Uncertainty-Aware Spatial Color Correlation for Low-Light Image Enhancement

Jin Kuang[1,2,3], **Dong Liu**[1,2*], **Yukuang Zhang**[4], **Shengsheng Wang**[4]
[1]School of Computer and Artificial Intelligence, Xiangnan University
[2]Hunan Engineering Research Center of Advanced Embedded Computing and Intelligent Medical Systems, Xiangnan University
[3]School of Geosciences Yangtze University
[4]College of Computer Science and Technology, Jilin University
gasking.stu@yangtzeu.edu.com, liudong@xnu.edu.com, zyk24@mails.jlu.edu.com, wss@jlu.edu.cn

Abstract

Most existing low-light image enhancement approaches primarily focus on architectural innovations, while often overlooking the intrinsic uncertainty within feature representations particularly under extremely dark conditions where degraded gradient and noise dominance severely impair model reliability and causal reasoning. To address these issues, we propose U^{2}CLLIE, a novel framework that integrates uncertainty-aware enhancement and spatial-color causal correlation modeling. From the perspective of entropy-based uncertainty, our framework introduces two key components: (1) An Uncertainty-Aware Dual-domain Denoise (*UaD*) Module, which leverages Gaussian-Guided Adaptive Frequency Domain Feature Enhancement (*G2AF*) to suppress frequency-domain noise and optimize entropy-driven representations. This module enhances spatial texture extraction and frequency-domain noise suppression/structure refinement, effectively mitigating gradient vanishing and noise dominance. (2) A hierarchical causality-aware framework, where a Luminance Enhancement Network (*LEN*) first performs coarse brightness enhancement on dark regions. Then, during the encoder-decoder phase, two asymmetric causal correlation modeling modules—Neighborhood Correlation State Space (*NeCo*) and Adaptive Spatial-Color Calibration (*AsC*)—collaboratively construct hierarchical causal constraints. These modules reconstruct and reinforce neighborhood structure and color consistency in the feature space. Extensive experiments demonstrate that U^{2}CLLIE achieves state-of-the-art performance across multiple benchmark datasets, exhibiting robust performance and strong generalization across various scenes.

Introduction

Low-light image enhancement remains challenging due to noise, color distortion, and detail loss from poor illumination, impairing both visual perception and high-level tasks like detection and nighttime driving.

Early methods like histogram equalization, gamma correction, and Retinex-based curve fitting struggle to jointly restore global illumination and preserve local details, often introducing artifacts and noise. Deep learning methods(Guo and Hu 2023)(Wang et al. 2022) improve performance by learning global semantics and local features, but still face limitations: (1) CNNs, constrained by local receptive fields,

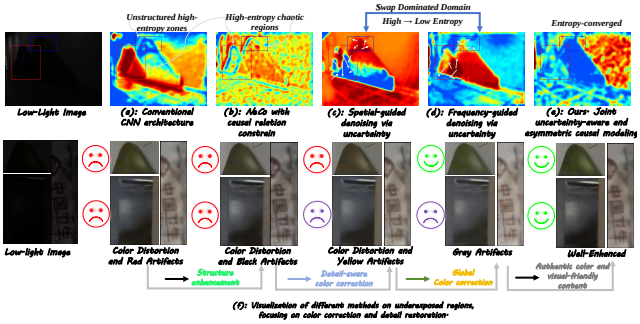

Figure 1: **Top row** compares entropy maps across five methods: (a) CNN baseline shows disordered entropy; (b) adding Mamba-based causal modeling reduces artifacts; (c) spatial-domain guidance enhances structural consistency; (d) frequency-domain cues improve color and structure fidelity; (e) our method yields the most coherent entropy and fine detail near edges. **Bottom row** shows corresponding enhancements, where the last column achieves artifact-free, locally consistent color and structure, highlighting the synergy of dual-domain uncertainty guidance and causal modeling.

fail to stably model underexposed regions and details, leading to chaotic features and color distortions (see Figure 1(a):

Red box shows unstable feature entropy causing representation chaos; blue box indicates blurred boundaries); (2) Transformers expand receptive fields via self-attention but suffer from high computational cost. Recent Mamba-based models(Gu and Dao 2023)(Liu et al. 2024) offer a balance between global modeling and efficiency via state space models (SSMs). Yet, directly applying Mamba to low-light enhancement is suboptimal due to weak local correlation modeling. Improvements like MambaLLIE(Weng et al. 2024) and BSMamba(Zhang et al. 2025a) introduce local bias or structure enhancement, but still lack hierarchical consistency across stages, resulting in artifacts (*e.g.*, Figure 1(f): 3rd column). Another key challenge lies in effective feature guidance. Some methods exploit frequency-aware fusion (FourLLIE(Wang, Wu, and Jin 2023), DMFourLLIE(Zhang et al. 2024)), semantic constraints (SCL-LLIE(Liang et al. 2022a)), or causal color modeling (SKF(Wu et al. 2023), SMGLLIE(Xu, Wang, and Lu 2023)), while others use

*corresponding author.

SNR-aware pixel enhancement(Xu et al. 2022). However, most overlook the role of uncertainty (*e.g.*, entropy) in low-light features. As shown in Figure 1(c–e), feature optimization can be seen as an entropy-reducing process, where lower entropy in dark regions leads to more stable features and better reconstruction (see Figure 1(f): last column).

To tackle these challenges, we propose U^{2}CLLIE (Uncertainty-Aware Spatial Color Correlation for Low-Light Image Enhancement), as shown in Figure 2. It addresses two core issues: **(1) Feature guidance collapse** under extreme darkness, where noise overwhelms gradients and disrupts causal color/detail correlation; **(2) Local perception modeling**, *i.e.*, effectively leveraging sparse low-light cues without strong priors to achieve consistent color and detail reconstruction for perceptual quality restoration.

