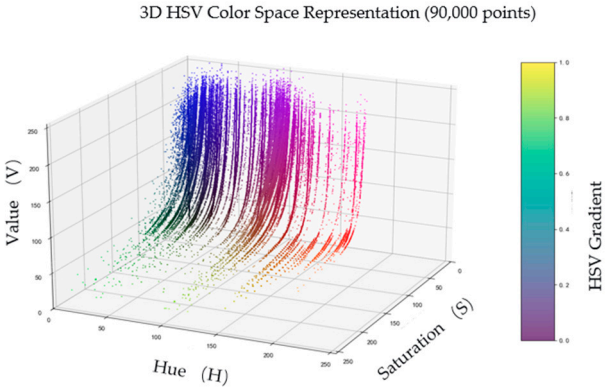
Input



Picture 1







Picture 2

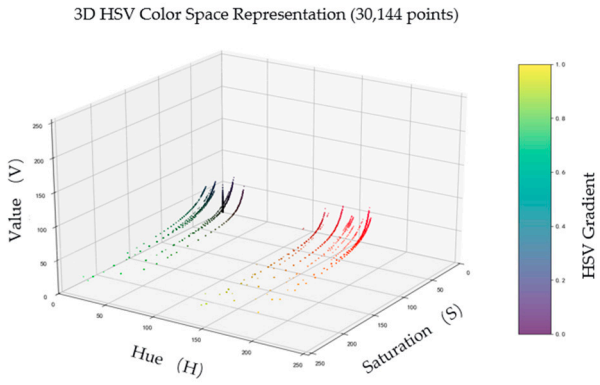


Figure 6. *Cont.*

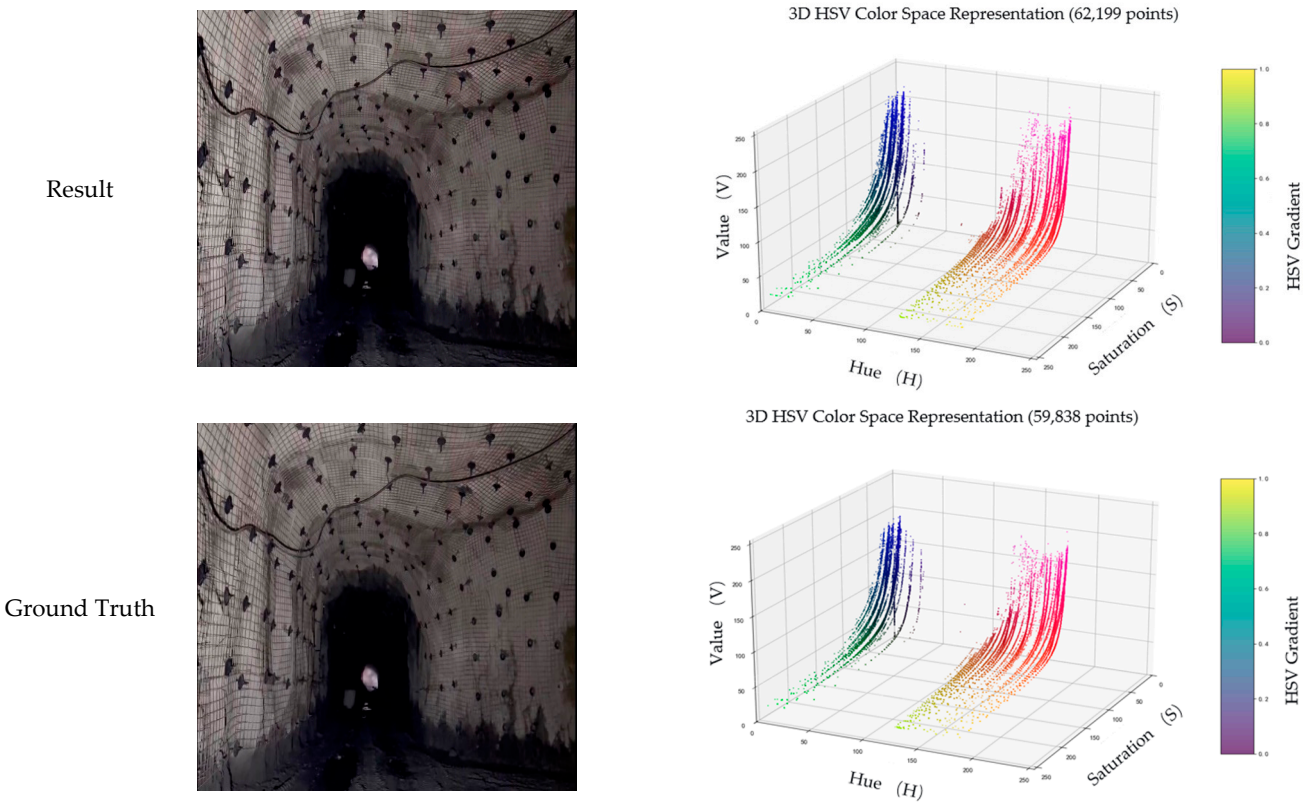


Figure 6. Comparison of 3D effect of HSV color space of enhanced image.

4.2. Quantitative Results

To ensure fair and reproducible comparisons, all competing methods were evaluated under identical conditions. For approaches that provide publicly available pre-trained models (including EnlightenGAN, KinD, URetinex-net, Restormer, PromptIR, and MIRNet), the official pre-trained weights are used in this study. For methods without available pre-trained models (such as SCI, Zero-DCE, Sparse, RUAS, NeRCO, and DRBN), models are re-trained on our roadway support low-light dataset training set until convergence. By comparing the proposed method against other excellent supervised/unsupervised low-light enhancement methods on the same test set, the quantitative results—including PSNR, SSIM, and additionally introduced perceptual metrics (NIQE, LPIPS, BRISQUE)—are presented in Table 2. The experimental results demonstrate that the PromptHDR algorithm achieves competitive performance across multiple evaluation metrics, with particularly outstanding performance in terms of PSNR and SSIM metrics.

