2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Diffusion-Based Low-Light Image Enhancement with Color and Luminance Priors

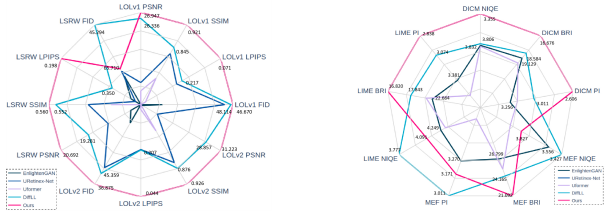
Xuanshuo Fu, Lei Kang† and Javier Vazquez-Corral

***Abstract*— Low-light images often suffer from low contrast, noise, and color distortion, degrading visual quality and impairing downstream vision tasks. We propose a novel conditional diffusion framework for low-light image enhancement that incorporates a Structured Control Embedding Module (SCEM). SCEM decomposes a low-light image into four informative components including illumination, illumination-invariant features, shadow priors, and color-invariant cues. These components serve as control signals that condition a U-Net–based diffusion model trained with a simplified noise-prediction loss. Thus, the proposed SCEM equipped Diffusion method enforces structured enhancement guided by physical priors. In experiments, our model is trained only on the LOLv1 dataset and evaluated without fine-tuning on LOLv2-real, LSRW, DICM, MEF, and LIME. The method achieves state-of-the-art performance in quantitative and perceptual metrics, demonstrating strong generalization across benchmarks. https://casted.github.io/scem/.**

I. INTRODUCTION

Low-light image enhancement (LLIE) aims to recover a clean, perceptually faithful normal-light image from severely underexposed, noisy input. Reliable LLIE is critical for nighttime photography, mobile vision, surveillance, and autonomous systems in which downstream perception degrades when contrast collapses, color shifts, and sensor noise dominate [1]. Real scenes further complicate the task: illumination varies spatially, shadows interact with mixed light sources, and color responses drift nonlinearly across the image [2]. Effective methods must therefore brighten selectively while preserving structure, texture, and plausible color.

Classical approaches manipulate pixel statistics or handcrafted illumination models. Histogram equalization expands dynamic range but readily amplifies noise and creates unnatural luminance [3]. Retinex-style decompositions separate illumination from reflectance and can restore detail, yet they are sensitive to parameterization and may introduce halos or color artifacts [4]. Point-wise nonlinear curves (gamma, sigmoid) lack spatial awareness and struggle under complex lighting.



(a) (b)

Fig. 1: Quantitative comparisons with state-of-the-art methods. (a) presents numerical scores for PSNR, SSIM, LPIPS, and FID on 3 datasets: LOLv1, LOLv2-real, and LSRW, which contain groundtruth normal-light images. (b) presents numerical scores for NIQE, BRISQUE and PI on 3 datasets: DICM, MEF, and LIME, which contain only low-light images without groundtruth normal-light counterparts. Note that our proposed methods are trained only on the LOLv1 training set and are directly evaluated on the remaining datasets. To enable intuitive comparison in the radar plot, we normalized all metrics and inverted those where lower is better (e.g., LPIPS, FID), so that higher values consistently indicate better performance.

Learning-based LLIE substantially improves fidelity but often remains weakly grounded in image formation. CNN regressors (e.g., U-Net, ResNet variants) map low-light to normal-light images directly, treating enhancement as black-box translation that can overfit and hallucinate colors [5], [6]. GAN formulations encourage realism but inherit adversarial instability and may globally remap appearance [7]. Diffusion models offer stronger generative coverage and training stability [2], yet vanilla conditional diffusion provides limited control over illumination consistency and color faithfulness when applied naively to LLIE.

We introduce a **Structured Control Embedding Module (SCEM)** that injects physically motivated, spatially aligned priors into a conditional diffusion backbone for LLIE. From each low-light input, SCEM computes four complementary guidance maps: (1) illumination to steer exposure balancing, (2) illumination-invariant features approximating reflectance for structure retention, (3) shadow priors to protect texture in dark and bright transitions, and (4) color-invariant cues that stabilize chromatic relationships. These maps are encoded and fused with U-Net features at every denoising step, making enhancement explicitly guided by scene lighting and

This work has been partially supported by the predoctoral program AGAUR-FI ajuts (2025 FI-2 00470) Joan Oró, which are backed by the Secretariat of Universities and Research of the Department of Research and Universities of the Generalitat of Catalonia, as well as the European Social Plus Fund; the Beatriu de Pinós del Departament de Recerca i Universitats de la Generalitat de Catalunya (2022 BP 00256); the Grant PID2024-162555OB-I00 funded by MCIN/AEI/10.13039/501100011033, ERDF "A way of making Europe" and by the Generalitat de Catalunya CERCA Program. JVC also acknowledges the 2025 Leonardo Grant for Scientific Research and Cultural Creation from the BBVA Foundation. The BBVA Foundation accepts no responsibility for the opinions, statements and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors.

Computer Vision Center, Cerdanyola del Vallès, Spain and Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain
{xuanshuo, lkang, javier.vazquez}@cvc.uab.cat
†denotes Corresponding Author.

color statistics rather than emergent from implicit regression. The design links Retinex-like reasoning with the expressive sampling power of diffusion, yielding more controllable and physically plausible results.

Trained solely on LOLv1, our SCEM-conditioned diffusion model generalizes without dataset-specific tuning to the public benchmarks LOLv2-real, LSRW, DICM, MEF, and LIME. It achieves state-of-the-art or highly competitive scores on the reference metrics PSNR and SSIM and on the perceptual metrics LPIPS, FID, NIQE, BRISQUE, and PI, yielding visibly sharper detail and more natural color reproduction.

Our contributions are threefold:

1) We propose SCEM, a structured control interface that embeds multi-channel illumination and appearance priors directly into a diffusion-based LLIE model, providing fine-grained, spatially aware guidance during denoising.
2) We operationalize a Retinex-informed decomposition jointly with shadow and color-invariance cues, enabling adaptive brightness boost while preserving texture and chromatic fidelity.
3) Extensive cross-dataset experiments show strong generalization from training on LOLv1 only and state-of-the-art performance on multiple LLIE benchmarks in both reference and no-reference metrics.

