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.

FSP-Net: A Frequency-Split Prior Network for Low-Light Image Enhancement in Underground Mines

Tianyi Zhong
Jilin University
Qingyi Xue
Jilin University
Pingping Liu
liupp@jlu.edu.cn
Jilin University
Yi Diao
Jilin University
Yuxin Wang
Jilin University

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11787244/v1

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

Additional Declarations: No competing interests reported.

FSP-Net: A Frequency-Split Prior Network for Low-Light Image Enhancement in Underground Mines

Tianyi Zhong[1†], Qingyi Xue[1†], Pingping Liu[2*], Yi Diao[1], Yuxin Wang[1]

[1]College of Computer Software, Jilin University, Changchun, China.
[2*]College of Computer Science and Technology, Jilin University, Changchun, China.

*Corresponding author(s). E-mail(s): liupp@jlu.edu.cn;
Contributing authors: zhongty5523@mails.jlu.edu.cn; xueqy5523@mails.jlu.edu.cn;
diaoyi5523@mails.jlu.edu.cn; alsk9389291@gmail.com;
†These authors contributed equally to this work.

Abstract

Low-light image enhancement (LLIE) is essential for extreme visual scenarios such as underground mines, where insufficient illumination, glare, and color cast often coexist and impair visual perception. Existing methods frequently neglect instance-level semantic consistency and frequency-aware reasoning, making them fragile under such coupled degradations. We propose FSP-Net, a Frequency-Split Prior Network that addresses these issues from a causal reasoning and wavelet-domain perspective. FSP-Net uses a compact encoder–decoder with wavelet-based frequency decoupling and red, green, and blue (RGB) residual learning. It mines structural and gradient priors from the low-pass band to guide decoder refinement. A causally guided contrastive objective promotes recovery of stable scene content across illumination and color changes. At the instance level, a Contrastive Language–Image Pretraining (CLIP) loss preserves regional consistency without changing inference. On the Low-Light (LOL) benchmarks, FSP-Net achieves a peak signal-to-noise ratio (PSNR) of 24.04 dB and a structural similarity index (SSIM) of 0.8538 on LOL-v1, and 23.04 dB and 0.8594 on LOL-v2-Real. On an unpaired underground mine dataset, it achieves the best Naturalness Image Quality Evaluator (NIQE) score of 4.1592 among the methods compared. These results indicate that FSP-Net can serve as a general LLIE backbone and support underground mine surveillance. The source code is available at https://github.com/zhongty5523/FSPNet.

Keywords: Low-light image enhancement, Underground mine monitoring, Wavelet decomposition, Prior-guided learning, Causal inference

Acknowledgements. The authors thank Dr. Tongshun Zhang for helpful discussions and guidance during this research.

1 Introduction

Low-light image enhancement (LLIE) is fundamental to computer vision, yet insufficient illumination, low contrast, and detail loss impair visual perception. Underground mines are particularly challenging: images captured under uneven or weak lighting also suffer from glare, dust, haze, and sensor noise. Robust enhancement therefore requires more than global brightening; it must

preserve scene layout, edges, and instance-level semantic fidelity.

Early pipelines based on histogram equalization or Retinex decomposition [1, 2] remain interpretable but struggle with non-uniform illumination and severe darkness. Deep networks [3, 4] learn end-to-end mappings and achieve strong results on natural-scene benchmarks. In mine images, underexposure, glare, and noise often co-occur, making pixel-level regression alone insufficient to recover illumination while preserving texture and layout.

Frequency-decoupled models [5–8] separate illumination-dominated low-frequency content from detail-dominated high-frequency content. Semantic-oriented methods use contrastive learning [9], segmentation knowledge [10], or vision–language priors [11] to improve regional plausibility. CWNet [5] uses counterfactual degradation and causally guided metric learning to recover stable scene content, with instance-level Contrastive Language–Image Pretraining (CLIP) scoring for semantic consistency. SPJFNet [7] uses self-mined gradient and structure cues to guide the wavelet branches. These approaches motivate an enhancer designed for coupled illumination, structure, and semantic degradations in underground tunnels.

FSP-Net combines a compact wavelet encoder–decoder with red, green, and blue (RGB) residual learning. Its gradient–structure guidance module mines priors from the wavelet low-pass band via adaptive gamma mapping and feeds them into the corresponding decoder branches. The training objective combines the counterfactual contrastive strategy of CWNet with a region-wise CLIP loss driven by mining-scene masks. Light and color degradation negatives synthesized from normal-light references discourage the network from relying on illumination or color shortcuts.

The main contributions of this paper are summarized as follows:

1. We mine adaptive structural and gradient priors once from the low-pass wavelet subband and inject them only into decoder blocks for frequency-specific refinement.
2. We combine light and color counterfactual negatives with a mining-scene instance-level CLIP loss to supervise global and regional restoration without adding inference-time modules.
3. We evaluate FSP-Net on paired low-light benchmarks and on unpaired natural and mine images, including ablations of the prior and training losses.

2 Related Work

Non-learning methods use physical models and handcrafted priors. LIME [12], for example, initializes the illumination map with the maximum RGB channel value at each pixel and refines it with a structure prior. Such methods often struggle to balance brightness correction, detail restoration, and natural appearance in complex scenes. RetinexNet [8] couples Retinex decomposition with supervised learning; Zero-DCE [13], Zero-DCE++ [14], and EnlightenGAN [15] learn without paired references. Semantic-guided methods [9, 10] impose region-aware constraints, while Transformer-based enhancers [4] improve global modeling.

The Horizontal/Vertical-Intensity (HVI) color space and the Color and Intensity Decoupling Network (CIDNet) reduce chromatic artifacts [16]. RetinexMamba and Wave-Mamba [17, 18] model long-range dependencies in Retinex- and wavelet-based networks. CWNet [5] combines a wavelet backbone with counterfactual metric learning and instance-level semantic scoring. Structure modeling and sparse gradient regularization [19, 20] help stabilize detail recovery. SPJFNet [7] injects self-mined structure and gradient priors across cascade stages; FSP-Net instead computes the priors once and injects them only during decoding.

