DEGRADATION-AWARE DYNAMIC ADAPTATION FOR UNPAIRED LOW-LIGHT IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Pretrained diffusion models provide powerful natural-image priors for low-light image enhancement (LLIE), offering a promising way to reduce the dependence on paired training data. However, existing unpaired diffusion-based methods typically employ a fixed conditioning mechanism shared by all inputs. This is suboptimal for real-world LLIE, where different images exhibit substantially different combinations of illumination deficiency, color distortion, and structural degradation. To address this limitation, we propose *Degradation-aware Dynamic Adaptation* (DDA), a one-step unpaired LLIE framework that adaptively exploits pretrained diffusion priors according to the degradation characteristics of each input. DDA contains two complementary modules. The *Degradation-aware Prior Generation* (DPG) module analyzes the illumination, color, and structure degradation of the input and produces an image-specific representation that identifies its restoration demands. The *Degradation-Driven State-Glimpse* (DDSG) adapter then translates these demands into layer-compatible dynamic updates for the image-conditioning pathway, allowing the diffusion prior to be injected differently for different low-light images. Extensive experiments on LOL-v2-real and LSRW, together with downstream classification, detection, and semantic segmentation evaluations, demonstrate that DDA consistently outperforms existing unsupervised LLIE methods. *Code will be released upon acceptance.*

1 INTRODUCTION

Low-light image enhancement (LLIE) aims to recover visually pleasing and semantically reliable images from insufficiently illuminated observations, benefiting both human perception and downstream vision systems (Zhang et al., 2026a;b). Real-world low-light degradation, however, goes far beyond reduced brightness, often involving non-uniform illumination, color shifts, and structural loss that jointly impair visual quality and high-level tasks (Zhu et al., 2026). Diffusion models, with their rich natural-image priors, offer a promising solution for reconstructing plausible details where underexposed regions provide little local evidence, and have thus attracted increasing attention for LLIE.

Existing diffusion-based LLIE methods fall into two categories: *task-specific diffusion training* and *pretrained diffusion adaptation*. The first trains a dedicated diffusion model for low-light restoration (Yi et al., 2023), either from aligned low-/normal-light pairs or through unsupervised objectives such as the latent Retinex priors in LightenDiffusion (Jiang et al., 2024); either way, the generative process is learned from scratch and inference relies on iterative sampling, incurring considerable training and inference costs. The second instead adapts pretrained diffusion models to LLIE, exploiting the generative knowledge of large-scale text-to-image models. Representative paradigms include the training-free guidance of AGLLDiff (Lin et al., 2025), the one-step unpaired translation of CycleGAN-Turbo (Parmar et al., 2024), and the semantically consistent unsupervised fine-tuning of GEFU (Wang et al., 2025), which substantially reduce the reliance on paired supervision and iterative inference. Despite this progress, existing pretrained-diffusion approaches inject visual conditions into the diffusion UNet through key/value projections that are fixed after training (Ye et al., 2023) (Fig. 1(a)). As a result, all inputs share the same conditioning pathway, whether the dominant degradation is insufficient illumination, chromatic distortion, or structural damage. Such sample-agnostic injection is suboptimal for real-world LLIE.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



Figure 1: (a) Existing image adapters (Ye et al., 2023; Wang et al., 2025) keep the conditioning weights fixed for all inputs. (b) Our adapter perceives the degradation state of each input and dynamically adjusts its weights accordingly, enabling input-specific enhancement.

Different degradation compositions demand different restoration behaviors (Zhang et al., 2019). To verify this, we analyze 100 real low-light images from LOL-v2-real and find that the relative proportions of illumination, color, and structure degradation vary substantially across samples (see the **Appendix** for computation details). This motivates an degradation-aware conditioning pathway that perceives the degradation composition of each input and adapts the image-conditioning projections accordingly (Fig. 1(b)). Realizing this design, however, poses two challenges. First, such compositions are not directly observable in the unpaired setting: without a normal-light reference, the restoration behavior an input demands remains unknown. Second, even with such demands identified, translating them into dynamic projection updates is non-trivial. Such updates must be compatible with the layer-specific base projection they modify; otherwise, they can corrupt the pretrained conditioning behavior.

To address these challenges, we propose *Degradation-aware Dynamic Adaptation (DDA)*, a one-step unpaired LLIE framework built upon a pretrained diffusion-based image-translation backbone. DDA consists of two complementary components. First, to recover the restoration demands without paired references, the *Degradation-aware Prior Generation (DPG)* module constructs multi-view illumination, color, and structure priors solely from the input image and dynamically integrates them through specialized fusion experts, producing an image-specific degradation-prior embedding that encodes these demands. Second, to translate these demands into layer-compatible adaptation, the *Degradation-Driven State-Glimpse (DDSG)* adapter combines this prior with a compact characterization of layer-specific conditioning projections to generate dynamic updates, tailoring image-condition injection to both the degradation of the current input and the projection state of each cross-attention layer.

Our contributions can be summarized as follows: (1) we identify sample-agnostic condition injection as a key limitation of pretrained diffusion-based LLIE and introduce degradation-aware dynamic adaptation, which tailors the image-conditioning pathway to the degradation composition of each input; (2) a DPG module that dynamically integrates multi-view illumination, color, and structure priors into an image-specific degradation-prior embedding, adapting the fusion process itself to each input; and (3) a DDSG adapter that translates input-specific degradation demands into compact dynamic updates by jointly conditioning on the degradation prior and the encoded state of each base projection, enabling sample-aware condition injection that remains compatible with the pretrained cross-attention mappings

2 Related Work

Unsupervised Low-light image enhancement. Based on the availability of normal-light references and task-specific training data, unsupervised LLIE methods are grouped into *zero-reference*, *zero-shot*, and *unpaired* approaches. *Zero-reference* and *zero-shot* approaches learn enhancement functions without normal-light references, typically relying on handcrafted non-reference constraints or image-specific priors. For example, Zero-DCE estimates enhancement curves with exposure, color, and spatial consistency losses (Guo et al., 2020), and Zero-DCE++ improves its efficiency and robustness (Li et al., 2021). Others introduce Retinex constraints, self-calibrated illumination learning, paired low-light instances, or pretrained priors to regularize enhancement (Liu et al., 2021; Ma et al., 2022; Fu et al., 2023; Wang et al., 2024; Shi et al., 2024). Although flexible, they offer limited guidance toward the normal-light target distribution. *Unpaired* approaches learn from independent low-light and normal-light image collections to capture target-domain characteristics without pixel-level alignment. EnlightenGAN introduces adversarial learning for unpaired enhancement (Jiang et al., 2021), while later methods exploit semantic guidance, prompt learning, implicit representations, and diffusion priors to improve perceptual quality and generalization (Zheng et al.,

2023; Liang et al., 2023; Yang et al., 2023; Jiang et al., 2024; Wang et al., 2025). Yet most existing unpaired methods employ a shared enhancement or conditioning mechanism across inputs, limiting adaptation to image-specific degradation compositions. In contrast, we combine degradation-aware prior generation with weight-aware dynamic adaptation to enable sample-specific enhancement under unpaired training.

Diffusion priors and adapter-based adaptation. Pretrained diffusion models have recently been exploited for LLIE due to their powerful generative priors (Jiang et al., 2024; Lin et al., 2025; Wang et al., 2025). Image-prompt adapters enable visual conditioning through trainable projection layers (Ye et al., 2023), which, however, remain fixed after training, yielding sample-agnostic injection. Beyond LLIE, adapting large pretrained models with input-dependent parameters has been explored along two lines: routing among pre-trained LoRA experts (Buehler & Buehler, 2024; Wu et al., 2024; He et al., 2026), and generating adapter parameters with hypernetworks (Lv et al., 2024). Both lines, however, fall short of degradation-aware adaptation on two sides. On the input side, routing-based methods either ignore degradation (Buehler & Buehler, 2024; Wu et al., 2024) or, as in MoR-DASR (He et al., 2026), estimate it via external CLIP-based scores under paired supervision, while hypernetwork-based ones rely on generic input features; none constructs an explicit degradation representation from the input itself. On the projection side, both lines determine the adaptation without access to the base projection state being modified, and routing is further confined to a fixed expert set. DDA closes both gaps by conditioning update generation on a degradation prior built from the input itself and a compact glimpse of the base projection state, realizing degradation-aware dynamic adaptation of the conditioning pathway for unpaired low-light image enhancement.

3 METHODOLOGY

3.1 PROBLEM STATEMENT

Let $\{x_i\}_{i=1}^{N_l} \subset \mathcal{X}_l$ and $\{y_j\}_{j=1}^{N_n} \subset \mathcal{X}_n$ denote unpaired collections of low-light and normal-light images, respectively, with no correspondence between samples. The goal is to learn an enhancement mapping $G : \mathcal{X}_l \to \mathcal{X}_n$ such that $\hat{y} = G(x)$ exhibits normal-light appearance while faithfully preserving the content of x. Without paired references, supervision is only available at the distribution level, and the degradation condition of each input remains unknown during both training and inference.

Overview. As shown in Fig. 2, the proposed framework, termed **DDA**, is built upon Stable Diffusion v2.1-Turbo (Rombach et al., 2022) and performs efficient one-step image-to-image translation for low-light enhancement. Our core idea is to make the diffusion conditioning process explicitly degradation-aware: we first perceive the input's degradation composition and then dynamically adapt the image-conditioning projections accordingly. To this end, we introduce two complementary modules. The *Degradation-aware Prior Generation* (DPG) module extracts illumination, color, and structure priors via multi-view construction, encodes them through multi-scale refinement, and fuses them via dynamic expert routing to produce an image-specific degradation-prior embedding c_x. The *Degradation-Driven State-Glimpse* (DDSG) adapter then generates sample-specific updates to the image-conditioning projections of each cross-attention layer by jointly considering the degradation prior c_x and the state of the projection weights being adapted, so that condition injection adapts to the input degradation while remaining compatible with the learned adapter mappings.

During training, DDA enhances each low-light sample x to $\hat{y} = G(x)$, which a symmetric degradation translator maps back to the low-light domain to close the cycle. Each normal-light sample y follows the reverse cycle symmetrically. The two mappings are optimized jointly under the CycleGAN-Turbo objectives (Parmar et al., 2024), which couple adversarial supervision in both translation directions with cycle-consistency reconstruction, so that the outputs align with the target domain while the input content is preserved. Within DDA, only the LoRA-adapted encoder–decoder, DPG, and DDSG are trainable, with all remaining pretrained weights frozen to retain the generative prior of the backbone and keep inference one-step (see the **Appendix** for details).