II. RELATED WORK

Traditional Enhancement Methods: Classical low-light enhancement relied on hand-crafted image processing. Histogram equalization (HE) [8] and derivatives rescale intensities to raise contrast but often blow out highlights and amplify noise in dark regions. Retinex theory models an image as illumination $\times$ reflectance [9]. MSR/MSRCR [10]) approximate this separation via multi-scale filtering, recovering shadow detail yet frequently introducing halos and color shifts. Subsequent refinements improved the filtering and paired Retinex with gamma correction [11] to better redistribute brightness. Simple non-linear tone mappings [12] provide additional contrast stretching. Despite being lightweight and sometimes effective, these rule-based schemes lack learned adaptability and struggle to suppress noise or correct complex color casts, motivating data-driven methods.

CNN-based Methods: Deep networks learn data-driven enhancement mappings. RetinexNet [5] jointly trains decomposition and enhancement subnets on paired low/normal-light data to produce pleasing results, but relies on supervised references and may not generalize. Zero-DCE [13] removes paired data by learning an image-adaptive tone curve via non-reference losses (spatial consistency, exposure, etc.), yet can falter under extreme illumination and introduce tonal artifacts. Multi-scale/attention architectures such as MIR-Net [6] fuse cross-resolution features to capture detail and context, yielding strong denoising and enhancement. Still, most CNNs only implicitly model illumination, risk dataset bias, and provide limited user control over brightness.

Transformer-based Methods: Uformer [14] extends a U-Net encoder–decoder with locally enhanced window self-attention and learnable multi-scale bias modulation to efficiently capture local and global context for low-light detail recovery. Restormer [15] employs channel-focused multi-Dconv head transposed attention and gated depthwise feed-forward blocks to restore high-resolution low-light images with reduced complexity. LLFormer [16] scales to ultra-high-definition inputs via axis-based self-attention and cross-layer feature fusion that lower attention cost while preserving fidelity. Despite strong context modeling and competitive results, Transformer models often demand careful tuning and large training corpora, and they still lack the generative flexibility and controllable sampling offered by recent generative approaches.

GAN-based Methods: GANs tackle low-light enhancement, especially when paired supervision is scarce. EnlightenGAN [7] trains an unpaired generator to brighten inputs while a global–local discriminator and self-regularized perceptual loss promote realistic results across domains, though adversarial training can be unstable and introduce color or smoothing artifacts. ReLLIE [17] casts enhancement as sequential pixel-wise curve adjustment via a reinforcement-learned policy driven by no-reference rewards; users can effectively halt when visually satisfied, but the policy lacks explicit illumination/reflectance separation, limiting interpretability.

Diffusion-based Methods: Denoising diffusion probabilistic models (DDPM) [18] and their accelerated variants like denoising diffusion implicit models (DDIM) [19] learn to reverse a gradual noise process. Conditioning diffusion models on inputs enables powerful image-to-image translation. Palette [20] demonstrates that a unified diffusion framework can outperform task-specific GANs across colorization, inpainting, and super-resolution. In low-light enhancement, CLE Diffusion [21] introduces user-controllable enhancement via illumination and semantic conditioning, while DiffLL [22] improves efficiency and texture fidelity using wavelet-domain diffusion and high-frequency refinement. GDP [23] treats enhancement as blind restoration, optimizing degradation parameters during sampling. Diff-Retinex [24] integrates Retinex-based decomposition with conditional diffusion to jointly address illumination, noise, and structural loss, enabling detailed content recovery and physically interpretable enhancement.

III. METHODOLOGY

We present a diffusion-based low-light image enhancement framework with explicit illumination modeling as shown in Fig. 2. The architecture employs four feature extractors to decompose input low-light images $\mathcal{I}$ into illumination T_{ref}, color invariance features $\Phi(x)$, illumination-invariant features R_c, and shadow priors S_{3ch}.

A. Structured Control Embedding Module

In this section, we present our method for decomposing a low-light image into its illumination and illumination-

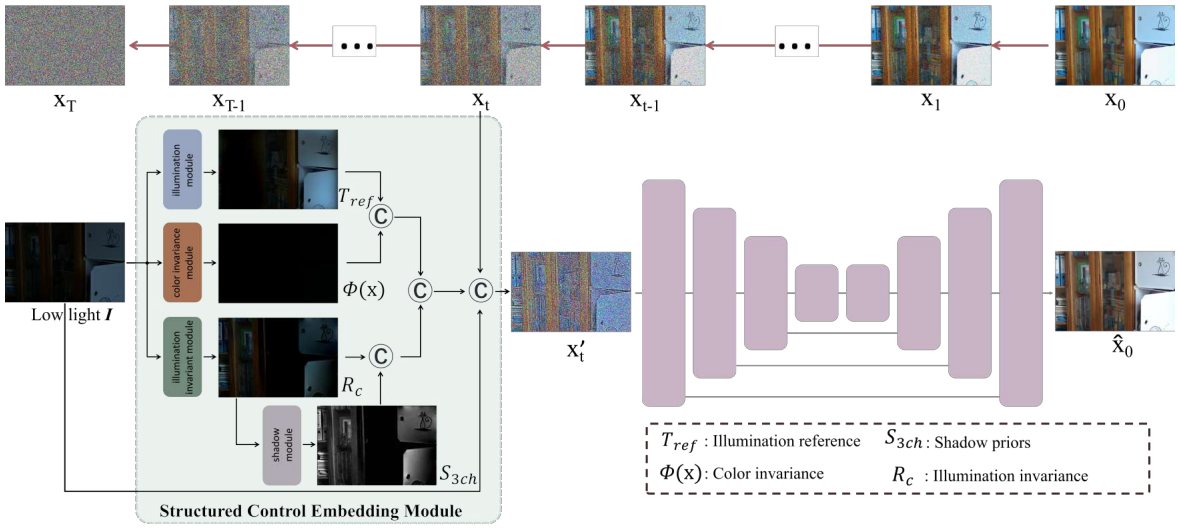


Fig. 2: The proposed architecture of SCEM equiped diffusion model for low-light image enhancement. During training, we extract four types of features from the input low-light image $\mathcal{I}$: illumination T_{ref}, color invariance features $\Phi(x)$, illumination-invariant features R_c, and shadow priors S_{3ch}. These features, along with the original low-light image, are concatenated with the randomly chosen t-th noised image X_t to form the input X'_t for the denoising training process. The diffusion model then generates the enhanced image $\hat{X}_0$. During inference, Gaussian noise X_T is concatenated with the same set of extracted features and the original low-light image I.

invariant information components by leveraging anisotropic structure priors and a Laplacian-regularized optimization framework. Our approach comprises two main stages. First, an illumination refinement procedure is applied using texture-aware weighting and frequency-domain regularization. Second, a post-processing step extracts shadow information from the intermediate results.

