Structure-aware joint low-light image enhancement and deblurring in the HVI space

Xingrui Liu
Zhejiang Sci-Tech University
Zhejiang, China
Siri_lxr@163.com

Ying Song*
Zhejiang Sci-Tech University
Zhejiang, China
ysong@zstu.edu.cn

Abstract

Images captured under low-light conditions often suffer from insufficient illumination and motion blur, severely degrading visual quality and hindering subsequent vision tasks. Existing approaches typically address enhancement and deblurring separately or model them jointly in the RGB space, but such strategies struggle to ensure both structural consistency and color fidelity, often leading to overexposure, residual blur, or color distortions. To address these issues, we propose a joint low-light enhancement and deblurring framework in the HVI color space, which effectively decouples luminance and chromatic channels, thereby improving the robustness of degradation restoration. Specifically, we design a Dynamic Feature-Aware Enhancement (DFAE) module that employs dynamic convolution kernels to adaptively handle spatially non-uniform blur, and a Structure-Aware Enhancement Subnetwork (SAES) that explicitly incorporates high-frequency and edge priors. Within SAES, a Structure-Aware Coordinate Attention Block (SCAB) facilitates effective fusion of structural features across channels. These components work collaboratively to enhance detail recovery while preserving global consistency. Our method achieves superior results on the LOL-Blur, LOLv1, LOLv2, and Real-LOLBlur datasets, delivering superior quantitative metrics and visual quality with sharper textures and more accurate colors.

***Keywords:** Low-light image enhancement, image deblurring, joint solution, HVI color space*

1. Introduction

Acquiring high-quality images under low-light conditions remains highly challenging. To compensate for insufficient illumination, imaging devices typically adopt long exposures, which are prone to motion blur from camera shake or object movement, or short exposures, which cause overall darkness. In real-world scenarios, nighttime images often suffer from compounded degradations of low-light and motion blur, which are more complex than single degradations.

For this complex degradation, a straightforward solution is to cascade low-light enhancement and deblurring methods. However, such sequential processing rarely yields satisfactory results. Low-light enhancement approaches [8, 28, 46, 2, 1] increase brightness and reduce noise but tend to amplify blur or lose spatial details. Conversely, deblurring methods [58, 5, 56, 18] trained under normal illumination perform poorly on low-light images with low contrast and noise. In both cascaded orders, errors are propagated and amplified throughout the cascade, leading to deficiencies in brightness, detail, and color. The fundamental limitation is that these methods treat low-light enhancement and deblurring as independent tasks, overlooking their strong coupling in real-world scenarios, and are thus inadequate for handling joint degradation effectively.

In recent years, several studies have attempted to jointly address low-light enhancement and deblurring within a unified framework. A representative example is LEDNet [64], which models both degradations through an end-to-end encoder–decoder architecture. This direction has since been extended by a series of joint approaches [31, 54, 26, 27], such as LLEDNet [27], which explores a Mamba-based solution. However, these methods still face notable limitations in detail restoration. Most operate in the sRGB color space, where luminance is strongly coupled with the three color channels—referred to as high color sensitivity in [22]—leading to noticeable color distortions in restored images [52]. Moreover, deep models tend to favor low-frequency components while exhibiting limited capacity to capture high-frequency details and edge structures, resulting in blur or contour distortions. Thus, effectively introducing structural guidance into joint tasks to compensate for deficiencies in detail and structure recovery remains an open and pressing challenge.

Our design is guided by two key insights. First, a low-light blurred image can be approximated as the outcome of low-pass filtering, where high-frequency components such

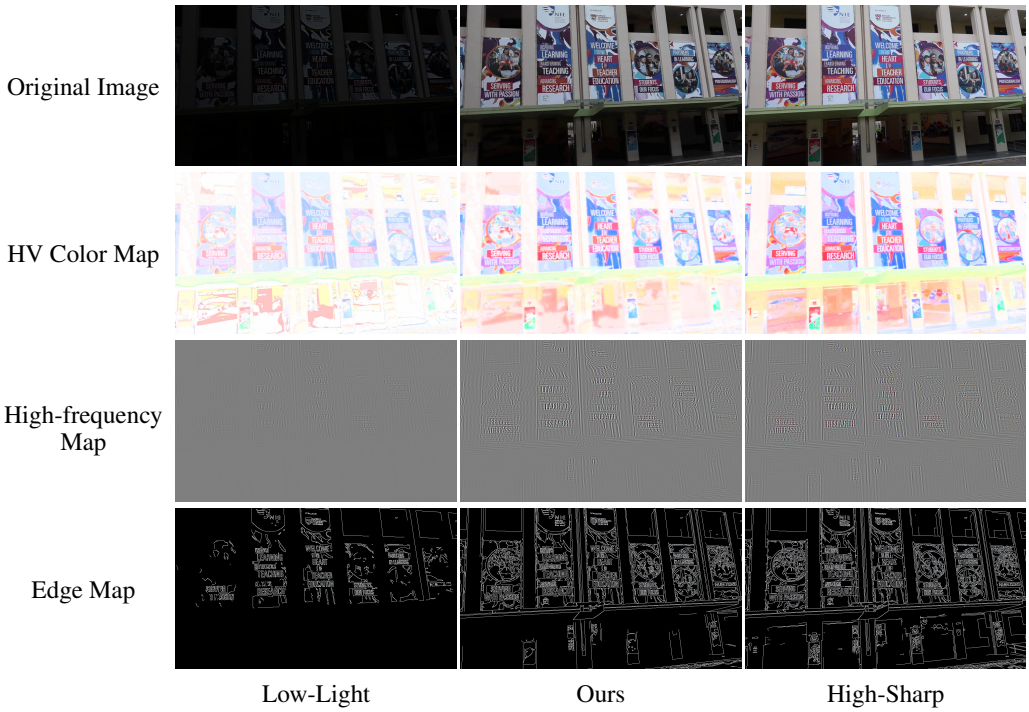


Figure 1. Comparison of low-light blurred image, our result, and the GT in terms of original image, HV color map, high-frequency map, and edge map. The low-light blurred image shows significant degradation in high-frequency and edge structures compared to the sharp image, while our method produces results more consistent with the structural and textural patterns of the sharp counterpart.

as textures and edges are suppressed or lost. As shown in the third and fourth rows of Fig. 1, low-light blurred images exhibit clear differences in high-frequency distribution and edge structures compared with their ground truth, and these differences are consistent with the overall structural discrepancies. This indicates that high-frequency and edge features better capture the structural patterns of sharp images, making structural guidance essential for restoring realistic textures and contours. Second, in the HVI color space, luminance and chromaticity are decoupled, with the HV plane preserving most of the color and structural information [52]. As illustrated in the second row of Fig. 1, these components are severely degraded under low-light blur. Motivated by this, we introduce a structure-enhancement mechanism that primarily operates on the HV plane, enabling direct modeling of chromatic–structural representations. This mitigates interference from luminance variations and strengthens the reconstruction of contours, edges, and textures.

Based on the above analysis, we propose a **S**tructure-aware **H**VI-based **I**mage e**N**hanc**E**ment Network (**SHINE-Net**), to jointly address low-light and blur degradations. The framework adopts a dual-branch architecture in which luminance and chromatic components are decoupled and processed separately: the HV-branch models chromatic and structural information, while the I-branch estimates and corrects luminance. To enhance structural fidelity, a structure-aware enhancement subnetwork is integrated into the HV-branch, leveraging high-frequency and edge cues to recover textures and contours lost under low-light blur. The structural features are fused through a structure-aware coordinate attention block, where high-frequency features emphasize local detail recovery and edge features reinforce global contours, complementing each other to improve overall image quality. In addition, a dynamic feature modeling mechanism exploits cross-layer interactions to generate adaptive representations, enabling finer restoration of non-uniform blur. With this design, SHINE-Net not only restores sharp details and structures but also preserves accurate colors.