Underground scenes couple extreme underexposure with localized strong light sources, dust, and noise. Yang et al. [21] combine global and local enhancement for mine personnel monitoring, and Yang et al. [22] use a band representation to balance signal fidelity and perceptual quality. The Underground Mine Generative Adversarial Network (UM-GAN) uses inverted grayscale inputs as a domain prior [23]. FSP-Net combines frequency-specific restoration and global dependency modeling with mine-oriented semantic supervision.

3 Method

3.1 Prior Guidance

Prior knowledge is incorporated into FSP-Net via two components: (i) endogenous gradient–structure priors mined from the wavelet low-pass band, and (ii) an instance-level CLIP consistency objective tailored to underground scenes.

3.1.1 Gradient–Structure Guidance Module

Following SPJFNet [7], the Gradient–Structure Guidance Module (GSGM) self-mines structural and gradient priors from the low-frequency component L_i, obtained by applying the discrete wavelet transform (DWT) twice to the input RGB image. Global average pooling on convolved features yields a descriptor $\mathbf{F}_a$, which a multilayer perceptron (MLP) with sigmoid activation maps to an adaptive gamma factor:

$$\Gamma = \sigma\big(\mathrm{MLP}(\mathbf{F}_a)\big), \quad \Gamma \in \mathbb{R}^{1\times1\times C}, \tag{1}$$

Two complementary priors are then mined from the same L_i:

$$S_i = L_i^{\Gamma}, \qquad G_i = \nabla(S_i), \tag{2}$$

where S_i guides global illumination-adjusted layout and G_i encodes edge-consistent structure. The low-frequency path is conditioned on S_i and the high-frequency path on G_i. Reference priors S_i^{gt} and G_i^{gt} are computed from the ground truth with the same pipeline, and a prior supervision loss encourages fidelity:

$$\begin{aligned}\mathcal{L}_s = \lambda_1 \|G_i - G_i^{\mathrm{gt}}\|_1 \\ + \lambda_2 \,\mathrm{MSE}\big(\mathcal{E}(S_i), \mathcal{E}(S_i^{\mathrm{gt}})\big),\end{aligned} \tag{3}$$

where MSE is the mean squared error, $\mathcal{E}(\cdot)$ denotes a binarized edge map, and $(\lambda_1, \lambda_2) = (1.0, 0.1)$.

Unlike SPJFNet, which recomputes and reinjects these priors at every cascade stage, FSP-Net mines them once on the quarter-resolution low-pass band and injects them only into decoder frequency-split enhancement blocks (FSEBs) via bilinear resizing and 1×1 projection. This keeps encoders prior-free, avoids stage-wise recomputation, and confines guidance naturally to the reconstruction path.

3.1.2 Instance-Level Semantic Consistency through CLIP

Mine LLIE requires region-wise semantic fidelity rather than global appearance matching, as downstream monitoring tasks (e.g., personnel detection, equipment inspection) rely on reliable object boundaries and layout consistency. We therefore introduce an instance-level CLIP consistency objective that supervises restoration only during training, without modifying the enhancer's architecture or requiring segmentation at inference.

Instance masks are extracted from the normal-light reference I^+ using a pre-trained High-Resolution Network (HRNet) segmenter [24] (Sect. 4.1). For each instance k, we construct a composite image I_{seg}^k by taking I^+ and replacing the pixels within the k-th instance mask with the corresponding pixels from the enhanced output $\hat{I} = \mathcal{F}(I)$. Thus, each composite contains a single "patched" instance from $\hat{I}$ within an otherwise pristine normal-light context. CLIP encoders then score whether this composite aligns better with the prompt "low-light" or "normal-light".

Let Φ_{image} and Φ_{text} denote the CLIP image and text encoders [11]. For each composite and text prompt T, let

$$s_k(T) = \cos\big(\Phi_{\mathrm{image}}(I_{\mathrm{seg}}^k), \Phi_{\mathrm{text}}(T)\big).$$

The aggregated semantic score is:

$$\hat{y} = \frac{1}{K}\sum_{k=1}^{K} \frac{e^{s_k(T_{\mathrm{normal}})}}{e^{s_k(T_{\mathrm{low}})} + e^{s_k(T_{\mathrm{normal}})}}, \tag{4}$$

where K is the number of valid instances. The instance-level semantic loss is binary cross-entropy:

$$\mathcal{L}_{\mathrm{clip}} = -\big(y \log \hat{y} + (1-y)\log(1-\hat{y})\big), \tag{5}$$

with $y = 1$ for normal-light references.

The loss supervises restoration only during training; neither HRNet nor CLIP is needed at test time.

3.2 Causal Inference Mechanism

Underground images often exhibit coupled illumination and chromatic degradations—extreme under-exposure, localized glare, dust-induced cast,

and lamp-specific color bias—while scene layout and object identity remain largely stable. Following CWNet [5], we treat content as the causal factor to be recovered and regard brightness as well as chromaticity as nuisance variables that may change without altering semantics. Pixel-level reconstruction alone cannot rule out brightness- or color-biased shortcuts, so we use a causally guided contrastive loss $\mathcal{L}_{\text{ca}}$ with counterfactual negatives synthesized from the normal-light reference I^+. Fig. 1 shows the contrast between the normal-light positive and the two nuisance axes.