Let low-light image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$ denote an RGB image captured under non-uniform low-light conditions. We assume that the image is formed as:

$$\mathbf{I}(x,y) = \mathbf{R}(x,y) \circ \mathbf{L}(x,y) \tag{1}$$

where $\mathbf{R}(x,y)$ is the intrinsic illumination-invariant information and $\mathbf{L}(x,y)$ is the illumination map; here, $\circ$ denotes element-wise multiplication.

Illumination and Illumination-invariant feature. An initial illumination estimate is computed by taking the maximum response over the three color channels:

$$T_{\text{ini}}(x,y) = \max_{c\in\{R,G,B\}} \mathbf{I}_c(x,y) + \delta, \quad \delta = 0.02 \tag{2}$$

This initialization guarantees stability even in darker regions. To preserve structural details while enforcing smoothness, we compute anisotropic weights based on local gradient statistics. The horizontal and vertical finite-difference operators are defined by f_1 and f_2, which approximate ∂_x and ∂_y, respectively. We then obtain the gradients:

$$\nabla_x T_{\text{ini}}(x,y) = (f_1 \bullet T_{\text{ini}})(x,y) = T_{\text{ini}}(x,y+1) - T_{\text{ini}}(x,y) \tag{3}$$

$$\nabla_y T_{\text{ini}}(x,y) = (f_2 \bullet T_{\text{ini}})(x,y) = T_{\text{ini}}(x+1,y) - T_{\text{ini}}(x,y) \tag{4}$$

where $\bullet$ denotes discrete convolution with reflective boundary handling. A global texture weight is computed as:

$$w_{\text{to}}(x,y) = \left(\max\left\{\frac{1}{3}\sum_{c=1}^{3}\sqrt{\left(\nabla_x T_{\text{ini},c}(x,y)\right)^2 + \left(\nabla_y T_{\text{ini},c}(x,y)\right)^2},\ \varepsilon_s\right\}\right)^{-1} \tag{5}$$

with $\varepsilon_s = 0.02$.

To refine the local structure, we smooth T_{ini} by applying a separable Gaussian filter $\mathcal{G}_\sigma$ with standard deviation σ, resulting in the smoothed image $\hat{T}(x,y) = \mathcal{G}_\sigma * T_{\text{ini}}(x,y)$.

The smoothed gradients $\widehat{\nabla}_x\hat{T}$ and $\widehat{\nabla}_y\hat{T}$ yield local weights:

$$w_{x,\text{local}}(x,y) = \left(\max\left\{\frac{1}{3}\sum_{c=1}^{3}\left|\widehat{\nabla}_{x,c}\hat{T}(x,y)\right|,\ \varepsilon\right\}\right)^{-1} \tag{6}$$

$$w_{y,\text{local}}(x,y) = \left(\max\left\{\frac{1}{3}\sum_{c=1}^{3}\left|\widehat{\nabla}_{y,c}\hat{T}(x,y)\right|,\ \varepsilon\right\}\right)^{-1} \tag{7}$$

where $\varepsilon = 0.001$. The final anisotropic weights become

$$\begin{aligned} w_x(x,y) &= w_{\text{to}}(x,y)\cdot w_{x,\text{local}}(x,y) \\ w_y(x,y) &= w_{\text{to}}(x,y)\cdot w_{y,\text{local}}(x,y) \end{aligned} \tag{8}$$

To refine the illumination, we solve the following energy minimization problem:

$$E(T) = \|T - T_{\text{ini}}\|_2^2 + \lambda\left(\|\nabla_x T\cdot w_x\|_2^2 + \|\nabla_y T\cdot w_y\|_2^2\right) \tag{9}$$

with regularization parameter λ. The corresponding Euler-Lagrange equation is discretized to yield the linear system $(\mathbf{I} + \lambda\,\mathbf{L}_w)\,\mathbf{t} = \mathbf{b}$, where $\mathbf{t}$ represents the vectorized refined illumination T_{ref}, and $\mathbf{b}$ is computed from T_{ini}, and $\mathbf{L}_w$ is an anisotropic Laplacian, where each pixel (x,y) is connected only to its four immediate neighbors through direction-specific weights w_x and w_y The diagonal entries of $\mathbf{L}_w$ satisfy the following condition:

$$D(x,y) = 1 - \left(w_x^{\text{right}}(x,y) + w_x^{\text{left}}(x,y) + w_y^{\text{up}}(x,y) + w_y^{\text{down}}(x,y)\right) \quad (10)$$

Since the system described above is entirely linear, it does not account for gamma correction. After obtaining T_{ref}, a pointwise gamma transformation is applied to remap the dynamic range. This nonlinear adjustment is used solely to enhance contrast in the final illumination map.

$$T_{\text{ref}}(x,y) \leftarrow (T_{\text{ref}}(x,y))^{1/\gamma} \quad (11)$$

T_{ref} is our illumination information for the model.

