Difficulty-Conditioned Attribute-Specific Restoration for Low-Light Image Enhancement

Mingzhu Zhang[1,2], Shuang Li[1,2], Jiaxu Leng[1,2], Long Sun[3], Can Zhang[1], Miaoqing Wang[1,2], and Xinbo Gao[1,2]†

[1] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[2] Chongqing Institute for Brain and Intelligence, Chongqing, 400064, China
[3] School of Artificial Intelligence and Robotics, Hunan University, Changsha, 410082, China
gaoxb@cqupt.edu.cn
† Corresponding author

Abstract. Low-light image enhancement (LLIE) aims to restore visibility and structural details from severely degraded images under insufficient illumination. However, real-world low-light images often exhibit spatially heterogeneous degradations, where mildly under-exposed areas are easy to recover while severely dark, noisy, or color-unstable regions remain challenging. Existing LLIE methods typically apply a globally uniform enhancement strategy, which often leads to suboptimal trade-offs such as over-smoothing textures in hard regions or over-enhancement in easier areas. To address this, we propose a Difficulty-Conditioned Attribute-Specific Restoration (DCASR) framework that explicitly models restoration difficulty and allocates enhancement capacity accordingly. DCASR incorporates a difficulty-aware enhancement network with two prior-driven modulators: a Global Intensity Modulator (GIM) for coherent illumination correction and a Structural Refinement Modulator (SRM) for detail recovery, together with an adaptive weighting scheme that emphasizes hard regions during optimization. Since residual-derived difficulty priors are unavailable at test time, we adopt a teacher–student scheme in which a privileged teacher derives difficulty priors from residual cues only during training, while a latent conditional diffusion student generates aligned priors conditioned on a single low-light input at inference. Experiments demonstrate that DCASR consistently outperforms state-of-the-art methods on multiple benchmarks. Code will be made publicly available at `github.com/mingzhuzhang1/DCASR-LLIE`.

Keywords: Low-light image enhancement · Difficulty-conditioned restoration · Attribute-specific enhancement · Diffusion-based prior learning

1 Introduction

Low-light image enhancement (LLIE) aims to recover visibility and natural appearance from images captured under insufficient illumination. Such images are

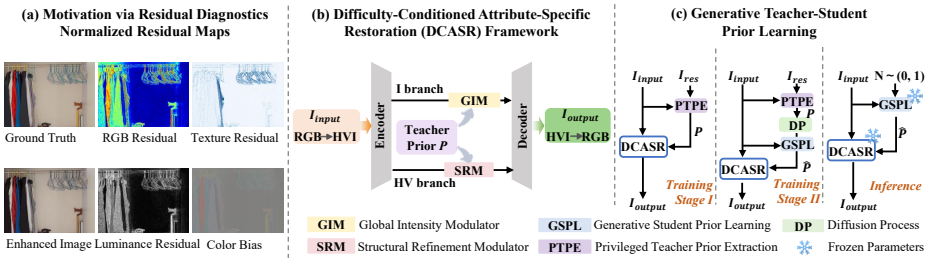

Fig. 1: Motivation via residual diagnostics and DCASR overview.

typically characterized by low contrast, amplified noise, chromatic distortions, and degraded details. Beyond perceptual improvement, reliable enhancement also provides higher-quality inputs for downstream recognition and understanding tasks [4, 8, 23, 26, 38, 62]. Owing to its practical importance and inherent challenges, LLIE has attracted increasing research attention in recent years.

Recent progress in LLIE has been largely driven by deep learning-based methods, which can be broadly categorized into regression-based and generation-based approaches that learn a mapping from low-light inputs to well-exposed images. Regression-based methods often incorporate physical priors, such as Retinex-based decomposition [1,2,46,47], or structural cues [43,49], while generation-based methods rely on generative frameworks, including GAN-based models [17] and diffusion-based models [16, 55]. Despite their methodological differences, most existing approaches apply a globally consistent enhancement strategy across the entire image. However, real low-light degradations exhibit strong spatial heterogeneity. Mildly under-exposed regions are relatively easy to recover, whereas severely dark, noisy, or chromatically distorted areas remain challenging. Consequently, uniform mappings lead to compromised trade-offs, such as over-smoothing fine textures or amplifying noise and color shifts in heavily degraded regions. As shown in Fig. 1(a), the high-residual regions exhibit structured, content-correlated patterns rather than random noise, indicating spatially heterogeneous degradation and the limitation of globally uniform enhancement.

This structured failure pattern suggests revisiting LLIE from a difficulty-aware perspective. High-error regions often correspond to intrinsically challenging areas, such as severely under-exposed textures, noise-dominated patches, or chromatically unstable regions. Learning an explicit difficulty prior from residual supervision enables restoration capacity to be allocated according to regional complexity, instead of enforcing a uniform enhancement strategy. However, naive integration of such priors is insufficient. Effective difficulty modeling must further consider the heterogeneous nature of degradation. Illumination degradation is typically globally coherent and low-frequency dominant, whereas structural and color degradations are spatially localized and high-frequency sensitive. As a result, applying a single difficulty signal uniformly in the RGB space is easily dominated by intensity variations and provides limited guidance for fine-grained structural refinement. Difficulty modulation should therefore be aligned with

the intrinsic statistical properties of different restoration attributes, separating global intensity calibration from detail-sensitive enhancement. In addition, residual-derived priors are unavailable during inference, making it necessary to infer difficulty representations from a single low-light input. However, this inference is challenging because restoration difficulty cannot be directly observed from the degraded image and may correspond to multiple plausible residual patterns. This motivates formulating difficulty prior learning as a conditional generative task, where a latent conditional diffusion model predicts difficulty-aware priors from low-light inputs.

To overcome these challenges, as illustrated in Fig. 1(b), we propose DCASR, a difficulty-aware prior-guided LLIE framework that performs attribute-aligned restoration in the HVI space. DCASR exploits the difficulty prior through two complementary mechanisms in the HVI space. In the intensity subspace, the difficulty-guided Global Intensity Modulator (GIM) converts the prior into channel-wise affine parameters and applies spatially invariant modulation to encourage globally consistent illumination correction. In the HV subspace, the difficulty-guided Structural Refinement Modulator (SRM) derives a high-frequency difficulty indicator via DCT-based analysis and uses it to sharpen spatial attention, prioritizing texture-sensitive regions during refinement. Beyond feature modulation, we further incorporate the prior into an adaptive weighting mechanism that reweights reconstruction supervision, progressively emphasizing hard regions while maintaining stability in easy ones. As shown in Fig. 1(c), to make such difficulty guidance available at inference time, we propose a two-stage teacher–student prior learning strategy. In the first stage, the teacher learns an oracle difficulty prior using residual cues available only during training. In the second stage, a conditional latent diffusion model predicts difficulty-aware priors directly from low-light inputs, enabling prior-guided enhancement without requiring ground truth at test time.