Light degradation negatives emulate tunnel exposure changes by rescaling an illumination map M [12] estimated from I^+ with $\gamma \in [2, 5]$:

$$I_\ell = \frac{I^+}{M} M^\gamma + \epsilon, \tag{6}$$

whereas color degradation negatives mimic dust and lamp color casts through bounded hue, saturation, and RGB perturbations (ΔH, ΔS, ΔK) that preserve geometry:

$$\begin{aligned} I_c = {} & \Delta H(I^+) + \Delta S(I^+) \\ & + \sum_{K \in \{R,G,B\}} \Delta K(I^+) + \epsilon. \end{aligned} \tag{7}$$

Here ϵ denotes the added noise. Each sample in $\{I_k^-\}$ shares scene content with I^+ but deviates along one nuisance axis. During training, $\mathcal{L}_{\text{ca}}$ attracts the enhanced output $\hat{I}$ toward I^+ while repelling it from the low-light input I and the synthesized negatives, so the enhancer learns to restore stable scene content rather than chase illumination- or color-driven appearance shifts. Together with wavelet-based frequency splitting, this mechanism helps suppress noise amplification and chromatic drift in complex mine scenes.

3.3 FSP-Net

Fig. 2 illustrates the overall pipeline. Given a low-light image $I_{\text{low}} \in \mathbb{R}^{H \times W \times 3}$, FSP-Net adopts a U-shaped encoder–decoder with stacked FSEBs. The encoder progressively downsamples features, and skip connections preserve shallow detail for decoder fusion. Decoder *FSEB*s upsample concatenated skip features and perform frequency-aware restoration; GSGM mines S_i and G_i once from the quarter-resolution low-pass band and injects them only into decoder blocks, while encoder blocks remain prior-free. Each FSEB decomposes its input with the DWT into a low-pass approximation and three oriented high-pass bands, applies the low-frequency enhancement module (LFEM) and high-frequency enhancement module (HFEM) guided by S_i and G_i, reconstructs the subbands with inverse DWT, and refines the result through gated spatial convolution with a local residual. The network prediction is obtained by a global RGB residual:

$$I_{\text{out}} = I + \text{Conv}_{1\times1}(\Phi(I)), \tag{8}$$

where Φ denotes the stacked encoder–decoder. At inference, H and W may be padded to multiples of four before Φ and cropped afterward.

3.3.1 High-Frequency Enhancement Module

After DWT, the high-frequency subbands X_{HL}, X_{LH}, and X_{HH} are organized into horizontal, vertical, and diagonal streams H^L, V^L, and D^L, respectively, and coupled with responses derived from the local approximation band L so that coarse illumination context guides detail recovery. When the gradient prior G_i is available, it is bilinearly resized and fused into each stream by channel-wise concatenation and 1×1 projection to sharpen edge-consistent structure while suppressing noise-dominated fluctuations in the wavelet domain. Following CWNet [5], we adopt orientation-specific two-dimensional state-space models (2D-SSMs): H-2D-SSM, V-2D-SSM, and D-2D-SSM scan H^L, V^L, and D^L along geometry-aligned paths, respectively:

$$\begin{aligned} \tilde{H}^L &= \text{H-2D-SSM}(H^L), \\ \tilde{V}^L &= \text{V-2D-SSM}(V^L), \\ \tilde{D}^L &= \text{D-2D-SSM}(D^L). \end{aligned} \tag{9}$$

Each directional block applies layer normalization, selective spatial scanning, and a gated feed-forward path, enabling long-range coupling among high-frequency coefficients with near-linear complexity. The enhanced bands are assembled with the refined low-pass map L via inverse wavelet transform; the reconstructed map is added to the stage residual and passed through lightweight spatial refinement.

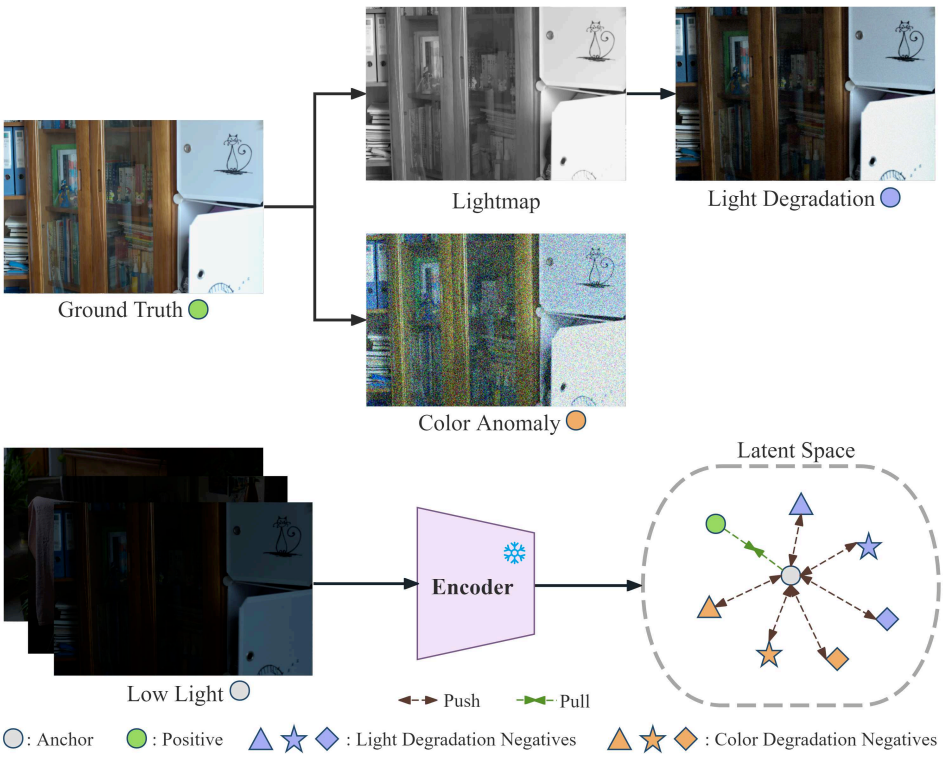

Fig. 1 Counterfactual learning for low-light enhancement. Light and color degradation negatives are synthesized from the normal-light reference. Contrastive learning pulls the enhanced representation toward the positive sample and pushes it away from the negatives, following CWNet [5]