The final enhanced illumination-invariant information $\mathbf{R}$ is then recovered by:

$$\mathbf{R}_c(x,y) = \frac{\mathbf{I}_c(x,y)}{T_{\text{ref}}(x,y)}, \quad c \in \{R,G,B\} \quad (12)$$

Shadow priors. The shadow information is extracted using the shadow extraction (SE) module. The SE module, illustrated in the gray box in Fig. 2, operates as follows:

We adopt a frequency-domain strategy based on a discrete Laplacian operator f_3. Let $\widehat{f_3}(u,v)$ be its 2D Fourier transform. Additionally, denote by $\widehat{f_1}(u,v)$ and $\widehat{f_2}(u,v)$ the Fourier transforms of f_1 and f_2 (f_1, f_2 as previously described), respectively. The frequency-domain update κ is then obtained by:

$$\mathcal{F}\{2\mathbf{R}\}(u,v) = \frac{\lambda|\widehat{f_3}(u,v)|^2\,\mathcal{F}\{\mathbf{2R}\}(u,v) + \beta\,\mathcal{F}\{\mathcal{N}_2\}(u,v)}{\lambda|\widehat{f_3}(u,v)|^2 + \beta\left(|\widehat{f_1}(u,v)|^2 + |\widehat{f_2}(u,v)|^2\right) + \varepsilon} \quad (13)$$

where β is an iteration-dependent parameter (typically defined as $\beta = 2^{(i-1)}/\tau$ for threshold τ), $\mathcal{N}_2$ aggregates the soft-thresholded gradient residuals of $2\mathbf{R}$ —where the factor 2 serves as an amplification coefficient, and ε ensures numerical stability. We apply the inverse Fourier transform to the frequency-domain update κ to obtain the spatial-domain projection S_1, where $S_1(x,y,c) \in [lb(x,y,c),\, ub(x,y,c)]$, with lb and ub representing the lower and upper bounds, respectively.

Thus, we decompose $2\mathbf{R}$ into a smooth structural component S_1 and a residual component S_2, where $S_2 = 2\mathbf{R} - S_1$. Finally, the output of the SE module is obtained after replicating S_2 across three channels: $S_{3\text{ch}} = \text{repmat}(S_2,\,1,\,1,\,3) \in \mathbb{R}^{H\times W\times 3}$, ensuring compatibility with standard RGB-based processing.

Color invariance. In order to make the generated luminance-enhanced image stable and invariant in color space; we construct a color representation invariant to global scaling of intensity, which is also a key requirement in tasks involving illumination decomposition, low-light enhancement, or intrinsic image recovery. We define a channel-wise affine-invariant mapping:

$$\Phi(x) = \left[\frac{x_r}{\|x_r\|_\infty}, \frac{x_g}{\|x_g\|_\infty}, \frac{x_b}{\|x_b\|_\infty}\right] \quad (14)$$

where $\|x_c\|_\infty = \max_{(i,j)} x_c(i,j)$ denotes the channel-wise ℓ_∞-norm, i.e., the maximum pixel intensity within each color channel.

The mapping Φ satisfies the following affine-invariance property:

$$\Phi(\alpha \cdot x) = \Phi(x), \quad \forall\, \alpha \in \mathbb{R}^+ \quad (15)$$

where α denotes a global positive scalar factor on the image x, applied uniformly for each pixel value of the image. Therefore exhibits scale invariance under global illumination changes. In geometric terms, $\Phi(x)$ projects the color vector at each pixel onto the canonical chromaticity subspace, removing dependency on per-channel intensity scaling.

In summary, our method leverages both spatial and frequency domain techniques to effectively separate the illumination and illumination-invariant layers in low-light images while also extracting complementary shadow information. We also extracted color invariant information to maintain color stability. The extracted information was utilized as a conditioning mechanism to control the noise process within the Diffusion model.

B. Fundamental Diffusion

Our approach leverages a diffusion framework built upon a U-Net backbone to tackle the low-light image enhancement problem. In our model, the forward diffusion process progressively adds Gaussian noise to a clean image x_0 over T timesteps. Specifically, at each timestep t, the noisy image x_t is defined as

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (16)$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ and each $\alpha_s \in (0,1)$ controls the noise schedule.

The reverse process aims to reconstruct x_0 from x_t by learning to predict the added noise. We parameterize this noise prediction with a U-Net architecture, which effectively preserves high-frequency details through its symmetric encoder-decoder structure and skip connections. Our network is conditioned on auxiliary information, including illumination, illumination-invariant features, shadow cues, and color invariance, which are concatenated to the input.

By incorporating the conditional signals into the diffusion process, our network is guided not only to denoise but also to enhance the relevant structures in low-light scenarios. In summary, our diffusion model with a U-Net backbone offers an effective mechanism for high-fidelity low-light image restoration, ensuring both pixel-level accuracy and perceptual coherence.

C. Loss

The loss function setting in this paper mainly refers to the loss function design of CLEDiffusion [21].

Diffusion Loss. The training objective is based on the simplified loss function [18], [41], denoted as L_{simple}, which directly measures the discrepancy between the true noise ϵ and its prediction $\epsilon_\theta(x_t,t,c)$:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t,x_0,\epsilon,c}\left[\|\epsilon - \epsilon_\theta(x_t,t,c)\|_2^2\right] \quad (17)$$

where c represents the conditional input comprising the aforementioned auxiliary cues. This formulation stabilizes

TABLE I: Quantitative evaluation on three datasets: LOLv1 [5], LOLv2-real [25], and LSRW [26]. All values are formatted with three decimal places. Higher PSNR and SSIM indicate better performance, while lower LPIPS and FID are preferable. Note that methods are trained only on the LOLv1 training set and are directly evaluated on the remaining datasets