Contributions. The main contributions of this work are as follows. **(i)** We propose DCASR, a difficulty-aware LLIE framework that exploits attribute decoupling in the HVI space for controllable restoration. **(ii)** We design complementary prior-guided mechanisms to handle illumination and detail-color recovery, and an adaptive weighting scheme to emphasize hard regions. **(iii)** We develop a two-stage teacher–student prior learning strategy that enables inference-time difficulty prior synthesis from a single low-light image using latent conditional diffusion. **(iv)** Extensive experiments demonstrate state-of-the-art performance and validate the effectiveness of each component through ablation studies.

2 Related Work

2.1 Deep Learning Based LLIE Methods

Similar to many image restoration tasks [3, 25, 29, 30, 35–37, 39, 41, 57, 64], most data-driven LLIE methods adopt an end-to-end learning paradigm to directly map low-light inputs to their normal-light counterparts. Under this unified framework, existing approaches can be broadly categorized into regression-based [1, 2,

28, 43, 46, 47] and generation-based methods [10, 16, 17, 40, 50, 55, 56]. Regression-based LLIE learns deterministic mappings via reconstruction losses, but globally uniform enhancement can struggle under severe degradation, often resulting in blurred textures and color shifts. To improve robustness, Retinex-inspired methods [1, 2, 28, 46, 47] perform illumination–reflectance decomposition, and CIDNet [53] further introduces color-invariant decomposition to mitigate chromatic distortion. Complementarily, auxiliary-cue-guided strategies leverage semantics [5, 49, 65] or structural constraints [34, 52, 66] to enhance perceptual quality. However, these approaches still largely depend on features extracted from degraded inputs, which may become unreliable in severely corrupted regions. Generation-based approaches pursue higher realism beyond deterministic regression. GAN-based methods [6, 12, 17] improve perceptual quality via adversarial learning, while diffusion models [10, 15, 16, 40] synthesize realistic details through iterative denoising. Despite improved visual fidelity, they typically incur higher computational cost and often still adopt globally consistent enhancement without explicitly modeling spatially varying degradation difficulty.

2.2 Difficulty-Guided Learning

Visual recognition and restoration often exhibit spatially non-uniform difficulty: regions with complex structures, severe degradation, or ambiguous semantics are harder to process. Difficulty-guided learning addresses this by explicitly estimating difficulty and allocating model capacity accordingly. For example, Li et al. [27] route harder regions to subsequent refinement stages, Nie et al. [33] emphasize anatomically ambiguous structures via difficulty-aware attention, and Leng et al. [21] refocus object detection on difficult instances. These efforts suggest that modeling difficulty can improve representation learning and prediction accuracy. Similar observations hold in low-level vision. In super-resolution, edge neighborhoods and spatially variant degradations are typically harder to reconstruct, and DDL-BSR [22] leverages difficulty cues to guide kernel estimation and reconstruction. However, difficulty-guided learning remains under-explored in LLIE. This motivates us to introduce difficulty guidance into LLIE by estimating a reconstruction difficulty map and using it for feature modulation, allocating more capacity to severely degraded regions while preserving stability in easier areas.

3 Proposed Method

3.1 Overview

To address spatially heterogeneous degradations in low-light images, we propose DCASR, a difficulty-conditioned attribute-specific restoration framework for LLIE that explicitly models restoration complexity and adaptively allocates enhancement capacity. As illustrated in Fig. 2, DCASR consists of a prior-guided enhancement network built upon an attribute-aligned dual-branch backbone

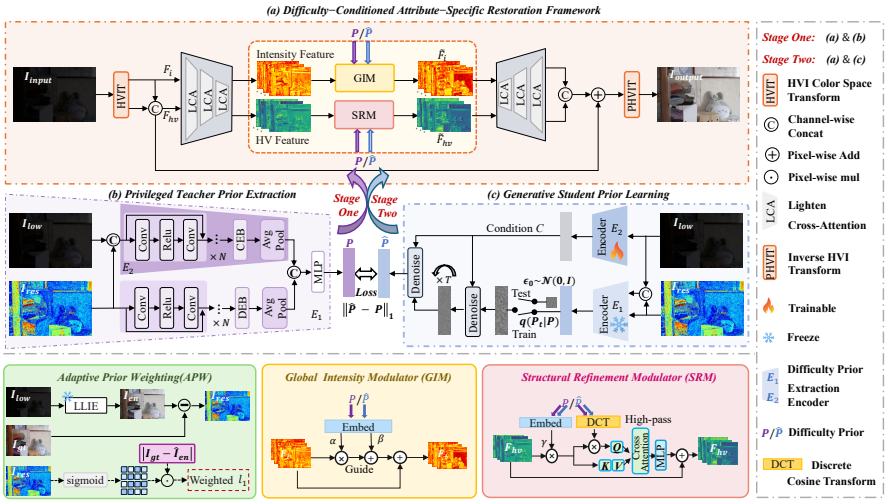

Fig. 2: Overall framework of DCASR. (a) Difficulty-Conditioned Attribute-Specific Restoration Framework: GIM for global illumination modulation and SRM for structural refinement. (b) Privileged Teacher Prior Extraction. (c) Generative Student Prior Approximation. The bottom row illustrates training-time residual construction, adaptive prior weighting (APW), and the GIM/SRM designs.

in the HVI space, and a two-stage prior learning strategy for inference-time difficulty estimation. In this dual-branch design, the intensity branch focuses on illumination-related restoration, while the HV branch is more sensitive to structural details. Built upon this HVI-based dual-branch decomposition, the prior-guided enhancement network injects difficulty priors into attribute-specific branches through two modulators: the Global Intensity Modulator (GIM) ensures globally coherent brightness calibration in the intensity branch, while the Structural Refinement Modulator (SRM) enhances detail restoration in the HV branch. To make difficulty guidance available without residual supervision at test time, we adopt a teacher–student strategy in which a privileged teacher learns oracle difficulty priors during training, and a latent conditional diffusion student predicts these priors from a single low-light input for residual-free enhancement.

3.2 Difficulty-Conditioned Attribute-Specific Restoration

We propose a difficulty-conditioned, attribute-decoupled restoration network that separates illumination calibration from structure- and color-sensitive refinement, while conditioning both processes on a learned difficulty prior. The network adopts a dual-branch design motivated by CIDNet [53], where the intensity branch is modulated by a difficulty-guided **Global Intensity Modulator (GIM)** for globally coherent brightness adjustment, and the complementary branch is refined by a difficulty-guided **Structural Refinement Modulator**

(SRM) to enhance structurally challenging regions. In addition, an **Adaptive Prior Weighting (APW)** reallocates supervision toward intrinsically difficult areas during optimization. Together, these components enable attribute-aware restoration with difficulty-conditioned behavior.

Global Intensity Modulator. Given that illumination degradation is predominantly characterized by low-frequency variations with strong global coherence [46, 63], intensity restoration benefits more from holistic rectification than from spatially fragmented adjustments. To this end, the Global Intensity Modulator (GIM) performs channel-wise global feature modulation by decoding a compact difficulty-aware prior P, rather than relying on explicit spatial guidance maps. This design enables stable and adaptive intensity calibration by leveraging the global degradation semantics encoded in P.