As can be clearly observed from the quantitative results in Table 2, the proposed PromptHDR algorithm comprehensively surpasses existing advanced methods in terms of image reconstruction quality. In the PSNR metric, PromptHDR achieved the highest score of 24.19 dB, demonstrating a slight yet consistent lead over the next best performers, MIRNet (24.14 dB) and PromptIR (23.62 dB), while realizing substantial performance gains compared to traditional methods such as SCI (14.78 dB) and Zero-DCE (16.76 dB). Regarding the SSIM metric, which measures structural similarity, PromptHDR also ranked first with the highest score of 0.839, outperforming not only MIRNet (0.830) and PromptIR (0.832) but also surpassing URetinex-net (0.835) and DRBN (0.834) by a significant margin.

Table 2. Performance of each algorithm on the roadway low-light image dataset.

Methods	PSNR	SSIM	NIQE	LPIPS	BRISQUE
SCI	14.78	0.646	1.80	0.3649	35.56
Zero-Dce	16.76	0.560	1.32	0.3187	38.62
Sparse	17.20	0.640	1.98	0.4135	41.39
ElightenGan	17.48	0.650	2.05	0.4716	43.78
RUAS	18.23	0.720	1.07	0.4595	42.17
NeRCO	19.701	0.771	0.77	0.4954	26.45
DRBN	19.86	0.834	0.89	0.3862	29.14
KinD	20.87	0.802	1.39	0.4379	33.95
URetinex-net	21.33	0.835	0.75	0.5132	17.83
Restormer	22.43	0.823	0.90	0.2945	23.57
PromptIR	23.62	0.832	0.78	0.2718	20.41
MIRNet	24.14	0.830	0.76	0.2536	18.92
PromptHDR	24.19	0.839	0.70	0.2384	16.59