3.3.2 Low-Frequency Enhancement Module

The low-frequency enhancement module (LFEM) operates on the approximation coefficient L produced by DWT at each *FSEB*. At half resolution, L mainly encodes slowly varying illumination, color bias, and coarse scene layout, whereas fine textures reside in the companion high-pass bands; decoupling enhancement in the wavelet domain therefore lets LFEM target global brightness and structure without competing with detail restoration. When the structural prior S_i is available, it is bilinearly resized and fused with L by channel-wise concatenation followed by 1×1 projection:

$$L_{\text{fused}} = \text{Conv}_{1\times1}\big([L, S_i^{\text{res}}]\big), \tag{10}$$

where $[\cdot,\cdot]$ denotes channel concatenation and S_i^{res} is the resized prior; otherwise, $L_{\text{fused}} = L$. This step injects illumination-aware structural context before spatial filtering. L_{fused} is then refined by two identical depthwise–pointwise stacks, each combining a 5×5 depthwise convolution for enlarged local context with a 1×1 convolution for cross-channel re-weighting, yielding stable global tone mapping on the half-resolution grid. The refined L guides the high-frequency branch, participates in inverse wavelet reconstruction, and closes the intra-stage residual path before full-resolution refinement.

3.3.3 Loss Function

Training follows a two-stage schedule. In **stage 1**, only the prior supervision in (3) is optimized to learn the gradient–structure mining module; the enhancement backbone is frozen:

$$\mathcal{L}^{(1)} = \mathcal{L}_s. \tag{11}$$

In **stage 2**, the learned prior module is loaded and kept fixed while the enhancement network is optimized with four loss terms:

$$\begin{aligned}\mathcal{L}^{(2)} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{ssim}} \\ + \lambda_{\text{ca}}\mathcal{L}_{\text{ca}} + \lambda_{\text{clip}}\mathcal{L}_{\text{clip}},\end{aligned} \tag{12}$$

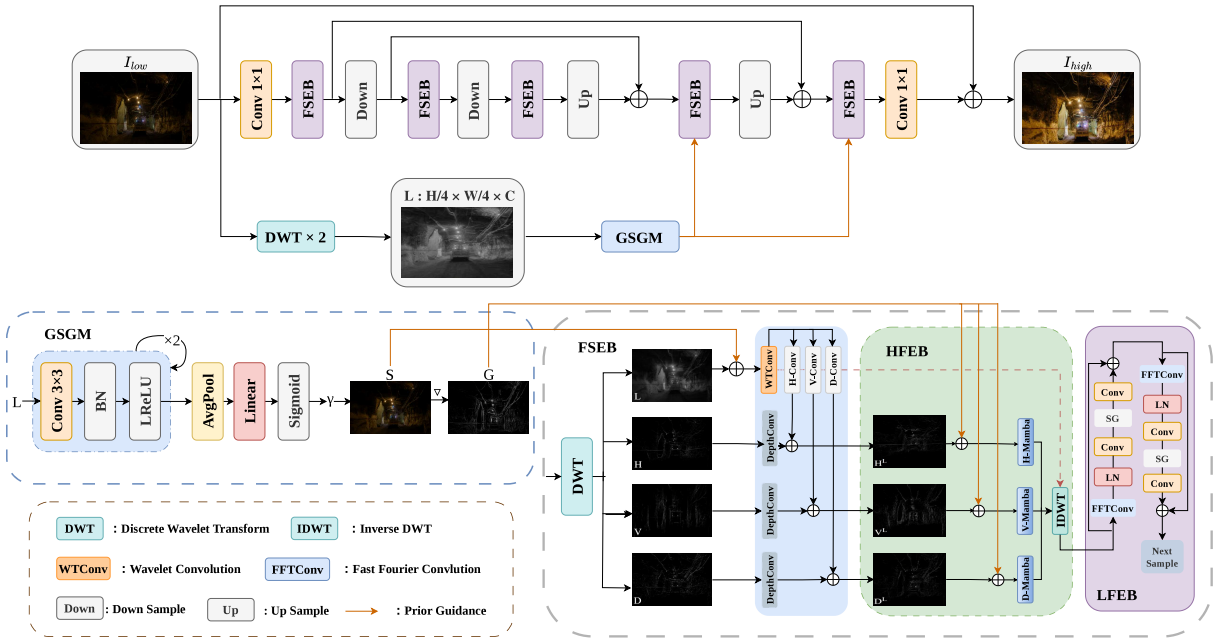

Fig. 2 Overall architecture of FSP-Net. Frequency-split enhancement blocks (FSEBs) process the low-light input through the encoder and decoder. The gradient–structure guidance module (GSGM) supplies low-pass priors to decoder FSEBs. HFEB and LFEB denote the high- and low-frequency modules

where λ denotes the loss weights; we set $(\lambda_{\text{rec}}, \lambda_{\text{ssim}}, \lambda_{\text{ca}}, \lambda_{\text{clip}}) = (1.0, 0.3, 0.01, 0.01)$. $\mathcal{L}_{\text{rec}}$ is the reconstruction loss, $\mathcal{L}_{\text{ssim}}$ is the structural similarity loss, $\mathcal{L}_{\text{ca}}$ is the causally guided contrastive loss, and $\mathcal{L}_{\text{clip}}$ is the instance-level CLIP consistency loss. $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{ssim}}$ constrain the enhanced output and ground truth end-to-end.