Our main contributions are summarized as follows:

- We propose SHINE-Net, a novel unified framework for joint low-light image enhancement and deblurring in the HVI color space.
- We design three novel modules: the Dynamic Feature-Aware Enhancement (**DFAE**) module for modeling spatially heterogeneous blur, the Structure-Aware Enhancement Subnetwork (**SAES**) for high-frequency and edge-aware feature extraction, and the Structure-Aware Coordinate Attention Block (**SCAB**) for adaptive structure-guided feature fusion.
- SHINE-Net achieves state-of-the-art results on the LOL-Blur and Real-LOLBlur datasets, outperforming

LEDNet by **+0.69 dB** in PSNR, while delivering superior quantitative performance and perceptual quality.

2. Related Works

2.1. Low-Light Image Enhancement

Low-light image enhancement (LLIE) aims to improve visual quality under insufficient illumination. Early methods, such as histogram equalization [37, 21] and Retinex-based approaches [20, 15], adjust pixel distributions or decompose illumination and reflectance, but often amplify noise and introduce color distortions in complex scenes.

With the advent of deep learning, LLIE has achieved significant progress. Representative methods include autoencoder-based LLNet [30], Retinex-inspired KinD models [62, 61], curve-estimation approaches such as Zero-DCE [8, 24], and adversarial frameworks like EnlightenGAN [14]. More recent Transformer- and Mamba-based architectures further improve global illumination modeling and detail recovery [1, 51]. Diffusion-based methods have also been introduced due to their strong generative capability under extremely low-light conditions [11, 66, 25], albeit at the cost of high inference complexity.

Beyond network architectures, several works explore color-space decoupling to reduce noise and chromatic artifacts. Bread [9] operates in YCbCr space for noise disentanglement, Diff-Retinex [55] decomposes illumination and reflectance in sRGB, and CIDNet [52] proposes the HVI space to better separate brightness and chromaticity. Recent studies further investigate multi-branch and cross-space designs, such as DMFourLLIE [60], which employs a dual-stage multi-branch framework in the Fourier domain, FusionNet [41], which fuses features across different color representations, and ICLR [50], which enhances interactions between luminance and chrominance branches in decoupled color spaces.

Nevertheless, most LLIE methods focus solely on illumination degradation and overlook structural corruption caused by motion blur, leading to suboptimal performance on real-world low-light blurry images. In contrast, our work explicitly addresses both illumination deficiency and non-uniform blur by incorporating structure-aware modeling.

2.2. Image Deblurring

Image deblurring is a long-standing problem in computer vision. Early studies formulated it as an inverse optimization task [4, 6, 23], relying on priors such as sparse representation [23], and the dark channel prior [36]. These handcrafted assumptions, however, are highly sensitive to blur types and imaging conditions, leading to limited robustness in real scenarios.

The availability of large paired datasets has driven the success of deep learning. CNN-based methods, such as SRN [42] with its multi-scale coarse-to-fine design and DMPHN [58] with hierarchical multi-patch design, significantly improved deblurring of spatially varying blur. Transformer further enhance global modeling [56, 48, 17], with Restormer [56] and MPRNet [57] achieving state-of-the-art performance. Recent works also explore lightweight modules [18] and frequency-domain representations [17], while prior-guided strategies employ high-frequency [29] or edge information [7, 63] to improve restoration.

Despite these advances, most approaches target well-lit conditions and struggle when low-light and non-uniform blur coexist, where maintaining structural and chromatic consistency is crucial. Moreover, many methods rely on single-type constraints, underutilizing richer priors. To address these limitations, we propose a joint enhancement framework in the HVI color space that integrates a structure-aware subnetwork to exploit both high-frequency and edge cues, yielding more robust restoration under low-light blur.

2.3. Joint Low-Light Enhancement and Deblurring

Compared to handling low-light enhancement or deblurring independently, joint solutions remain limited. Early attempts simply cascaded separate LLIE and deblurring networks, but such pipelines are ineffective for real-world cases where illumination deficiency couples with non-uniform blur. To overcome this, several unified frameworks have emerged. LEDNet [64] employs an encoder–decoder design for joint restoration, while FourierDiff [31] adopts a zero-shot diffusion strategy to recover details and enhance both brightness and structure. Ye *et al*. [54] integrate Transformer modules with structural priors, PDHAT [26] introduces perceptual decoupling via auxiliary tasks, and LIEDNet [27] leverages a Mamba-based framework to fuse global and local features for end-to-end restoration.

Nevertheless, most existing methods are constrained to the RGB space, where inter-channel coupling hampers the joint preservation of color fidelity and structural consistency, often causing cross-distortions. In contrast, our approach operates in the HVI space, where separating luminance and chromaticity mitigates channel interference. Combined with structural guidance, this enables more effective blur removal and illumination enhancement.

3. Method

This paper presents a novel method to address the joint degradation of low-light and blur in images. As shown in Fig. 2(a), given a low-light blurry image in the RGB space, the objective is to restore a clear image under normal illumination. Prior studies [52] have shown that enhancing low-light images in the HVI color space is more effective than in RGB or HSV, as it decouples luminance and chromaticity while reducing artifacts. Based on this, our method adopts

the HVI space and employs its standard RGB–HVI conversion scheme. We introduce the structure-aware HVI-based image enhancement network (SHINE-Net), whose backbone is dual-branched, separately modeling the HV-plane and I-axis information with HVI three-channel and I single-channel inputs. The overall design follows a U-Net–like architecture [40], consisting of an encoder, a decoder, and a structure-aware enhancement subnetwork. The encoder extracts and enhances features, while the decoder further refines them. Since Lighten Cross-Attention (LCA) [52] has been shown to effectively strengthen interactions between brightness and color branches, we incorporate LCA at every resolution level of both branches.

To enhance the network's ability for restoring severely degraded structures and textures in low-light blurry images, we design a Structure-Aware Enhancement Subnetwork (SAES). This subnetwork captures richer high-frequency and edge information and improves the structural and textural fidelity of reconstructed images by guiding multi-feature fusion, thereby alleviating non-uniform blur. Within SAES, we propose a Structure-aware Coordinate Attention Block (SCAB), which adaptively exploits image features and two types of structural features to optimize the reconstruction of clear images under normal illumination. Furthermore, in the HV branch, we introduce a Dynamic Feature-Aware Enhancement (DFAE) module, which employs skip connections to transform enhanced encoder features into dynamic convolutional kernels. These kernels guide the decoder to adaptively reconstruct blurred regions, enabling fine-grained feature recovery.

3.1. Dynamic Feature-Aware Enhancement Module

Blurred images often exhibit pronounced spatial variations. In dynamic scenes, blur typically results from local object motion or depth-of-field differences. For example, moving objects may produce motion trails while static backgrounds remain sharp. In such cases, blur kernels are strongly spatially dependent, rendering conventional static convolution kernels ineffective in handling this heterogeneity. To enhance structural recovery under low-light conditions with complex mixed blur, we incorporate a novel Dynamic Feature-Aware Enhancement (DFAE) module into the HV branch of the proposed network. Specifically, three DFAE units are deployed at different hierarchical levels of the HV branch, as illustrated in Fig. 2. Inspired by [65] and LEDNet [64], the DFAE module leverages skip connections to transform enhanced encoder features into dynamic convolution kernels, which guide the decoder to adaptively reconstruct structures in blurred regions, thus achieving fine-grained feature restoration. Unlike static convolution, the dynamic kernels in DFAE are explicitly conditioned on the encoder features, enabling the module to perceive local blur severity and semantic context, and to adapt the enhancement strategy accordingly at each spatial location.