On no-reference image quality assessment metrics, PromptHDR likewise exhibited performance. For the NIQE index, PromptHDR achieved the lowest value of 0.70, significantly outperforming URetinex-net (0.75), MIRNet (0.76), and PromptIR (0.78), indicating that its enhancement results are closer to high-quality images in terms of naturalness and visual comfort. In the perceptual similarity metric LPIPS, PromptHDR led with the lowest error score of 0.2384, compared to MIRNet (0.2536) and PromptIR (0.2718), revealing its advantage in preserving image semantic content and perceptual quality. Furthermore, for the blind image quality assessment metric BRISQUE, PromptHDR achieved the lowest score of 16.59, markedly superior to URetinex-net (17.83) and MIRNet (18.92), further validating its effectiveness in suppressing noise and artifacts while enhancing overall visual quality. These results demonstrate that PromptHDR not only improves the signal-to-noise ratio of images but also more effectively preserves structural information, achieving state-of-the-art performance in both pixel-level fidelity and perceptual structural similarity.

To further evaluate the practical utility of the proposed method, the model complexity and computational efficiency of PromptHDR are compared against other competing methods. All experiments were conducted under the same hardware environment (see Table 3), with a uniform input image resolution of 128 $\times$ 128. Number of parameters (Params) and floating-point operations (GFLOPs) are used to measure model complexity, and used frames processed per second (FPS) as the indicator for runtime efficiency.

Table 3. Model complexity and efficiency comparison.

Methods	Params	GFLOPs	FPS
SCI	0.08	0.12	65.2
Zero-Dce	0.08	0.15	58.7
Sparse	0.15	0.20	50.3
ElightenGan	8.64	2.31	42.5
RUAS	0.10	0.18	55.1
NeRCO	2.50	5.00	25.0
DRBN	4.32	7.65	20.5
KinD	8.16	12.45	15.8
URetinex-net	1.43	3.87	28.3
Restormer	26.13	15.82	12.1
PromptIR	15.73	7.15	24.6
MIRNet	31.96	18.27	10.5
PromptHDR	12.45	5.83	30.2

4.3. Ablation Study

To rigorously evaluate the contribution of each core component within the complex PromptHDR architecture, a comprehensive ablation study was conducted. Eight variant models are structured by incrementally removing or adjusting the key proposed modules—namely the Lighting Extraction Module (LEM), the Sampling Module (comprising HWD and DySample), and the Mamba-based Prompt Block. All variants were trained and evaluated on the same Roadway Support Low-light Dataset under identical settings. The quantitative results, summarized in Table 4, unequivocally demonstrate the necessity and effectiveness of each component, with the complete PromptHDR model achieving the highest performance.

Table 4. Results of ablation experiments.

Model	Illumination Estimate	Sampling	Prompt	PSNR	SSIM
Architecture 1	×	×	×	23.45	0.829
Architecture 2	√	×	×	23.63	0.833
Architecture 3	×	√	×	23.98	0.832
Architecture 4	×	×	√	23.58	0.826
Architecture 5	√	√	×	23.99	0.835
Architecture 6	√	×	√	23.78	0.835
Architecture 7	×	√	√	23.46	0.834
Architecture 8	√	√	√	24.14	0.839

The quantitative results in Table 4 provide compelling evidence for the contribution of each module. The baseline model (Architecture 1), which lacks all three proposed components, serves as a reference point. The performance increment observed in Architectures 2, 3, and 4 reveals the individual impact of the LEM, Sampling, and Prompt modules, respectively. For instance, the introduction of the Lighting Extraction Module (Architecture 2) alone brought a notable gain in PSNR, underscoring its role in mitigating color distortion by providing explicit illumination priors. The Sampling Module (Architecture 3) contributed the most significant single-module performance jump, highlighting its critical function in multi-scale feature fusion for detail preservation. The incorporation of the Prompt Block (Architecture 4) also provided a clear benefit, validating its ability to guide the enhancement process using illumination features.