4 Experiments

4.1 Datasets and Experimental Settings

For paired evaluation, we train and test on the official Low-Light (LOL) splits: LOL-v1 [8] (485 training and 15 test pairs) and LOL-v2-Real [25] (689 training and 100 test pairs). For no-reference evaluation, we train on the single image contrast enhancement (SICE) dataset [26], which comprises 589 sequences and 4,413 images. Each scene has 3–18 exposures, including a well-exposed reference, covering a wider illumination range than conventional low-/normal-light pairs. We test the SICE-trained model on a low-light subset of MIT-Adobe FiveK [27] and the Coal Mine Underground Drilling Site Object Detection Dataset [28]. We measure paired quality by peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). We measure no-reference quality by the Naturalness Image Quality Evaluator (NIQE) [29] and the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [30], for which lower scores are better.

We pre-train an HRNet-W48 segmenter on mining scenes with six category labels: miner, hydraulic support guard plate, large coal, towline, miner behavior, and mine safety helmet. We freeze the segmenter during FSP-Net training and use its predictions on normal-light references to construct the masks for $\mathcal{L}_{\text{clip}}$ (Sect. 3.1.2). The segmenter is not used at inference.

We implement FSP-Net in PyTorch and run the experiments on one NVIDIA GeForce RTX 4090 with 24 GB memory. We randomly crop 256×256 patches and augment them with horizontal and vertical flips and rotations. Adam is used with $\beta_1 = 0.9$, $\beta_2 = 0.99$, an initial learning rate of 4×10^{-4}, and a batch size of 8. On each LOL split, each stage runs for 250,000 iterations, halving the learning rate after 130,000, 180,000, and 220,000 iterations. On SICE, each stage runs for 36,000 iterations, with learning rate decays after 16,000, 22,000, and 28,000 iterations. Stage 1 updates only the prior module. Stage 2 loads and freezes that

Table 1 Paired results on LOL-v1. Best in bold; second best underlined

Method	PSNR (dB)	SSIM
LIME	16.76	0.4440
NPE	16.97	0.5928
RetinexNet	16.77	0.5600
EnlightenGAN	17.48	0.6500
KinD	20.87	0.7995
KinD++	17.97	0.8042
FourLLIE	20.99	0.8071
DMFourLLIE	22.98	0.8273
SNR-Aware	23.93	0.8460
RetinexMamba	23.15	0.8210
Retinexformer	**25.16**	0.8450
Wave-Mamba	22.76	0.8419
CWNet	23.60	<u>0.8496</u>
FSP-Net	<u>24.04</u>	**0.8538**

Table 2 Paired results on LOL-v2-Real. Best in bold; second best underlined

Method	PSNR (dB)	SSIM
LIME	15.24	0.4190
NPE	17.13	0.4642
RetinexNet	15.47	0.5670
EnlightenGAN	18.23	0.6170
KinD	17.54	0.6695
KinD++	19.08	0.8176
FourLLIE	22.34	0.8450
DMFourLLIE	22.71	0.8583
SNR-Aware	21.48	0.8478
RetinexMamba	21.73	0.8290
Retinexformer	22.80	0.8400
CWNet	21.65	**0.8604**
HVI-CID	**24.11**	0.8110
FSP-Net	<u>23.04</u>	<u>0.8594</u>

module and optimizes the enhancement network as described in Sect. 3.3.3.

4.2 Comparison with Current Methods

We compare with LIME [12], Naturalness Preserved Enhancement (NPE) [31], RetinexNet [8], EnlightenGAN [15], KinD [32], KinD++ [33], FourLLIE [6], DMFourLLIE [34], SNR-Aware [35], RetinexMamba [17], Retinexformer [4], Wave-Mamba [18], CWNet [5], and HVI-CID [16]. For paired comparisons, learned methods use the corresponding LOL training split. We calculate PSNR and SSIM without matching the prediction to the ground-truth mean (GT-Mean). Tables 1 and 2 show that FSP-Net achieves 24.04 dB PSNR and the best SSIM of 0.8538 on LOL-v1, and 23.04 dB PSNR and 0.8594 SSIM on LOL-v2-Real.

Fig. 3 compares enhancement results on LOL-v2-Real. Several baselines brighten the image but leave color drift, residual noise, or insufficient local contrast. FSP-Net recovers clearer textures and more even illumination in the example shown.

4.3 No-Reference Evaluation

On FiveK and the unpaired mine dataset, we compare FSP-Net with Zero-DCE [13], EnlightenGAN [15], RetinexNet [8], CWNet [5], HVI-CID [16], and LIME [12]. Table 3 reports NIQE and BRISQUE. EnlightenGAN leads on both FiveK metrics, while FSP-Net remains close on BRISQUE. On the unpaired mine set, FSP-Net improves NIQE from 6.2582 for the raw input to 4.1592, the best value among the methods compared. HVI-CID and CWNet achieve lower mine BRISQUE but higher NIQE.

Fig. 4 shows a FiveK example with Expert C's rendition as a visual reference. Figs. 5 and 6 show two underground scenes. In these examples, RetinexNet amplifies noise, while Zero-DCE and LIME can leave dark pockets. FSP-Net recovers dark regions without strongly overexposing highlights and preserves textures in the drilling areas.

4.4 Ablation Studies

We remove the gradient–structure guidance module, instance-level CLIP consistency loss, and causally guided contrastive loss in turn (Table 4). On the mine set, BRISQUE rises from 18.45 for the full model to 24.78, 28.37, and 30.73, respectively. The full model also achieves the best FiveK BRISQUE (18.70) among the ablation settings. Some variants have slightly lower mine NIQE, but their BRISQUE is consistently worse. On LOL-v1, removing the prior module reduces PSNR by 0.44 dB and SSIM by 0.042. Adding an HVI color-space conversion front end [16] also reduces PSNR (23.87 dB versus 24.04 dB), so FSP-Net retains end-to-end RGB processing.