Specifically, U^{2}CLLIE first employs a Luminance Enhancement Network (*LEN*) for global brightness adjustment. A Neighborhood Correlation State Space Module (*NeCo*) in the encoder models causal dependencies via learnable neighborhood-aware weights, while an Adaptive Spatial-Color Calibration Module (*AsC*) in the decoder captures object-level color correlations. These asymmetric modules impose progressive causal constraints to enhance detail and color reconstruction. To suppress structural noise and recover texture, an Uncertainty-Aware Dual-domain Denoising Module (*UaD*) is introduced, which guides enhancement via dual-domain (frequency/spatial) modeling under an entropy minimization perspective. Specifically, the Gaussian-Guided Adaptive Frequency Domain Feature Enhancement (*G2AF*) module performs adaptive frequency-domain denoising with Gaussian kernels. Our main contributions are:

1) Reformulating low-light enhancement as an uncertainty entropy minimization problem, with a *UaD* module enabling dual-domain guidance for noise suppression and detail recovery.
2) Designing an asymmetric causal modeling strategy via *NeCo* (feature-level correlation) and *AsC* (pixel-level correlation), enforcing hierarchical consistency.
3) Achieving state-of-the-art performance and strong generalization across benchmark datasets and downstream tasks.

Related work

Low-Light Image Enhancement

Traditional methods. Histogram equalization(Pisano et al. 1998), gamma correction(Wang, Liang, and Liu 2009) directly enhance brightness but ignore the interplay between luminance, noise, and details, often resulting in artifacts and limited effectiveness in real-world scenarios.

Retinex-based methods. (Wang et al. 2013)(Zhang et al. 2021) decompose images into illumination and reflection to jointly model brightness and detail. While methods like RetinexFormer(Cai et al. 2023), RetinexMamba(Bai et al. 2024), and Diff-Retinex++(Yi et al. 2025) improve local/-global modeling through semantic priors, SSMs, or diffusion models, their performance is limited by the illumination estimation network and increased computational cost, with insufficient exploration of stable causal dependencies.

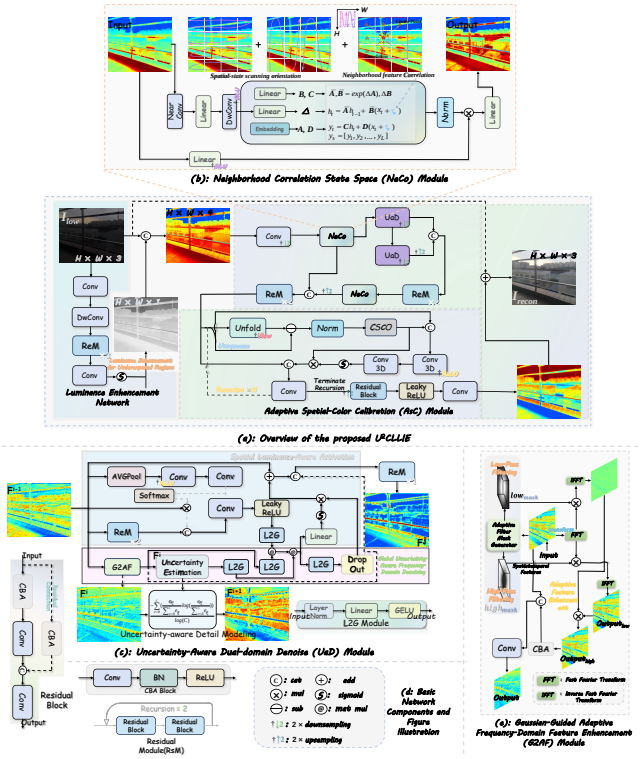

Figure 2: The proposed U^{2}CLLIE. (a) Overview of U^{2}CLLIE, where *LEN* enhances dark regions and suppresses overexposure; *AsC* models color-wise causal correlations with low complexity. (b) *NeCo* captures causal consistency via multi-directional neighborhood scanning. (c) *UaD* refines textures (spatial branch) and suppresses noise (frequency branch) under uncertainty guidance. (d) Basic network structure. (e) *G2AF* adaptively enhances frequency features using Gaussian masks. (*Please zoom in to the best view.*)

Deep learning-based methods. (Jiang et al. 2021)(Wang et al. 2024b)(Yang et al. 2023)(Liang et al. 2022b)(Wang et al. 2024a)(Yan et al. 2025)(Zhang et al. 2025b) leverage CNNs for enhanced local-global feature modeling. To address CNNs' limited receptive fields, Transformer-based approaches(Chen et al. 2021)(Zamir et al. 2022) use local window attention for long-range interaction. Frequency-domain models (*e.g.*, FourLLIE, DMFourLLIE) improve structural consistency by enhancing amplitude features. More recently, SSM-based methods (*e.g.*, Wave-Mamba(Zou et al. 2024), BSMamba, MambaLLIE) offer a balance between efficiency and modeling capacity, yet still struggle with causal consistency in extremely dark regions and adjacent structures.

Uncertainty-Aware Entropy-Guided Modulation

Uncertainty entropy reflects model confidence and feature stability, and has shown strong representational power in tasks like semi-supervised learning(Springenberg 2015), clustering(Jain et al. 2017), and domain adaptation(Hu et al. 2025)(Lee et al. 2024). High-entropy regions often correspond to noisy, low-SNR areas with unstable features, while

low-entropy regions are typically more reconstructable. In low-light enhancement, entropy has been used to guide learning paradigms—from spatial regression to confidence-aware optimization(Zhou et al. 2003) to improve perceptual quality and reduce noise. However, most methods rely on uncertainty cues in a single domain, limiting their ability to capture spatial and frequency dependencies—particularly in high-entropy regions prone to texture loss and color shifts.

Methodology

Motivation and Overview of the U^{2}CLLIE

Most low-light enhancement methods adopt residual regression between low-light inputs and normal-light targets. However, for severely underexposed regions, this paradigm suffers from two core problems:

Problem 1: Feature degradation in dark regions. Including gradients degradation (***Limitation 1***) and noise dominated (***Limitation 2***).

Problem 2: Breakdown of neighborhood context modeling. ***See the supplement for formal analysis of these issues.***

Our key motivation is to avoid relying on raw noisy inputs and instead reconstruct robust features via uncertainty-guided and causally-constrained modeling. Accordingly, as shown in Figure 2, U^{2}CLLIE adopts a U-shaped architecture with following core designs:

1. Uncertainty-aware entropy guidance (addresses Problem 1):

- *LEN* enhances luminance and provides contextual cues for underexposed regions;
- *UaD* reconstructs denoised, modelable features under dual-domain (spatial/frequency) guidance; *G2AF* adaptively suppresses frequency-domain noise.