Methods	Reference	LOL v1				LOLv2-real				LSRW			
		PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
NPE [27]	TIP' 13	16.970	0.484	0.400	104.057	17.333	0.464	0.396	100.025	16.188	0.384	0.440	90.132
SRIE [28]	CVPR' 16	11.855	0.495	0.353	88.728	14.451	0.524	0.332	78.834	13.357	0.415	0.399	69.082
LIME [29]	TIP' 16	17.546	0.531	0.387	117.892	17.483	0.505	0.428	118.171	17.342	0.520	0.471	75.595
RetinexNet [5]	BMVC' 18	16.774	0.462	0.417	126.266	17.715	0.652	0.436	133.905	15.609	0.414	0.454	108.350
DSLR [30]	TMM' 20	14.816	0.572	0.375	104.428	17.000	0.596	0.408	114.306	15.259	0.441	0.464	84.930
DRBN [31]	CVPR' 20	16.677	0.730	0.345	98.732	18.466	0.768	0.352	89.085	16.734	0.507	0.457	80.727
Zero-DCE [13]	CVPR' 20	14.861	0.562	0.372	87.238	18.059	0.580	0.352	80.449	15.867	0.443	0.411	63.320
MIRNet [6]	ECCV' 20	24.138	0.830	0.250	69.179	20.020	0.820	0.233	49.108	16.470	0.477	0.430	93.811
EnlightenGAN [7]	TIP' 21	17.606	0.653	0.372	94.704	18.676	0.678	0.364	84.044	17.106	0.463	0.406	69.033
ReLLIE [17]	ACM MM' 21	11.437	0.482	0.375	95.510	14.400	0.536	0.334	79.838	13.685	0.422	0.404	65.221
RUAS [32]	CVPR' 21	16.405	0.503	0.364	101.971	15.351	0.495	0.395	94.162	14.271	0.461	0.501	78.392
DDIM [19]	ICLR' 21	16.521	0.776	0.376	84.071	15.280	0.788	0.387	76.387	14.858	0.486	0.495	71.812
CDEF [33]	TMM' 22	16.335	0.585	0.407	90.620	19.757	0.630	0.349	74.055	16.758	0.465	0.399	62.780
SCI [34]	CVPR' 22	14.784	0.525	0.366	78.598	17.304	0.540	0.345	67.624	15.242	0.419	0.404	56.261
URetinex-Net [35]	CVPR' 22	19.842	0.824	0.237	52.383	21.093	0.858	0.208	49.836	18.271	0.518	0.419	66.871
SNRNet [36]	CVPR' 22	24.610	0.842	0.233	55.121	21.480	0.849	0.237	54.532	16.499	0.505	0.419	65.807
Uformer [14]	CVPR' 22	19.001	0.741	0.354	109.351	18.442	0.759	0.347	98.138	16.591	0.494	0.435	82.299
Restormer [15]	CVPR' 22	20.614	0.797	0.288	72.998	24.910	0.851	0.264	58.649	16.303	0.453	0.427	69.219
Palette [20]	SIGGRAPH'22	11.771	0.561	0.498	108.291	14.703	0.692	0.333	83.942	13.570	0.476	0.479	73.841
UHDFour2× [37]	ICLR' 23	23.093	0.821	0.259	56.912	21.785	0.854	0.292	60.837	17.300	0.529	0.443	62.032
WeatherDiff [38]	TPAMI' 23	17.913	0.811	0.272	73.903	20.009	0.829	0.253	59.670	16.507	0.487	0.431	96.050
GDP [23]	CVPR' 23	15.896	0.542	0.421	117.456	14.290	0.493	0.435	102.416	12.887	0.362	0.412	76.908
DiffLL [22]	ACM TOG'23	26.336	0.845	0.217	48.114	28.857	0.876	0.207	45.359	19.281	0.552	0.350	**45.294**
QuadPrior [39]	CVPR' 24	22.849	0.800	0.201	–	20.592	0.811	0.202	–	–	–	–	–
Lightendiffusion [40]	ECCV' 24	20.188	0.814	0.316	85.930	22.697	0.853	0.306	75.582	18.397	0.534	0.428	67.801
Ours	–	**26.947**	**0.921**	**0.071**	**46.670**	**31.223**	**0.926**	**0.044**	**36.875**	**20.692**	**0.560**	**0.198**	65.710

training and helps the model learn to invert the diffusion process more precisely.

Illumination Alignment Loss. To enforce global brightness consistency, we define

$$\mathcal{L}_{\text{illum}} = \|G(\hat{x}_0) - G(x_0)\|_1 \tag{18}$$

where $G(\cdot)$ converts an RGB image to grayscale. This loss ensures that the enhanced output maintains similar overall luminance to the ground truth.

Chromatic Fidelity Loss. To reduce color distortion, we minimize the angular difference between normalized RGB vectors:

$$\mathcal{L}_{\text{chrom}} = \sum_{i=1}^{H\times W} \left(1 - \frac{\hat{x}_0^{(i)} \cdot x_0^{(i)}}{\|\hat{x}_0^{(i)}\|_2 \, \|x_0^{(i)}\|_2}\right) \tag{19}$$

where the index i runs over all pixels, promoting accurate chromatic alignment.

Structural Similarity Loss. To preserve local texture and structure, we adopt an SSIM-based loss:

$$\mathcal{L}_{\text{SSIM}} = 1 - \frac{(2\mu_{x_0}\mu_{\hat{x}_0} + c_1)(2\sigma_{x_0\hat{x}_0} + c_2)}{(\mu_{x_0}^2 + \mu_{\hat{x}_0}^2 + c_1)(\sigma_{x_0}^2 + \sigma_{\hat{x}_0}^2 + c_2)} \tag{20}$$

with μ and σ denoting local means and standard deviations (computed over a sliding window), $\sigma_{x_0\hat{x}_0}$ the covariance, and constants c_1, c_2 ensuring numerical stability.

Deep Feature Consistency Loss. To further align high-level semantic details, we employ a VGG-based perceptual loss:

$$\mathcal{L}_{\text{feat}} = \sum_{l\in L} \frac{1}{N_l} \left\|\phi_V^l(\hat{x}_0) - \phi_V^l(x_0)\right\|_2^2 \tag{21}$$

where $\phi_V^l(\cdot)$ extracts features from the l-th layer of VGGNet and $N_l = H_l \times W_l \times C_l$ normalizes for feature dimension.