The DFAE module introduces a dynamic kernel generation mechanism, where position-dependent convolution kernels are derived from the encoder's enhanced features at multiple scales. This design results in spatially variant and feature-adaptive convolution kernels, where each kernel is specialized for the local structure and degradation pattern at its corresponding spatial position. As shown in Fig. 3, given multi-scale features E from the encoder, a lightweight convolutional stack (three 3×3 convolutions followed by one 1×1 convolution) predicts dynamic kernels $K \in \mathbb{R}^{C \times H \times W \cdot d^2}$. Each spatial position (x, y) produces a kernel $K_{(x,y)} \in \mathbb{R}^{C \times d \times d}$, where d is set to 5 following [65]. For the decoder features D, the DFAE module extracts local patches $D^{\text{patch}} \in \mathbb{R}^{C \times d \times d}$ around each position and applies the corresponding dynamic kernel:

$$D'_{(x,y)} = \sum_{i=1}^{C} \sum_{u=1}^{d} \sum_{v=1}^{d} K_{(x,y)}(i,u,v) \cdot D^{\text{patch}}_{(x,y)}(i,u,v). \quad (1)$$

By modulating decoder features with kernels dynamically generated from encoder representations, DFAE enables feature-aware enhancement aligned with region-specific blur characteristics. The module outputs enhanced features D', preserving spatial resolution while adaptively refining structures and edges. In this way, DFAE performs content-driven dynamic adaptation in the HV branch, improving the restoration of spatially variant structures.

3.2. Structure-Aware Enhancement Subnet

Although joint methods for low-light enhancement and deblurring have advanced, most still rely on a single structural prior and operate in the RGB space. Owing to the strong inter-channel coupling in RGB, these approaches often fail to recover severely degraded structures and textures in low-light blurred images. To improve reconstruction quality, inspired by [54], we design a Structure-Aware Enhancement Subnetwork (SAES) in the HV branch of our framework. The SAES captures richer high-frequency and edge structural information, which complement each other and, when fused with image features, enhance both fine detail recovery and structural integrity in the reconstructed image. In this work, structure refers to spatially salient patterns such as edges, contours, and high-frequency texture boundaries that are critical for accurate deblurring and visual fidelity.

In the HV branch, we introduce the Structure-Aware Enhancement Subnetwork (SAES) between the encoder and decoder to reconstruct intermediate features under structural guidance. Operating at a lower spatial resolution via downsampling, SAES reduces computational cost while preserving reconstruction capacity. As illustrated in Fig. 2(b), SAES takes the deep HV features F_{enc} as input and

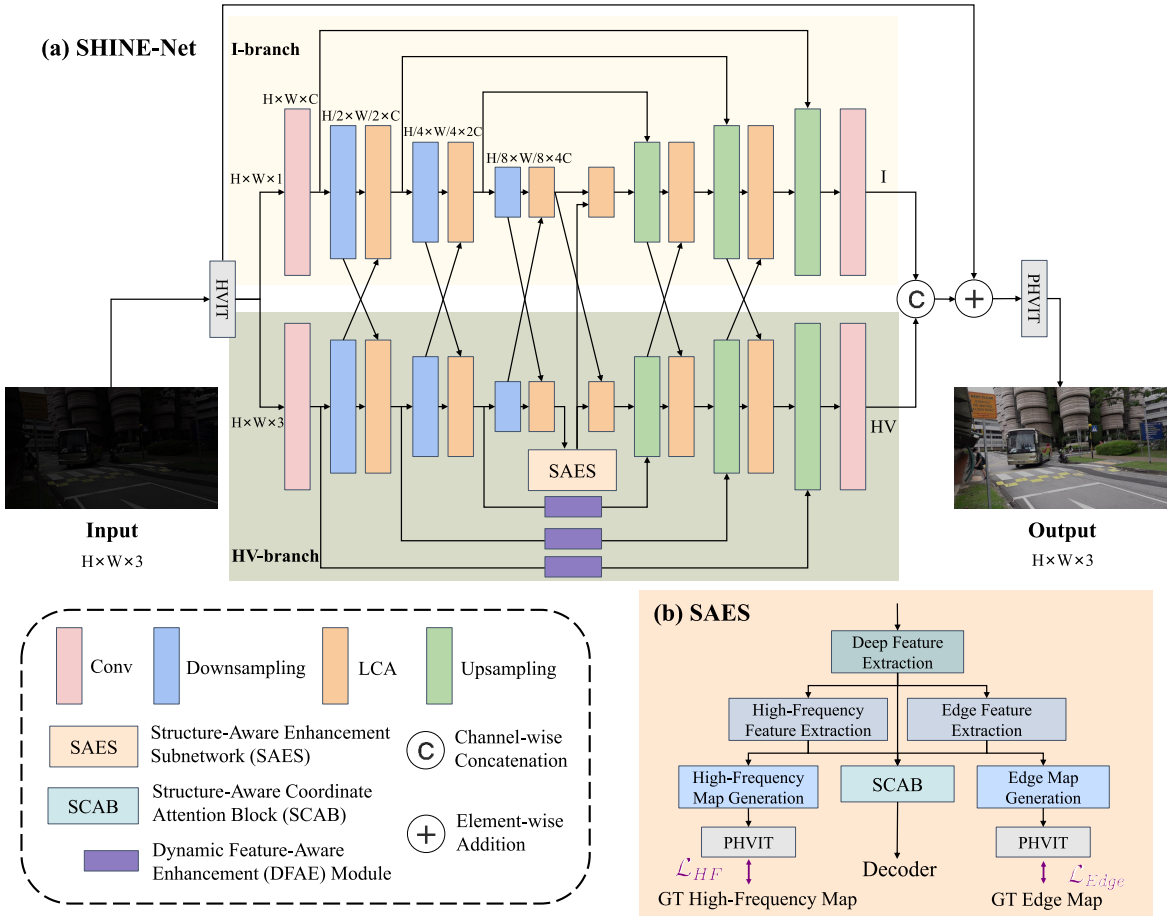


Figure 2. The overview of SHINE-Net. (a) SHINE-Net adopts HVIT to transform an sRGB image into HV and I maps, which are processed by a dual-branch architecture with LCA, SAES, and DFAE, and then fused and converted by PHVIT to generate an enhanced sRGB image. (b) The SAES extracts high-frequency and edge features and fuses them via SCAB to guide structural and textural restoration.

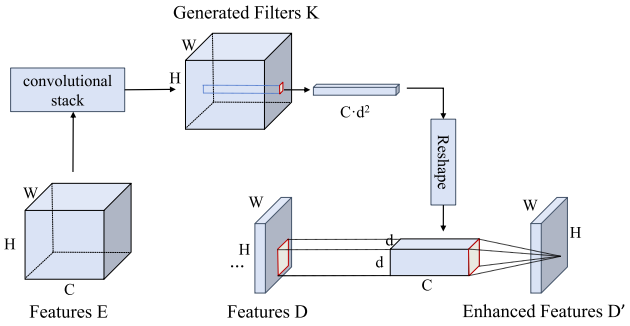


Figure 3. Overview of the DFAE module. Position-dependent dynamic kernels are generated from encoder features and applied to decoder features for adaptive structural restoration.

consists of three components: deep feature extraction, high-frequency/edge modeling branches, and SCAB for guided feature fusion. First, the input features are processed by five stacked residual blocks to obtain the deep representation:

$$F_{\text{deep}} = \text{ResBlock}_5(F_{\text{enc}}). \tag{2}$$

Then, F_{deep} is refined through two parallel branches, generating high-frequency features F_{hf} and edge features F_{ed} with residual block sequences:

$$F_{hf}, F_{ed} = \text{ResBlock}_{hf}(F_{\text{deep}}),\ \text{ResBlock}_{ed}(F_{\text{deep}}). \tag{3}$$

Through these parallel branches, SAES explicitly disentangles and models complementary structural cues, enabling the network to perceive and preserve fine-grained spatial details under low-light blur. To supervise these branches, the ground-truth (GT) image is downsampled by a factor of 8 in RGB space, from which high-frequency and edge maps are extracted as regression targets. To avoid color-space mismatch, the predicted features are converted back to RGB before loss computation. An L_1 loss is applied for normalization, ensuring accurate learning of statistical priors during training. Additional visual comparisons of the in-

termediate high-frequency and edge representations learned by SAES are provided in the supplementary material.