Specifically, GIM projects the prior $P \in \mathbb{R}^D$ into channel-wise affine parameters via a lightweight mapping network $\mathcal{M}_{gain}$, implemented as a single linear layer that outputs a $2C'$-dimensional vector:

$$[\boldsymbol{\alpha}, \boldsymbol{\beta}] = \mathcal{M}_{gain}(P), \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{C' \times 1 \times 1}. \tag{1}$$

These parameters modulate the latent intensity feature map $F_i \in \mathbb{R}^{C' \times H' \times W'}$ through a residual channel-wise affine transformation:

$$\tilde{F}_i = F_i \odot \big(\mathbf{1} + \boldsymbol{\alpha}\big) + \boldsymbol{\beta}, \tag{2}$$

where $\odot$ denotes element-wise multiplication, $\mathbf{1} \in \mathbb{R}^{C' \times 1 \times 1}$ is an all-ones tensor, and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are broadcast along spatial dimensions $H' \times W'$.

This spatially invariant modulation injects the difficulty prior in a structured and efficient manner. By enforcing channel-wise global adjustment, GIM preserves illumination consistency and reduces local artifacts that may arise from dense spatial prediction. Meanwhile, $\boldsymbol{\alpha}$ primarily controls global contrast (gain) and $\boldsymbol{\beta}$ controls global brightness (bias), allowing the network to adapt intensity restoration according to the overall difficulty encoded in P.

Structural Refinement Modulator. Complementary to the global rectification in the I-branch, the HV-branch focuses on suppressing noise and chromatic artifacts while recovering fine-grained details. Although the prior P captures the overall degradation severity, directly projecting this global code is often insufficient to explicitly steer high-frequency refinement. We therefore propose the Structural Refinement Modulator (SRM), which leverages frequency-domain analysis [18] to derive a structure-aware gating signal and guide detail restoration.

We first map P into channel-wise affine embeddings to coarsely modulate the HV-branch feature map $F_{hv} \in \mathbb{R}^{C' \times H' \times W'}$:

$$F_{coarse} = F_{hv} \odot \boldsymbol{\gamma} + \boldsymbol{\delta} + F_{hv}, \tag{3}$$

where $\boldsymbol{\gamma}, \boldsymbol{\delta} \in \mathbb{R}^{C' \times 1 \times 1}$ are learnable embeddings predicted from P (e.g., via a linear projection), and $\odot$ denotes element-wise multiplication.

To derive a high-frequency structural response for detail-sensitive refinement, we treat the difficulty prior $P \in \mathbb{R}^D$ as a 1D prior embedding and apply the orthonormal Discrete Cosine Transform (DCT) to obtain a frequency-inspired decomposition. A high-pass mask is then applied to retain the components above a cutoff frequency τ, and the filtered response is reconstructed by the corresponding inverse transform:

$$\mathcal{X}_k = c_k \sum_{n=0}^{D-1} P_n \cos\left[\frac{\pi}{D}\left(n + \frac{1}{2}\right)k\right], \quad k = 0, \ldots, D-1, \tag{4}$$

$$\mathcal{X}_h[k] = m_\tau[k]\mathcal{X}_k, \quad m_\tau[k] = \begin{cases} 1, & k \geq \tau, \\ 0, & k < \tau, \end{cases} \tag{5}$$

$$P_{hf}[n] = \left| \sum_{k=0}^{D-1} c_k \mathcal{X}_h[k] \cos\left[\frac{\pi}{D}\left(n + \frac{1}{2}\right)k\right] \right|, \quad n = 0, \ldots, D-1, \tag{6}$$

where $c_0 = \sqrt{1/D}$ and $c_k = \sqrt{2/D}$ for $k > 0$. Here, $\mathcal{X}_k$ denotes the k-th DCT coefficient of P, m_τ is a high-pass mask with cutoff frequency τ, and $P_{hf} \in \mathbb{R}^D$ denotes the reconstructed high-frequency structural response of the difficulty prior. Unlike a scalar statistic, P_{hf} preserves high-frequency variations encoded in the prior embedding and serves as a structural modulation signal.

Conditioning the Query projection on P_{hf} allows SRM to adjust the attention response according to high-frequency difficulty cues:

$$\tilde{F}_{hv} = \mathrm{FFN}\Big(\mathrm{Attn}\big(\phi_q\big(\bar{F} \odot (1 + \lambda P_{hf})\big), \phi_k(\bar{F}), \phi_v(\bar{F})\big) + F_{coarse}\Big) + F_{hv}, \tag{7}$$

where $\bar{F} = \mathrm{Norm}(F_{coarse}) \in \mathbb{R}^{C' \times H' \times W'}$, P_{hf} is used in a broadcast-compatible form, λ controls the modulation strength, ϕ_q, ϕ_k, and ϕ_v are linear projections, $\mathrm{Attn}(\cdot)$ denotes scaled dot-product attention, $\mathrm{FFN}(\cdot)$ is a feed-forward network, and $\tilde{F}_{hv} \in \mathbb{R}^{C' \times H' \times W'}$ is the refined output. By modulating the Query with the DCT-derived structural response, SRM adaptively adjusts attention for detail-sensitive refinement. Since the spatial layout is provided by the HV features, this modulation encourages structurally informative regions encoded in the HV features to receive stronger refinement under complex degradations.

Adaptive Prior Weighting. To effectively utilize the learned prior (either P or $\hat{P}$) for optimization, we construct a spatial attention mechanism that dynamically re-weights supervision. Given the residual prior X_{res} (or its estimation), we first compute the raw difficulty magnitude M_{diff} via channel-wise aggregation. To ensure scale invariance, we normalize it to $\tilde{M}_{diff} \in [0, 1]$ via per-image min-max scaling. To ensure training stability, a base importance map $\tilde{W}$ is generated

via a scaled Sigmoid function with a stop-gradient operation:

$$\tilde{W} = \mathrm{sg}\left[\frac{\sigma\Big(\omega(\tilde{M}_{diff} - 0.5)\Big)}{\mathbb{E}[\sigma\Big(\omega(\tilde{M}_{diff} - 0.5)\Big)] + \epsilon}\right], \tag{8}$$

where $\sigma(\cdot)$ denotes the Sigmoid function, $\omega = 6.0$ controls the focus sharpness, $\mathbb{E}[\cdot]$ represents spatial averaging, and $\mathrm{sg}[\cdot]$ denotes the stop-gradient operation.