More importantly, the combinations of these modules (Architectures 5–7) show that their effects are complementary. The synergy is most potent in the full model (Architecture 8), which integrates all components and achieves the peak performance (PSNR: 24.19, SSIM: 0.839). The fact that no incomplete architecture could match the full model's performance strongly suggests that each module addresses a distinct and critical aspect of the low-light enhancement challenge underground—illumination non-uniformity, detail loss at different scales, and context-aware feature refinement—and that their co-design is essential for optimal results.

The visual comparisons in Figure 7 provide an intuitive understanding of how each module progressively refines the output. The variant employing only the Lighting Extraction Module (Architecture 2) establishes a baseline with improved color uniformity, yet it still exhibits blurred details and insufficient local contrast around rock bolt features. Subsequently, the introduction of the Mamba-based Prompt Block (Architecture 6) leads to a perceptible recovery of finer textures and edges, as the model adaptively focuses on critical local details guided by the illumination features. Ultimately, the complete PromptHDR model (Architecture 8), which integrates the Sampling Module, produces the most visually pleasing result by successfully balancing global brightness with local

sharpness. This final output clearly reveals rock bolt structures (e.g., nuts and washers) while maintaining natural color rendition and suppressing noise. The observed progressive visual improvement aligns with the quantitative metrics and solidifies the claim that each module is indispensable for achieving the final high-quality enhancement.

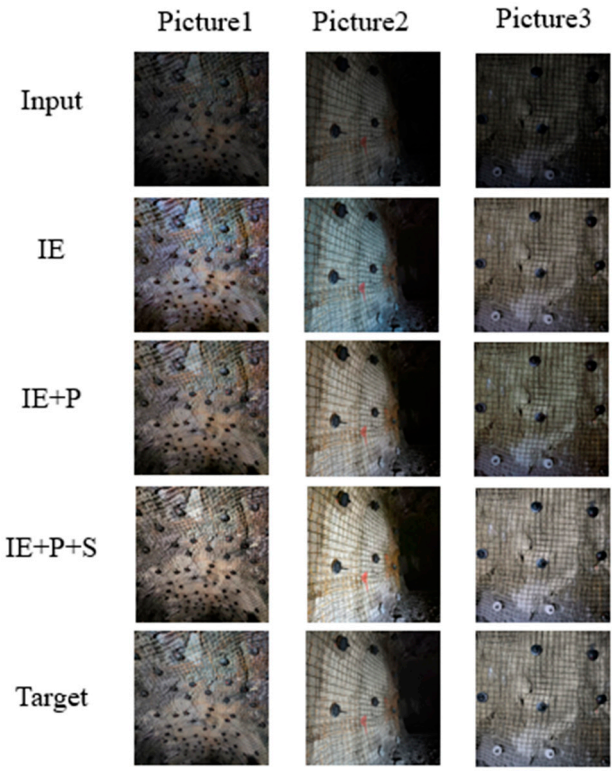


Figure 7. Comparative effects of the ablation experiment pictures.

An internal ablation study on the Sampling Module itself further justifies its design. As quantified in Table 5, using only HWD downsampling or only DySample upsampling results in a performance drop compared to their joint use. This empirically confirms that the two components form a cohesive unit: effective downsampling is crucial for creating a multi-scale feature hierarchy and reducing computational burden, while high-fidelity upsampling is equally critical for reconstructing a high-resolution output with preserved details. Their combination is fundamental to the success of the U-shaped network in our task.

Table 5. Sampling module ablation experiments.

Sampling	Upsampling	Downsampling	Overall
PSNR	23.76	23.88	24.14
SSIM	0.837	0.831	0.839

4.4. Bolt Detection Performance Evaluation

The aforementioned ablation studies have verified the effectiveness of the core components internally. However, the ultimate value of our work lies in its improvement of practical application tasks. Therefore, we next proceed to the downstream bolt detection experiment to form a complete chain of reasoning from methodology to application.

We employed the high-performing and widely adopted YOLOv8 model as the benchmark detector. The specific experimental settings are as follows.