2. Hierarchical causal modeling (addresses Problem 2):

- *NeCo* (encoder) captures directional neighborhood causality;
- *AsC* (decoder) models structured color-wise correlations for consistent reconstruction.

These components collaboratively enhance SNR, suppress noise/artifacts, and recover structural and color consistency in extremely dark scenes(see Figure 1(f): last column).

Luminance Enhancement Network (*LEN*)

As the front-end module, *LEN* enhances dark regions to facilitate subsequent feature modeling. It first projects the input $I_{\text{low}} \in \mathbb{R}^{H\times W\times 3}$ into feature space $\tilde{F}_{\text{ill}} \in \mathbb{R}^{H\times W\times c_1}$ via a 1×1 convolution (c_1=32), followed by a 3×3 depthwise separable convolution for local detail extraction, and three residual blocks (*ReM*). A Sigmoid activation then produces a spatial luminance-aware map $\hat{F}_{ill} \in \mathbb{R}^{H\times W\times 1}$:

$$\begin{aligned}\tilde{F}_{\text{ill}} &= Conv(I_{\text{low}}),\\ \hat{F}_{\text{ill}} &= \sigma\Big(\text{Conv}\big(\text{ReLU}(\text{ReM}_{\times 3}(\text{DepthConv}(\tilde{F}_{\text{ill}})))\big)\Big).\end{aligned} \tag{1}$$

LEN is trained to predict the luminance difference $\hat{I}_{diff} = \frac{I^{ycrcb}_{high} - I^{ycrcb}_{low}}{I^{ycrcb}_{high}}$ between low-light and normal images in the *Y* channel of YCrCb space (ablation comparisons in HSV, Lab, and HVI will be presented in experiment). Its loss($\mathcal{L}_{len}$) is $\frac{1}{B\cdot H\cdot W}\sum_{n=1}^{B}\sum_{i,j}^{H,W}(\hat{I}^{(n,i,j)}_{diff} - \hat{F}^{(n,i,j)}_{\text{ill}})^2$.

Uncertainty Estimation

Uncertainty-Aware Dual-domain Denoise (*UaD*) Module. *UaD* employs entropy-guided suppression of high-uncertainty regions to reduce noise, while enhancing low-entropy, high-frequency details for robust feature extraction in extreme darkness. Given input features $F^{i-1} \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times c_2}$, the spatial branch generates luminance-sensitive features F_{spa}, while the frequency branch produces entropy-guided features F^i. Specifically, F^{i-1} is first enhanced by the *G2AF* module, followed by uncertainty entropy computation, yielding F^{i+1}. The features undergoes cross-attention and nonlinear to yield F_{fre}. Subsequently, F^{i-1} and F_{fre} are added pixel-wise($\oplus$) and concatenated with the denoised feature F^i, resulting in $F^j \in \mathbb{R}^{\frac{H}{2}\times\frac{H}{2}\times c_2}$. This facilitates multi-domain collaboration to suppress noise and artifacts in high-entropy uncertain regions.

$$\begin{aligned}\vec{F} &= \text{Softmax}(\text{Conv}^2(\text{AvgPool}(F^{i-1})))\otimes F^{i-1},\\ F_{\text{spa}} &= \text{LeakyReLU}(\text{Conv}(\text{cat}(\vec{F}, \text{ReM}(F^{i-1})))),\\ \mathcal{H}(x) &= -\tfrac{1}{\log c}\textstyle\sum_{c=1}^{C} p_c(x)\log p_c(x),\\ F^i &= G2AF(F^{i-1}), F^{i+1} = \mathcal{H}(F^i),\\ \text{Attn}(Q,K,V) &= \text{Softmax}(QK^\top/\sqrt{d_{\text{head}}})V,\\ F_u &= \text{Attn}(F_{\text{spa}}, F^i, F^{i+1}),\\ F_{\text{fre}} &= \sigma(\text{Linear}(F_u))\otimes\text{Dropout}(\text{L2G}(F_u)),\\ F^j &= \text{ReM}(\text{cat}(F^{i-1}\oplus F_{\text{fre}}, F^i)).\end{aligned} \tag{2}$$

Here, $Softmax$, ($\sigma : Sigmoid$), and $LeakyReLU$ serve as a weight normalization function, an independent distribution activation, and a nonlinear activation functions, respectively; $AvgPool$ denotes average pooling; dim_{head} is the shared channel dimension of *Q*, *K*, and *V*; cat indicates concatenation along dimension dim; $DropOut$ is stochastic neuron dropout; $Conv^i$ denotes convolution applied i times;

Gaussian-Guided Adaptive Frequency-Domain Feature Enhancement (*G2AF*) Module. We propose the *G2AF* module, which jointly regulates amplitude and phase while adaptively generating Gaussian mask radii based on frequency feature distributions. Specifically, input feature F^{i-1} is transformed via *FFT* yielding $\tilde{F}^{i-1}_{\text{fre}}$, followed by the *Adaptive Filter Mask Generator* (see Eq. 3 and *pseudocode in the supplementary*) to produce frequency-aware Gaussian masks for low- and high-frequency components ($low_{\text{mask}}, high_{\text{mask}} \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times 1}$). This process can be described as follows:

$$\begin{aligned}\text{dist}_x &= \text{linspace}(-1,1,H).\text{view}(1,H,1),\\ \text{dist}_y &= \text{linspace}(-1,1,W).\text{view}(1,1,W),\\ \text{dist} &= \sqrt{\text{dist}_x^2 + \text{dist}_y^2},\\ r_{\text{low}} &\leftarrow r_{\text{low}}\cdot\sigma(\tilde{F}^{i-1}_{\text{fre}}), r_{\text{high}} \leftarrow r_{\text{high}}\cdot\sigma(\tilde{F}^{i-1}_{\text{fre}}),\\ low_{\text{mask}} &= \exp\left(-\text{dist}^2/(2r^2_{\text{low}}+\epsilon)\right),\\ high_{\text{mask}} &= \exp\left(-\text{dist}^2/(2r^2_{\text{high}}+\epsilon)\right).\end{aligned} \tag{3}$$