Total Loss : The total loss is a weighting of the above losses:

$$\begin{aligned}\mathcal{L}_{\text{Total}} =& \mathcal{L}_{\text{simple}} + \omega_{\text{illum}}\mathcal{L}_{\text{illum}} + \omega_{\text{chrom}}\mathcal{L}_{\text{chrom}} \\ &+ \omega_{\text{SSIM}}\mathcal{L}_{\text{SSIM}} + \omega_{\text{feat}}\mathcal{L}_{\text{feat}}\end{aligned} \tag{22}$$

where ω_{illum}, ω_{chrom}, ω_{ssim}, ω_{feat} are the weights of the corresponding losses and we use the same values as in [21].

IV. EXPERIMENT

A. Experimental Settings

Implementation Details. Our models are trained on NVIDIA A40 GPUs. The AdamW [42] optimizer was used and the learning rate was set to 5×10^{-5} with a learning rate decay coefficient of 1×10^{-4}. The batch size and patch size are set to 8 and 256×256 respectively. The backbone network of Diffusion is the commonly used U-Net network structure [43]. The time step T is set to 1000 for the training phase and the implicit sampling step is set to 100 for both the training and inference phases.

Datasets. Our model is trained on the LOLv1 [5] real image dataset. The LOLv1 real image dataset contains 500 pairs of real images, of which 485 pairs are used for training and 15 pairs are used for evaluation. We also evaluated our model directly on the real image portion of the LOLv2 [25] dataset as well as the LSRW [26] dataset without training and fine-tuning. Our model is tested directly on 100 pairs of evaluation images for LOLv2-real and 50 pairs of evaluation images for LSRW. In addition to verify the generalization ability of our model we add tests on the DICM dataset [44], the MEF dataset [45], and the LIME dataset [29].

Metrics. In our experiments we use two evaluation metrics commonly used in low-light enhancement tasks, PSNR and

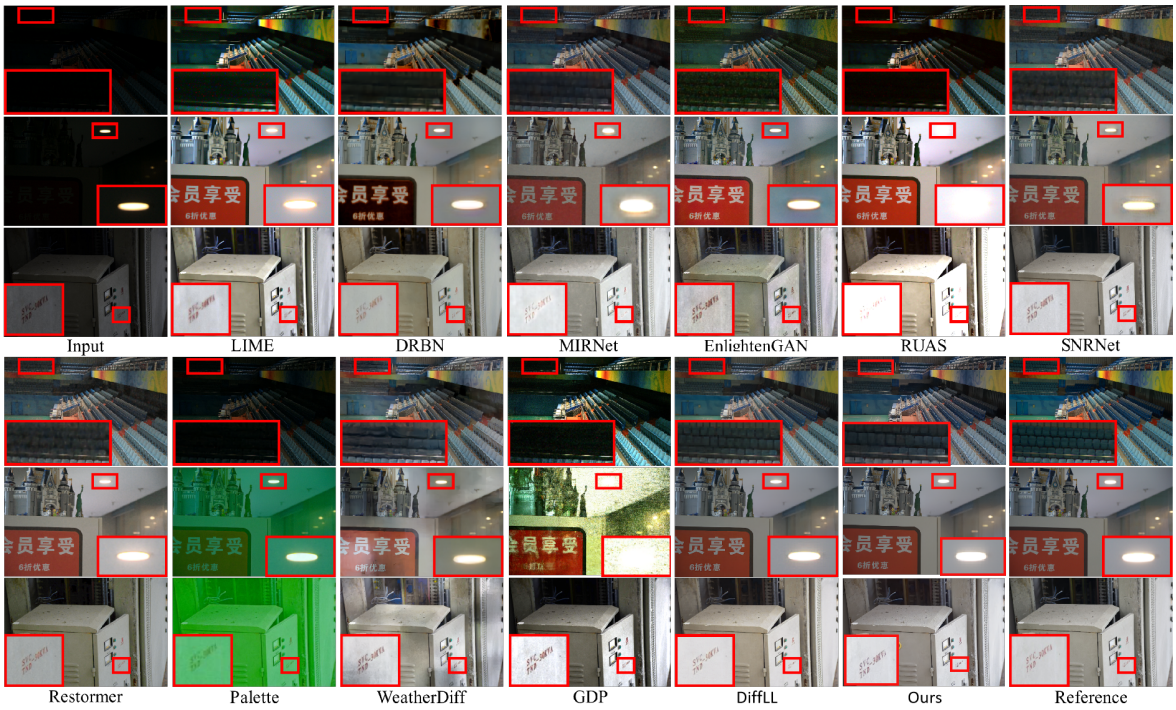


Fig. 3: Visual comparisons of our approach with competing methods. The input image is from datasets LOLv1, LOLv2-real, LSRW for the first, second and third rows, respectively.

SSIM [46], for evaluation. We also use two perceptual evaluation metrics LPIPS [47] as well as FID [48] to evaluate the images. In this paper LPIPS uses AlexNet [49] as the backbone. In verifying the generalization ability of the model, we use NIQE [50], BRISQUE [51] and PI [52] as evaluation metrics for those datasets that do not contain paired images.

B. Comparison Methods

Quantitative Comparison. To validate our diffusion-based enhancement framework with structured illumination priors, we performed experiments on LOLv1, LOLv2-real and LSRW (Table I), training only on LOLv1 and evaluating off-the-shelf elsewhere. Our method outperforms prior work on all metrics and datasets. On LOLv1, it achieves PSNR 26.947 and SSIM 0.921—surpassing DiffLL (26.336/0.845) and SNRNet (24.610/0.842)—and sets a new LPIPS record of 0.071 (vs. 0.201). On LOLv2-real, it generalizes strongly with PSNR 31.223 and SSIM 0.926 (+2.366 dB, +0.050 vs. DiffLL), while securing the lowest FID 36.875 and LPIPS 0.044. On LSRW, despite diverse scenes and lighting, our framework maintains the best PSNR 20.692, LPIPS 0.198, and SSIM 0.560; being still competitive on FID. Thus, our perceptual consistency, driven by explicit illumination, shadow, and color-invariant conditioning, is superior to previous SOTA methods.

As shown in Table II, we validate the generalization ability of our proposed model. Our method achieves the best perceptual quality on DICM across all metrics (NIQE, BRI, PI), on LIME for both BRI and PI, and on MEF in terms of BRI. Moreover, our approach ranks second in terms of NIQE on LIME. Compared with DiffLL, which consistently performs well, our approach further reduces distortion and artifacts, delivering sharper, more natural-looking results. These findings demonstrate the robustness and generalization ability of our method across diverse low-light scenarios.