4.4.1. Dataset and Evaluation Metrics

To ensure the acquisition of authentic and undistorted visual features of bolts in the detection model, a detection dataset was constructed using 400 original normal-illumination images with corresponding bolt annotations. Data augmentation was deliberately avoided to maintain precise correspondence between bounding boxes and image content, as well as to preserve the authenticity of illumination conditions. The dataset was partitioned into training, validation, and test sets following an 8:1:1 ratio for training a YOLOv8m model with adequate bolt recognition capability. The training procedure extended over 300 epochs with input images resized to 640 × 640 pixels, resulting in a stable and converged model. Detailed system configurations employed during training are provided in Table 1.

Core evaluation metrics in object detection were documented as comprehensive indicators for algorithm performance assessment, including Precision, Recall, and mean Average Precision (mAP). Precision quantifies the prediction accuracy of detected instances, while Recall measures the model's coverage capability of true positive cases, representing the proportion of correctly identified positives among all actual positives. The mean Average Precision (mAP), calculated as the macro-average of Average Precision (AP) across all categories, serves as a crucial benchmark for model evaluation—higher mAP values indicate superior overall detection accuracy.

$$\mathrm{mAP} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{AP_i} \tag{20}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{21}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{22}$$

where N represents the total number of all categories; TP denotes the number of correctly predicted positive samples; FP indicates the number of negative samples incorrectly predicted as positive; and FN refers to the number of positive samples incorrectly predicted as negative.

4.4.2. Comparative Experimental Results and Analysis

To focus on the most representative comparisons and clearly present the core differences in task performance, the pre-trained YOLOv8 model weights were held fixed during the evaluation, and a set of five key conditions was established for performance recording. These include original low-light images (serving as the performance lower-bound baseline), normal-light images (serving as the performance upper-bound reference), along with test set images enhanced by two high-performing mainstream algorithms, KinD and PromptIR, respectively, and finally, those enhanced by the proposed PromptHDR method. This setup forms a comprehensive and efficient evaluation framework, spanning from baseline and competitor methods to our proposed approach. The quantitative results of the downstream task validation are presented in Table 6.

Table 6. Detection performance comparison of various image enhancement algorithms.

Methods	mAP@0.5	Precision	Recall
Origion picture + yolov8	80.60	0.856	0.652
KinD + yolov8	81.46	0.907	0.712
PromptIR + yolov8	82.50	0.918	0.765
PromptHDR + yolov8	87.97	0.940	0.799
Normal picture + yolov8	88.61	0.952	0.815

As shown in Table 6, compared to the various evaluation metrics obtained from the original low-light images, all image enhancement methods, when used as a preprocessing step, improve bolt detection performance to varying degrees. This confirms the necessity of image enhancement in low-light environments. Furthermore, among all the compared algorithms, the proposed PromptHDR algorithm achieves dominantly superior performance, with an mAP of 87.97%, while maintaining the highest precision and a competitive recall rate. Compared to the second-best performer, PromptIR, PromptHDR achieves an absolute improvement of 5.4% in mAP, fully demonstrating that the visual features restored by PromptHDR-enhanced images are more aligned with the requirements of the target detection model. Finally, it is noteworthy that the detection performance on low-light images processed by PromptHDR most closely approaches the performance upper bound obtained directly on ideal normal-light images (in terms of mAP). This indicates that our method effectively bridges the gap between low-light conditions and ideal visual perception.

5. Conclusions

This study addresses the challenges of low illumination, detail blur, and color distortion in rock bolt images captured in dark underground mining environments by proposing PromptHDR, a low-light image enhancement algorithm based on the Transformer architecture. The algorithm employs a lighting extraction module to explicitly model illumination distribution, effectively suppressing color distortion caused by non-uniform lighting; introduces a prompt module that integrates Mamba with illumination features to enhance the model's contextual understanding of tunnel scenes and its ability to preserve bolt details; and constructs a sampling module comprising DySample and HWD components to achieve high-quality image reconstruction through multi-scale feature fusion. Experiments on a real-world underground tunnel low-light dataset demonstrate that PromptHDR achieves PSNR and SSIM values of 24.19 dB and 0.839, respectively, outperforming several mainstream methods and exhibiting superior visual enhancement effects.