3.2 DEGRADATION-AWARE PRIOR GENERATION

The *Degradation-aware Prior Generation* (DPG) module consists of two components: a *Multi-view Prior Extraction Block* and a *Dynamic Expert Fusion Block*. The former is introduced because

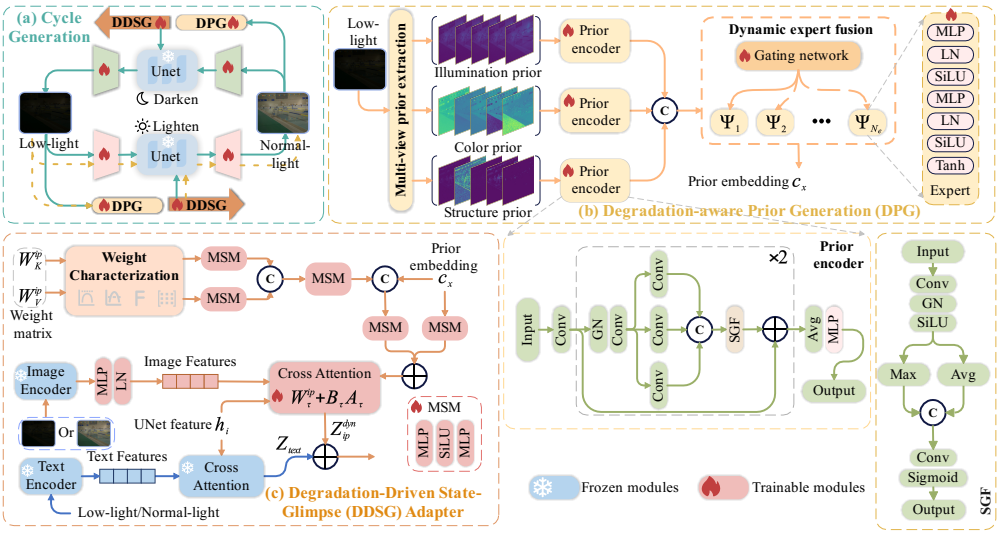


Figure 2: Overview of the proposed degradation-aware dynamic adaptation framework. (a) Overall architecture for unpaired low-light image enhancement. (b) The DPG module extracts multi-view illumination, color, and structure priors and dynamically fuses them into an image-specific degradation-prior embedding. (c) The DDSG combines the degradation prior with the current K/V projection-weight states to predict low-rank residual updates for dynamically adapting the K/V projection matrices.

illumination-, color-, and structure-related degradations are entangled in the RGB space and are difficult to characterize using a single unified representation. It therefore constructs explicit priors from these three perspectives and encodes them into type-specific degradation features. The latter is designed to fuse these features into a unified degradation-prior embedding c_x. Since the importance and interaction of the three degradation types vary across input images, a fixed fusion rule cannot provide an appropriate combination for every sample. The block therefore predicts sample-dependent routing scores from the joint degradation representation and selects a corresponding fusion expert to generate c_x. The reverse normal-light-to-low-light translation follows the same DPG pipeline to obtain a direction-specific prior embedding.

3.2.1 Multi-view prior extraction.

Within each degradation type, a single image statistic captures only a partial manifestation of the underlying degradation. For example, illumination degradation may involve global darkness, uneven exposure, or channel-dependent intensity variations, whereas structural degradation may appear as weakened edges, reduced local contrast, or texture smoothing. We therefore characterize the illumination, color, and structure priors using multiple complementary views and encode each prior type into a compact degradation feature.

Given an input image $x \in \mathbb{R}^{3\times H\times W}$, the multi-view prior tensor for degradation type $t \in \{\text{illum}, \text{color}, \text{struct}\}$ is constructed as

$$P_t = \text{Concat}\left[\phi_1^t(x), \phi_2^t(x), \ldots, \phi_{M_t}^t(x)\right] \in \mathbb{R}^{C_t\times H\times W}, \tag{1}$$

where $\phi_m^t(\cdot)$ denotes the m-th view-construction function for degradation type t, and M_t is the number of views. C_t denotes the total number of channels after concatenating all views of type t, and $\text{Concat}[\cdot]$ denotes channel-wise concatenation.

In our implementation, we set $M_{\text{illum}} = 6$, $M_{\text{color}} = 5$, and $M_{\text{struct}} = 5$. The illumination prior comprises mean intensity, Rec. 709 luminance, maximum-channel response, lightness, YCgCo luminance, and channel-energy response. The color prior comprises YUV chrominance, YCbCr chrominance, opponent-color representation, normalized chromaticity, and saturation. The structure prior comprises gradient magnitude, Laplacian response, relative total variation, local structure-

tensor response, and local contrast. Detailed definitions of all view-construction functions are provided in **Appendix**.

Each prior tensor P_t is processed by a lightweight prior encoder. The encoders for different prior types share the same architecture but use independent parameters, allowing their heterogeneous channel compositions and spatial responses to be mapped into a common feature space:

$$F_t^0 = \mathrm{Conv}_t^{\mathrm{in}}(P_t) \in \mathbb{R}^{C\times H\times W}, \tag{2}$$

where C denotes the unified feature dimension.

To aggregate degradation responses at different spatial scales, the encoder applies two stacked multi-scale refinement blocks. In the k-th block, $k \in \{1, 2\}$, the input feature is first normalized and projected:

$$F_{t,\mathrm{p}}^k = \mathrm{Conv}_t^k\left(\mathrm{GN}(F_t^{k-1})\right). \tag{3}$$

Parallel depthwise convolutions are then used to extract responses with different receptive fields:

$$F_{t,\mathrm{ms}}^k = \mathrm{Concat}\left[\mathrm{DWConv}_{s\times s}^d\left(F_{t,\mathrm{p}}^k\right) \,\middle|\, (s,d) \in \mathcal{K}\right], \tag{4}$$

where $\mathcal{K} = \{(3, 1), (5, 1), (3, 2)\}$ denotes the kernel-size and dilation-rate pairs. Since the concatenated multi-scale responses may contain spatially redundant or weakly relevant activations, a spatial gated fusion unit is used to emphasize degradation-sensitive regions. A residual connection is further introduced to preserve the original prior responses during refinement:

$$F_t^k = F_t^{k-1} + \mathrm{SGF}_t^k\left(F_{t,\mathrm{ms}}^k\right). \tag{5}$$

Here, $\mathrm{SGF}(\cdot)$ consists of convolution, group normalization, SiLU activation, and Avg/Max-pooling-based spatial attention followed by convolution and sigmoid gating. After two refinement blocks, the type-specific degradation feature is obtained as

$$f_t = \mathrm{MLP}_t\left(\mathrm{GAP}(F_t^2)\right), \tag{6}$$

where $\mathrm{GAP}(\cdot)$ denotes global average pooling.

3.2.2 Dynamic expert fusion.

The three type-specific features are first concatenated as $f_{\mathrm{all}} = \mathrm{Concat}[f_{\mathrm{illum}}, f_{\mathrm{color}}, f_{\mathrm{struct}}]$ so that their cross-prior relations can be considered jointly. A fixed fusion mapping would apply the same interaction pattern to all inputs, although the relative contributions of illumination, color, and structure cues vary across images. We therefore parameterize the fusion operation using N_e independently learned experts, each representing a distinct cross-prior fusion pattern and consisting of two sequential MLP–LN–SiLU (MLS) blocks followed by a $\tanh$ activation, i.e., $\Psi_e(\cdot) = \tanh(MLS(MLS(\cdot)))$ for $e = 1, \ldots, N_e$.

The fusion expert is selected according to the degradation composition of the current input. An MLP-based gating network $\mathcal{G}$ predicts the routing probabilities $\boldsymbol{p} = \mathrm{Softmax}(\mathcal{G}(f_{\mathrm{all}}))$, where p_e denotes the routing probability of the e-th expert. Rather than averaging the outputs of different fusion experts, top-1 routing selects $e^* = \arg\max_{e\in\{1,\ldots,N_e\}} p_e$, which encourages each expert to specialize in distinct degradation compositions. The selected expert then produces the degradation-prior embedding $c_x = \Psi_{e^*}(f_{\mathrm{all}})$, which serves as the degradation condition for the subsequent DDSG module. We set $N_e = 4$ in all experiments, and a straight-through estimator is used during training to enable gradient propagation through the discrete top-1 routing decision.

3.3 Degradation-Driven State-Glimpse Adapter

The *Degradation-Driven State-Glimpse (DDSG) adapter* performs degradation-adaptive image-condition injection by modulating the image-adapter projections in the UNet cross-attention layers. To align the condition with the translation direction, we use a gamma-transformed view of the input x: the lightened view $x^{1/2.2}$ for low-light enhancement, and the darkened view $x^{1.2}$ for the reverse translation. The view is encoded by a pretrained CLIP image encoder with an MLP and layer normalization, serving as the image token z_{ip}.

Given the intermediate UNet feature h_i and the image token z_{ip}, a conventional image adapter generates the image-conditioned response by projecting z_{ip} into keys and values:

$$Z_{\text{ip}} = \text{Softmax}\left(\frac{Q_i K_{\text{ip}}^{\text{T}}}{\sqrt{d}}\right) V_{\text{ip}}, \tag{7}$$

where $Q_i = h_i W_q$, $K_{\text{ip}} = z_{\text{ip}}(W_k^{\text{ip}})^{\text{T}}$, and $V_{\text{ip}} = z_{\text{ip}}(W_v^{\text{ip}})^{\text{T}}$. Since W_k^{ip} and W_v^{ip} are shared across samples, the image-condition injection is sample-agnostic and cannot accommodate diverse low-light degradations. To achieve sample-adaptive K/V projections, DDSG must jointly predict the updated projection parameters from both the degradation prior c_x and the current layer's base projections $W_k^{\text{ip}}, W_v^{\text{ip}}$. However, feeding c_x together with the full projection matrices into a prediction function would incur prohibitive computational cost. We therefore address this challenge through two coordinated components. The *weight-state encoding* block compresses the high-dimensional W_k^{ip} and W_v^{ip} into lightweight state representations s_k, s_v, enabling subsequent prediction at an acceptable computational cost. The *state-glimpse dynamic projection* block then takes c_x along with s_k, s_v as input, and predicts residual updates via a dual-path hypernetwork with low-rank factorization; these residuals are added to the original projections to obtain the dynamic K/V projections. These two components are detailed below.