Subsequently, F^{i-1} and $low_{mask}, high_{mask}$ are multiplied pixel-wise and then transformed the spatial domain via *iFFT* to obtain $\tilde{F}_{low}^{i-1}$ and $\tilde{F}_{high}^{i-1}$. Then, $\tilde{F}_{low}^{i-1}, \tilde{F}_{high}^{i-1}$ are multiplied($\otimes$) and convolved for local frequency aggregation. The result is concatenated with input feature F^{i-1} and refined via convolution to produce the final feature $F^i \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times c}$.

$$\begin{aligned} \tilde{F}_{\text{low}}^{i-1} &= \lambda \cdot \text{iFFT}(F^{i-1} \otimes low_{\text{mask}}), \\ \tilde{F}_{\text{high}}^{i-1} &= (1-\lambda) \cdot \text{iFFT}(F^{i-1} \otimes high_{\text{mask}}), \\ F^i &= \text{Conv}\big(\text{cat}(\text{CBA}(\tilde{F}_{\text{low}}^{i-1} \oplus \tilde{F}_{\text{high}}^{i-1}), F^{i-1})\big). \end{aligned} \tag{4}$$

In Equations (3–4), $linspace(\cdot,\cdot,\cdot)$ generates evenly spaced values; $view(\cdot,\cdot,\cdot)$ reshapes feature dimensions; λ is a hyperparameter (set to 0.5); r_{low}, r_{high} are learnable parameters initialized to 0.3 and 0.1; ϵ is set to 10^{-6}.

How *UaD* Addresses Problem 1:

- **Overcoming Gradient Vanishing (Limitation 1):** Through the cross-attention mechanism, the *V* term is associated with the uncertainty entropy F^{i+1}, leveraging the entropy distribution to address the gradient vanishing issue in dark regions. Specifically, in extremely dark areas (high entropy: $F^{i+1} \to 1$), the modulation factor $(\hat{I}_{\text{recon}} - I_{\text{high}})$ effectively prevents gradient vanishing:

$$\frac{\partial \mathcal{L}_{\text{recon}}(x)}{\partial \theta} \propto F^{i+1} \cdot (\hat{I}_{\text{recon}} - I_{\text{high}}) \cdot \frac{\partial F_{\text{fre}}}{\partial \theta}. \tag{5}$$

- **Mitigating Noise Dominance (Limitation 2):** In noise-concentrated regions, the following operation: $F_{\text{fre}} = \phi(\mathcal{H}(F^i))$(where $\phi(\cdot)$ denotes a network operation, and $\mathcal{H}(\cdot)$ is defined in Eq. 2).

$$\begin{aligned} &\frac{\partial F_{\text{fre}}}{\partial \theta} \propto \frac{\partial \phi(\mathcal{H}(F^i))}{\partial \mathcal{H}} \cdot \frac{\partial \mathcal{H}(F^i)}{\partial F^i} \cdot \frac{\partial F^i}{\partial \theta} \\ &\Rightarrow \frac{\partial \mathcal{L}_{\text{recon}}(x)}{\partial \theta} \propto F^{i+1} \cdot (\hat{I}_{\text{recon}} - I_{\text{high}}) \cdot \frac{\partial F_{\text{fre}}}{\partial \theta}, \\ &\Rightarrow \underbrace{F^{i+1}}_{\in(0,1)} \cdot \underbrace{(\hat{I}_{\text{recon}} - I_{\text{high}})}_{>0} \cdot \underbrace{\frac{\partial \phi(\mathcal{H}(F^i))}{\partial \mathcal{H}} \cdot \frac{\partial \mathcal{H}(F^i)}{\partial F^i} \cdot \frac{\partial F^i}{\partial \theta}}_{\neq 0} \\ &\Rightarrow \mathcal{E}_{\text{noise}} \leq \gamma \cdot \left\| \frac{\partial F^i}{\partial \theta} \right\|_{Upper\ Bound}, \text{ with } \gamma \in (0,1) \end{aligned} \tag{6}$$

In Eq. 6, the left side reflects the network fitting noise, while the right side demonstrates that uncertainty entropy guides the noise toward a likelihood upper bound. The uncertainty entropy steers parameter updates toward easier directions, thus mitigating noise fitting in challenging low-light regions. *See supplementary for derivations and symbol definitions.*

Causal Relationship-Aware Context Modeling

U^{2}CLLIE introduces an asymmetric encoder-decoder: the encoder models neighborhood causality (*NeCo*), and the decoder enhances local color consistency (*AsC*), effectively constraining pixel-wise color features via adjacent causal cues to resolve Problem 2.

Neighborhood Correlation State Space (*NeCo*) Module. Neighborhood convolution is first applied to input feature $F_1 \in \mathbb{R}^{H \times W \times c}$ to extract local context. Meanwhile, during multi-directional scanning, a learnable offset ($\tau_p \in \mathbb{R}^{4 \times dim}$) enhances SSM local states via adjacent causal cues. This forms the basis of our *NeCo* module (see Figure 2):

$$h_t = \bar{\mathbf{A}} h_{t-1} + \bar{\mathbf{B}}(x_t + \tau_p), y_t = \mathbf{C} h_t + \mathbf{D}(x_t + \tau_p). \tag{7}$$

Here, τ_p encodes local causal correlations with spatial states; $\bar{\mathbf{A}}, \bar{\mathbf{B}}, \mathbf{C}, \mathbf{D}$ are detailed in Figure 2(b).