Validation experiments based on YOLOv8 demonstrate that images enhanced by PromptHDR attain a bolt detection mAP of 87.97%, while maintaining the highest precision and a competitive recall rate, significantly outperforming all compared methods. This successfully establishes a complete, closed-loop argument from 'image quality enhancement' to 'detection performance improvement,' thereby providing compelling evidence for the algorithm's practical value in the intelligent construction of mines

Although PromptHDR performs well in low-light image enhancement, it still has certain limitations when dealing with the complex and variable real-world underground conditions. In environments filled with dust and haze, light undergoes spatially varying degradation patterns due to multiple scattering and absorption, leading to decreased overall image contrast and increased noise. As the current model does not explicitly model such physical degradation processes, the enhancement effectiveness might be constrained. Furthermore, the algorithm is relatively sensitive to motion blur caused by equipment vibration or movement, lacking explicit modeling and compensation mechanisms for motion blur, which could affect the accurate identification of bolt structures in dynamic scenarios.

Looking forward, we will proceed with our research from the following aspects: On the one hand, we will explore integrated optimization frameworks that combine dehazing, deblurring, and low-light enhancement to improve the model's robustness and generalization in scenarios with composite degradations. On the other hand, we will focus on the lightweight design and engineering optimization of the algorithm, promoting its integration into embedded platforms and inspection robotic systems. This aims to achieve real-time and reliable visual enhancement in underground low-light environments, ultimately serving the goals of intelligent mine construction and safe, efficient production.

Author Contributions: Conceptualization, W.Y. and F.Q.; methodology, Y.W., J.X., J.L. and L.Z.; validation, W.Y. and Y.W.; investigation, J.X., J.L. and L.Z.; data curation, W.Y., Y.W. and J.X.; writing—original draft preparation, W.Y., Y.W. and J.X.; writing—review and editing, F.Q., Y.W. and L.Z.; supervision, W.Y. and F.Q.; project administration, W.Y. and F.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article; further inquiries can be directed to the corresponding author.