3.3.1 WEIGHT-STATE ENCODING

We characterize each projection weight matrix $W_\tau^{\text{ip}} \in \mathbb{R}^{d_o \times d_i}$, $\tau \in \{k, v\}$, using compact statistics that capture its response distribution and global magnitude, where d_i and d_o denote the input and output dimensions, respectively. Specifically, the input-channel energy $p_e(W_\tau^{\text{ip}}) = \text{diag}((W_\tau^{\text{ip}})^{\text{T}} W_\tau^{\text{ip}}) \in \mathbb{R}^{d_i}$ has its j-th element as the squared ℓ_2 norm of the j-th input-channel column, indicating the response strength of the corresponding input dimension, while the global statistics $p_g(W_\tau^{\text{ip}}) = [\|W_\tau^{\text{ip}}\|_F, \max(|W_\tau^{\text{ip}}|), \text{mean}(W_\tau^{\text{ip}})] \in \mathbb{R}^3$ characterize the overall scale, peak magnitude, and signed offset of the projection matrix. The two descriptors are concatenated and mapped to a compact weight-state embedding:

$$p_\tau = \text{Concat}\left[p_e(W_\tau^{\text{ip}}), p_g(W_\tau^{\text{ip}})\right] \in \mathbb{R}^{d_i+3}, \tag{8}$$

$$s_\tau = \text{MLP}_\tau\left(\text{SiLU}\left(\text{MLP}_\tau(p_\tau)\right)\right). \tag{9}$$

The key and value projections are encoded separately into s_k and s_v, which characterize the base projections to ensure compatibility of the weight updates.

3.3.2 STATE-GLIMPSE DYNAMIC PROJECTION

Given the weight-state embeddings s_k and s_v, DDSG first fuses them into a joint weight-state embedding $s_w = \text{MSM}(\text{Concat}[s_k, s_v])$, where $\text{MSM}(\cdot)$ is a lightweight MLP–SiLU–MLP block. The joint embedding s_w summarizes the state of the base K/V projections, allowing the update generator to account for the projection weights it modifies.

To generate sample-adaptive projection updates, we adopt a dual-path hypernetwork conditioned on both the degradation prior and the weight state. The first path predicts degradation-driven parameters $\theta_{\text{base}} = \text{MSM}(c_x)$ from the degradation-prior embedding c_x, serving as the primary driver of sample-specific adaptation, while the second path computes $\theta_{\text{state}} = \text{MSM}(\text{Concat}[c_x, s_w])$ to calibrate the update according to the current K/V projection state, improving its compatibility with the base projections. The final dynamic parameter vector is obtained by residual combination as $\theta = \theta_{\text{base}} + \theta_{\text{state}}$.

Instead of generating full projection matrices, DDSG reshapes θ into low-rank factors $\{B_k, A_k, B_v, A_v\}$ with $B_\tau \in \mathbb{R}^{d_o \times r}$ and $A_\tau \in \mathbb{R}^{r \times d_i}$, where $r = 4$ denotes the rank of the dynamic update, and adds the low-rank residuals to the base image-adapter projections as $W_\tau^{\text{ip,dyn}} = W_\tau^{\text{ip}} + B_\tau A_\tau$. This formulation efficiently adapts the image-conditioned K/V projections while keeping the update space compact. The dynamic projections then produce sample-adaptive keys and values, $K_{\text{ip}}^{\text{dyn}} = z_{\text{ip}}(W_k^{\text{ip,dyn}})^{\text{T}}$ and $V_{\text{ip}}^{\text{dyn}} = z_{\text{ip}}(W_v^{\text{ip,dyn}})^{\text{T}}$, which are substituted into the to yield the dynamic image-conditioned residual $Z_{\text{ip}}^{\text{dyn}}$, allowing each low-light input to receive degradation-specific enhancement modulation.

Table 1: Quantitative comparison on the LOL-v2-real and LSRW datasets. The best results are highlighted in **bold**. * denotes training on the LSRW dataset.

Method	LOL-v2-real			LSRW		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
EnlightenGAN *TIP'21*	18.68	0.6728	0.3013	17.59	0.4867	0.3117
Zero-DCE *CVPR'20*	18.06	0.5736	0.3125	15.86	0.4536	0.3176
Zero-DCE++ *TPAMI'21*	17.37	0.4373	0.3118	16.21	0.4571	0.3266
RUAS *CVPR'21*	15.33	0.4876	0.3097	14.33	0.4841	0.4800
SCI *CVPR'22*	17.30	0.5335	0.3079	15.24	0.4240	0.3218
PairLIE *CVPR'23*	19.88	0.7777	0.2341	17.60	0.5118	0.3290
SADG *AAAI'23*	16.93	0.5372	0.3513	16.32	0.4564	0.3471
CLIP-LIT *ICCV'23*	15.18	0.5290	0.3689	13.47	0.3572	0.4089
NeRCo* *ICCV'23*	19.66	0.7172	0.2705	**19.46**	**0.5506**	0.3052
QuadPrior *CVPR'24*	20.30	0.7909	0.1858	16.90	0.5429	0.3459
ZERO-IG *CVPR'24*	18.13	0.7455	0.2478	16.44	0.5033	0.3744
LightenDiffusion *ECCV'24*	18.27	0.7944	0.2457	16.92	0.5250	0.3824
Li et al. *ICLR'25*	20.22	0.7922	0.2600	16.10	0.5126	0.3928
SPNet *TMM'25*	17.98	0.7900	–	16.63	0.5200	–
GEFU *ICCV'25*	21.32	0.8073	0.1928	18.41	0.5341	0.2974
Ours	**22.17**	**0.8201**	**0.1743**	19.08	0.5422	**0.2871**

Table 2: Comparison on high-level vision tasks: classification on CODaN, detection on DARK FACE, and semantic segmentation on ACDC. Inference time is measured on an RTX 3090 GPU.

Method	CODaN		DARK FACE		ACDC	
	Top-1 (%)↑	Time	mAP (%)↑	Time	mIoU (%)↑	Time
Baseline	53.24	–	10.8	–	11.4	–
EnlightenGAN *TIP'21*	56.42	0.10s	14.2	0.68s	12.7	0.74s
Zero-DCE *CVPR'20*	57.76	0.02s	15.9	0.05s	13.8	0.08s
Zero-DCE++ *TPAMI'21*	59.88	0.04s	15.2	0.42s	15.4	0.08s
RUAS *CVPR'21*	51.60	0.06s	14.0	0.48s	15.8	0.30s
SCI *CVPR'22*	58.84	0.02s	14.7	0.08s	15.6	0.16s
PairLIE *CVPR'23*	52.29	0.02s	16.0	1.04s	15.8	0.13s
SADG *AAAI'23*	56.80	0.02s	14.9	0.05s	14.3	0.08s
CLIP-LIT *ICCV'23*	54.64	0.01s	14.1	0.39s	16.1	0.07s
NeRCo *ICCV'23*	54.15	0.10s	12.4	1.28s	15.6	1.54s
QuadPrior *CVPR'24*	59.48	2.44s	15.7	3.41s	14.4	4.47s
ZERO-IG *CVPR'24*	47.60	0.03s	15.6	0.34s	14.8	0.51s
LightenDiffusion *ECCV'24*	57.40	0.87s	16.3	1.42s	16.2	1.69s
GEFU *ICCV'25*	60.92	0.18s	16.9	0.91s	14.9	0.94s
Ours	**61.32**	0.14s	**17.3**	0.96s	**16.7**	1.15s
Δ *vs. Baseline*	*+8.08*	–	*+6.5*	–	*+5.3*	–

Meanwhile, the text-conditioned branch of SD2.1-Turbo remains frozen and provides a semantic residual Z_{text} through the pretrained text cross-attention pathway. The final conditioning residual $Z_i = Z_{\text{text}} + Z_{\text{ip}}^{\text{dyn}}$ is added to the intermediate UNet feature for degradation-aware semantic modulation. The reverse normal-to-low-light translation follows the same pipeline, using the darkened view as the directional image condition.

4 Experiment

4.1 Experiment Setup

Datasets and implementation details. We train all enhancement models on the unpaired EnlightenGAN (Jiang et al., 2021) dataset, containing 1,014 low-light and 1,102 normal-light images. Low-level performance is evaluated on the official LOL-v2-real test set (100 images) (Yang et al., 2021) and two LSRW subsets (Hai et al., 2023) captured by Huawei (30 images) and Nikon (20 images). For downstream evaluation, enhanced images are fed into fixed task-specific models: ResNet-18 (He et al., 2016) trained on the CODaN (Lengyel et al., 2021) daytime split for classification on its dark test split, YOLOv3 (Redmon & Farhadi, 2018) trained on WIDER FACE (Yang et al., 2016) for detection on DARK FACE (Yang et al., 2020), and RefineNet (Lengyel et al., 2021) trained on Cityscapes (Cordts et al., 2016) for segmentation on the dark ACDC (Sakaridis et al., 2021) validation set. Detailed descriptions of the datasets are provided in the Appendix. Following the evaluation protocol of GEFU (Wang et al., 2025), the enhancement quality is assessed by PSNR, SSIM, and LPIPS, whereas the performance of classification, detection, and segmentation is measured by Top-1 accuracy, mAP, and mIoU, respectively. Due to space limitations, detailed implementation settings are provided in the **Appendix**.

4.2 Comparison with State-of-the-Art Methods

Fifteen unsupervised methods are compared: EnlightenGAN Jiang et al. (2021), Zero-DCE (Guo et al., 2020), Zero-DCE++ (Li et al., 2021), RUAS (Liu et al., 2021), SCI (Ma et al., 2022), PairLIE (Fu et al., 2023), SADG (Zheng et al., 2023), CLIP-LIT (Liang et al., 2023), NeRCo (Yang et al., 2023), QuadPrior (Wang et al., 2024), Zero-IG (Shi et al., 2024), LightenDiffusion (Jiang et al., 2024), Li et al. (Li et al., 2025), SPNet (Luo et al., 2025) and GEFU (Wang et al., 2025).

Quantitative comparison. From Tab. 1 and Tab. 2, we draw the following findings: **1) Our method achieves state-of-the-art overall performance on both datasets.** On LOL, it ranks first on all three metrics, surpassing GEFU by 0.85 dB PSNR and 0.0128 SSIM while reducing LPIPS by 0.0185. On LSRW, it achieves the lowest LPIPS (0.2871) with highly competitive PSNR and SSIM. Although NeRCo reports slightly higher PSNR and SSIM, it is trained directly on LSRW, whereas

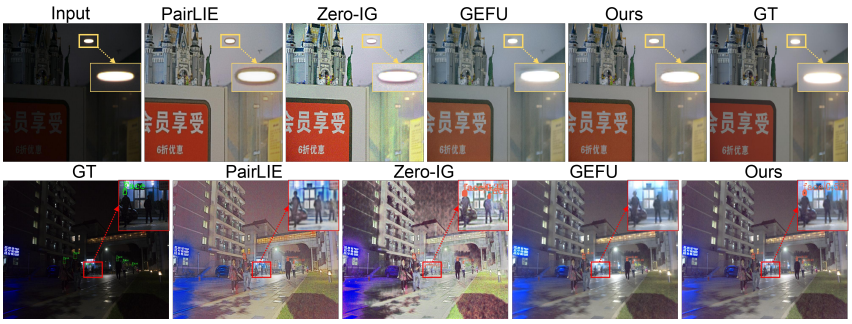


Figure 3: Qualitative comparisons of LLIE (top) and downstream object detection (bottom).