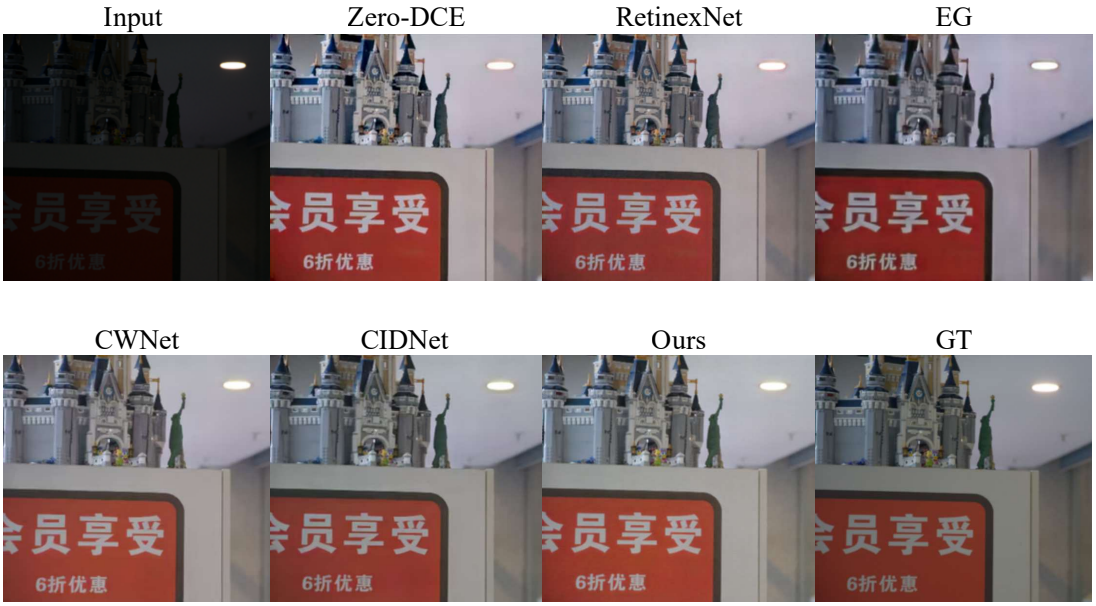

Fig. 3 Visual comparison on LOL-v2-Real. EG denotes EnlightenGAN, CIDNet denotes the HVI-CID result, and GT denotes the normal-light ground truth from LOL-v2-Real [25]

Table 3 No-reference results on the FiveK subset and mine dataset. Best in bold; second best underlined

Method	MIT-Adobe FiveK		Coal mine	
	NIQE ↓	BRISQUE ↓	NIQE ↓	BRISQUE ↓
LIME	4.5517	20.1668	4.8107	18.2148
Input	5.1912	24.2532	6.2582	72.1132
Zero-DCE	4.3789	21.4621	5.0955	19.2549
EnlightenGAN	**3.9416**	**17.2549**	4.8630	19.8211
RetinexNet	4.5207	<u>18.4012</u>	5.1578	26.1116
CWNet	5.1029	22.6752	<u>4.6988</u>	<u>13.4651</u>
HVI-CID	<u>4.2211</u>	20.2124	5.4231	**12.6216**
FSP-Net	4.4236	18.7014	**4.1592**	18.4518

5 Conclusion

FSP-Net combines wavelet-based frequency splitting with gradient–structure priors, instance-level CLIP supervision, and counterfactual contrastive learning for low-light image enhancement. It is competitive on the paired LOL-v1 and LOL-v2-Real benchmarks and achieves the lowest NIQE among the compared methods on unpaired mine images. These results support the use of frequency-specific restoration for low-light images in which preserving local structures and scene content matters.

Statements and Declarations

Funding. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests. The authors have no relevant financial or non-financial interests to disclose.

Author Contributions. Tianyi Zhong and Qingyi Xue contributed equally to this work. Tianyi Zhong: Methodology, Software, Investigation, Writing – original draft. Qingyi Xue: Methodology, Software, Investigation, Writing – original draft. Pingping Liu: Conceptualization, Supervision, Project administration, Writing – review & editing. Yi Diao: Investigation, Data

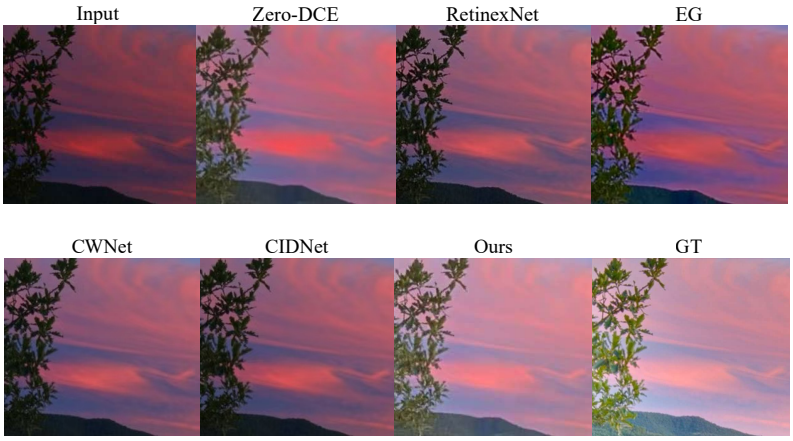

Fig. 4 Visual comparison on MIT-Adobe FiveK. GT denotes the rendition by Expert C [27], included as a visual reference