We further introduce a learnable scalar $\xi \in [0, 1]$ to regulate the intensity of this guidance. The final spatial weight map W is formulated as $W = \xi \cdot \tilde{W} + (1-\xi)$. To explicitly enforce this spatial attention during restoration, we formulate the Prior-Guided Loss $\mathcal{L}_{apw}$ as the weighted mean absolute error:

$$\mathcal{L}_{apw} = \left\| W \odot (\hat{I} - I_{gt}) \right\|_1, \tag{9}$$

where $\odot$ denotes the Hadamard product. This mechanism enables the network to transition from uniform learning ($\xi = 0$) to hard-example mining ($\xi \to 1$), providing a robust basis for our optimization.

3.3 Generative Teacher–Student Prior Learning

The effectiveness of our framework relies on acquiring a reliable difficulty-aware prior P. While oracle priors can be derived from residual cues during training, they are unavailable at inference. We therefore formulate prior acquisition as a generative teacher–student learning problem. Inferring P from a single low-light input is inherently ill-posed, since restoration difficulty cannot be uniquely determined from degraded observations. Instead of deterministic regression, we model P as a conditional distribution. A privileged teacher first learns an oracle prior using training-only residual information, and a generative student is trained to approximate this distribution conditioned solely on the low-light input. We adopt a latent conditional diffusion model [13,50] to capture the high-dimensional prior distribution, enabling reliable prior synthesis at inference for fully residual-free enhancement.

Privileged Teacher Prior Extraction. In this stage, designated as the Privileged Teacher Phase, we aim to extract an oracle difficulty-aware prior $P \in \mathbb{R}^D$ using privileged information to encode spatially varying enhancement difficulty. Specifically, we first employ a frozen pre-trained enhancement model Φ_{llie} to generate the residual map $I_{res} \in \mathbb{R}^{3\times H\times W}$. The residual calculation is formulated as

$$I_{res} = |I_{gt} - \Phi_{llie}(I_{low})|, \tag{10}$$

where $|\cdot|$ denotes the element-wise absolute difference. After that, we design a dual-branch encoder architecture to process these cues. The context branch $\mathcal{E}_1$ takes the channel-wise concatenation of I_{low} and I_{res} to capture content-dependent context, while the difficulty branch $\mathcal{E}_2$ processes I_{res} to extract pure

error patterns. Both $\mathcal{E}_1$ and $\mathcal{E}_2$ are constructed primarily of stacked Residual Blocks [11]. The feature extraction process is expressed as

$$F_c = \mathcal{E}_1(I_{low} \oplus I_{res}), \quad F_d = \mathcal{E}_2(I_{res}), \tag{11}$$

where $\oplus$ denotes concatenation. After passing through the encoder layers, the resulting spatial features are aggregated into compact vectors via Global Average Pooling (GAP), which maps a spatial tensor $F \in \mathbb{R}^{C\times H'\times W'}$ to a channel descriptor in $\mathbb{R}^C$:

$$\mathbf{z}_c = \mathrm{GAP}(F_c) \in \mathbb{R}^D, \quad \mathbf{z}_d = \mathrm{GAP}(F_d) \in \mathbb{R}^D, \tag{12}$$

where f denotes the base channel width (we use D=256 in our implementation). These descriptors are fused and mapped by a multi-layer perceptron (MLP) to obtain the latent prior:

$$P = \mathcal{M}_{prior}(\mathbf{z}_c \oplus \mathbf{z}_d), \quad \mathcal{M}_{prior} : \mathbb{R}^{2D} \to \mathbb{R}^D, \quad P \in \mathbb{R}^D. \tag{13}$$

In our implementation, $\mathcal{M}_{prior}$ is realized by stacked linear layers with LeakyReLU activations. Simultaneously, this prior P is injected into the enhancement backbone to modulate the GIM and SRM modules for adaptive restoration.

Generative Student Prior Learning. In this stage, designated as the Generative Student Phase (see Fig. 2), we aim to achieve residual-free inference by training a student model to hallucinate a pseudo-prior $\hat{P}$ solely from low-light inputs. Inspired by the high-fidelity generation of DDPM [13] and the efficiency of latent-space diffusion [50], we conduct the prior learning within the compact latent space to reduce computational costs and result randomness.

To establish the supervision, we freeze the pre-trained encoder E_1 from Stage I as the Privileged Teacher to extract the oracle latent prior P from the ground-truth data. This P serves as the target distribution for the forward diffusion process, where we progressively inject Gaussian noise over timestep t:

$$q(P_t|P) = \mathcal{N}(P_t; \sqrt{\bar{\alpha}_t}P, (1-\bar{\alpha}_t)\mathbf{I}), \tag{14}$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$, and β_t controls the noise variance.

In the reverse denoising phase, we employ a student encoder E_2 to extract conditional features C from the input image I_{low}. Unlike direct prior prediction, we train a denoising network ϵ_θ to estimate the noise in P_t conditioned on C. Consequently, the reverse sampling step from P_t to P_{t-1} is formulated as

$$P_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(P_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(P_t, C, t)\right) + \sqrt{1-\alpha_t}\epsilon_t, \tag{15}$$

where $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$ represents the random noise for stochastic sampling. During inference, where the privileged information is unavailable, we sample a random Gaussian noise $\epsilon_0 \sim \mathcal{N}(0, \mathbf{I})$ and iteratively denoise it using the trained student

network and condition C. After T iterations, we obtain the generated prior $\hat{P}$, which is subsequently injected into the backbone to guide the residual-free enhancement.

We further impose a teacher–student prior alignment loss to encourage the student-generated prior to match the teacher oracle prior:

$$\mathcal{L}_{cdp} = \|\hat{P} - P\|_1, \tag{16}$$

where P denotes the oracle difficulty-aware prior extracted by the privileged teacher using training-only residual information, and $\hat{P}$ is the prior generated by the student diffusion model conditioned on I_{low}.

3.4 Optimization

In the first stage, we optimize DCASR using reconstruction losses defined in both the RGB and HVI domains. Following CIDNet [53], the training objective in the first stage consists of an $\mathcal{L}_1$ loss together with structural and perceptual constraints:

$$\mathcal{L}_{stage1} = \mathcal{L}_1 + \mathcal{L}_{ssim} + \mathcal{L}_{edge} + \mathcal{L}_{perc}. \tag{17}$$

where $\mathcal{L}_1$, $\mathcal{L}_{ssim}$, $\mathcal{L}_{edge}$, and $\mathcal{L}_{perc}$ respectively enforce pixel-wise reconstruction, structural similarity, edge consistency, and perceptual supervision.

In the second stage, the standard $\mathcal{L}_1$ loss is replaced by the proposed adaptive prior-weighted loss $\mathcal{L}_{apw}$ to emphasize difficult regions. Meanwhile, the diffusion-based student model is trained to predict the difficulty prior using the prior alignment loss $\mathcal{L}_{cdp}$. The training objective in the second stage is:

$$\mathcal{L}_{stage2} = \eta(\mathcal{L}_{apw} + \mathcal{L}_{ssim} + \mathcal{L}_{edge} + \mathcal{L}_{perc}) + \mathcal{L}_{cdp}, \tag{18}$$

where $\eta = 0.01$ balances reconstruction and prior learning.