Table 3: Ablation study on the LOL-v2-real dataset.

Module	Setting	PSNR↑	SSIM↑
Overall	w/ IP-adapter	20.08	0.7016
	w/o DPG	21.20	0.7967
DPG	w/o Prior Encoder	21.37	0.8011
	w/o Expert	21.66	0.8033
	w/ Single Expert	21.55	0.7968
	w/o Gating	21.87	0.8022
DDSG	w/o Weight-state Encoding	OOM	OOM
	w/ Constant Weight State	21.62	0.7838
	w/ Layer-ID Embedding	21.49	0.7944
	w/o θ_{state}	21.51	0.7954
	w/o θ_{base}	21.36	0.7888
Full Model		**22.17**	**0.8201**

Table 4: Descriptor ablation.

Descriptor setting	PSNR ↑	SSIM ↑
w/o Input-channel Energy	21.77	0.8023
w/o Frobenius Norm	21.64	0.8007
w/o MAV	21.69	0.8095
w/o Mean Value	21.82	0.8020
Full Descriptors	**22.17**	**0.8201**

Table 5: Fusion-expert ablation.

Number of experts N_e	PSNR ↑	SSIM ↑
2	21.76	0.8074
3	21.94	0.8183
4	**22.17**	**0.8201**
5	22.16	0.8105

our model is trained only on the unpaired EnlightenGAN dataset and evaluated on LSRW without fine-tuning, demonstrating strong cross-dataset generalization. **2) Our method consistently improves downstream high-level vision.** As shown in Tab. 2, it achieves the best results on all three benchmarks, improving classification accuracy by 0.40 points over GEFU, detection mAP by 0.4 points, and segmentation mIoU by 0.5 points over LightenDiffusion. These gains indicate that our framework enhances perceptual quality while preserving task-relevant structures and semantics across diverse low-light scenarios. **3) The dynamic adaptation introduces only modest overhead.** Compared with GEFU, which shares the same backbone, our method maintains comparable inference efficiency while achieving consistently better performance, showing a favorable effectiveness–efficiency trade-off.

Visual analysis. Figure 3 shows representative results for low-level enhancement and downstream detection. Due to space limitations, more comparisons are provided in the **Appendix**. Our method produces more natural illumination and color while reducing ghosting artifacts, as highlighted by the yellow boxes. On DARK FACE, it also reduces missed and false detections of dim and small-scale faces, as indicated by the red boxes, demonstrating improvements in both restoration quality and downstream task reliability.

4.3 Ablation Study and Further Analysis

Overall design. We first evaluate the overall effectiveness of the proposed dynamic adaptation framework (Tab. 3). Replacing DPG and DDSG with the original static image adapter (IP-adapter) reduces PSNR by 2.09 dB, confirming the necessity of sample-specific dynamic conditioning. Removing DPG and directly feeding generic image features to DDSG also degrades performance, demonstrating the benefit of explicit degradation-prior modeling.

DPG design. We further ablate the main components of DPG. In *w/o Prior Encoder*, replacing the multi-view prior encoders with MLPs reduces performance, showing that spatial and multi-scale

encoding is important for extracting discriminative degradation cues. The degradation of *w/o Expert* verifies the benefit of expert-based cross-prior fusion, while *Single Expert* further indicates that a shared fusion pattern cannot adequately model diverse degradation compositions. Although *w/o Gating* outperforms the other DPG variants, it remains inferior to the full model, confirming that input-dependent routing is necessary to select a suitable fusion expert for each image.

DDSG design. For *w/o Weight-state Encoding*, directly encoding the flattened K/V projection matrices causes out-of-memory (OOM) due to the excessive input dimensionality. For *w/ Constant State*, we retain the θ_{state} branch but replace the weight-state embedding with a layer-shared constant vector. We further introduce *w/ Learnable Layer-ID Embedding*, which assigns each projection layer an independent learnable embedding while keeping the branch architecture unchanged. Its performance degradation confirms that the gain comes from projection-weight information rather than additional model capacity. Removing θ_{state} or θ_{base} further verifies the complementary roles of projection-state-aware calibration and degradation-driven adaptation.

We further conduct ablation experiments on the description of weight states within the weight-state encoding block. As reported in Table 4, removing any single descriptor degrades both metrics, confirming that all four components contribute to the weight-state characterization. The Frobenius norm is the most influential, causing the largest PSNR drop of 0.53 dB, which suggests that the global scale of the base projection is the most informative cue for calibrating the low-rank updates; the maximum absolute value (MAV) contributes the least to SSIM, with a reduction of 0.0106. These results support combining channel-wise response energy with complementary global statistics in the weight-state embedding.

Degradation-aware analysis. We verify that the dynamic updates are driven by input degradation rather than image content. On LOL-v2-real, we synthesize multiple degraded variants of each test image with the content fixed, and measure the cosine distance between the induced updates W_V^{θ} of paired variants. For illumination degradation (gamma transformation), the distance increases monotonically with the illumination gap, rising from near zero for small gaps to 0.51 for large ones. For color degradation (HSV saturation scaling), a similar trend is observed, with the distance growing from 0.01 to 0.26. For structure degradation (additive Gaussian noise), the distance again grows monotonically with the noise-level gap, rising from 0.01 to 0.16 for large ones. These results indicate that DPG faithfully perceives illumination, color, and structure degradations and that the induced updates vary accordingly (details in **Appendix**).

Effect of the number of fusion experts. We vary the number of experts N_e in the Dynamic Expert Fusion Block, and train all variants under identical settings. As shown in Table 5, performance varies within a moderate range across expert counts, indicating that DPG is not highly sensitive to N_e. Reducing N_e to two or three degrades both metrics, as fewer experts cannot cover the diverse cross-prior fusion patterns, while increasing N_e to five brings no gain, with nearly identical PSNR but lower SSIM, suggesting that the fusion capacity saturates at four experts. We therefore adopt $N_e = 4$ in the full model.

Due to space limitations, the supplementary material provides detailed formulations of the multi-view degradation priors and degradation compositions, implementation and evaluation details, extensive ablation and statistical analyses across different degradation conditions, comparisons with alternative adaptation methods, additional real-world and qualitative evaluations, and a discussion of limitations and future directions.

5 Conclusion

We presented DDA, which addresses the sample-agnostic condition injection of pretrained diffusion-based LLIE by adapting the image-conditioning pathway to the degradation composition of each input. DPG perceives multi-view degradation priors from the input alone and fuses them adaptively, while DDSG translates the perceived degradation into layer-compatible low-rank updates, allowing a frozen backbone to deliver sample-specific enhancement without paired supervision. Experiments across benchmarks and downstream tasks show consistent gains, and the induced updates vary continuously with input degradation, confirming that the adaptation is genuinely degradation-driven. This work highlights degradation-aware dynamic adaptation as a practical route for deploying pretrained generative priors in real-world low-light image enhancement.

REFERENCES

Eric L. Buehler and Markus J. Buehler. X-lora: Mixture of low-rank adapter experts, a flexible framework for large language models with applications in protein mechanics and molecular design. *APL Machine Learning*, 2(2):026119, 2024.

Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.

Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF conference on CVPR*, pp. 22252–22261, 2023.

Theo Gevers, Arjan Gijsenij, Joost Van de Weijer, and Jan-Mark Geusebroek. *Color in computer vision: fundamentals and applications*. John Wiley & Sons, 2012.

Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on CVPR*, pp. 1780–1789, 2020.

Xiaojie Guo, Yu Li, and Haibin Ling. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017.

Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90:103712, 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Xiao He, Zhijun Tu, Kun Cheng, Mingrui Zhu, Jie Hu, Nannan Wang, and Xinbo Gao. Mixture of ranks with degradation-aware routing for one-step real-world image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 4681–4689, 2026.

International Telecommunication Union. Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. Recommendation ITU-R BT.601-7, 2011.

International Telecommunication Union. Parameter values for the HDTV standards for production and international programme exchange. Recommendation ITU-R BT.709-6, 2015.

Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *ECCV*, pp. 161–179. Springer, 2024.

Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021.

Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Transactions on Image Processing*, 22(12):5372–5384, 2013. doi: 10.1109/TIP.2013.2284059.

Attila Lengyel, Sourav Garg, Michael Milford, and Jan C Van Gemert. Zero-shot day-night domain adaptation with a physics prior. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4399–4409, 2021.

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

Huaqiu Li, Xiaowan Hu, and Haoqian Wang. Interpretable unsupervised joint denoising and enhancement for real-world low-light scenarios. *The Thirteenth International Conference on Learning Representations*, 2025.

Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF ICCV*, pp. 8094–8103, 2023.

Yunlong Lin, Tian Ye, Sixiang Chen, Zhenqi Fu, Yingying Wang, Wenhao Chai, Zhaohu Xing, Wenxue Li, Lei Zhu, and Xinghao Ding. Aglldiff: Guiding diffusion models towards unsupervised training-free real-world low-light image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5307–5315, 2025.

Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on CVPR*, pp. 10561–10570, 2021.

Yu Luo, Xuanrong Chen, Jie Ling, Chao Huang, Wei Zhou, and Guanghui Yue. Unsupervised low-light image enhancement with self-paced learning. *IEEE Transactions on Multimedia*, 27: 1808–1820, 2025. doi: 10.1109/TMM.2024.3521752.

Chuancheng Lv, Lei Li, Shitou Zhang, Gang Chen, Fanchao Qi, Ningyu Zhang, and Hai-Tao Zheng. Hyperlora: Efficient cross-task generalization via constrained low-rank adapters generation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 16376–16393, 2024.

Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. doi: 10.1109/TIP.2015.2442920.

Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on CVPR*, pp. 5637–5646, 2022.

Henrique S. Malvar and Gary J. Sullivan. YCoCg-R: A color space with RGB reversibility and low dynamic range. In *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Doc. JVT-I014r3*, 2003.

Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-step image translation with text-to-image models. *arXiv preprint arXiv:2403.12036*, 2024.

Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. Adapterfusion: Non-destructive task composition for transfer learning. pp. 487–503, 2021.

Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022.

Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10765–10775, 2021.

Yiqi Shi, Duo Liu, Liguo Zhang, Ye Tian, Xuezhi Xia, and Xiaojing Fu. Zero-ig: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images. In *Proceedings of the IEEE/CVF conference on CVPR*, pp. 3015–3024, 2024.

Koen E. A. van de Sande, Theo Gevers, and Cees G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010.

Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77(8):9211–9231, 2018. doi: 10.1007/s11042-017-4783-x.