TABLE II: Quantitative comparison of different methods on the DICM [44], MEF [45], and LIME [29] datasets. The best results are highlighted in **bold** and the second-best in underlined, and BRI represents BRISQUE [51].

Methods	DICM			MEF			LIME		
	NIQE↓	BRI↓	PI↓	NIQE↓	BRI↓	PI↓	NIQE↓	BRI↓	PI↓
$LIME_{1001[29]}$	4.476	27.375	4.216	4.744	39.095	5.160	5.045	32.842	4.859
$DRBN_{1001[31]}$	4.369	30.708	3.800	4.869	44.669	4.711	4.562	34.564	3.973
$Zero\text{-}DCE_{1001[13]}$	3.951	23.350	3.149	3.500	29.359	**2.989**	4.379	26.054	3.239
$MIRNet_{1001[6]}$	4.021	22.104	3.691	4.202	34.499	3.756	4.378	28.623	3.398
$EnlightenGAN_{1001[7]}$	3.832	19.129	3.256	3.556	26.799	3.270	4.249	22.664	3.381
$RUAS_{1001[32]}$	7.306	46.882	5.700	5.435	42.120	4.921	5.322	34.880	4.581
$DDIM_{1001[19]}$	3.899	19.787	3.213	3.621	28.614	3.376	4.399	24.474	3.459
$SCI_{1001[34]}$	4.519	27.922	3.700	3.608	26.716	3.286	4.463	25.170	3.376
$URetinex\text{-}Net_{1001[35]}$	4.774	24.544	3.565	4.231	34.720	3.665	4.694	29.022	3.713
$SNRNet_{1001[36]}$	3.804	19.459	3.285	4.063	28.331	3.753	4.597	29.023	3.677
$Uformer_{1001[14]}$	3.847	19.657	3.180	3.935	25.240	3.582	4.300	21.874	3.565
$Restormer_{1001[15]}$	3.964	19.474	3.152	3.815	25.322	3.436	4.365	22.931	3.292
$Palette_{1001[20]}$	4.118	18.732	3.425	4.459	25.602	4.205	4.485	20.551	3.579
$UHDFour2\times_{1001[37]}$	4.575	26.926	3.684	4.231	29.538	4.124	4.430	20.263	3.813
$WeatherDiff_{1001[38]}$	3.773	20.387	3.130	3.753	30.480	3.312	4.312	28.090	3.424
$GDP_{1001[23]}$	4.358	19.294	3.552	4.609	34.859	4.115	4.891	27.460	3.694
$DiffLL_{1001[22]}$	3.806	18.584	3.011	**3.427**	24.165	3.011	**3.777**	19.843	3.074
Ours	**3.355**	**16.676**	**2.606**	3.827	**21.092**	3.171	4.091	**16.820**	**2.838**

The results validate the robustness and superiority of our method. By integrating illumination decomposition, shadow priors, and color constraints into the conditional diffusion process, we effectively guide the denoising trajectory and facilitate the generation of visually plausible and structurally faithful enhanced images.

Qualitative Comparison. Fig. 3 shows a qualitative comparison of three different images. Our method recovers both local and global structures under challenging low-light conditions and accurately captures subtle boundaries, restores

TABLE III: Ablation study for our proposed SCEM module on LOLv1.

SCEM	PSNR ↑	SSIM ↑	LPIPS ↓
–	22.220	0.810	0.220
✓	**26.947**	**0.921**	**0.071**

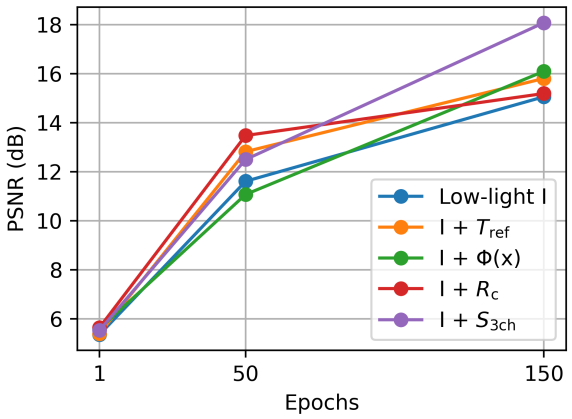


Fig. 4: Ablation study comparing PSNR for different input configurations: (1) low-light image $\mathcal{I}$ only, (2) $\mathcal{I}$ with illumination T_{ref}, (3) $\mathcal{I}$ with color-invariant features $\Phi(x)$, (4) $\mathcal{I}$ with illumination-invariant features R_c, and (5) $\mathcal{I}$ with shadow priors S_{3ch}.

the fine texture, and mitigates the color shift throughout. In contrast, for the other methods, certain foreground regions appear overenhanced or remain too dim, with incomplete artifact removal and lost edge definition.

C. Ablation Study

We utilize LOLv1 for our ablation study, focusing on the four main components of the proposed SCEM: illumination T_{ref}, color-invariant features $\Phi(x)$, illumination-invariant features R_c, and shadow priors S_{3ch}. The default input is the low-light image $\mathcal{I}$, and we evaluate the contribution of each component to the LLIE process by concatenating $\mathcal{I}$ with each component, as illustrated in Fig. 4. All variants show similar performance in the first epoch (PSNR 5.3–5.6 dB, SSIM 0.03–0.05), indicating that auxiliary inputs do not yield significant benefit under extremely limited training.

At epoch 50, the baseline attains 11.61 dB / 0.3414. Adding color-invariant features slightly degrades performance (11.07 dB / 0.3012), underscoring the stage-dependent value of different priors. Reference illumination raises PSNR to 12.82 dB (SSIM 0.3328), while illumination-invariant inputs achieve 13.48 dB / 0.5182. Shadow priors similarly boost SSIM (0.5139) with PSNR 12.50 dB. By epoch 150, the baseline reaches 15.06 dB/0.7035. Color-invariant features yield the highest PSNR (16.09 dB) at the expense of SSIM (0.6397); illumination cues continue to aid enhancement; and shadow priors deliver the top PSNR (18.08 dB) but moderate SSIM (0.5718).