The model maintains local consistency during multi-directional scanning while suppressing redundant correlations. Input $F_1 \in \mathbb{R}^{H \times W \times c}$ is processed in two branches: one extracts local constraints using neighborhood convolution, projection, depthwise convolution, and *SiLU* activation, followed by fine-scale convolutions with learnable offsets; the other applies a linear layer and *SiLU*. The outputs are multiplied and projected back to the original dimension:

$$\begin{aligned} \tilde{F}_1 =& \text{Norm}(\text{2DSSM}(\text{SiLU}(\\ & \text{DwConv}(\text{Linear}(\text{NearConv}(F_1)))))), \\ \tilde{F}_2 =& \text{SiLU}(\text{Linear}(F_1)), \tilde{F}_{\text{out}} = \text{Linear}(\tilde{F}_1 \otimes \tilde{F}_2). \end{aligned} \tag{8}$$

Here, *Norm* is LayerNorm layer; *SiLU* is the activation function; *DwConv* is the depthwise separable convolution; *Linear* is the linear projection; *2DSSM* is our SSM-based scanner(see Eq. 8);

Adaptive Spatial-Color Calibration (*AsC*) Module. Efficient neighborhood causal modeling is vital in decoding for capturing color consistency. The proposed *AsC* module operates as follows: Given feature map $\hat{F}$, we extract local patches via *Unfold* to obtain $\hat{F}_{\text{patch}}$, and map $\hat{F}$ to a central feature $\hat{F}_c$. After normalizing distances, the top-*k* (in this work, *k*=8) nearest neighbors are selected to form $\hat{F}_{\text{neighbor}}$. Concatenating $\hat{F}_c$ and $\hat{F}_{\text{neighbor}}$, a 3D convolution and Sigmoid yield correlation weights $\hat{F}_{\text{cluster}}$. Then, $\hat{F}_{\text{cluster}}, \hat{F}_{\text{neighbor}}$ are element-wise multiplied and summed along the feature-distance dimension to obtain $\hat{F}_{\text{relation}}$. To mitigate sparsity, $\hat{F}_{\text{relation}}$ is fused with $\hat{F}$ via a 1×1 convolution, yielding the final output $\hat{F}_{asc} \in \mathbb{R}^{H \times W \times c}$ with enhanced color-aware causal modeling:

$$\begin{aligned} \hat{F}_{\text{patch}} &= \text{view}(\text{Unfold}(\hat{F})), \hat{F}_c = \text{unsqueeze}(\hat{F}), \\ \hat{F}_{\text{neighbor}} &= \text{CscO}(\text{Norm}(\hat{F}_{\text{patch}} - \hat{F}_{\text{neighbor}})), \\ \hat{F}_{\text{cluster}} &= \sigma(\text{Conv3D}^2(\text{cat}((\hat{F}_c, \hat{F}_{\text{neighbor}})))), \\ \hat{F}_{\text{relation}} &= \sum_{c=0}^{C} (\hat{F}_{\text{cluster}}^c \cdot \hat{F}_{\text{neighbor}}^c), \\ \hat{F}_{\text{asc}} &= Conv(cat(\hat{F}_{\text{relation}}, \hat{F})). \end{aligned} \tag{9}$$

Here, *view(·)* reshapes features; *unsqueeze* adds a dimension; $Conv3D^i$ denotes the *i*-th 3D convolution operation. *CscO denotes Color feature Space Correlation Optimization (further details are provided in the supplementary material).*

How *NeCo* combined with *AsC* addresses Problem 2: The asymmetric collaboration between *NeCo* and *AsC* effectively resolves the breakdown of neighborhood context

modeling in low-light images. (1) *NeCo* introduces learnable causal weights to model brightness and texture dependencies within local regions, enhancing feature-level neighborhood constraints. (2) *AsC* adaptively computes similarity within feature blocks to model object-level color correlations, ensuring high intra-object color consistency. (3) *NeCo* guides *AsC*'s alignment, while *AsC* reinforces *NeCo*'s structural modeling. This closed-loop feedback enhances the capture of stable causal dependencies under noise and weak signals. (see Figure 1(f)).

Loss Function

The total loss function of U^{2}CLLIE consists of seven terms:

$$\mathcal{L}_{recon} = \lambda_1\mathcal{L}_{mse} + \lambda_2\mathcal{L}_{ssim} + \lambda_3\mathcal{L}_{per} + \lambda_4\mathcal{L}_{global} + \lambda_5\mathcal{L}_{color} + \lambda_6\mathcal{L}_{grad} + \lambda_7\mathcal{L}_{len}. \quad (10)$$

Where, $\lambda_{1:7} = [0.95, 0.01, 0.01, 0.1, 0.5, 0.1, 0.1]$. $\mathcal{L}_{mse}$: L2 pixel-wise reconstruction loss; $\mathcal{L}_{ssim}$: Structural similarity loss; $\mathcal{L}_{per}$: Perceptual loss (L1 on VGG features); $\mathcal{L}_{global}$: Global histogram alignment via *KL divergence* after *Softmax*; $\mathcal{L}_{color}$: Color consistency (cosine similarity on normalized $\hat{I}_{recon}/I_{high}$); $\mathcal{L}_{grad}$: Boundary detail loss (L1 on high-frequency gradients); $\mathcal{L}_{len}$: Lighting difference loss from the *LEN* module. Derivations for $\mathcal{L}_{global}$, $\mathcal{L}_{color}$, and $\mathcal{L}_{grad}$ are in the **supplementary material**.

Experiment

Benchmark Datasets and Settings

Datasets. We evaluate our method on 8 publicly available benchmark datasets: ***1) Reference datasets:*** LOLv2-Real(Yang et al. 2021) (689 real-world training/100 testing pairs); LOLv2-Synthetic(Yang et al. 2021) (900 synthetic training/100 testing pairs); LSRW-Huawei(Hai et al. 2023) (2,450 training/30 testing pairs, Huawei P40 captured); ***2) Unpaired datasets:*** DICM(Lee, Lee, and Kim 2013) (69 images), LIME (10 images), MEF(Ma, Zeng, and Wang 2015) (17 images), NPE(Wang et al. 2013) (85 images), and VV(24 images) are used to evaluate perceptual quality in the absence of ground-truth references.

Implementation Details. All experiments are conducted on an Intel 6326 CPU with a NVIDIA A100 GPU under Ubuntu 20.04.4 LTS. We use PyTorch 1.12.1 with CUDA 12.4. Input images are randomly cropped to 384×384 and augmented via random horizontal and vertical flipping. The batch size is 4. We use the Adam optimizer (β_1=0.95, β_2=0.99) with an initial learning rate of 2.5×10^{-4}, adjusted via a multi-step scheduler. The model is trained for 1.5×10^5 iterations.

Evaluation Metrics. For reference datasets, we report PSNR, SSIM(Wang et al. 2004), and LPIPS (with AlexNet backbone). For unpaired datasets, perceptual quality is assessed using NIQE(Mittal, Soundararajan, and Bovik 2012).