Sen Wang, Shao Zeng, Tianjun Gu, Zhizhong Zhang, Ruixin Zhang, Shouhong Ding, Jingyun Zhang, Jun Wang, Xin Tan, Yuan Xie, et al. From enhancement to understanding: Build a generalized bridge for low-light vision via semantically consistent unsupervised fine-tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13804–13814, 2025.

Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. doi: 10.1109/TIP.2013.2261309.

Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *Proceedings of the IEEE/CVF conference on CVPR*, pp. 26057–26066, 2024.

Joachim Weickert. *Anisotropic Diffusion in Image Processing*. B. G. Teubner, 1998.

Xun Wu, Shaohan Huang, and Furu Wei. Mixture of lora experts. *arXiv preprint arXiv:2404.13628*, 2024.

Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia. Structure extraction from texture via relative total variation. *ACM Transactions on Graphics*, 31(6):1–10, 2012.

Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5525–5533, 2016.

Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF ICCV*, pp. 12918–12927, 2023.

Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29:5737–5752, 2020.

Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021.

Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.

Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12302–12311, 2023.

Tongshun Zhang, Pingping Liu, Zijian Zhang, and Qiuzhan Zhou. Spjfnet: self-mining prior-guided joint frequency enhancement for ultra-efficient dark image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12798–12806, 2026a.

Tongshun Zhang, Pingping Liu, Zixuan Zhong, Zijian Zhang, and Qiuzhan Zhou. Beyond illumination: Fine-grained detail preservation in extreme dark image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12789–12797, 2026b.

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pp. 1632–1640, 2019.

Naishan Zheng, Jie Huang, Man Zhou, Zizheng Yang, Qi Zhu, and Feng Zhao. Learning semantic degradation-aware guidance for recognition-driven unsupervised low-light image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 3678–3686, 2023.

Xiyu Zhu, Wei Wang, Xin Yuan, and Xiao Wang. Slcformer: Spectral-local context transformer with physics-grounded flare synthesis for nighttime flare removal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 13988–13996, 2026.

AI USE STATEMENT

During the preparation of this manuscript, the authors used large language models (LLMs), including GPT-5 and Kimi, for (i) language polishing and editing of the text (e.g., grammar checking, phrasing improvements, and stylistic refinement) and (ii) code implementation assistance (e.g., writing utility scripts and debugging). The LLMs were not used for any other substantive research tasks, including but not limited to generating research ideas, designing experiments, deriving proofs, analyzing data, or interpreting results. All code was reviewed, tested, and verified by the authors, and all technical content, scientific claims, and conclusions were conceived, produced, and validated by the authors, who take full responsibility for the accuracy, originality, and integrity of this work.

A APPENDIX

This supplementary material provides a comprehensive account of the proposed framework, including detailed formulations of the *multi-view degradation priors*, the computation of *degradation compositions*, and *implementation details*. It also presents dataset and evaluation protocols, extensive *ablation studies*, *degradation-aware* and *statistical significance analyses*, a *comparison with alternative dynamic adaptation methods*, additional *evaluations on real-world low-light datasets*, additional *qualitative results*, and a discussion of *limitations and future directions*. Together, these materials provide further methodological clarification and experimental evidence to support the findings reported in the main paper.

B ADDITIONAL DETAILS OF THE PROPOSED METHOD

B.1 DETAILS OF MULTI-VIEW PRIOR CONSTRUCTION

This section details the view-construction functions $\phi_m^t(\cdot)$ introduced in the multi-view prior extraction block of the main paper. Given an input image $x \in \mathbb{R}^{3\times H\times W}$ with red, green, and blue channels R, G, and B normalized to $[0, 1]$, we construct the illumination, color, and structure priors as follows. For each degradation type t, the views are listed in the order of $\phi_1^t, \phi_2^t, \ldots, \phi_{M_t}^t$, and all view maps preserve the spatial resolution of the input. Since the views have heterogeneous value ranges (e.g., signed chrominance responses versus normalized ratios), each view map is linearly rescaled to $[0, 1]$ before channel-wise concatenation. Unless otherwise specified, $\varepsilon = 10^{-6}$ is a small constant for numerical stability.

B.2 SELECTION BOUNDARY

The adopted prior set is not intended to exhaust all possible handcrafted image descriptors. Instead, we restrict the selection to priors that satisfy three requirements: they should characterize a distinct manifestation of low-light degradation, preserve spatial correspondence with the input image, and be computed deterministically with limited overhead. Under this boundary, the illumination views describe complementary intensity and luminance responses, the color views isolate chromatic variations under different standardized color representations, and the structure views characterize first-order edges, second-order responses, local regularity, geometric structure, and contrast. Although individual views within each type may be partially correlated (e.g., the YUV and YCbCr chrominance responses), they differ in their chrominance axes and dynamic ranges, providing complementary sensitivities to chromatic distortions. This design avoids an unrestricted expansion of re-

dundant descriptors while retaining complementary degradation cues that can be directly processed by the degradation-aware prior generation (DPG) module.

B.3 Illumination Prior

The illumination prior comprises $M_{\text{illum}} = 6$ views: mean intensity I_{mean}, Rec. 709 luminance I_{rec} (International Telecommunication Union, 2015), maximum-channel response $I_{\max}$ (Guo et al., 2017), HSL lightness I_{light}, YCgCo luminance I_{ycgco} (Malvar & Sullivan, 2003), and channel-energy response I_{l2}:

$$P_{\text{illum}} = \text{Concat}\left[I_{\text{mean}}, I_{\text{rec}}, I_{\max}, I_{\text{light}}, I_{\text{ycgco}}, I_{\text{l2}}\right]. \tag{10}$$

Let $R, G, B \in [0, 1]$ denote the red, green, and blue channel intensities of the input image, respectively. The six views are defined as

$$I_{\text{mean}} = \frac{R+G+B}{3}, \tag{11}$$

$$I_{\text{rec}} = 0.2126R + 0.7152G + 0.0722B, \tag{12}$$

$$I_{\max} = \max\{R, G, B\}, \tag{13}$$

$$I_{\text{light}} = \frac{\max\{R, G, B\} + \min\{R, G, B\}}{2}, \tag{14}$$

$$I_{\text{ycgco}} = \frac{1}{4}R + \frac{1}{2}G + \frac{1}{4}B, \tag{15}$$

$$I_{\text{l2}} = \sqrt{\frac{R^2 + G^2 + B^2}{3}}. \tag{16}$$

Together, these complementary views cover global, perceptual, and channel-dependent illumination statistics.

B.4 Color Prior

The color prior comprises $M_{\text{color}} = 5$ views: YUV chrominance C_{YUV} (International Telecommunication Union, 2011), YCbCr chrominance C_{YCbCr} (International Telecommunication Union, 2011), opponent-color representation C_{opp} (van de Sande et al., 2010), normalized chromaticity C_{rg} (Gevers et al., 2012), and saturation C_{sat}:

$$P_{\text{color}} = \text{Concat}\left[C_{\text{YUV}}, C_{\text{YCbCr}}, C_{\text{opp}}, C_{\text{rg}}, C_{\text{sat}}\right]. \tag{17}$$

The YUV chrominance view is defined as

$$C_{\text{YUV}} = [U, V], \quad \begin{aligned} U &= -0.147R - 0.289G + 0.436B, \\ V &= 0.615R - 0.515G - 0.100B. \end{aligned} \tag{18}$$

The YCbCr chrominance view is defined as

$$C_{\text{YCbCr}} = [C_b, C_r], \quad \begin{aligned} C_b &= -0.169R - 0.331G + 0.500B, \\ C_r &= 0.500R - 0.419G - 0.081B. \end{aligned} \tag{19}$$

The remaining three views are defined as

$$C_{\text{opp}} = \left[\frac{R-G}{\sqrt{2}}, \frac{R+G-2B}{\sqrt{6}}\right], \tag{20}$$

$$C_{\text{rg}} = \left[\frac{R}{I_\Sigma + \varepsilon}, \frac{G}{I_\Sigma + \varepsilon}\right], \quad I_\Sigma = R + G + B, \tag{21}$$

$$C_{\text{sat}} = \frac{\max\{R, G, B\} - \min\{R, G, B\}}{\max\{R, G, B\} + \varepsilon}. \tag{22}$$

The blue chromaticity channel is omitted in C_{rg} because it is approximately determined by the red and green components. The YUV, YCbCr, and opponent-color views preserve the direction of chromatic deviations, whereas normalized chromaticity and saturation suppress the influence of absolute illumination.

B.5 Structure Prior

We construct the structure prior from the Rec. 709 luminance map $I \triangleq I_{\mathrm{rec}}$ in Eq. equation 12. It comprises $M_{\mathrm{struct}} = 5$ views: gradient magnitude S_{grad}, Laplacian response S_{lap}, relative total variation S_{rtv} (Xu et al., 2012), local structure-tensor response S_{eig} (Weickert, 1998), and local contrast S_{contrast}:

$$P_{\mathrm{struct}} = \mathrm{Concat}\left[S_{\mathrm{grad}}, S_{\mathrm{lap}}, S_{\mathrm{rtv}}, S_{\mathrm{eig}}, S_{\mathrm{contrast}}\right]. \tag{23}$$

The gradient-magnitude view is defined as

$$S_{\mathrm{grad}} = \sqrt{g_x^2 + g_y^2 + \varepsilon}, \qquad g_x = K_x * I, \;\; g_y = K_y * I, \tag{24}$$

where $*$ denotes convolution, and K_x and K_y are the normalized 3×3 Sobel kernels:

$$K_x = \frac{1}{8}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad K_y = \frac{1}{8}\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

The Laplacian-response view is defined as

$$S_{\mathrm{lap}} = \left|K_{\mathrm{lap}} * I\right|, \qquad K_{\mathrm{lap}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{25}$$

For the remaining views, let $\mathcal{N}(p)$ denote a local neighborhood centered at pixel p. The relative-total-variation view (Xu et al., 2012) is defined as

$$S_{\mathrm{rtv}}(p) = \frac{D_x(p)}{L_x(p) + \varepsilon} + \frac{D_y(p)}{L_y(p) + \varepsilon}, \tag{26}$$

where, for $\delta \in \{x, y\}$,

$$\begin{aligned} w_{p,q} &= \frac{\exp\left(-\|p - q\|_2^2/(2\sigma_r^2)\right)}{\sum_{q' \in \mathcal{N}(p)} \exp\left(-\|p - q'\|_2^2/(2\sigma_r^2)\right)}, \\ D_\delta(p) &= \sum_{q \in \mathcal{N}(p)} w_{p,q} |g_\delta(q)|, \\ L_\delta(p) &= \left| \sum_{q \in \mathcal{N}(p)} w_{p,q} g_\delta(q) \right|. \end{aligned}$$

The structure-tensor response is computed from the smoothed tensor

$$J = G_{\sigma_s} * \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{bmatrix},$$

where G_{σ_s} is a Gaussian smoothing kernel. The eigenvalue-ratio view is then defined as

$$S_{\mathrm{eig}} = \frac{\lambda_2}{\lambda_1 + \varepsilon}, \qquad \lambda_{1,2} = \frac{T \pm \sqrt{(J_{11} - J_{22})^2 + 4J_{12}^2}}{2}, \tag{27}$$

where $T = J_{11} + J_{22}$ is the trace of J, and the plus and minus signs correspond to λ_1 and λ_2, respectively, with $\lambda_1 \geq \lambda_2 \geq 0$. This response characterizes local structural isotropy: it is small for strongly oriented structures and becomes larger when substantial variations occur along multiple directions.