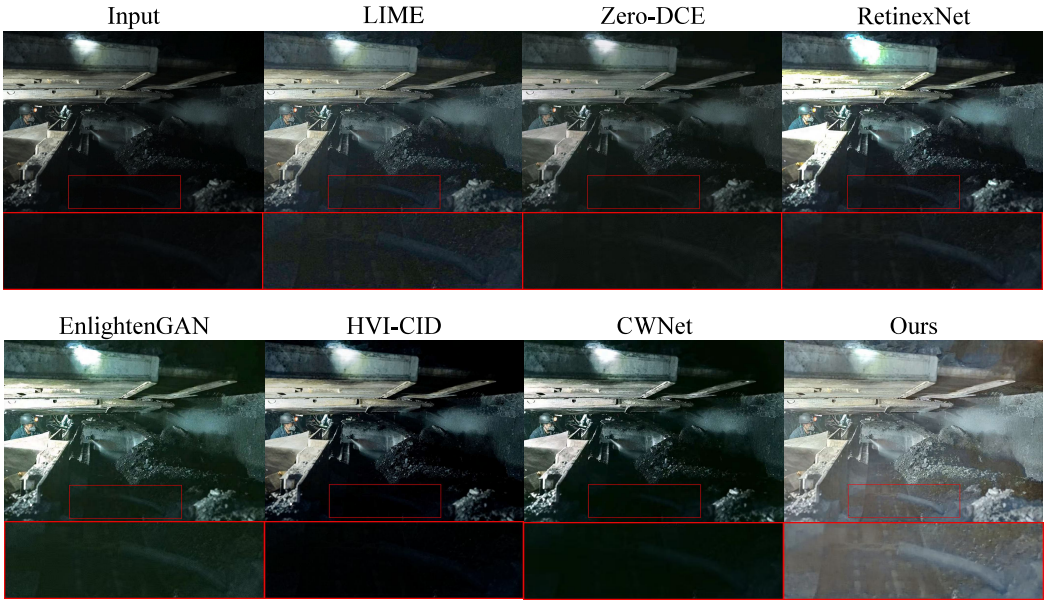

Fig. 5 Visual comparison on underground mine scene 1, with enlarged regions showing local details. Input from the drilling-site dataset [28]

curation, Writing – review & editing. Yuxin Wang: Investigation, Data curation, Writing – review & editing. All authors read and approved the final manuscript.

Data Availability. The public datasets used for enhancement are available from their original sources: SICE [26], LOL-v1 [8], LOL-v2-Real [25], and MIT-Adobe FiveK [27]. The unpaired drilling-site images are available from the Science Data Bank [28]. Derived results are available from the corresponding author upon reasonable request.

Code Availability. The FSP-Net source code is publicly available at https://github.com/zhongty5523/FSPNet.

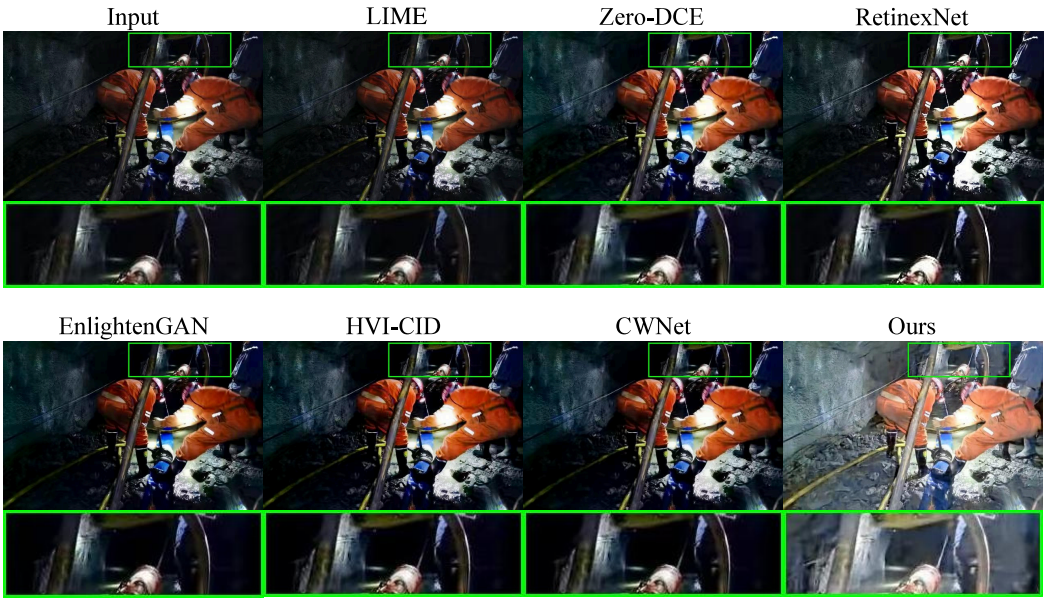

Fig. 6 Visual comparison on underground mine scene 2, with enlarged regions showing local texture and contrast. Input from the drilling-site dataset [28]

Table 4 Ablation results on the FiveK subset and mine dataset. Best in bold; second best underlined

Ablation setting	MIT-Adobe FiveK		Coal mine	
	NIQE ↓	BRISQUE ↓	NIQE ↓	BRISQUE ↓
Without GSGM	4.6225	19.9352	<u>4.1174</u>	<u>24.7773</u>
Without CLIP loss	4.4436	<u>19.3021</u>	**4.0788**	28.3699
Without causal loss	**4.3009**	23.2350	4.4679	30.7289
FSP-Net	<u>4.4236</u>	**18.7014**	4.1592	**18.4518**

References

[1] Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Comput. Vis. Graph. Image Process. 39(3), 355–368 (1987). https://doi.org/10.1016/S0734-189X(87)80186-X

[2] Jobson, D.J., Rahman, Z., Woodell, G.A.: A multiscale Retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans. Image Process. 6(7), 965–976 (1997). https://doi.org/10.1109/83.597272

[3] Lore, K.G., Akintayo, A., Sarkar, S.: LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognit. 61, 650–662 (2017). https://doi.org/10.1016/j.patcog.2016.06.008

[4] Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage Retinex-based Transformer for low-light image enhancement. In: Proc. ICCV, pp. 12504–12513 (2023). https://doi.org/10.1109/ICCV51070.2023.01149

[5] Zhang, T., Liu, P., Lu, Y., Cai, M., Zhang, Z., Zhang, Z., Zhou, Q.: CWNet: Causal wavelet network for low-light image enhancement. In: Proc. ICCV, pp. 8789–8799 (2025). https://doi.org/10.1109/ICCV51701.2025.00822