Finally, the original features are fused with the high-frequency and edge features, and the subnetwork outputs structure-aware enhanced HV representations, which are subsequently fed into the decoder. By incorporating SAES, the model gains a markedly improved capacity to perceive and restore edge structures and textures in low-light blurred regions.

3.3. Structure-Aware Coordinate Attention Block

Since image, high-frequency, and edge features represent three independent information, simple concatenation cannot fully exploit their complementarity. Furthermore, while conventional channel attention mechanisms (*e.g.* SE [13] and ECA [44]) are effective for feature fusion, they neglect positional information, which is essential for constructing spatial attention maps. This omission often leads to suboptimal deblurring performance [12]. To overcome this limitation, we propose a Structure-aware Coordinate Attention Block (SCAB), which explicitly encodes spatial information during feature fusion. SCAB integrates structure-aware cues into the attention generation process, enabling spatially important regions, such as edges and texture boundaries, to receive stronger responses. By jointly emphasizing texture details from the high-frequency branch and contour information from the edge branch, SCAB adaptively fuses complementary structural features to improve overall restoration quality.

Specifically, as illustrated in Fig. 4, the deep feature F_{deep} is first combined with the high-frequency and edge features:

$$F_{sum}^{hf} = F_{deep} + F_{hf}, \quad F_{sum}^{ed} = F_{deep} + F_{ed}. \tag{4}$$

Both F_{sum}^{hf} and F_{sum}^{ed} are processed by the coordinate attention mechanism [12]. Taking the high-frequency branch as an example, the feature F_{sum}^{hf} is aggregated along horizontal and vertical directions via average pooling to generate two context vectors:

$$\begin{aligned} F_{avg}^{x} &= \frac{1}{W}\sum_{i=1}^{W} F_{sum}^{hf}(:,:,i) \in \mathbb{R}^{C\times H\times 1}, \\ F_{avg}^{y} &= \frac{1}{H}\sum_{j=1}^{H} F_{sum}^{hf}(:,j,:) \in \mathbb{R}^{C\times 1\times W}. \end{aligned} \tag{5}$$

These vectors are concatenated and transformed by a 1×1 Conv–BN–ReLU layer to obtain an intermediate embedding:

$$F_{int}^{hf} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{1\times 1}([F_{avg}^{x}; F_{avg}^{y}]))), \tag{6}$$

where $r = 8$ is the reduction ratio. F_{int}^{hf} is split into F_{x}^{hf} and F_{y}^{hf}, each passed through a 1×1 convolution and sigmoid activation to produce horizontal and vertical attention maps:

$$\begin{aligned} W_{x}^{hf} &= \sigma(\mathrm{Conv}_{1\times 1}(F_{x}^{hf})) \in \mathbb{R}^{C\times H\times 1}, \\ W_{y}^{hf} &= \sigma(\mathrm{Conv}_{1\times 1}(F_{y}^{hf})) \in \mathbb{R}^{C\times 1\times W}. \end{aligned} \tag{7}$$

The attention maps reweight F_{sum}^{hf}, yielding the enhanced high-frequency feature:

$$F_{att}^{hf} = F_{sum}^{hf} \odot W_{x}^{hf} \odot W_{y}^{hf}. \tag{8}$$

Similarly, the edge branch produces F_{att}^{ed}. The final fused representation is

$$\begin{aligned} F_{fused} &= F_{deep} + F_{att}^{hf} + F_{att}^{ed}, \\ F_{out} &= \mathrm{ResGroup}(F_{fused}), \end{aligned} \tag{9}$$

where $\mathrm{ResGroup}(\cdot)$ denotes a stack of 10 residual blocks that further enhance structure-aware features before passing them into the decoder.

In practice, the contributions of DFAE and SAES vary across spatial regions. In areas dominated by motion blur, DFAE plays a more prominent role by adapting convolutional behaviors to local degradation patterns. Conversely, in regions with clearer edges and textures, SAES and SCAB emphasize structural cues to reinforce boundary sharpness and texture continuity. This complementary interaction enables the network to flexibly balance dynamic adaptation and structural guidance across diverse low-light blur scenarios.

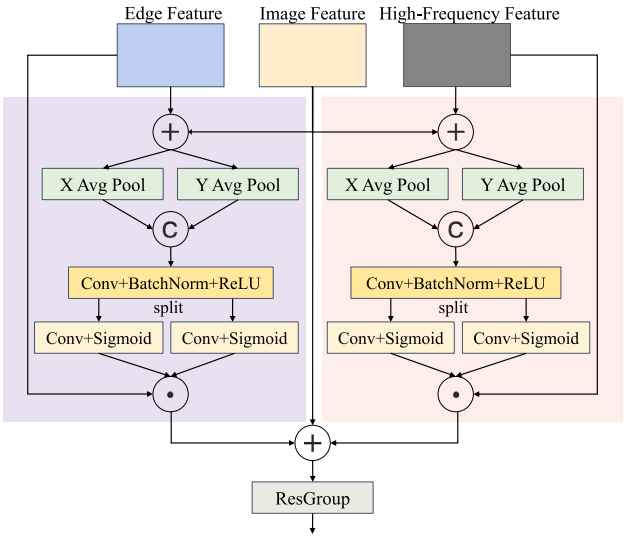


Figure 4. The proposed structure-aware coordinate attention block. High-frequency and edge features are summed with image features, passed through coordinate attention, and then refined by a residual group.

3.4. Loss Functions

To achieve effective enhancement of low-light blurred images, we design a composite loss function with four com-

ponents. Two of them enforce color-space consistency, namely the HVI-space consistency loss and the RGB-space consistency loss, while the other two ensure structural fidelity, including the high-frequency loss and the edge loss. The overall training objective is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{RGB} + \lambda_{HVI} \cdot \mathcal{L}_{HVI} + \lambda_{struct} \cdot \left(\mathcal{L}_{HF} + \mathcal{L}_{Edge}\right), \tag{10}$$

where λ_{HVI} and λ_{struct} are hyperparameters balancing the contributions of each term. Specifically, λ_{HVI} regulates the influence of color consistency, while λ_{struct} controls the strength of structural supervision. The effect of λ_{struct} is further analyzed through an ablation study, with detailed results provided in the supplementary material.

Color Space Consistency Loss. To enhance the accuracy and continuity of image distribution in the HVI color space, optimize color transitions, reduce artifacts, and ensure pixel-level fidelity in the sRGB space, we introduce the HVI-space consistency loss $\mathcal{L}_{HVI}$ and the RGB-space consistency loss $\mathcal{L}_{RGB}$. Following the design in [52], the key intuition is that the enhanced outputs in both spaces should remain consistent with their corresponding ground truth (GT). Specifically, the enhanced HVI map should approximate the HVI representation converted from the GT image, while the restored RGB image should closely match the GT sRGB image. Formally, the losses are defined as:

$$\mathcal{L}_{HVI} = l(\hat{I}_{HVI}, I_{HVI}), \quad \mathcal{L}_{RGB} = l(\hat{I}_{RGB}, I_{RGB}), \tag{11}$$

where $\hat{I}_{HVI}$ and $\hat{I}_{RGB}$ denote the enhanced results in the HVI and RGB spaces, respectively, and I_{HVI} and I_{RGB} are the corresponding ground-truth targets. The function $l(\cdot)$ is a hybrid loss combining L1 loss, SSIM loss, Laplacian pyramid loss, and perceptual loss.