Finally, the local-contrast view is defined as

$$S_{\mathrm{contrast}}(p) = \frac{|I(p) - \mu_{\mathcal{N}}(I)(p)|}{\mu_{\mathcal{N}}(I)(p) + \varepsilon}, \tag{28}$$

where $\mu_{\mathcal{N}}(I)(p)$ denotes the mean luminance within the local neighborhood centered at p.

The illumination, color, and structure priors thus comprise $M_{\mathrm{illum}} = 6$, $M_{\mathrm{color}} = 5$, and $M_{\mathrm{struct}} = 5$ views, and the concatenated prior tensors have $C_{\mathrm{illum}} = 6$, $C_{\mathrm{color}} = 9$, and $C_{\mathrm{struct}} = 5$ channels, consistent with the definitions of M_t and C_t in the main paper. Each two-channel chromatic view is counted as a single view in M_t, while both of its channels contribute to C_t. Table 6 summarizes all views.

Table 6: Summary of the multi-view priors in the degradation-aware prior generation module. All operations are performed pixel-wise unless a local operator is explicitly specified.

Prior type	View	Symbol	Definition
Illumination	Mean intensity	I_{mean}	$(R+G+B)/3$
	Rec. 709 luminance	I_{rec}	$0.2126R + 0.7152G + 0.0722B$
	Maximum-channel response	$I_{\max}$	$\max\{R, G, B\}$
	Lightness	I_{light}	$(\max\{R, G, B\} + \min\{R, G, B\})/2$
	YCgCo luminance	I_{ycgco}	$R/4 + G/2 + B/4$
	Channel-energy response	I_{l2}	$\sqrt{(R^2 + G^2 + B^2)/3}$
Color	YUV chrominance	C_{YUV}	$[U, V]$, $U = -0.147R - 0.289G + 0.436B$, $V = 0.615R - 0.515G - 0.100B$
	YCbCr chrominance	C_{YCbCr}	$[C_b, C_r]$, $C_b = -0.169R - 0.331G + 0.500B$, $C_r = 0.500R - 0.419G - 0.081B$
	Opponent-color representation	C_{opp}	$[(R-G)/\sqrt{2},\ (R+G-2B)/\sqrt{6}]$
	Normalized chromaticity	C_{rg}	$[R/(I_\Sigma + \varepsilon),\ G/(I_\Sigma + \varepsilon)]$, $I_\Sigma = R+G+B$
	Saturation	C_{sat}	$(\max\{R, G, B\} - \min\{R, G, B\})/(\max\{R, G, B\} + \varepsilon)$
Structure	Gradient magnitude	S_{grad}	$\sqrt{g_x^2 + g_y^2 + \varepsilon}$, $g_x = K_x * I$, $g_y = K_y * I$
	Laplacian response	S_{lap}	$\lvert K_{\text{lap}} * I \rvert$
	Relative total variation	S_{rtv}	$D_x/(L_x + \varepsilon) + D_y/(L_y + \varepsilon)$
	Local structure-tensor response	S_{eig}	$\lambda_2/(\lambda_1 + \varepsilon)$, $\lambda_1 \geq \lambda_2$: eigenvalues of J
	Local contrast	S_{contrast}	$\lvert I - \mu_{\mathcal{N}}(I) \rvert / (\mu_{\mathcal{N}}(I) + \varepsilon)$

B.6 Computation of Degradation Proportions

We analyze all $N = 100$ paired test images from LOL-v2-real, with the resulting statistics summarized in Fig. 4. Let x_i and y_i denote the i-th low-light image and its normal-light reference, respectively. Before prior extraction, each image is converted to RGB with values in $[0, 1]$ and resized to 256×256.

For each degradation type $t \in \mathcal{T} = \{\text{illum}, \text{color}, \text{struct}\}$, we construct the prior tensors $P_t(x_i)$ and $P_t(y_i)$, where C_t denotes the total number of concatenated prior channels and $[\cdot]_c$ denotes the c-th channel. The raw discrepancy between the low-light image and its reference in the corresponding prior space is defined as

$$D_i^t = \frac{1}{C_t HW} \sum_{c=1}^{C_t} \left\| [P_t(x_i)]_c - [P_t(y_i)]_c \right\|_1, \tag{29}$$

where H and W are the spatial dimensions after resizing, and the ℓ_1 norm is taken over all spatial locations. The absolute discrepancy is computed separately for each channel before averaging, thereby preventing differences of opposite signs across channels from canceling and assigning equal weight to each concatenated prior channel. Accordingly, the color prior is evaluated over its nine concatenated channels rather than its five semantic views.

Since the three prior types have different response scales, the raw discrepancies are calibrated by their dataset-level means:

$$\mu_t = \max\left(\frac{1}{N} \sum_{j=1}^{N} D_j^t,\ \epsilon \right), \qquad \widetilde{D}_i^t = \frac{D_i^t}{\mu_t}, \tag{30}$$

where ϵ is used for numerical stability. The relative proportion of degradation type t in image i is then defined as

$$r_i^t = \frac{\widetilde{D}_i^t}{\max\left(\sum_{t' \in \mathcal{T}} \widetilde{D}_i^{t'},\ \epsilon \right)}. \tag{31}$$

The three proportions of each image thus sum to one; we refer to $\{r_i^t\}_{t \in \mathcal{T}}$ as the degradation composition of image i in the main paper. Its dominant degradation mode is determined by

$$t_i^\star = \arg\max_{t \in \mathcal{T}} r_i^t. \tag{32}$$

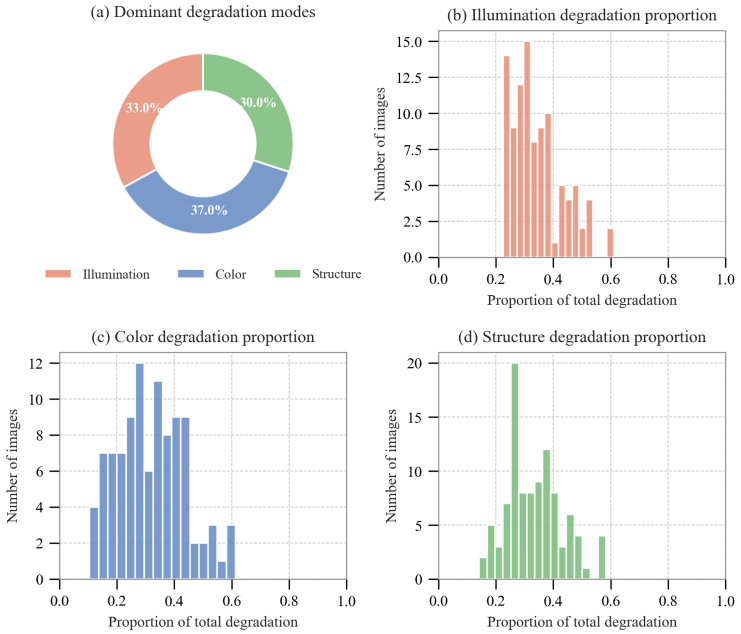


Figure 4: Statistical analysis of degradation compositions on the LOL-v2-real test set (100 images). (a) Frequency of each dominant degradation mode (Eq. 33). (b–d) Histograms of the per-image degradation proportions r_i^t (Eq. 31) for illumination, color, and structure; the horizontal axis denotes the proportion of total degradation.

Fig. 4(a) reports the frequency of each dominant mode:

$$\pi_t = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}[t_i^\star = t]\,, \tag{33}$$

where $\mathbb{I}[\cdot]$ is the indicator function. Fig. 4(b–d) further visualize the distributions of $\{r_i^t\}_{i=1}^N$ for illumination, color, and structure, respectively, each using 16 equal-width bins over $[0, 1]$ and reporting image counts. The wide spread of these distributions confirms that degradation compositions vary substantially across real low-light images. The paired normal-light references are used only for this offline statistical analysis, not as supervision during unpaired training.

B.7 Implementation Details

We build the proposed model upon the CycleGAN-Turbo training framework (Parmar et al., 2024) and adopt the backbone configuration used in GEFU (Wang et al., 2025). Skip connections are introduced between the corresponding encoder and decoder stages to retain fine-grained spatial information during translation. The pretrained SD-Turbo backbone is kept frozen throughout training; only the degradation-aware prior generation module, the degradation-driven state-glimpse adapter, and the trainable LoRA adapters in the VAE encoder and decoder are optimized. Adversarial supervision is provided by two CLIP-based discriminators following the Vision-Aided GAN design (Parmar et al., 2024).