Main Results on Benchmarks

Quantitative Comparison. As shown in Tables 1 and 2, U^{2}CLLIE achieves the highest PSNR and SSIM across all three reference datasets, and the best LPIPS on LOLv2-Synthetic. Our entropy-guided feature-level approach outperforms not only the pixel-level SNR-based SNR-Aware,

Method	PSNR↑	SSIM↑	LPIPS↓	#Params (M)	#FLOPs (G)
NPE	17.08	0.391	0.230	-	-
LIME	17.00	0.382	0.207	-	-
Kind	16.56	0.569	0.226	8.02	34.99
MIRNet	19.98	0.609	0.215	31.76	785.00
SGM	15.43	0.570	0.237	2.31	-
FECNet	21.09	0.612	0.234	0.15	5.82
HDMNet	20.81	0.607	0.238	2.32	-
SNR-Aware	20.67	0.591	0.192	39.12	26.35
FourLLIE	21.11	0.626	0.183	0.12	4.07
UHDFour	19.39	0.601	0.247	17.54	4.78
Retinexformer	<u>21.23</u>	<u>0.631</u>	0.170	1.61	15.57
UHDFormer	20.64	0.624	0.181	0.34	3.24
RetinexMamba	20.88	0.630	0.169	3.59	34.76
CIDNet	20.47	0.613	**0.138**	1.88	7.57
Ours	**21.35**	**0.633**	<u>0.162</u>	0.48	6.18

Table 1: Quantitative comparison on LSRW-Huawei. The **best** and <u>second-best</u> results are shown in **bold** and <u>underline</u> respectively. *Note: no GT-Mean strategy was used.*

Figure 3: Qualitative results on LSRW-Huawei show that U2CLLIE achieves superior global brightness and local structure preservation, effectively avoiding oversmoothing. (*Please zoom in to the best view.*)

but also recent methods such as DiffUIR, GSAD, UHDFormer and CIDNet. Under 256×256 input, U^{2}CLLIE maintains a lightweight design (0.48M params, 6.18 GFLOPs), significantly surpassing heavier models such as MIRNet (31.76M), SNR-Aware (39.12M), EnlightenGAN (114.35M), and CLE Diffusion (37.02M), striking a compelling balance between quality and efficiency.

Qualitative Comparison. Figure 3 and 4 show visual compared results. Existing methods often fail to recover fine details or maintain color consistency in extremely dark regions, leading to noticeable noise and artifacts (*e.g.*, gray/black shifts in the lamp area of Figure 4). In contrast, U^{2}CLLIE benefits from entropy-guided feature extraction for robust representation in low-light regions, and asymmetric causal modeling that balances luminance enhancement with local structure and color fidelity.

Quantitative Evaluation on *Unpaired* Datasets. Table 3 reports NIQE on five unpaired datasets. U^{2}CLLIE achieves the lowest (*i.e.*, best) scores across all cases, demonstrating superior perceptual quality.

Method	Venue	LOLv2-Real			LOLv2-Synthetic			Complexity (M/G)	
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	Params	FLOPs
NPE(Wang et al. 2013)	TIP 2013	17.33	0.464	0.2359	16.60	0.778	0.1079	-	-
LIME(Guo, Li, and Ling 2016)	TIP 2016	15.24	0.419	0.2203	16.88	0.758	0.1041	-	-
MIRNet(Zamir et al. 2020)	ECCV 2020	20.02	0.820	-	21.94	0.876	-	31.76	785.00
SGM(Yang et al. 2021)	TIP 2021	20.06	0.816	0.0727	22.05	0.909	0.4841	2.31	-
EnlightenGAN(Jiang et al. 2021)	TIP 2021	18.23	0.617	0.3090	16.57	0.734	0.2200	114.35	61.01
IPT(Chen et al. 2021)	CVPR 2021	19.80	0.813	-	18.30	0.811	-	115.31	-
FECNet(Huang et al. 2022)	ECCV 2022	20.67	0.795	0.0995	22.57	0.894	0.0699	0.15	5.82
SNR-Aware(Xu et al. 2022)	CVPR 2022	21.48	0.848	0.0740	24.13	0.927	0.0318	39.12	26.35
Restormer(Zamir et al. 2022)	CVPR 2022	19.94	0.827	-	21.41	0.830	-	26.13	-
Bread(Guo and Hu 2023)	IJCV 2023	20.83	0.847	0.1740	17.63	0.919	0.0910	2.02	19.85
SNR-SKF(Wu et al. 2023)	CVPR 2023	20.66	0.812	0.0757	17.21	0.774	0.0731	39.44	27.88
CLE Diffusion(Yin et al. 2023)	MM 2023	21.99	0.798	0.1910	20.28	0.852	0.1020	37.02	3102.41
UHDFour(Li et al. 2023)	ICLR 2023	19.42	0.790	0.1151	23.64	0.900	0.0341	17.54	4.78
FourLLIE(Wang, Wu, and Jin 2023)	MM 2023	22.34	0.847	**0.0573**	24.65	0.919	0.0389	0.12	4.07
GSAD(Hou et al. 2023)	NeurIPS 2024	20.15	0.846	0.1130	24.47	0.929	0.0510	17.36	442.02
DiffUIR(Zheng et al. 2024)	CVPR 2024	19.71	0.825	0.2110	19.61	0.863	0.1600	36.26	9.88
UHDFormer(Wang et al. 2024a)	AAAI 2024	19.71	0.832	0.0758	24.48	0.928	0.0310	0.34	3.24
Ours	—	**22.63**	**0.851**	0.1005	**25.15**	**0.931**	**0.0308**	0.48	6.18

Table 2: Quantitative comparison on LOLv2-Real, LOLv2-Synthetic. The **best** and second-best results are shown in **bold** and underline respectively. *Please note that we did not use the GT-Mean strategy.*

Figure 4: Qualitative comparisons with state-of-the-art methods on the LOLv2-Real and LOLv2-Synthetic datasets demonstrate the superiority of our approach. Our U^{2}CLLIE restores local details more accurately, suppresses artifacts and noise effectively, and preserves consistent color transitions.(*Please zoom in to the best view.*)

Ablation Study

We conduct ablation studies on the LSRW-Huawei dataset under the same settings as *"Implementation Details"*.