Structural Detail Loss. To strengthen the network's ability to restore edge structures and textures, we further introduce the high-frequency loss $\mathcal{L}_{HF}$ and the edge loss $\mathcal{L}_{Edge}$ as supervisory terms for the structure-aware enhancement subnetwork (SAES). For the high-frequency loss, we employ the Discrete Cosine Transform (DCT) [38] to extract high-frequency components. Let F_{pred} denote the predicted high-frequency map obtained from SAES via the perceptual-inverse HVI transformation (PHVIT) [52], and F_{GT} the ground-truth map derived by applying DCT to the downsampled ($\times 8$) normal-light sharp image. The loss is defined as:

$$\mathcal{L}_{HF} = \|F_{pred} - F_{GT}\|_1. \tag{12}$$

For the edge loss, we adopt the Canny operator [3] to extract edge information. Specifically, E_{pred} denotes the predicted edge map generated by SAES through PHVIT, and E_{GT} is the ground-truth edge map obtained by applying the Canny operator to the downsampled ($\times 8$) normal-light sharp image. The loss is defined as:

$$\mathcal{L}_{Edge} = \|E_{pred} - E_{GT}\|_1. \tag{13}$$

4. Experiments

In this section, we first present the datasets and evaluation metrics, followed by implementation details, comparisons with state-of-the-art methods, and ablation studies. In addition, we report the computational complexity, including the number of parameters, FLOPs, and inference time, in the supplementary material.

4.1. Evaluation Datasets and Metrics

For low-light deblurring, models were trained on LOL-Blur [64] and tested on both LOL-Blur and the Real-LOL-Blur dataset [64]. For low-light enhancement, we used LOL-v1 [49], LOL-v2 synthetic [53], and five real-world datasets: DICM [21], LIME [10], MEF [33], NPE [45], and VV [43]. LOL-Blur provides 10,200 training pairs (1120×640) and 1,800 test pairs, while Real-LOL-Blur includes 240 real low-light blurred images and 160 nighttime samples from RealBlur [39]. The five real-world datasets (NPE, LIME, MEF, DICM, and VV) include 8, 10, 17, 64, and 24 images, respectively, with varying resolutions.

For datasets with ground truth (LOL-Blur, LOL), we report PSNR and SSIM [47], as well as LPIPS [59] using AlexNet [19]. For Real-LOL-Blur, we use no-reference metrics MUSIQ [16], NRQM [32], and NIQE [35]. For real-world low-light datasets (NPE, LIME, MEF, DICM, VV), we use NIQE and BRISQUE [34]. MUSIQ, trained on KonIQ-10k, is particularly sensitive to contrast and sharpness.

4.2. Experimental Settings

Our model is implemented in PyTorch and trained on an NVIDIA RTX 4090. We use the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of 1×10^{-4}, decayed to 1×10^{-7} by cosine annealing. Training lasts 400 epochs with batch size 4 on randomly cropped 256×256 patches. Loss weights are set to $\lambda_{HVI} = 1$ and $\lambda_{struct} = 0.01$. For evaluations on real-world datasets, the pretrained model on synthetic data is directly applied without fine-tuning.

4.3. Joint Low-Light Enhancement and Deblurring

We compare our model against 13 representative baselines, categorized into five groups. (1) Cascaded methods of enhancement followed by deblurring: Zero-DCE [8] → MIMO [5] and RUAS [28] → MIMO [5]. (2) Cascaded methods of deblurring followed by enhancement: MIMO [5] → Zero-DCE [8] and Restormer [56] → LLFlow [46]. (3) Low-light enhancement models:

Method	Color Model	PSNR↑	SSIM↑	LPIPS↓
Zero-DCE [8] → MIMO [5]	RGB	17.68	0.542	0.510
RUAS [28] → MIMO [5]	RGB	17.81	0.569	0.523
MIMO [5] → Zero-DCE [8]	RGB	17.52	0.570	0.498
Restormer [56] → LLFlow [46]	RGB	21.89	0.772	0.347
KinD++ [61]	Retinex	21.26	0.753	0.359
DMPHN [58]	RGB	22.20	0.817	0.301
MIMO [5]	RGB	22.41	0.835	0.262
Restormer [56]	RGB	23.63	0.841	0.247
LLFlow [46]	RGB	24.48	0.846	0.235
RetinexFormer [2]	Retinex	25.34	0.826	0.214
LEDNet [64]	RGB	26.18	0.874	0.179
LIEDNet [27]	RGB	27.48	0.884	0.132
CIDNet [52]	HVI	26.05	0.876	0.107
SHINE-Net (Ours)	HVI	26.87	0.902	0.093

Table 1. Quantitative results of different methods on the LOL-Blur dataset. The best performance is in red, the second best in blue, and the third best in violet.

KinD++ [61], LLFlow [46], RetinexFormer [2], and CIDNet [52]. (4) Deblurring models: DMPHN [58], MIMO [5], and Restormer [56]. (5) Joint enhancement and deblurring methods: LEDNet [64] and LIEDNet [27]. For fair comparison, we retrained the low-light enhancement and deblurring networks on the LOL-Blur dataset.

Evaluation on the LOL-Blur Dataset. Table 1 reports the quantitative results of different methods on the LOL-Blur dataset. Our method achieves the best performance in terms of SSIM and LPIPS, demonstrating its strong capability in handling the combined degradations of low light and blur. Overall, end-to-end models consistently outperform cascaded frameworks, highlighting the advantage of joint restoration for low-light blurred images. In addition, models operating in the HVI color space show clear superiority over those in the sRGB domain, confirming the potential of HVI-based enhancement. Furthermore, by effectively integrating high-frequency and edge information, our method achieves even greater improvements, surpassing the second-best approach by 0.018 in SSIM and 0.014 in LPIPS.

Figure 5 presents qualitative comparisons on the LOL-Blur dataset. Most existing methods primarily enhance image brightness but struggle with blur removal, leading to suboptimal restoration quality. In contrast, our method leverages the HVI color space to refine textures and edge structures, achieving superior recovery of structural details while simultaneously improving illumination. These results further demonstrate the effectiveness of SHINE-Net. More qualitative results are presented in the supplementary material.

Evaluation on Real-World Dataset. To further assess the generalizability of the proposed method, we conduct experiments on the Real-LOL-Blur dataset. Table 2 summarizes the results in terms of MUSIQ, NRQM, and NIQE. Our model achieves the best performance on NRQM and NIQE, demonstrating superior perceptual quality with sharper details and better color contrast. Consistent with the synthetic results, cascaded frameworks perform worse, and methods constrained to the sRGB space also yield suboptimal outcomes. In contrast, our approach leverages the HVI space to jointly restore illumination, texture, and structure, enabling adaptive optimization of degraded images and stronger generalization to real-world scenarios.

Method	MUSIQ↑	NRQM↑	NIQE↓
Zero-DCE [8] → MIMO [5]	39.36	5.206	4.459
RUAS [28] → MIMO [5]	34.39	3.322	6.812
MIMO [5] → Zero-DCE [8]	28.36	3.697	6.892
Restormer [56] → LLFlow [46]	35.42	5.011	4.982
KinD++ [61]	31.74	3.854	7.299
DMPHN [58]	35.08	4.470	5.190
MIMO [5]	35.37	5.140	4.851
Restormer [56]	36.65	5.497	5.093
LLFlow [46]	34.87	5.312	5.202
RetinexFormer [2]	35.60	4.414	5.441
LEDNet [64]	43.48	5.729	4.509
LIEDNet [27]	44.74	6.069	4.289
CIDNet [52]	45.89	6.477	4.150
SHINE-Net (Ours)	45.78	6.478	4.134

Table 2. Quantitative results of different methods on the Real-LOL-Blur dataset. The best performance is in red, the second best in blue, and the third best in violet.