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

References

1. Wang, W.; Liu, X.; Liu, X.; Rong, L.; Hao, L.; He, Q.; Liao, F.; Tang, H. Carbon Decoupling of the Mining Industry in Mineral-Rich Regions Based on Driving Factors and Multi-Scenario Simulations: A Case Study of Guangxi, China. *Processes* **2025**, *13*, 2474. [CrossRef]
2. Jia, X.; Zhan, Y.; Tian, X. Recent Advances in Treatment Technologies and Resource Utilization of Mine Tailings in Hunan Province, China. *Processes* **2025**, *13*, 2957. [CrossRef]
3. Miao, D.; Feng, Q.; Zeng, W. Research on Cooling and Dust Removal Technology of Circulating Airflow in Metal Mine Working Face. *Processes* **2025**, *13*, 1374. [CrossRef]
4. Jiang, B.; Xin, Z.; Zhang, X.; Deng, Y.; Wang, M.; Li, S.; Ren, W. Mechanical properties and influence mechanism of confined concrete arches in high-stress tunnels. *Int. J. Min. Sci. Technol.* **2023**, *33*, 829–841. [CrossRef]
5. Zuo, J.; Wang, J.; Jiang, Y. Macro/meso failure behavior of surrounding rock in deep roadway and its control technology. *Int. J. Coal Sci. Technol.* **2019**, *6*, 301–319. [CrossRef]
6. Liu, S.; Liu, H.; Xie, K.; Shan, L.; Xiao, F.; Wang, B.; Wang, Y. Stability Analysis and Support Optimization for a Coal Mine Roadway Subjected to High Horizontal Stress. *Appl. Sci.* **2025**, *15*, 2276. [CrossRef]
7. Isleyen, E.; Duzgun, S.; Carter, M.R. Roof fall hazard detection with convolutional neural networks using transfer learning. *arXiv* **2020**, arXiv:2012.03681. [CrossRef]
8. Su, Y.; Wang, J.; Wang, X.; Hu, L.; Yao, Y.; Shou, W.; Li, D. Zero-reference deep learning for low-light image enhancement of underground utilities 3D reconstruction. *Autom. Constr.* **2023**, *152*, 104930. [CrossRef]
9. 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]
10. Tian, Z.; Qu, P.; Li, J.; Sun, Y.; Li, G.; Liang, Z.; Zhang, W. A survey of deep learning-based low-light image enhancement. *Sensors* **2023**, *23*, 7763. [CrossRef]
11. Guo, J.; Ma, J.; García-Fernández, Á.F.; Zhang, Y.; Liang, H. A survey on image enhancement for Low-light images. *Heliyon* **2023**, *9*, e14558. [CrossRef]
12. Hai, J.; Hao, Y.; Zou, F.; Lin, F.; Han, S. Advanced retinexnet: A fully convolutional network for low-light image enhancement. *Signal Process. Image Commun.* **2023**, *112*, 116916. [CrossRef]
13. Ye, L.; Wei, W.; Wang, Y.; Guo, Y.; Yu, H.; Wang, Y. Research on data acquisition technology of intelligent laboratory based on retinex-net image recognition. *J. Phys. Conf. Ser.* **2023**, *2474*, 012059. [CrossRef]
14. Feng, W.; Wu, G.; Zhou, S.; Li, X. Low-light image enhancement based on Retinex-Net with color restoration. *Appl. Opt.* **2023**, *62*, 6577–6584. [CrossRef]
15. Liu, C.; Wang, Z.; Birch, P.; Wang, X. Efficient retinex-based framework for low-light image enhancement without additional networks. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *35*, 4896–4909. [CrossRef]
16. Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; Zhang, Y. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 12504–12513.
17. Zhao, X.; Li, L. A low-light-level image enhancement algorithm combining Retinex and Transformer. In Proceedings of the International Conference on Remote Sensing, Mapping, and Image Processing (RSMIP 2024), Xiamen, China, 19–21 January 2024; pp. 704–712.
18. Liu, S.; Zhang, H.; Li, X.; Yang, X. Retinexformer+: Retinex-Based Dual-Channel Transformer for Low-Light Image Enhancement. *Comput. Mater. Contin.* **2025**, *82*, 1969. [CrossRef]
19. Li, C.; Guo, J.; Porikli, F.; Pang, Y. LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognit. Lett.* **2018**, *104*, 15–22. [CrossRef]

20. 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, Nice, France, 21–25 October 2019; pp. 1632–1640.
21. 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]
22. Wang, R.; Jiang, B.; Yang, C.; Li, Q.; Zhang, B. MAGAN: Unsupervised low-light image enhancement guided by mixed-attention. *Big Data Min. Anal.* **2022**, *5*, 110–119. [CrossRef]
23. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5637–5646.
24. Mi, A.; Luo, W.; Qiao, Y.; Huo, Z. Rethinking zero-DCE for low-light image enhancement. *Neural Process. Lett.* **2024**, *56*, 93. [CrossRef]
25. Li, C.; Guo, C.; Loy, C.C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4225–4238. [CrossRef] [PubMed]
26. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10561–10570.
27. Zhang, Z.; Shi, B.; Zhang, H.; Xu, H.; Zhang, Y.; Wu, Y.; Dong, B.; Zheng, Q. NerCo: A Contrastive Learning Based Two-Stage Chinese NER Method. In Proceedings of the IJCAI, Macao, China, 19–25 August 2023; pp. 5287–5295.
28. Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3063–3072.
29. Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; Jiang, J. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5901–5910.
30. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
31. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Learning enriched features for fast image restoration and enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1934–1948. [CrossRef] [PubMed]
32. Potlapalli, V.; Zamir, S.W.; Khan, S.H.; Shahbaz Khan, F. Promptir: Prompting for all-in-one image restoration. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 71275–71293.

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.