Following CycleGAN-Turbo (Parmar et al., 2024), the total loss is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cycle}} + \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}} + \lambda_{\text{idt}}\mathcal{L}_{\text{idt}}, \tag{34}$$

where the cycle-consistency loss enforces round-trip reconstruction in both translation directions with the shared one-step translator, the adversarial loss $\mathcal{L}_{\text{GAN}}$ is computed with the vision-aided discriminators on the translated outputs of both directions, and the identity loss $\mathcal{L}_{\text{idt}}$ regularizes the translator to preserve inputs from both domains. All reconstruction terms combine an L1 loss with an LPIPS-based perceptual loss (Parmar et al., 2024): the L1 weights of the cycle and identity terms

Algorithm 1 Forward pass of DDA

Require: Input image x; translation direction $u \in \{\text{light}, \text{dark}\}$; degradation-prior embedding module DPG; base image-adapter projections W_k^{ip} and W_v^{ip}

function DDAFORWARD(x, u)
  Construct P_{illum}, P_{color}, and P_{struct} from x
  $c_x \leftarrow \text{DPG}\big(P_{\text{illum}}, P_{\text{color}}, P_{\text{struct}}\big)$
  if $u = \text{light}$ **then**
    $x^u \leftarrow x^{1/2.2}$
  else
    $x^u \leftarrow x^{1.2}$
  end if
  $z_{\text{ip}} \leftarrow \text{LN}\big(\text{MLP}(E_{\text{clip}}(x^u))\big)$
  Encode x into the latent space using the VAE encoder
  Execute the one-step UNet forward pass
  for each cross-attention layer i equipped with DDSG **do**
    Encode W_k^{ip} and W_v^{ip} into s_k and s_v
    $s_w \leftarrow \text{MSM}\big(\text{Concat}(s_k, s_v)\big)$
    $\theta_{\text{base}} \leftarrow \text{MSM}(c_x)$
    $\theta_{\text{state}} \leftarrow \text{MSM}\big(\text{Concat}(c_x, s_w)\big)$
    $\theta \leftarrow \theta_{\text{base}} + \theta_{\text{state}}$
    Reshape θ into $\{B_k, A_k, B_v, A_v\}$
    $W_k^{\text{ip,dyn}} \leftarrow W_k^{\text{ip}} + B_k A_k, \quad W_v^{\text{ip,dyn}} \leftarrow W_v^{\text{ip}} + B_v A_v$
    $K_{\text{ip}}^{\text{dyn}} \leftarrow z_{\text{ip}}\big(W_k^{\text{ip,dyn}}\big)^\top, \quad V_{\text{ip}}^{\text{dyn}} \leftarrow z_{\text{ip}}\big(W_v^{\text{ip,dyn}}\big)^\top$
    Compute $Z_{\text{ip}}^{\text{dyn}}$ using $K_{\text{ip}}^{\text{dyn}}$ and $V_{\text{ip}}^{\text{dyn}}$
    $Z_i \leftarrow Z_{\text{text}} + Z_{\text{ip}}^{\text{dyn}}$
    Add Z_i to the intermediate UNet feature
  end for
  Decode the denoised latent using the VAE decoder
  return the translated image
end function

are 1, and the corresponding LPIPS weights are 10 and 1, respectively. We set $\lambda_{\text{GAN}} = 0.5$ and $\lambda_{\text{idt}} = 1$ for all experiments.

For data augmentation, each training image is randomly flipped horizontally and randomly cropped to 256×256. We optimize the trainable parameters using AdamW with a fixed learning rate of 1×10^{-5} and a batch size of 1. Our model is implemented with PyTorch 2.0.1 under Python 3.10 and trained for 50K iterations on a single NVIDIA RTX 4090 GPU. We fix the random seed to 42, consistent with GEFU, and the reported results are reproducible under the same environment across multiple GPUs.

B.8 Optimization and Inference Procedure

Algorithm 1 details the forward pass of DDA, and Algorithm 2 summarizes the bidirectional optimization and inference procedures. During training, low-light and normal-light images are sampled independently to perform bidirectional translation. During inference, only the low-to-normal branch is retained.

C Additional Experiments

C.1 Dataset Details

Low-level vision datasets. The enhancement model is trained using the EnlightenGAN dataset (Jiang et al., 2021), which provides two independently collected domains containing 1,014 low-light images and 1,102 normal-light images. For evaluation, we use the official test split of LOL-v2-real (Yang et al., 2021), consisting of 100 low-light images and their corresponding normal-light references. The reference images are used exclusively for quantitative evaluation and do not

Algorithm 2 Training and inference procedures of DDA

Require: Unpaired low-light and normal-light domains $\mathcal{X}_l$ and $\mathcal{X}_n$; pretrained one-step diffusion backbone (frozen)

procedure TRAINING
  for each training iteration **do**
    Sample $x \sim \mathcal{X}_l$ and $y \sim \mathcal{X}_n$
    $\hat{y} \leftarrow \text{DDAFORWARD}(x, \text{light})$
    $\tilde{x} \leftarrow \text{DDAFORWARD}(\hat{y}, \text{dark})$
    $\hat{x} \leftarrow \text{DDAFORWARD}(y, \text{dark})$
    $\tilde{y} \leftarrow \text{DDAFORWARD}(\hat{x}, \text{light})$
    Compute $\mathcal{L}_{\text{GAN}}$, $\mathcal{L}_{\text{cycle}}$, and $\mathcal{L}_{\text{idt}}$ following CycleGAN-Turbo
    Update DPG, DDSG, and the LoRA adapters using $\mathcal{L}_{\text{total}}$
  end for
end procedure

procedure INFERENCE(x)
  $\hat{y} \leftarrow \text{DDAFORWARD}(x, \text{light})$
  return $\hat{y}$
end procedure

provide paired supervision during optimization. We further evaluate cross-dataset performance on the Huawei and Nikon subsets of LSRW (Hai et al., 2023), which contain 30 and 20 test images, respectively. These two subsets are captured using different camera devices and are evaluated directly by the model trained on EnlightenGAN, without additional training or fine-tuning on LSRW. PSNR, SSIM, and LPIPS are computed against the corresponding normal-light references to measure distortion-based, structural, and perceptual enhancement quality.

Downstream task datasets. For downstream recognition, we consider image classification, face detection, and semantic segmentation under low-illumination conditions. CODaN (Lengyel et al., 2021) contains 10,000 daytime training images from 10 semantic categories, together with a daytime test split and a dark test split of 2,500 images each. Following its day-to-night evaluation setting, ResNet-18 (He et al., 2016) is trained only on the daytime training split and tested on enhanced images from the dark split, with Top-1 accuracy as the evaluation metric. DARK FACE (Yang et al., 2020) contains 10,000 face images collected in real low-light environments. Following the adopted evaluation protocol, 600 images are used for testing, while YOLOv3 (Redmon & Farhadi, 2018), trained on WIDER FACE (Yang et al., 2016), is kept fixed and evaluated using mAP at an IoU threshold of 0.5. For semantic segmentation, we use the dark validation split of ACDC (Sakaridis et al., 2021) and a RefineNet model trained on Cityscapes (Cordts et al., 2016); segmentation accuracy is reported using mIoU. In all three evaluations, the enhancement output is directly passed to the corresponding pretrained task model, and neither the enhancement network nor the downstream model is adapted on the target test set.

C.2 Ablation of Multi-view Degradation Priors

Table 7 reports both prior-type and view-level ablations on LOL-v2-real. The full model achieves a PSNR of 22.17 dB and an SSIM of 0.8201. For each view-level variant, only the corresponding view is removed, while all remaining priors and network components are kept unchanged.

Effect of the three prior types. Removing any complete prior type causes a clear performance degradation, indicating that illumination, color, and structure cues play complementary roles. Removing the illumination prior reduces the performance to 21.20 dB and 0.7926, while removing the color prior produces the largest overall deterioration, reaching only 21.16 dB and 0.7601. Without the structure prior, the model obtains 21.45 dB and 0.7981. Although the structure removal causes a relatively smaller decrease than the other two variants, its performance remains substantially below that of the full model. These results indicate that no individual prior type alone suffices to characterize the heterogeneous degradations encountered in real low-light images.

Table 7: Ablation study of multi-view degradation priors on the LOL-v2-real test set. Each "w/o" variant removes only the specified prior or semantic view while keeping all other components unchanged. The full model retains all illumination, color, and structure views. Best results are shown in bold.

Prior-type and illumination-view ablations				**Color- and structure-view ablations**			
Group	Variant	PSNR ↑	SSIM ↑	Group	Variant	PSNR ↑	SSIM ↑
Prior type	w/o P_{illum}	21.20	0.7926	Color	w/o C_{rg}	21.78	0.8058
	w/o P_{color}	21.16	0.7601		w/o C_{YCbCr}	21.87	0.8076
	w/o P_{struct}	21.45	0.7981		w/o C_{YUV}	21.88	0.8073
Illumination	w/o I_{mean}	22.16	0.8001		w/o C_{opp}	21.88	0.8062
	w/o I_{rec}	21.96	0.8049		w/o C_{sat}	21.70	0.7956
	w/o I_{max}	22.05	0.8102	Structure	w/o S_{grad}	21.69	0.8069
	w/o I_{light}	21.89	0.8048		w/o S_{lap}	21.99	0.8128
	w/o I_{ycgco}	22.00	0.7941		w/o S_{rtv}	21.77	0.8064
	w/o I_{l2}	21.93	0.8130		w/o S_{eig}	22.02	0.8192
					w/o S_{contrast}	21.61	0.8034
Full model		PSNR: **22.17**		SSIM: **0.8201**			

Illumination priors. The illumination views capture different definitions of image brightness and channel response. Among them, removing I_{light} produces the largest PSNR decrease, reducing the result to 21.89 dB. Removing I_{ycgco} gives the lowest SSIM of 0.7941, suggesting that its luminance response is not fully compensated by the other intensity views. The effects of I_{rec}, I_{max}, and I_{l2} are also evident, with their removal resulting in PSNR values of 21.96, 22.05, and 21.93 dB, respectively. Notably, removing I_{mean} only slightly changes PSNR (22.16 dB) but reduces SSIM to 0.8001. The different responses of PSNR and SSIM indicate that these illumination statistics are partially correlated but not interchangeable, motivating their joint use rather than selecting a single brightness measure.

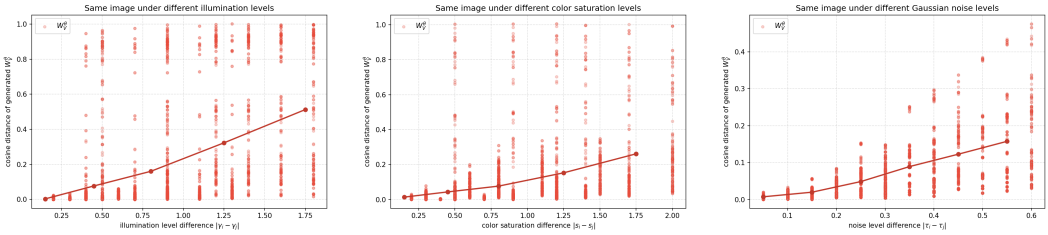
Color priors. The color-prior ablations show that multiple chromatic representations are required to characterize the diverse color casts in real low-light images. Removing C_{sat} leads to the largest degradation within this group, yielding 21.70 dB and 0.7956. Removing normalized chromaticity C_{rg} also causes a pronounced PSNR reduction to 21.78 dB. The YUV, YCbCr, and opponent-color views produce similar but consistently lower results when removed: $21.88/0.8073$, $21.87/0.8076$, and $21.88/0.8062$, respectively. Their relatively close ablation results reflect shared chrominance information, whereas their consistent deterioration relative to the full model shows that the corresponding representations remain complementary rather than strictly redundant.

Structure priors. The structure views characterize different aspects of edge visibility, local detail, and spatial regularity. Removing local contrast S_{contrast} causes the largest performance decrease in this group, resulting in 21.61 dB and 0.8034, and removing gradient magnitude S_{grad} leads to the second-largest drop, reducing the performance to 21.69 dB and 0.8069. Removing S_{rtv} yields 21.77 dB and 0.8064. The effects of the Laplacian and structure-tensor views are comparatively smaller but remain measurable: removing S_{lap} and S_{eig} gives $21.99/0.8128$ and $22.02/0.8192$, respectively.