Figure 6 shows the visual comparisons on the Real-LOL-Blur dataset. As observed, cascaded methods often introduce severe noise, artifacts, and illumination imbalance, demonstrating that simple combinations of low-light enhancement and deblurring are insufficient to handle the coexisting degradations. In low-light regions, most existing methods fail to recover texture details, resulting in noticeable artifacts and residual blur. In contrast, our network effectively decouples brightness and color to restore a natural illumination, while simultaneously modeling high-

Figure 5. Qualitative comparison on the LOL-Blur dataset. Compared with existing methods, SHINE-Net achieves better illumination enhancement and more faithful recovery of structural details.

frequency and edge information. By adaptively incorporating these structural cues into the enhancement process, our method reconstructs richer and sharper structural details. More qualitative results are presented in the supplementary material.

4.4. Low-Light Enhancement

For the low-light image enhancement (LLIE) task, we compare our method against ten representative approaches, including one traditional method LIME [10], one unsupervised learning method RUAS [28], and eight supervised learning methods: RetinexNet [49], KinD [62], Uformer [48], Restormer [56], SNR-Aware [51], RetinexFormer [2], RetinexMamba [1], and CIDNet [52].

Evaluation on Synthetic LLIE Datasets. Table 3 presents the results on the LOL-v1 and LOL-v2 synthetic datasets. SHINE-Net consistently achieves superior performance across both benchmarks, with SSIM scores of 0.868 on LOL-v1 and 0.941 on LOL-v2 synthetic, surpassing state-of-the-art methods. These results suggest that our model restores images with higher structural similarity, benefiting from its adaptive recovery of texture and edge structures.

Figures 7 show the qualitative results on LOL-v1 dataset. Previous methods often suffer from color distortions, exposure imbalance, and detail loss, which compromise the overall restoration quality. In contrast, our method delivers results with more consistent brightness and color, along with sharper structural details, achieving superior visual quality that is perceptually closer to the ground truth. More qualitative results are presented in the supplementary material.

Evaluation on Real-World Datasets. In addition to the synthetic LLIE datasets, we further evaluate our method on five real-world datasets, with the results summarized in Table 4. Overall, SHINE-Net consistently outperforms competing approaches across different datasets. For instance, on the DICM dataset, our method achieves a NIQE score of 3.849, representing improvements of 1.038, 0.386, and 0.166 over LEDNet, LIEDNet, and CIDNet, respectively. Moreover, SHINE-Net attains the best NIQE scores on the DICM and VV datasets, as well as the best BRISQUE scores on the LIME and MEF datasets. These results demonstrate the robustness and generalizability of SHINE-Net in handling real-world LLIE tasks. Qualitative visual comparisons on real-world datasets are provided in the supplementary material, further validating the effectiveness of our method under diverse lighting conditions.

Method	Color Model	LOL-v1		LOL-v2 Synthetic	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑
LIME [10]	Retinex	16.34	0.624	17.19	0.764
RUAS [28]	RGB	16.40	0.503	16.55	0.652
Retinex-Net [49]	Retinex	17.19	0.589	17.13	0.798
KinD [62]	Retinex	20.35	0.813	13.29	0.578
Uformer [48]	RGB	18.55	0.721	19.66	0.871
SNR-Aware [51]	SNR+RGB	24.61	0.842	24.14	0.928
Restormer [56]	RGB	22.37	0.816	21.41	0.830
RetinexFormer [2]	Retinex	23.93	0.831	25.67	0.930
RetinexMamba [1]	Retinex	24.03	0.827	25.89	0.935
CIDNet [52]	HVI	23.81	0.857	25.13	0.939
SHINE-Net (Ours)	HVI	24.26	0.868	25.69	0.941

Table 3. Quantitative comparison on the LOL-v1 and LOL-v2 Synthetic datasets. The best performance is in red, the second best in blue, and the third best in violet.

4.5. Ablation Study

To validate the effectiveness of the proposed method, we conduct ablation experiments on the LOL-Blur dataset. The evaluation focuses on three aspects: (1) the effectiveness of the DFAE module, (2) the effectiveness of the SCAB, and (3) the contribution of structural guidance.

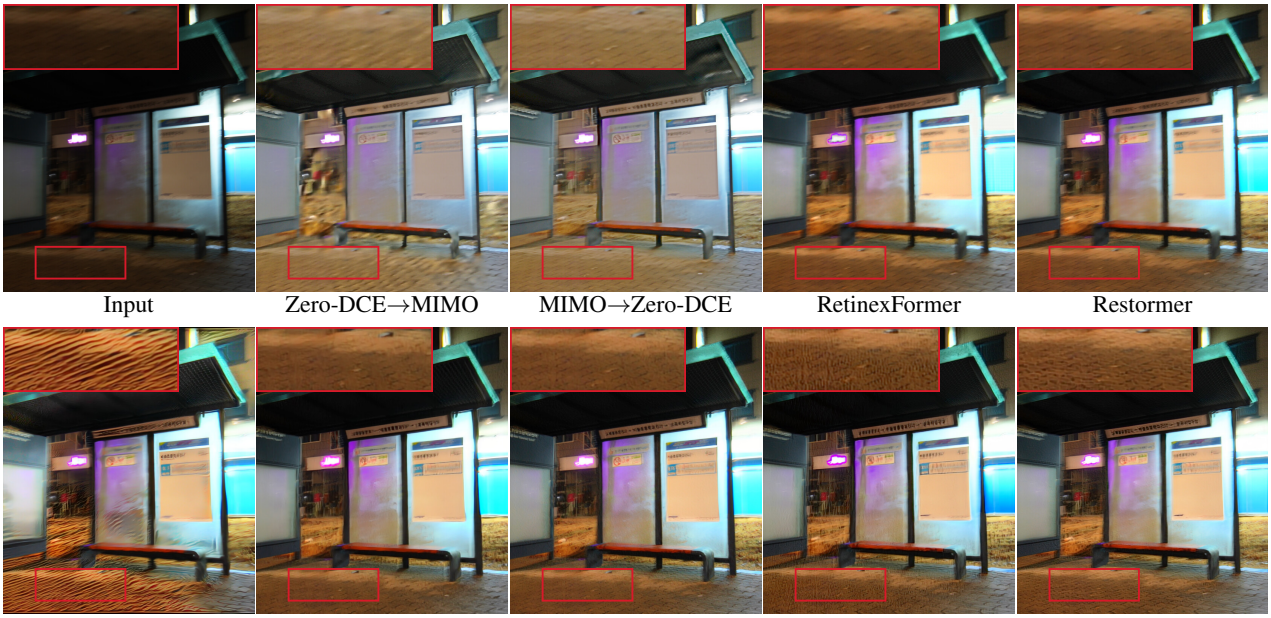


Figure 6. Qualitative comparison with previous methods on the Real-LOL-Blur dataset. Our method effectively suppresses noise and artifacts while recovering natural illumination and richer structural details.