[6] Wang, C., Wu, H., Jin, Z.: FourLLIE: Boosting low-light image enhancement by Fourier frequency information. In: Proc. ACM MM, pp. 7459–7469 (2023). https://doi.org/10.1145/3581783.3611909

[7] Zhang, T., Liu, P., Zhang, Z., Zhou, Q.: SPJFNet: Self-mining prior-guided joint frequency enhancement for ultra-efficient dark image restoration. Proc. AAAI Conf. Artif. Intell. 40(15), 12798–12806 (2026). https://doi.org/10.1609/aaai.v40i15.38277

[8] Wei, C., Wang, W., Yang, W., Liu, J.: Deep Retinex decomposition for low-light enhancement. In: Proc. BMVC (2018). https://doi.org/10.48550/arXiv.1808.04560

[9] Liang, D., Li, L., Wei, M., Yang, S., Zhang, L., Yang, W., Du, Y., Zhou, H.: Semantically contrastive learning for low-light image enhancement. Proc. AAAI Conf. Artif. Intell. 36(2), 1555–1563 (2022). https://doi.org/10.1609/aaai.v36i2.20046

[10] Wu, Y., Pan, C., Wang, G., Yang, Y., Wei, J., Li, C., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: Proc. CVPR, pp. 1662–1671 (2023). https://doi.org/10.1109/CVPR52729.2023.00166

[11] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: Proc. ICML, pp. 8748–8763 (2021). https://doi.org/10.48550/arXiv.2103.00020

[12] Guo, X., Li, Y., Ling, H.: LIME: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 26(2), 982–993 (2017). https://doi.org/10.1109/TIP.2016.2639450

[13] Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proc. CVPR, pp. 1780–1789 (2020). https://doi.org/10.1109/CVPR42600.2020.00185

[14] 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. 44(8), 4225–4238 (2022). https://doi.org/10.1109/TPAMI.2021.3063604

[15] 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. 30, 2340–2349 (2021). https://doi.org/10.1109/TIP.2021.3051462

[16] Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., Zhang, Y.: HVI: A new color space for low-light image enhancement. In: Proc. CVPR, pp. 5678–5687 (2025). https://doi.org/10.1109/CVPR52734.2025.00533

[17] Bai, J., Yin, Y., He, Q., Li, Y., Zhang, X.: RetinexMamba: Retinex-based Mamba for low-light image enhancement. In: Neural Information Processing (ICONIP 2024), pp. 427–442. Springer, Singapore (2025). https://doi.org/10.1007/978-981-96-6596-9_30

[18] Zou, W., Gao, H., Yang, W., Liu, T.: Wave-Mamba: Wavelet state space model for ultra-high-definition low-light image enhancement. In: Proc. ACM MM, pp. 1534–1543 (2024). https://doi.org/10.1145/3664647.3681580

[19] Xu, X., Wang, R., Lu, J.: Low-light image enhancement via structure modeling and guidance. In: Proc. CVPR, pp. 9893–9903 (2023). https://doi.org/10.1109/CVPR52729.2023.00954

[20] Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep Retinex network for robust low-light image enhancement. IEEE Trans. Image Process. 30, 2072–2086 (2021). https://doi.org/10.1109/TIP.2021.3050850

[21] Yang, W., Wang, S., Wu, J., Chen, W., Tian, Z.: A low-light image enhancement method for personnel safety monitoring in underground coal mines. Complex Intell. Syst. 10, 4019–4032 (2024). https://doi.org/10.1007/s40747-024-01387-2

[22] Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. IEEE Trans. Image Process. 30, 3461–3473 (2021). https://doi.org/10.1109/TIP.2021.3062184

[23] Han, W., Xiao, Y., Yin, Y.: UM-GAN: Underground mine GAN for underground mine low-light image enhancement. IET Image Process. 18(8), 2154–2160 (2024). https://doi.org/10.1049/ipr2.13092

[24] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B.: Deep high-resolution representation learning for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 43(10), 3349–3364 (2021). https://doi.org/10.1109/TPAMI.2020.2983686

[25] 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: Proc. CVPR, pp. 3063–3072 (2020). https://doi.org/10.1109/CVPR42600.2020.00313

[26] Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Trans. Image Process. 27(4), 2049–2062 (2018). https://doi.org/10.1109/TIP.2018.2794218

[27] Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: Proc. CVPR, pp. 97–104 (2011). https://doi.org/10.1109/CVPR.2011.5995413

[28] Zhou, W., Dong, L., Ye, O., She, X., Duan, X., Peng, Z., Wang, S., Zhao, N., Guo, X.: Coal Mine Underground Drilling Site Object Detection Dataset, V1. Science Data Bank (2024). https://doi.org/10.57760/sciencedb.j00001.01020

[29] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. 20(3), 209–212 (2013). https://doi.org/10.1109/LSP.2012.2227726

[30] Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21(12), 4695–4708 (2012). https://doi.org/10.1109/TIP.2012.2214050

[31] Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process. 22(9), 3538–3548 (2013). https://doi.org/10.1109/TIP.2013.2261309

[32] Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: Proc. ACM MM, pp. 1632–1640 (2019). https://doi.org/10.1145/3343031.3350926

[33] Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. Int. J. Comput. Vis. 129, 1013–1037 (2021). https://doi.org/10.1007/s11263-020-01407-x

[34] Zhang, T., Liu, P., Zhao, M., Lv, H.: DMFourLLIE: Dual-stage and multi-branch Fourier network for low-light image enhancement. In: Proc. ACM MM, pp. 7434–7443 (2024). https://doi.org/10.1145/3664647.3681083

[35] Xu, X., Wang, R., Fu, C.W., Jia, J.: SNR-aware low-light image enhancement. In: Proc. CVPR, pp. 17714–17724 (2022). https://doi.org/10.1109/CVPR52688.2022.01719