Overall, the ablations support the adopted design boundary: the selected views form a compact set that covers complementary degradation characteristics, while indiscriminately introducing additional descriptors would increase redundancy and computational cost without demonstrated benefit. We therefore do not claim that the chosen set is globally exhaustive; rather, it provides an effective and reproducible balance between degradation coverage, complementarity, and efficiency.

C.3 Degradation-Aware Analysis

To verify that the dynamic updates of DDA are genuinely driven by input degradation rather than image content, we conduct controlled analyses on the LOL-v2-real test set. Specifically, we synthesize multiple variants of each test image by applying one controlled degradation perturbation



(a) Illumination degradation. (b) Color degradation. (c) Structural degradation.

Figure 5: Degradation-aware adaptation of DDA-generated V-projection updates W_V^θ under different degradation levels. Each point represents a (degradation-level pair, layer) sample, and the curves show the mean cosine distance per degradation-gap bin. (a) Gamma-induced illumination variations. (b) HSV saturation scaling variations. (c) Additive Gaussian noise variations. The cosine distance increases monotonically with the degradation gap while image content remains fixed, demonstrating degradation-aware adaptation to illumination, color, and structural degradation.

Table 8: Paired statistical tests on LOL-v2-real against GEFU. ↑/↓ indicate higher/lower is better.

Metric	Diff.	95% CI	d	p-value
PSNR↑	$+0.85$	$[+0.37, +1.33]$	$+0.354$	1.5×10^{-4}
SSIM↑	$+0.0128$	$[+0.0092, +0.0165]$	$+0.702$	4.7×10^{-9}
LPIPS↓	-0.0185	$[-0.0220, -0.0150]$	-1.046	8.4×10^{-14}

at a time while preserving the underlying scene content: gamma transformations with different γ values ($\gamma < 1$ brightens, $\gamma > 1$ darkens) for illumination, HSV saturation scaling with factors $s \in \{0.2, 0.5, 0.8, 1.0, 1.3, 1.7, 2.2\}$ for color, and additive Gaussian noise with standard deviation $\tau \in \{0, 0.05, 0.10, 0.20, 0.30, 0.45, 0.60\}$ (normalized to the intensity range $[0, 1]$) for structure. Each variant is fed into DDA, and the induced low-rank updates W_V^θ are extracted at every adapter layer. We then measure the pairwise cosine distance between updates induced by different variants of the same image. The K-projection updates W_K^θ exhibit the same trend and are omitted for clarity.

Illumination-driven adaptation. Fig. 5(a) presents the illumination analysis, where the horizontal axis measures the gamma gap $|\Delta\gamma| = |\gamma_i - \gamma_j|$ and the vertical axis measures the cosine distance between the induced updates W_V^θ. The binned mean distance increases monotonically from 0.08 for $|\Delta\gamma| \in [0.3, 0.6]$ to 0.51 for $|\Delta\gamma| \in [1.5, 2.0]$. The larger distances and wider variation at large gaps indicate that the generated updates respond continuously to illumination degradation in an image- and layer-dependent manner.

Color-driven adaptation. Fig. 5(b) presents the same analysis for color, with the horizontal axis measuring the saturation gap $|\Delta s| = |s_i - s_j|$. A similar monotonic trend is observed: the binned mean distance grows from 0.014 for small gaps to 0.26 for large ones, indicating that the update space is also continuously organized by color degradation.

Structure-driven adaptation. Fig. 5(c) presents the same analysis for structure. We use additive Gaussian noise as a controlled form of structural degradation: increasing the noise level progressively disrupts local spatial correlations, obscuring edges, contours, and texture patterns, while preserving the underlying image content. Accordingly, we compare pairs of noise levels (τ_i, τ_j) and measure their gap as $|\Delta\tau| = |\tau_i - \tau_j|$. The binned mean cosine distance increases monotonically with this gap, from 0.007 for small gaps to 0.16 for large ones, indicating that the update space continuously adapts to the degree of structural degradation.

By comparing variants of the same image, we control for scene content, such that changes in the induced updates primarily reflect the applied degradation perturbations. These results demonstrate that DPG captures degradation-sensitive priors for sample-specific adaptation.

Table 9: Comparison with alternative dynamic adaptation methods on the LOL-v2-real test set.

Method	PSNR ↑	SSIM ↑
MoLE	20.57	0.7075
Multiple adapters	20.47	0.7340
Ours	**22.17**	**0.8201**

Table 10: No-reference quality comparison on five real-world low-light datasets. ↑ indicates higher is better. Best results are in **bold**.

Dataset	MUSIQ ↑		CLIP-IQA ↑	
	GEFU	Ours	GEFU	Ours
DICM	58.54	**59.57**	0.5205	**0.5268**
LIME	59.33	**60.78**	0.4766	**0.4896**
MEF	63.34	**64.82**	0.5052	**0.5212**
NPE	66.32	**68.15**	**0.5396**	0.5144
VV	32.38	**34.17**	0.3402	**0.3451**
Avg.	55.98	**57.50**	0.4764	**0.4794**

C.4 Statistical Significance Analysis

To verify that our improvements over GEFU are statistically significant, we conduct paired statistical tests on all 100 LOL-v2-real test images, where the paired difference (Diff.) is defined as our per-image score minus GEFU's. We use the non-parametric Wilcoxon signed-rank test as the primary paired test and report the corresponding p-values. We additionally report the 95% confidence interval (CI) of the mean paired difference and Cohen's d_z, computed by standardizing the mean paired difference by the standard deviation of the paired differences. As shown in Table 8, our method significantly outperforms GEFU on all three metrics at the 0.05 level, with effect-size magnitudes ranging from small-to-medium to large, and all CIs excluding zero. The parametric paired t-test yields consistent conclusions, indicating that the observed gains are robust across the two testing procedures.

C.5 Comparison with Alternative Dynamic Adaptation Methods.

We further compare DDA with two alternative methods for realizing input-dependent adaptation. Specifically, the complete DDA pathway, including both DPG and DDSG, is replaced by either a MoLE-based method (Wu et al., 2024) or a multiple-adapter method (Pfeiffer et al., 2021) at the same insertion points in the backbone. The backbone, training data, optimization settings, and evaluation protocol remain unchanged across all variants. The MoLE-based method maintains a bank of trainable LoRA experts and routes each input to a single expert through input-dependent top-1 gating. The multiple-adapter method employs several independently parameterized adapters and combines their outputs with predicted fusion weights. All adapters and the corresponding fusion module are jointly optimized during training, following the general adapter-fusion formulation. In contrast, DDA obtains the degradation-aware representation c_x through DPG and uses DDSG to generate input-specific low-rank updates conditioned jointly on c_x and the weight states of the base K/V projections.

C.6 Testing on Unsupervised Datasets.

To further evaluate the performance of our method in real-world scenarios, we compare our method with GEFU on five commonly used real-world low-light datasets without ground-truth references, including LIME Guo et al. (2017), NPE Wang et al. (2013), MEF Ma et al. (2015), DICM Lee et al. (2013), and VV Vonikakis et al. (2018). We employ two no-reference image quality assessment metrics, CLIP-IQA and MUSIQ-IQA, to quantitatively evaluate the perceptual quality of the enhanced images. As shown in Table 10, our method consistently achieves higher MUSIQ scores than GEFU

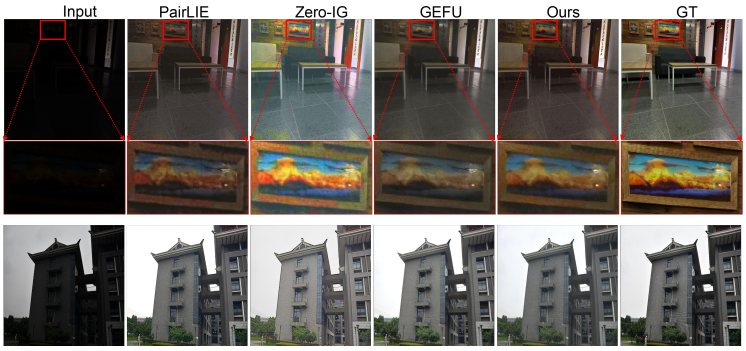


Figure 6: Visual comparisons on two real-world scenes from LSRW (Hai et al., 2023): a severely underexposed indoor scene (top) and an outdoor backlit scene (bottom). Zoomed insets highlight local color and detail restoration



Figure 7: Face detection results on images enhanced by different methods under three low-light scenes with increasing illumination levels. Compared with PairLIE (Fu et al., 2023), Zero-IG (Shi et al., 2024), and GEFU (Wang et al., 2025), our enhancement produces fewer false positives and missed faces across all scenes, better preserving the semantic structures required by downstream perception.

across all five datasets, improving the average score from 55.98 to 57.50. In terms of CLIP-IQA, our method outperforms GEFU on four out of the five datasets, with the average score increasing from 0.4764 to 0.4794. These results demonstrate that our method achieves favorable perceptual quality, as assessed by no-reference quality metrics, across diverse real-world low-light scenes.

C.7 Additional Visual Results

Figure 6 presents visual comparisons on two representative real-world scenes from LSRW (Hai et al., 2023). In the severely underexposed indoor scene (first row), PairLIE and GEFU brighten the image but leave the colors dull and hazy, whereas Zero-IG over-enhances the scene and introduces obvious color distortions, such as the greenish floor and the oversaturated curtain. As highlighted by the zoomed inset, our method restores the painting with vivid and faithful colors, achieving the closest appearance to the normal-light reference. In the outdoor backlit scene (second row), Zero-IG overexposes the sky and washes out the facade contrast, PairLIE produces a grayish tone, and GEFU yields a slightly flat appearance, while our result best matches the reference in both illumination and architectural details. These comparisons show that our method produces more realistic enhancements across diverse real-world conditions.

Figure 7 presents the face detection results on the enhanced images under three representative low-light scenes with different illumination levels. Detection on the images enhanced by PairLIE (Fu et al., 2023), Zero-IG (Shi et al., 2024), and GEFU (Wang et al., 2025) suffers from varying de-

grees of false positives and missed faces, whereas the images enhanced by our method lead to more accurate face detection across all three scenes. These results indicate that our enhancement better preserves the semantic structures required by downstream perception, beyond improving visual quality alone.

D LIMITATIONS AND FUTURE WORK

DDA characterizes low-light degradation using predefined illumination, color, and structure views. Although these priors capture representative degradation factors and yield consistent improvements across different benchmarks, they may not fully describe complex or previously unseen degradations in real-world scenes, such as spatially entangled noise, flare, and severe sensor-specific artifacts. Future work will explore data-driven prior discovery and learnable degradation representations that preserve the interpretability of the current design while improving its coverage and generalization to more diverse restoration scenarios.