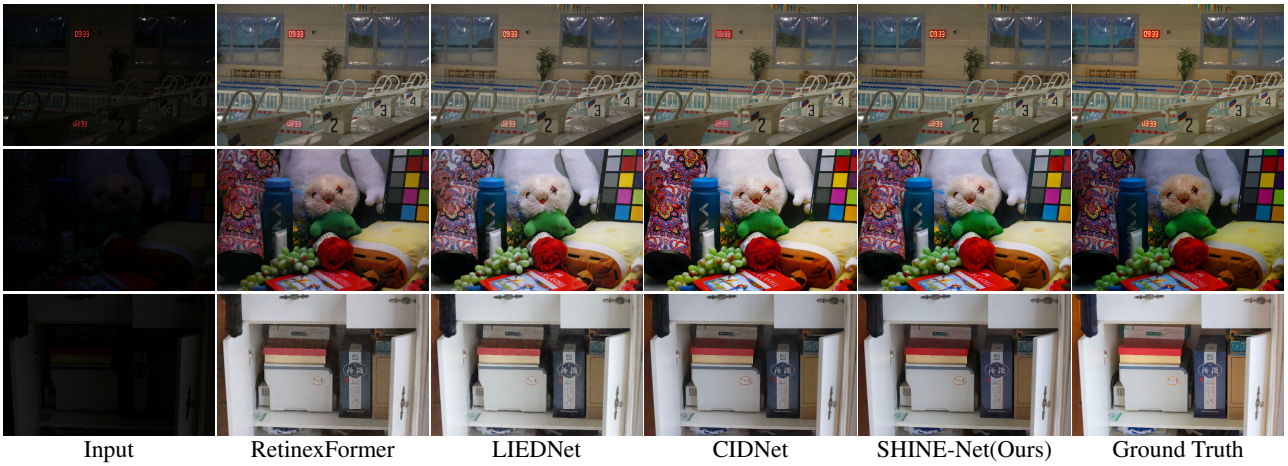


Figure 7. Qualitative comparison with previous methods on the LOL-v1 datasets. Our SHINE-Net produces results with more consistent brightness, accurate colors, and sharper details, closely resembling the ground truth.

Effectiveness of the DFAE Module. Table 5 reports the experimental results. Applying the DFAE module only to the I-branch yields a PSNR of 26.29 dB, improving the baseline by 0.25 dB. When extending the DFAE module to both the HV-branch and the I-branch, the performance increases marginally compared to applying it only to the I-branch, but remains inferior to the HV-only configuration. This phenomenon can be attributed to the distinct roles of the HV-plane and the I-plane in the HVI color space. The HV-plane primarily encodes chromaticity and structural cues, where dynamic feature modulation effectively enhances texture and edge-aware representations. In contrast, applying DFAE to the I-branch, which mainly represents global illumination, may introduce redundant or conflicting dynamic adjustments that interfere with stable luminance restoration. As a result, jointly applying DFAE to both branches does not yield further gains. In our final design, applying the DFAE module solely to the HV-branch achieves the best trade-off, reaching 26.87 dB PSNR with relatively low overhead. As illustrated in Figure 8, this design leads to clear improvements in detail restoration, color stability, and overall sharpness, validating the effectiveness

Method	DICM		LIME		MEF		NPE		VV	
	NIQE↓	BRISQUE↓	NIQE↓	BRISQUE↓	NIQE↓	BRISQUE↓	NIQE↓	BRISQUE↓	NIQE↓	BRISQUE↓
LEDNet [64]	4.887	33.618	4.631	29.694	4.892	35.825	5.490	27.491	4.047	33.514
LIEDNet [27]	4.235	21.269	4.113	19.457	4.455	23.209	4.223	15.810	3.363	24.227
CIDNet [52]	4.015	30.814	3.903	16.134	3.461	13.100	3.639	16.753	3.243	29.551
SHINE-Net (Ours)	3.849	26.708	4.004	15.209	3.740	11.675	3.880	17.902	3.121	26.681

Table 4. Quantitative results on five real-world low-light datasets in terms of NIQE and BRISQUE (lower is better).

Method	PSNR↑	SSIM↑	LPIPS↓
without DFAE	26.04	0.875	0.105
DFAE only applied to I-branch	26.29	0.879	0.107
DFAE applied to I-branch and HV-branch	26.44	0.883	0.100
Ours	26.87	0.892	0.093

Table 5. Ablation study on the effectiveness of DFAE.

Method	PSNR↑	SSIM↑	LPIPS↓
Direct Fusion	26.59	0.885	0.096
Direct Fusion with an Attention	26.35	0.887	0.095
Cascade Fusion (High-Frequency→Edge)	26.31	0.885	0.101
Cascade Fusion (Edge→High-Frequency)	26.07	0.882	0.101
Ours	26.87	0.892	0.093

Table 6. Ablation study on different feature fusion strategies.

of the DFAE module for joint low-light deblurring.

Effectiveness of the SCAB. We further evaluate several strategies for fusing image, high-frequency, and edge information, as summarized in Table 6. The first strategy concatenates the three feature types along the channel dimension and applies a 3×3 convolution for fusion, denoted as Direct Fusion. To enhance adaptability, we introduce a coordinate attention (CA) module [12] that jointly models channel and positional information, referred to as Direct Fusion with an Attention. We also design two cascade-style baselines: Cascade Fusion (High-Frequency→Edge), which first fuses image and high-frequency features followed by edge features, and Cascade Fusion (Edge→High-Frequency), which reverses the order. Table 6 reports the quantitative results, showing that Direct Fusion with an Attention outperforms Direct Fusion, while both cascade approaches yield limited improvements. In contrast, the proposed SCAB achieves consistent advantages across all metrics by effectively leveraging high-frequency and edge cues, thereby producing sharper textures and clearer structures. A visual example in Figure 9 further demonstrates that integrating SCAB yields more natural illumination, richer details, and fewer artifacts in restored normal-light images.

Contribution of Structural Guidance. We evaluate the effectiveness of structural guidance in Table 7. In SAES, incorporating either high-frequency or edge information via feature modeling, together with the corresponding supervision, consistently improves all metrics, confirming that both cues are beneficial for restoration. By jointly integrating high-frequency and edge information, the model achieves the best performance, with a PSNR improvement of 0.53 dB over the baseline. As shown in Fig. 10, high-frequency cues enhance fine details, while edge cues improve global structural consistency. Notably, applying structural guidance to the HV-branch yields consistent gains, indicating that the HV plane provides a stable representation for structural modeling, as it captures chromatic and spatial variations while being less sensitive to global illumination changes.

Method	PSNR↑	SSIM↑	LPIPS↓
without frequency and edge	26.34	0.883	0.102
with frequency	26.53	0.884	0.098
with edge	26.71	0.886	0.101
Ours	26.87	0.892	0.093

Table 7. Ablation study on the effectiveness of frequency and edge information.

5. Limitations and Future Work

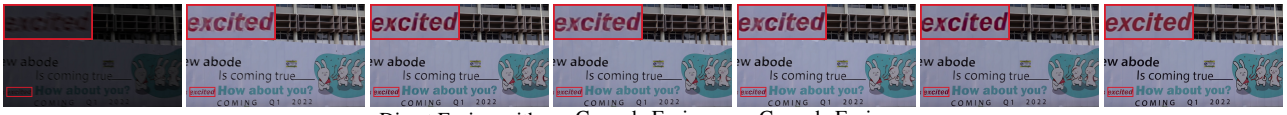
Although SHINE-Net achieves strong performance on both synthetic and real-world datasets, its effectiveness degrades under extreme conditions. As shown in Fig. 11, extremely low illumination combined with severe motion blur may lead to over-enhancement or unclear reconstructions, as recoverable structural signals become insufficient. In such cases, high-frequency and edge priors become unreliable due to corrupted or ambiguous structures, which explains the observed failure cases. Such extreme degradations remain challenging for existing enhancement and deblurring approaches. Future work will explore more robust structural modeling and extend SHINE-Net to lightweight video enhancement by jointly exploiting temporal and spatial information.

6. Conclusion

This paper presents a unified restoration framework for addressing coupled low-light and blur degradations in images. Unlike prior methods that operate in the RGB space, we decouple luminance and chromatic information in the HVI space and incorporate structure-guided priors to enhance the synergy between illumination enhancement and structural recovery. The framework integrates three key modules: DFAE for dynamic feature modeling, SAES for leveraging high-frequency and edge information, and SCAB for structure-aware multi-feature fusion. Experi-

Input | without DFAE | DFAE only applied to I-branch | DFAE applied to HV&I-branch | Ours | Ground Truth

Figure 8. Visual comparisons of the DFAE module ablation study.