In summary, shadow priors maximize final PSNR, illumination-invariant representations most effectively accelerate convergence and structural fidelity, color-invariant features trade structure for color consistency, and explicit illumination cues provide modest acceleration, guiding the design of hybrid inputs for brightness versus structure objectives.

In Table III, our ablation study on the LOLv1 dataset demonstrates that the inclusion of the SCEM conditional input dramatically improves the model's performance. Without this conditional mechanism, the model achieves a PSNR of 22.220 dB, SSIM of 0.810, and LPIPS of 0.220. With our proposed SCEM module, PSNR rises to 26.947 dB, SSIM to 0.921, and LPIPS drops to 0.071. These improvements indicate that the conditional mechanism plays a vital role in guiding the enhancement process, leading to significantly lower reconstruction errors, improved structural fidelity, and superior perceptual quality, thereby validating its critical importance in our low-light image enhancement framework.

V. CONCLUSION

We present a novel diffusion-based low-light image enhancement method that is conditioned on color and luminance priors. More in detail, we propose to use four specific priors: illumination, illumination-invariant, shadow, and color invariance priors. These priors are introduced into the diffusion process thanks to our proposed Structured Control Embedding Module. Experimental results demonstrate that our model achieves state-of-the-art performance and exhibits strong generalization across diverse datasets without fine-tuning. These findings underscore the robustness and effectiveness of our method in restoring visual quality under challenging lighting conditions.

REFERENCES

[1] X. Xu, R. Wang, and J. Lu, "Low-light image enhancement via structure modeling and guidance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9893–9903.

[2] H. Zhou, W. Dong, X. Liu, Y. Zhang, G. Zhai, and J. Chen, "Low-light image enhancement via generative perceptual priors," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 10, 2025, pp. 10 752–10 760.

[3] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE transactions on consumer electronics*, vol. 53, no. 2, pp. 593–600, 2007.

[4] B. Li, S. Wang, and Y. Geng, "Image enhancement based on retinex and lightness decomposition," in *2011 18th IEEE International Conference on Image Processing*. IEEE, 2011, pp. 3417–3420.

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

[6] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*. Springer, 2020, pp. 492–511.

[7] 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*, vol. 30, pp. 2340–2349, 2021.

[8] J. R. Jebadass and P. Balasubramaniam, "Low light enhancement algorithm for color images using intuitionistic fuzzy sets with histogram equalization," *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 8093–8106, 2022.

[9] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Journal of the Optical society of America*, vol. 61, no. 1, pp. 1–11, 1971.

[10] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE transactions on image processing*, vol. 6, no. 3, pp. 451–462, 1997.

[11] J. J. Jeon, J. Y. Park, and I. K. Eom, "Low-light image enhancement using gamma correction prior in mixed color spaces," *Pattern Recognition*, vol. 146, p. 110001, 2024.

[12] M.-x. Yang, G.-j. Tang, X.-h. Liu, L.-q. Wang, Z.-g. Cui, and S.-h. Luo, "Low-light image enhancement based on retinex theory and dual-tree complex wavelet transform," *Optoelectronics Letters*, vol. 14, no. 6, pp. 470–475, 2018.

[13] 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*, 2020, pp. 1780–1789.

[14] 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*, 2022, pp. 17 683–17 693.

[15] 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*, 2022, pp. 5728–5739.

[16] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 3, 2023, pp. 2654–2662.

[17] R. Zhang, L. Guo, S. Huang, and B. Wen, "Rellie: Deep reinforcement learning for customized low-light image enhancement," in *Proceedings of the 29th ACM international conference on multimedia*, 2021, pp. 2429–2437.

[18] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[19] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[20] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in *ACM SIGGRAPH 2022 conference proceedings*, 2022, pp. 1–10.

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

[22] H. Jiang, A. Luo, H. Fan, S. Han, and S. Liu, "Low-light image enhancement with wavelet-based diffusion models," *ACM Transactions on Graphics (TOG)*, vol. 42, no. 6, pp. 1–14, 2023.

[23] B. Fei, Z. Lyu, L. Pan, J. Zhang, W. Yang, T. Luo, B. Zhang, and B. Dai, "Generative diffusion prior for unified image restoration and enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 9935–9946.

[24] 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*, 2023, pp. 12 302–12 311.

[25] 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*, vol. 30, pp. 2072–2086, 2021.

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

[27] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE transactions on image processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[28] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2782–2790.

[29] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[30] S. Lim and W. Kim, "Dslr: Deep stacked laplacian restorer for low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 23, pp. 4272–4284, 2020.

[31] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3063–3072.

[32] 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*, 2021, pp. 10 561–10 570.

[33] X. Lei, Z. Fei, W. Zhou, H. Zhou, and M. Fei, "Low-light image enhancement using the cell vibration model," *IEEE Transactions on Multimedia*, vol. 25, pp. 4439–4454, 2022.

[34] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5637–5646.

[35] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5901–5910.

[36] 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*, 2022, pp. 17 714–17 724.

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

[38] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 10 346–10 357, 2023.

[39] W. Wang, H. Yang, J. Fu, and J. Liu, "Zero-reference low-light enhancement via physical quadruple priors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 26 057–26 066.

[40] H. Jiang, A. Luo, X. Liu, S. Han, and S. Liu, "Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models," in *European Conference on Computer Vision*. Springer, 2024, pp. 161–179.

[41] D. Kingma, T. Salimans, B. Poole, and J. Ho, "Variational diffusion models," *Advances in neural information processing systems*, vol. 34, pp. 21 696–21 707, 2021.

[42] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[43] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[44] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE transactions on image processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[45] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.

[46] 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*, vol. 13, no. 4, pp. 600–612, 2004.

[47] 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*, 2018, pp. 586–595.

[48] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.

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

[50] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[51] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[52] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 pirm challenge on perceptual image super-resolution," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.