4 Experiments

4.1 Datasets and Settings

Datasets and Metrics. We evaluate the proposed method on the LOL benchmark, encompassing LOL-v1 [46] and LOL-v2 [54], where LOL-v2 is partitioned into LOL-v2-real and LOL-v2-synthetic subsets. Following the official protocols, LOL-v1 comprises 485 training and 15 testing images, while LOL-v2-real and LOL-v2-synthetic contain 689 and 900 training pairs, respectively, each with 100 samples for evaluation. To assess real-world generalization without paired supervision, we further test on four widely recognized unpaired datasets: DICM [20], LIME [9], MEF [31], and NPE [42]. For paired benchmarks, we report PSNR, SSIM [45], and LPIPS [59] (with an AlexNet [19] backbone); for unpaired scenarios, we adopt NIQE [32] as the primary metric. Higher PSNR and SSIM values signify superior reconstruction fidelity, whereas lower scores in LPIPS and NIQE denote better perceptual quality and naturalness.

Table 1: Quantitative comparisons on the LOL benchmark. [Key: **Best**, Second Best, ↑: Higher is better, ↓: Lower is better].

Method	References	LOL-v1			LOL-v2-real			LOL-v2-synthetic		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
MIRNet [58]	ECCV'20	21.512	0.788	0.222	21.648	0.810	0.313	22.059	0.894	0.122
SGM [54]	TIP'21	17.810	0.765	0.219	20.060	0.819	0.163	22.050	0.903	0.090
RUAS [28]	CVPR'21	14.630	0.573	0.285	17.970	0.646	0.329	16.940	0.648	0.222
EnlightenGAN [17]	TIP'21	17.480	0.658	0.314	18.230	0.617	0.309	16.570	0.734	0.221
Restormer [57]	CVPR'22	21.710	0.750	0.218	22.430	0.759	0.268	20.700	0.803	0.195
SNR-Net [51]	CVPR'22	23.005	0.824	0.164	21.103	0.839	0.169	24.173	0.924	0.064
Retinexformer [2]	CVPR'23	23.830	0.832	0.141	21.272	0.841	0.163	25.281	0.928	0.064
UHDFormer [24]	ICLR'23	22.880	0.835	0.137	18.570	0.646	0.417	17.460	0.747	0.269
QuadPrior [43]	CVPR'24	18.330	0.768	0.205	20.590	0.760	0.200	17.100	0.746	0.317
LightenDiffusion [16]	ECCV'24	19.120	0.749	0.195	21.710	0.789	0.187	19.940	0.826	0.170
DMFourLLIE [61]	ACM MM'24	22.430	0.801	0.124	18.860	0.821	0.122	24.890	0.935	–
CWNet [60]	ICCV'25	23.601	0.850	0.123	21.647	0.860	0.136	25.501	0.936	**0.044**
URetinex++ [48]	TPAMI'25	23.830	0.819	0.115	23.920	0.798	0.187	18.990	0.744	0.210
CIDNet [53]	CVPR'25	23.809	**0.857**	0.086	23.904	0.866	0.122	25.129	0.939	0.045
Ours	–	**24.042**	0.850	**0.082**	**24.053**	**0.874**	**0.109**	**25.610**	**0.940**	0.045

Training Protocol and Implementation Details. All experiments are implemented in PyTorch and trained on a single NVIDIA RTX A6000 (48GB) GPU. We use random 256×256 crops with a batch size of 8, and optimize the network using Adam with $(\beta_1, \beta_2) = (0.9, 0.99)$. The learning rate is initialized to 1×10^{-4} and smoothly decayed to 1×10^{-7} using cosine annealing. We adopt a two-stage training schedule for a total of 1000 epochs. Stage I (500 epochs) jointly trains the enhancement network and the privileged teacher prior extractor to obtain oracle difficulty priors. Stage II (500 epochs) trains the student prior generator and fine-tunes the enhancement network with the teacher fixed. Our model contains 6.70M parameters and runs at 0.148 s/image for full inference on 256×256 inputs with batch size 1.

4.2 Comparisons with State-of-the-Art Methods

We compare DCASR with representative state-of-the-art LLIE approaches, including MIRNet [58], SGM [54], RUAS [28], PairLIE [7], EnlightenGAN [17], Restormer [57], LLFlow [44], SNR-Net [51], Retinexformer [2], UHDFormer [24], QuadPrior [43], LightenDiffusion [16], DMFourLLIE [61], CWNet [60], URetinex [47], URetinex++ [48], RetinexNet [46], LL-SKF [49], GSAD [14], and CIDNet [53]. For competing methods, we evaluate officially released checkpoints when available; otherwise, we retrain them under identical preprocessing and evaluation protocols to ensure a fair comparison.

Quantitative Results on the LOL Benchmark. Tab. 1 reports quantitative comparisons on the LOL benchmark. DCASR achieves the highest PSNR across the three settings, reaching 24.04 dB on LOL-v1, 24.05 dB on LOL-v2-Real, and 25.61 dB on LOL-v2-Synthetic. Compared with CIDNet, which serves as our primary baseline, DCASR provides consistent yet steady improvements in PSNR while maintaining competitive SSIM and LPIPS performance. Although the numerical gains are moderate, the improvements are stable across different subsets,

Table 2: Robustness testing experiments. All methods are trained on LOL-v1 and tested on LOL-v2-synthetic. Unsup. and Sup. denote unsupervised and supervised modes, respectively.

Method	Mode	PSNR↑	SSIM↑	LPIPS↓
RUAS [28]	Unsup.	15.326	0.488	0.458
PairLIE [7]	Unsup.	**19.074**	0.797	0.230
EnlightenGAN [17]	Unsup.	16.183	0.734	0.220
LLFlow [44]	Sup.	17.119	0.812	0.224
Retinexformer [2]	Sup.	16.570	0.769	0.252
GSAD [14]	Sup.	15.854	0.748	0.243
CIDNet [53]	Sup.	17.653	0.801	0.231
Ours	Sup.	18.790	**0.814**	**0.217**

Table 3: Quantitative comparisons on real-world unpaired datasets. These NIQE (↓) results are obtained either from original papers or via testing with their optimal LOL [46] weights.

Methods	MEF	LIME	DICM	NPE	Mean↓
SNR [51]	4.14	5.51	4.62	4.36	4.66
URetinex [47]	3.79	**3.86**	4.15	4.69	4.12
LLFlow [44]	3.92	5.29	3.78	4.16	4.29
LL-SKF [49]	4.03	5.15	3.70	4.08	4.24
RetinexNet [46]	4.90	4.91	4.31	4.39	4.63
CIDNet [53]	3.60	4.18	4.18	4.29	4.06
Ours	**3.37**	4.13	**3.62**	**3.92**	**3.76**

suggesting that the proposed difficulty-aware guidance contributes positively to restoration fidelity without compromising perceptual quality.