Input | Direct Fusion | Direct Fusion with an Attention | Cascade Fusion (Fre→Edge) | Cascade Fusion (Edge→Fre) | Ours | Ground Truth

Figure 9. Visual comparisons of the SCAB ablation study.

Input | without frequency and edge | with frequency | with edge | Ours | Ground Truth

Figure 10. Visual comparisons of structural guidance ablation study.

Figure 11. Failure cases under extreme degradations. Our method struggles with extremely low-light and heavily blurred scenes, leading to incomplete brightness enhancement and loss of details.

ments on both synthetic and real-world datasets show that our method consistently surpasses existing approaches, delivering sharper details, more accurate colors, and superior overall visual quality.

Acknowledgement

This study was funded by Zhejiang Provincial Key R&D Plan No. 2023C01047 & 2023C01041.

References

[1] J. Bai, Y. Yin, Q. He, Y. Li, and X. Zhang. Retinexmamba: Retinex-based mamba for low-light image enhancement. In *International Conference on Neural Information Processing*, pages 427–442. Springer, 2024. 1, 3, 9

[2] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12504–12513, 2023. 1, 8, 9

[3] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 2009. 7

[4] S. Cho and S. Lee. Fast motion deblurring. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–8. ACM, 2009. 3

[5] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 1, 7, 8

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *Acm Siggraph 2006 Papers*, pages 787–794. ACM, 2006. 3

[7] Z. Fu, Y. Zheng, T. Ma, H. Ye, J. Yang, and L. He. Edge-aware deep image deblurring. *Neurocomputing*, 502:37–47, 2022. 3

[8] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1780–1789, 2020. 1, 3, 7, 8

[9] X. Guo and Q. Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1):48–66, 2023. 3

[10] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 7, 9

[11] J. Hou, Z. Zhu, J. Hou, H. Liu, H. Zeng, and H. Yuan. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36:79734–79747, 2023. 3

[12] Q. Hou, D. Zhou, and J. Feng. Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13713–13722, 2021. 6, 11

[13] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 6

[14] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021. 3

[15] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 3

[16] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 7

[17] L. Kong, J. Dong, J. Ge, M. Li, and J. Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023. 3

[18] L. Kong, J. Dong, J. Tang, M.-H. Yang, and J. Pan. Efficient visual state space model for image deblurring. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12710–12719, 2025. 1, 3

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 7

[20] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical society of America*, 61(1):1–11, 1971. 3

[21] C. Lee, C. Lee, and C.-S. Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013. 3, 7

[22] J. Lee, J. Park, S. Baik, and K. M. Lee. Rethinking rgb color representation for image restoration models. *arXiv preprint arXiv:2402.03399*, 2024. 1

[23] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE conference on computer vision and pattern recognition*, pages 1964–1971. IEEE, 2009. 3

[24] C. Li, C. Guo, and C. C. Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE transactions on pattern analysis and machine intelligence*, 44(8):4225–4238, 2021. 3

[25] J. Li, B. Li, Z. Tu, X. Liu, Q. Guo, F. Juefei-Xu, R. Xu, and H. Yu. Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15205–15215, 2024. 3

[26] Y. Li, R. Xu, Y. Niu, W. Guo, and T. Zhao. Perceptual decoupling with heterogeneous auxiliary tasks for joint low-light image enhancement and deblurring. *IEEE Transactions on Multimedia*, 26:6663–6675, 2024. 1, 3

[27] M. Liu, Y. Cui, W. Ren, J. Zhou, and A. C. Knoll. Liednet: A lightweight network for low-light enhancement and deblurring. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. 1, 3, 8, 11

[28] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo. 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*, pages 10561–10570, 2021. 1, 7, 8, 9

[29] Y. Liu, F. Fang, T. Wang, J. Li, Y. Sheng, and G. Zhang. Multi-scale grid network for image deblurring with high-frequency guidance. *IEEE Transactions on Multimedia*, 24:2890–2901, 2021. 3

[30] K. G. Lore, A. Akintayo, and S. Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 3

[31] X. Lv, S. Zhang, C. Wang, Y. Zheng, B. Zhong, C. Li, and L. Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25378–25388, 2024. 1, 3

[32] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017. 7

[33] K. Ma, K. Zeng, and Z. Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. 7

[34] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012. 7

[35] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 7

[36] J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Deblurring images via dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 40(10):2315–2328, 2017. 3

[37] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. 3

[38] K. R. Rao and P. Yip. *Discrete cosine transform: algorithms, advantages, applications*. Academic press, 2014. 7

[39] J. Rim, H. Lee, J. Won, and S. Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *European conference on computer vision*, pages 184–201. Springer, 2020. 7

[40] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4

[41] K. Shi, Y. Feng, T. Hu, Y. Cao, P. Wu, Y. Liang, Y. Zhang, and Q. Yan. Fusionnet: Multi-model linear fusion framework for low-light image enhancement. *arXiv preprint arXiv:2504.19295*, 2025. 3

[42] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8174–8182, 2018. 3

[43] V. Vonikakis, R. Kouskouridas, and A. Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77(8):9211–9231, 2018. 7

[44] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020. 6

[45] S. Wang, J. Zheng, H.-M. Hu, and B. Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9):3538–3548, 2013. 7

[46] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 2604–2612, 2022. 1, 7, 8

[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 7

[48] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 3, 9

[49] C. Wei, W. Wang, W. Yang, and J. Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 7, 9

[50] X. Xu, H. Liu, W. Liu, W. Wang, J. Wu, and K. Jiang. Iclr: Inter-chrominance and luminance interaction for natural color restoration in low-light image enhancement. *arXiv preprint arXiv:2511.13607*, 2025. 3

[51] X. Xu, R. Wang, C.-W. Fu, and J. Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 3, 9

[52] Q. Yan, Y. Feng, C. Zhang, G. Pang, K. Shi, P. Wu, W. Dong, J. Sun, and Y. Zhang. Hvi: A new color space for low-light image enhancement. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5678–5687, 2025. 1, 2, 3, 4, 7, 8, 9, 11

[53] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 7

[54] J. Ye, L. Yang, C. Qiu, and Z. Zhang. Joint low-light enhancement and deblurring with structural priors guidance. *Expert Systems with Applications*, 249:123722, 2024. 1, 3, 4

[55] X. Yi, H. Xu, H. Zhang, L. Tang, and J. Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12302–12311, 2023. 3

[56] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 1, 3, 7, 8, 9

[57] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 3

[58] H. Zhang, Y. Dai, H. Li, and P. Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5978–5986, 2019. 1, 3, 8

[59] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 7

[60] T. Zhang, P. Liu, M. Zhao, and H. Lv. Dmfourllie: Dual-stage and multi-branch fourier network for low-light image enhancement. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 7434–7443, 2024. 3

[61] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. 3, 8

[62] Y. Zhang, J. Zhang, and X. Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 3, 9

[63] S. Zheng, Z. Zhu, J. Cheng, Y. Guo, and Y. Zhao. Edge heuristic gan for non-uniform blind deblurring. *IEEE Signal Processing Letters*, 26(10):1546–1550, 2019. 3

[64] S. Zhou, C. Li, and C. Change Loy. Lednet: Joint low-light enhancement and deblurring in the dark. In *European conference on computer vision*, pages 573–589. Springer, 2022. 1, 3, 4, 7, 8, 11

[65] S. Zhou, J. Zhang, J. Pan, H. Xie, W. Zuo, and J. Ren. Spatio-temporal filter adaptive network for video deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2482–2491, 2019. 4

[66] Z. Zhu, J. Hou, H. Liu, H. Zeng, and J. Hou. Learning efficient and effective trajectories for differential equation-based image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 3