Module Component. Based on the results in Table 4, the baseline with residual blocks yields 20.15 dB PSNR, 0.627 SSIM, and 0.177 LPIPS. Adding *LEN* improves PSNR to 20.26 dB, showing the benefit of luminance-aware enhancement. *NeCo* further raises PSNR to 21.14 dB and SSIM to 0.637, highlighting the effectiveness of neighborhood structural modeling. *UaD* brings a 0.72 dB gain over the baseline by enhancing perceptual feature quality via uncertainty guidance. Removing *AsC* or *UaD-AsC* from the full model causes clear performance drops, especially without *AsC*. The complete U^{2}CLLIE achieves the best PSNR (21.35 dB), SSIM (0.633) and LPIPS (0,162), confirming the joint benefit of entropy-guided and causal-aware design.

Loss Function. As shown in Table 4, the Baseline ($\mathcal{L}_{mse} + \mathcal{L}_{ssim}$) yields unsatisfactory perceptual quality (PSNR: 20.85 dB, SSIM: 0.628, LPIPS: 0.173). Introducing L1 and L2, which incorporate feature-level consistency (Bseline + $\mathcal{L}_{per}$) and luminance distribution alignment (L1 + $\mathcal{L}_{global}$), significantly improves visual performance. L3 adds local color consistency($\mathcal{L}_{color}$), yielding the best LPIPS (0.161). The full model achieves the highest PSNR (21.35 dB) and SSIM (0.633), and the second-best LPIPS (0.162), balancing perceptual quality and reconstruction accuracy.

Method	LIME	VV	DICM	NPE	MEF	AVG
Kind	4.77	3.84	3.61	4.18	4.82	4.24
MIRNet	6.45	4.74	4.04	5.24	5.50	5.19
SGM	5.45	4.88	4.73	5.21	5.75	5.21
FECNet	6.04	3.35	4.14	4.50	4.71	4.55
HDMNet	6.40	4.46	4.77	5.11	5.99	5.35
Bread	4.72	3.30	4.18	4.16	5.37	4.35
Retinexformer	3.44	3.71	4.01	3.89	3.73	3.76
FourLLIE	4.40	3.17	**3.37**	3.91	4.36	3.84
DMFourLLIE	3.23	3.30	3.61	3.56	3.57	3.46
CWNet	3.58	3.74	4.50	4.54	4.76	3.70
Ours	**3.19**	**2.97**	3.48	**3.31**	**3.18**	**3.23**

Table 3: NIQE scores (↓) of state-of-the-art methods on five unpaired datasets. The **best** and second-best results are shown in **bold** and underlined, respectively. 'AVG' denotes the average NIQE score over all five datasets. (All methods pre-trained on LSRW-Huawei.)

Method	Metrics		
	PSNR↑	SSIM↑	LPIPS↓
Module Component			
Baseline	20.15	0.627	0.177
Ours w/ *LEN* Only	20.26	0.626	0.173
Ours w/ *NeCo* Only	21.14	0.637	0.166
Ours w/ *UaD* Only	20.87	0.630	0.170
Ours w/ *AsC* Only	20.28	0.623	0.172
Ours w/o *AsC*&*UaD*	21.07	0.631	0.164
Ours w/o *AsC*	21.11	0.632	0.171
Loss Function			
Baseline: ($\mathcal{L}_{mse}$+$\mathcal{L}_{ssim}$)	20.85	0.628	0.173
L1: Baseline &$\mathcal{L}_{per}$	21.00	0.630	0.167
L2: L1 &$\mathcal{L}_{global}$	20.97	0.631	0.166
L3: L2 &$\mathcal{L}_{color}$	21.05	0.631	**0.161**
Ours	**21.35**	**0.633**	0.162

Table 4: Ablation study on U^{2}CLLIE. Effectiveness of core modules and Loss components.

Module Replacement. To validate the architectural superiority of U^{2}CLLIE, we perform controlled replacements of key components with those from representative methods. As shown in Table 5, replacing *LEN* with the illumination branches from Retinexformer and CIDNet reduces PSNR to 20.77 dB and 20.98 dB, respectively. While CIDNet outperforms Retinexformer by 0.21 dB due to its HVI-aware spatial modeling, our design—using only a simple residual structure—achieves the best PSNR, SSIM and LPIPS scores, confirming that U^{2}CLLIE attains strong performance without architectural complexity.

Ablation Settings	Metrics		
	PSNR↑	SSIM↑	LPIPS↓
Module Replacement			
LEN → Retinexformer	20.77	0.629	0.164
LEN → CIDNet	20.98	0.630	0.171
Uncertainty Entropy Dominated Domain			
w/ Spatial Domain	21.21	0.628	0.164
w/ Frequency Domain *G2AF* → DMFourLLIE	21.30	0.629	0.163
w/ Frequency Domain *G2AF* → WaveMamba	21.24	0.631	0.163
Ours	**21.35**	**0.633**	**0.162**

Table 5: Comprehensive Ablation studies on U^{2}CLLIE framework design.

Uncertainty Entropy Dominated Domain. To evaluate the effectiveness of frequency-domain uncertainty guidance, we design four variants. As shown in Table 5, frequency-domain methods consistently outperform spatial-domain ones. Replacing our G2AF with other frequency modules leads to PSNR, SSIM, and LPIPS drops, color distortion, and loss of detail (see Figure 5 for more details).

Additional ablations and U^{2}CLLIE's generalization to downstream tasks are in the supplementary.

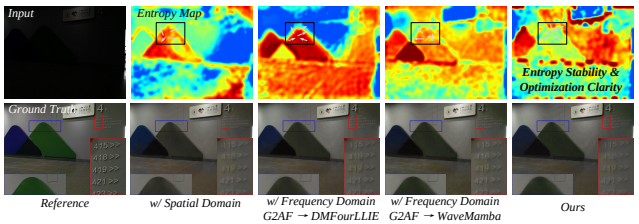

Figure 5: Uncertainty entropy distribution and visualization of color/detail reconstruction. (Arrow direction: high entropy → low entropy). U^{2}CLLIE exhibits stable entropy, clear optimization direction, consistent color (artifact-free), and enhanced detail (text clearly visible). *Please zoom in to the best view.*

Conclusion

This paper presents U^{2}CLLIE, a novel low-light enhancement framework guided by frequency-domain uncertainty entropy. The *UaD* module quantifies feature entropy to regulate gradients, alleviating noise dominance and optimization disorder in dark regions. Asymmetric causal modeling (*NeCo* and *AsC*) restores local consistency via neighborhood reasoning. Extensive experiments demonstrate that

U^2CLLIE surpasses state-of-the-art methods in both metrics and perceptual quality, validating the effectiveness of entropy-guided enhancement.