Generalization from LOL-v1 to LOL-v2. To further evaluate cross-dataset robustness under distribution shift, we train all supervised methods on LOL-v1 and directly test them on LOL-v2-Synthetic. As summarized in Tab. 2, DCASR achieves the best PSNR among supervised methods and obtains the best SSIM and LPIPS among all compared methods. Specifically, DCASR surpasses CIDNet and LLFlow by 1.14 dB and 1.67 dB in PSNR, respectively. These results indicate that modeling degradation difficulty improves robustness to domain discrepancies across both real and synthetic low-light degradation distributions.

Unpaired Real-World Dataset Evaluation. To assess real-world generalization without paired supervision, we report NIQE results on four unpaired datasets (DICM, LIME, MEF, and NPE) in Tab. 3. Lower NIQE values indicate more natural enhancement quality. DCASR achieves the best average NIQE and obtains competitive or superior scores across diverse in-the-wild scenes, suggesting consistent zero-shot generalization to real-world scenarios.

Visualization Comparison. Fig. 3 provides qualitative comparisons between our approach and several representative state-of-the-art methods. While competing methods improve global brightness, they often suffer from local artifacts, noise residuals, or color inconsistencies in severely degraded regions. In contrast, DCASR produces clearer and more natural results, with better detail preservation and color consistency. The improvements are attributed to the proposed difficulty-aware allocation strategy, which dynamically distributes restoration capacity according to degradation severity, ensuring balanced illumination and structural fidelity across the entire image. In Fig. 4, we further show residual diagnostic comparisons, including RGB, luminance, texture, and color bias residuals. Compared with Retinexformer, DCASR produces lower residual responses and reduced chromatic bias in challenging regions, indicating improved detail restoration and color consistency under spatially heterogeneous degradations. Fig. 5 further visualizes the gain brought by prior guidance over the w/o prior

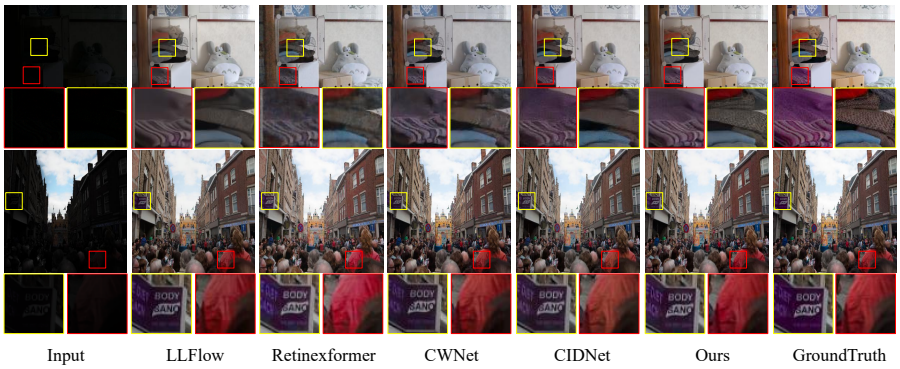

Fig. 3: Qualitative comparison on LOL-v1 and LOL-v2. Zoomed-in patches show that ours yields more coherent illumination (brighter yet non-overexposed), preserves garment textures and text details, and maintains more faithful colors with fewer casts.

Table 4: Ablation of DCASR components on LOL-v2-real. [Key: Best, Second Best, ↑: Higher is better, ↓: Lower is better].

Exp.	GIM	SRM	$\mathcal{L}_{apw}$	PSNR↑	SSIM↑	LPIPS↓
(a)	✗	✗	✗	23.498	0.857	0.123
(b)	✓	✗	✗	23.625	0.868	0.113
(c)	✗	✓	✗	23.765	0.868	0.111
(d)	✓	✓	✗	23.894	0.870	0.112
(e)	✓	✓	✓	**24.053**	**0.874**	**0.109**

Table 5: Design ablation of modulators on LOL-v2-real. I, HV denote branch modulators. **CA**: Channel Attention used for substitution.

Category	I	HV	PSNR↑	SSIM↑	LPIPS↓
Homog.	GIM	GIM	22.923	0.844	0.200
	SRM	SRM	22.241	0.840	0.183
Placement	SRM	GIM	22.662	0.850	0.171
Substitution	CA	CA	23.012	0.864	0.131
Proposed	**GIM**	**SRM**	**23.894**	**0.870**	**0.112**

baseline, where green indicates residual reduction and red indicates residual increase. The observed improvements mainly appear in structurally complex and severely degraded regions, supporting the effectiveness of difficulty-aware prior guidance.

4.3 Ablation Study

We conduct ablation studies on the LOL-v2-Real dataset to validate DCASR. We evaluate the difficulty-prior learning strategy, the branch-specific effects of GIM and SRM, and the contribution of the adaptive prior-weighted loss $\mathcal{L}_{apw}$.

Component and Optimization Ablation. As summarized in Tab. 4, we evaluate the incremental contribution of each module. Starting from a standard U-Net baseline in Exp. (a), the integration of GIM and SRM in Exp. (d) yields a significant PSNR gain of 0.396 dB, validating the efficacy of our dual-branch modulation.The introduction of the prior-driven supervision $\mathcal{L}_{apw}$ in Exp. (e) further enhances the PSNR to 24.053 dB, highlighting its ability to stabilize restoration in challenging regions.

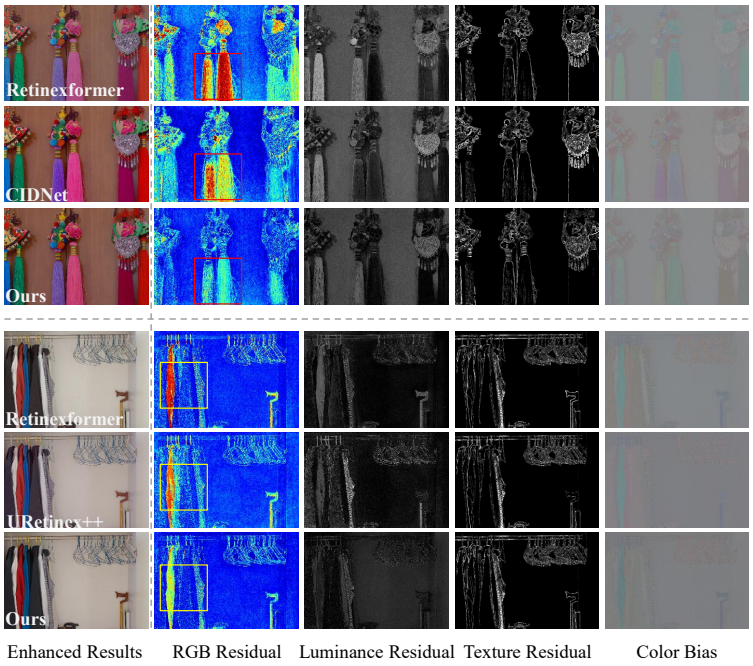

Fig. 4: Residual diagnostic comparison with Retinexformer [2]. DCASR yields lower residual responses and reduced color bias, especially in challenging regions.

Attribute-aligned Design and Placement. We justify our attribute-aligned configuration in Tab. 5. Replacing our specialized modulators with standard Channel Attention (CA) leads to a 0.882 dB PSNR drop, indicating that generic attention is less effective for intensity and structural restoration. Moreover, using homogeneous modulators, such as dual GIM or dual SRM, or swapping the positions of GIM and SRM, causes substantial performance degradation. Notably, swapping GIM and SRM decreases PSNR by 1.232 dB, from 23.894 dB to 22.662 dB. These results verify the necessity of assigning GIM to the I branch for global intensity modulation and SRM to the HV branch for spatially-aware structural refinement.

Analysis on Prior Learning and Strategy. Tab. 6 analyzes different prior sources and learning strategies. Introducing priors consistently improves over the no-prior baseline, while our residual-derived difficulty prior achieves the best performance among all guidance strategies. Unlike generic cues such as SNR, brightness, attention, and confidence maps, which mainly describe degradation statistics, our prior is supervised by restoration residuals and thus directly captures where restoration fails.

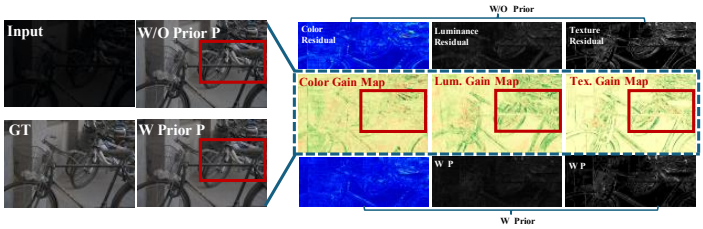

Fig. 5: Prior gain maps relative to w/o prior: green = reduced residuals (improvement), red = increased (degradation).

Table 6: Ablation study of prior prediction and guidance strategies on LOL-v2-Real. [Key: **Best**, Second Best, ↑: Higher is better, ↓: Lower is better].

(a) Prior prediction				(b) Prior guidance			
Pred.	PSNR↑	SSIM↑	LPIPS↓	**Guide**	PSNR↑	SSIM↑	LPIPS↓
MLP	23.490	0.868	0.115	w/o prior	23.498	0.857	0.123
CNN	23.431	0.855	0.116	SNR	23.129	0.867	0.112
Resblock	23.534	0.870	0.114	Brightness	23.516	0.865	0.120
Multi-scale	23.416	0.865	0.117	Attention	23.616	0.868	0.111
Transformer	23.611	0.869	0.110	Confidence	23.768	0.862	0.126
Diffusion (Ours)	**24.053**	**0.874**	**0.109**	Ours	**24.053**	**0.874**	**0.109**

5 Conclusion

In this paper, we propose DCASR, a difficulty-conditioned framework for low-light image enhancement that explicitly models the spatially heterogeneous nature of real-world degradations. By introducing a learned difficulty prior, DCASR allocates restoration capacity adaptively across regions with different enhancement complexity. To enable prior guidance at inference, we further develop a teacher–student strategy that distills residual-derived oracle priors into a latent conditional diffusion model, allowing reliable prior synthesis from a single low-light input. Experiments on multiple benchmarks demonstrate that DCASR achieves consistent improvements in illumination naturalness, detail preservation, and color stability. In future work, we plan to explore integrating difficulty-aware enhancement with downstream vision tasks in challenging environments.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants No. 62221005, 62472060, U22A2096, and U23A20318, in part by the Science and Technology Innovation Key R&D Program of Chongqing under Grant No. CSTB2023TIAD-STX0016, in part by the Natural Science Foundation of Chongqing under Grant No. CSTB2024NSCQ-QCXMX0060, Grant No. CSTB2023NSCQ-LZX0061, and in part by the Chongqing Postgraduate Research and Innovation Project under Grant No. CYB25260.

References

1. Bai, J., Yin, Y., He, Q., Li, Y., Zhang, X.: Retinexmamba: Retinex-based mamba for low-light image enhancement. In: International conference on neural information processing. pp. 427–442. Springer (2024)
2. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 12504–12513 (2023)
3. Chen, X., Li, H., Li, M., Pan, J.: Learning a sparse transformer network for effective image deraining. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5896–5905 (2023)
4. Cui, Z., Zhu, Y., Gu, L., Qi, G.J., Li, X., Zhang, R., Zhang, Z., Harada, T.: Exploring resolution and degradation clues as self-supervised signal for low quality object detection. In: European Conference on Computer Vision. pp. 473–491. Springer (2022)
5. Fan, M., Wang, W., Yang, W., Liu, J.: Integrating semantic segmentation and retinex model for low-light image enhancement. In: Proceedings of the 28th ACM international conference on multimedia. pp. 2317–2325 (2020)
6. Fu, Y., Hong, Y., Chen, L., You, S.: Le-gan: Unsupervised low-light image enhancement network using attention module and identity invariant loss. Knowledge-Based Systems **240**, 108010 (2022)
7. Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.K.: Learning a simple low-light image enhancer from paired low-light instances. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22252–22261 (2023)
8. Guo, Q., Feng, W., Gao, R., Liu, Y., Wang, S.: Exploring the effects of blur and deblurring to visual object tracking. IEEE Transactions on Image Processing **30**, 1812–1824 (2021)
9. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on image processing **26**(2), 982–993 (2016)
10. He, C., Fang, C., Zhang, Y., Tang, L., Huang, J., Li, K., Guo, Z., Li, X., Farsiu, S.: Reti-Diff: Illumination degradation image restoration with Retinex-based latent diffusion model. In: Int. Conf. Learn. Represent. (2025)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
12. He, M., Wang, R., Zhang, M., Lv, F., Wang, Y., Zhou, F., Bian, X.: Swinlight-gan a study of low-light image enhancement algorithms using depth residuals and transformer techniques. Scientific Reports **15**(1), 12151 (2025)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
14. Hou, J., Zhu, Z., Hou, J., Liu, H., Zeng, H., Yuan, H.: Global structure-aware diffusion process for low-light image enhancement. Advances in Neural Information Processing Systems **36**, 79734–79747 (2023)
15. Jiang, H., Luo, A., Fan, H., Han, S., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. ACM Transactions on Graphics (ToG) **42**(6), 1–14 (2023)
16. Jiang, H., Luo, A., Liu, X., Han, S., Liu, S.: Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In: European Conference on Computer Vision. pp. 161–179. Springer (2024)

17. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing **30**, 2340–2349 (2021)
18. Khayam, S.A.: The discrete cosine transform (dct): theory and application. Michigan State University **114**(1), 31 (2003)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)
20. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation of 2d histograms. IEEE transactions on image processing **22**(12), 5372–5384 (2013)
21. Leng, J., Mo, M., Zhou, Y., Gao, C., Li, W., Gao, X.: Pareto refocusing for drone-view object detection. IEEE Transactions on Circuits and Systems for Video Technology **33**(3), 1320–1334 (2022)
22. Leng, J., Wang, J., Mo, M., Gan, J., Lu, W., Gao, X.: Difficulty-guided variant degradation learning for blind image super-resolution. IEEE Transactions on Neural Networks and Learning Systems **36**(7), 13080–13093 (2024)
23. Leng, J., Wang, Z., Li, S., Gao, X.: Dynamic-static collaboration for unsupervised domain adaptive video-based visible-infrared person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 5955–5963 (2026)
24. Li, C., Guo, C.L., Zhou, M., Liang, Z., Zhou, S., Feng, R., Loy, C.C.: Embedding fourier for ultra-high-definition low-light image enhancement. In: ICLR (2023)
25. Li, L., Zhang, Y., Yuan, L., Gao, X.: Plgnet: prior-guided local and global interactive hybrid network for face super-resolution. IEEE Transactions on Circuits and Systems for Video Technology **34**(10), 10166–10181 (2024)
26. Li, S., Leng, J., Kuang, C., Tan, M., Gao, X.: Video-level language-driven video-based visible-infrared person re-identification. IEEE Transactions on Information Forensics and Security (2025)
27. Li, X., Liu, Z., Luo, P., Change Loy, C., Tang, X.: Not all pixels are equal: Difficulty-aware semantic segmentation via deep layer cascade. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3193–3202 (2017)
28. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: 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. pp. 10561–10570 (2021)
29. Liu, X., Shi, Z., Wu, Z., Chen, J., Zhai, G.: Griddehazenet+: An enhanced multi-scale network with intra-task knowledge transfer for single image dehazing. IEEE Transactions on Intelligent Transportation Systems **24**(1), 870–884 (2022)
30. Lu, L., Xiong, Q., Xu, B., Chu, D.: Mixdehazenet: Mix structure block for image dehazing network. In: 2024 International Joint Conference on Neural Networks (IJCNN). pp. 1–10. IEEE (2024)
31. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. IEEE Transactions on Image Processing **24**(11), 3345–3356 (2015)
32. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters **20**(3), 209–212 (2012)
33. Nie, D., Wang, L., Xiang, L., Zhou, S., Adeli, E., Shen, D.: Difficulty-aware attention network with confidence learning for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 1085–1092 (2019)

34. Rana, D., Lal, K.J., Parihar, A.S.: Edge guided low-light image enhancement. In: 2021 5th international conference on intelligent computing and control systems (ICICCS). pp. 871–877. IEEE (2021)
35. Sun, L., Dong, J., Tang, J., Pan, J.: Spatially-adaptive feature modulation for efficient image super-resolution. In: ICCV (2023)
36. Sun, L., Pan, J., Tang, J.: ShuffleMixer: An efficient convnet for image super-resolution. In: Advances in Neural Information Processing Systems (2022)
37. Sun, Z., Zhang, C., Li, J., Zhang, M., Gao, X.: Wsformer: Wavelet-based sparse transformer for blind image restoration. IEEE Transactions on Image Processing (2026)
38. Tang, H., Li, Z., Zhang, D., He, S., Tang, J.: Divide-and-conquer: Confluent triple-flow network for RGB-T salient object detection. IEEE Trans. Pattern Anal. Mach. Intell. **47**(3), 1958–1974 (2025)
39. Tsai, F.J., Peng, Y.T., Lin, Y.Y., Tsai, C.C., Lin, C.W.: Stripformer: Strip transformer for fast image deblurring. In: European conference on computer vision. pp. 146–162. Springer (2022)
40. Wan, F., Xu, B., Pan, W., Liu, H.: Psc diffusion: patch-based simplified conditional diffusion model for low-light image enhancement. Multimedia Systems **30**(4), 187 (2024)
41. Wang, M., Leng, J., Li, S., Kuang, C., Sun, L.: Portraitsr: Artist-inspired prior learning for progressive face super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 9984–9992 (2026)
42. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE transactions on image processing **22**(9), 3538–3548 (2013)
43. Wang, W., Yang, H., Fu, J., Liu, J.: Zero-reference low-light enhancement via physical quadruple priors. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 26057–26066 (2024)
44. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.P., Kot, A.: Low-light image enhancement with normalizing flow. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 2604–2612 (2022)
45. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
46. Wei, C., Wang, W., Yang, W., Liu, J.: Deep Retinex decomposition for low-light enhancement. In: British Machine Vision Conference (2018)
47. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: 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. pp. 5901–5910 (2022)
48. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: Interpretable Optimization-Inspired Unfolding Network for Low-Light Image Enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(4), 2545–2562 (2025)
49. Wu, Y., Pan, C., Wang, G., Yang, Y., Wei, J., Li, C., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1662–1671 (2023)
50. Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Van Gool, L.: Diffir: Efficient diffusion model for image restoration. In: Int. Conf. Comput. Vis. pp. 13095–13105 (2023)

51. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17714–17724 (2022)
52. Xu, X., Wang, R., Lu, J.: Low-light image enhancement via structure modeling and guidance. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9893–9903 (2023)
53. Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., Zhang, Y.: Hvi: A new color space for low-light image enhancement. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 5678–5687 (2025)
54. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing **30**, 2072–2086 (2021)
55. Yi, X., Xu, H., Zhang, H., Tang, L., Ma, J.: Diff-retinex++: Retinex-driven reinforced diffusion model for low-light image enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025)
56. Yin, Y., Xu, D., Tan, C., Liu, P., Zhao, Y., Wei, Y.: Cle diffusion: Controllable light enhancement diffusion model. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 8145–8156 (2023)
57. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
58. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: European conference on computer vision. pp. 492–511. Springer (2020)
59. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
60. Zhang, T., Liu, P., Lu, Y., Cai, M., Zhang, Z., Zhang, Z., Zhou, Q.: Cwnet: Causal wavelet network for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8789–8799 (2025)
61. Zhang, T., Liu, P., Zhao, M., Lv, H.: Dmfourllie: Dual-stage and multi-branch fourier network for low-light image enhancement. In: Proceedings of the 32nd ACM international conference on multimedia. pp. 7434–7443 (2024)
62. Zhang, Y., Zhang, Y., Zhang, Z., Zhang, M., Tian, R., Ding, M.: Isp-teacher: Image signal process with disentanglement regularization for unsupervised domain adaptive dark object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 7387–7395 (2024)
63. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: Proceedings of the 27th ACM international conference on multimedia. pp. 1632–1640 (2019)
64. Zheng, M., Sun, L., Dong, J., Pan, J.: Efficient video super-resolution for real-time rendering with decoupled g-buffer guidance. In: CVPR (2025)
65. Zheng, S., Gupta, G.: Semantic-guided zero-shot learning for low-light image/video enhancement. In: Proceedings of the IEEE/CVF Winter conference on applications of computer vision. pp. 581–590 (2022)
66. Zhu, M., Pan, P., Chen, W., Yang, Y.: Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 13106–13113 (2020)