References

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

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

Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; and Gao, W. 2021. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12299–12310.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

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

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

Hai, J.; Xuan, Z.; Yang, R.; Hao, Y.; Zou, F.; Lin, F.; and Han, S. 2023. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712.

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

Hu, Z.; Hu, Y.; Li, X.; Tang, S.; and Duan, L.-Y. 2025. Beyond Entropy: Region Confidence Proxy for Wild Test-Time Adaptation. *arXiv preprint arXiv:2505.20704*.

Huang, J.; Liu, Y.; Zhao, F.; Yan, K.; Zhang, J.; Huang, Y.; Zhou, M.; and Xiong, Z. 2022. Deep fourier-based exposure correction network with spatial-frequency interaction. In *European Conference on Computer Vision*, 163–180. Springer.

Jain, H.; Zepeda, J.; Pérez, P.; and Gribonval, R. 2017. Subic: A supervised, structured binary code for image search. In *Proceedings of the IEEE international conference on computer vision*, 833–842.

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

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

Lee, J.; Jung, D.; Lee, S.; Park, J.; Shin, J.; Hwang, U.; and Yoon, S. 2024. Entropy is not enough for test-time adaptation: From the perspective of disentangled factors. *arXiv preprint arXiv:2403.07366*.

Li, C.; Guo, C.-L.; Zhou, M.; Liang, Z.; Zhou, S.; Feng, R.; and Loy, C. C. 2023. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*.

Liang, D.; Li, L.; Wei, M.; Yang, S.; Zhang, L.; Yang, W.; Du, Y.; and Zhou, H. 2022a. Semantically contrastive learning for low-light image enhancement. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 1555–1563.

Liang, Y.; Wang, B.; Ren, W.; Liu, J.; Wang, W.; and Zuo, W. 2022b. Learning hierarchical dynamics with spatial adjacency for image enhancement. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2767–2776.

Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; and Liu, Y. 2024. Vmamba: Visual state space model. *Advances in neural information processing systems*, 37: 103031–103063.

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

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

Pisano, E. D.; Zong, S.; Hemminger, B. M.; DeLuca, M.; Johnston, R. E.; Muller, K.; Braeuning, M. P.; and Pizer, S. M. 1998. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11(4): 193.

Springenberg, J. T. 2015. Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv preprint arXiv:1511.06390*.

Wang, C.; Pan, J.; Wang, W.; Fu, G.; Liang, S.; Wang, M.; Wu, X.-M.; and Liu, J. 2024a. Correlation matching transformation transformers for uhd image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 5336–5344.

Wang, C.; Wu, H.; and Jin, Z. 2023. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7459–7469.

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

Wang, W.; Yang, H.; Fu, J.; and Liu, J. 2024b. Zero-reference low-light enhancement via physical quadruple priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 26057–26066.

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

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

Wang, Z.-G.; Liang, Z.-H.; and Liu, C.-L. 2009. A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for PDP. *Displays*, 30(3): 133–139.

Weng, J.; Yan, Z.; Tai, Y.; Qian, J.; Yang, J.; and Li, J. 2024. Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. *Advances in Neural Information Processing Systems*, 37: 27440–27462.

Wu, Y.; Pan, C.; Wang, G.; Yang, Y.; Wei, J.; Li, C.; and Shen, H. T. 2023. Learning semantic-aware knowledge guidance for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1662–1671.

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

Xu, X.; Wang, R.; and Lu, J. 2023. Low-light image enhancement via structure modeling and guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9893–9903.

Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; and Zhang, Y. 2025. Hvi: A new color space for low-light image enhancement. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 5678–5687.

Yang, S.; Ding, M.; Wu, Y.; Li, Z.; and Zhang, J. 2023. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, 12918–12927.

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

Yi, X.; Xu, H.; Zhang, H.; Tang, L.; and Ma, J. 2025. Diff-Retinex++: Retinex-Driven Reinforced Diffusion Model for Low-Light Image Enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yin, Y.; Xu, D.; Tan, C.; Liu, P.; Zhao, Y.; and Wei, Y. 2023. Cle diffusion: Controllable light enhancement diffusion model. In *Proceedings of the 31st ACM International Conference on Multimedia*, 8145–8156.

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

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2020. Learning enriched features for real image restoration and enhancement. In *European conference on computer vision*, 492–511. Springer.

Zhang, T.; Liu, P.; Cai, M.; Zhang, Z.; Lu, Y.; and Zhou, Q. 2025a. BSMamba: Brightness and Semantic Modeling for Long-Range Interaction in Low-Light Image Enhancement. *arXiv preprint arXiv:2506.18346*.

Zhang, T.; Liu, P.; Lu, Y.; Cai, M.; Zhang, Z.; Zhang, Z.; and Zhou, Q. 2025b. CWNet: Causal Wavelet Network for Low-Light Image Enhancement. *arXiv preprint arXiv:2507.10689*.

Zhang, T.; Liu, P.; Zhao, M.; and Lv, H. 2024. DMFourLLIE: dual-stage and multi-branch fourier network for low-light image enhancement. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 7434–7443.

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

Zheng, D.; Wu, X.-M.; Yang, S.; Zhang, J.; Hu, J.-F.; and Zheng, W.-S. 2024. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 25445–25455.

Zhou, D.; Bousquet, O.; Lal, T.; Weston, J.; and Schölkopf, B. 2003. Learning with local and global consistency. *Advances in neural information processing systems*, 16.

Zou, W.; Gao, H.; Yang, W.; and Liu, T. 2024. Wave-mamba: Wavelet state space model for ultra-high-definition low-light image enhancement. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 1534–1543.