REPRIFUSION: RETINEX-BASED LOW-LIGHT IMAGE ENHANCEMENT VIA PRIOR-SPECIFIC FUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Low-light image enhancement increasingly benefits from physical and pretrained visual priors. However, existing methods often rely on a single prior or integrate multiple representations through uniform fusion, limiting the effective use of their complementary information. We propose **RePriFusion**, a low-light image enhancement framework that integrates learned retinex-based prior features and vision foundation model features through selected restoration stages and prior-specific fusion modules. First, we design a cascaded Retinex estimator to progressively extract illumination and reflectance prior features and constructs an intermediate reconstruction. Second, we design prior-specific fusion modules that introduce illumination, reflectance, and the foundation model priors into selected U-Net stages using dedicated fusion operators tailored to their restoration roles. Finally, mixed-dataset training and noise augmentation are further adopted to improve robustness to diverse low-light distributions. Experiments on five paired LLIE benchmarks and eight unseen datasets show strong enhancement quality and zero-shot generalization.

1 INTRODUCTION

Low-light image enhancement (LLIE) aims to improve the visibility of under-exposed images while preserving natural appearance, structural details, and color fidelity. Early approaches mainly rely on hand-crafted operations, including histogram equalization, gamma correction, and illumination adjustment. Although these methods can increase overall brightness, they often amplify noise and introduce over-enhancement or color distortions in complex low-light scenes Guo et al. (2017); Wang et al. (2013).

With the rapid development of deep neural networks, learning-based methods have largely replaced hand-crafted enhancement pipelines Wei et al. (2018); Guo et al. (2020); Zhang et al. (2019); Zamir et al. (2022). Among them, Retinex-guided approaches have become a prominent direction, as the illumination–reflectance formulation provides useful physical prior guidance for low-light restoration Wei et al. (2018); He et al. (2025); Liu et al. (2021b); Brateanu et al. (2026). However, many methods Liu et al. (2021b); Ma et al. (2022); Cai et al. (2023) rely primarily on illumination prior guidance, leaving complementary appearance cues from reflectance underused. Although some methods Yi et al. (2023); He et al. (2025) incorporate both illumination and reflectance priors, they usually represent them as feature maps and use them only for feature-level guidance. Consequently, the Retinex relation between the two components is not explicitly reflected at the image level. Moreover, the two priors are often treated uniformly across the restoration network, despite their distinct properties and stage-dependent roles. Such uniform fusion may introduce mismatched or redundant guidance at unsuitable stages and incur unnecessary computation.

Another major challenge of LLIE is cross-dataset generalization. Most existing methods are trained and evaluated on a single benchmark, making them susceptible to dataset-specific degradation patterns. However, real low-light images exhibit substantial distribution shifts caused by variations in cameras, sensor noise, illumination conditions, and scene content Liu et al. (2021a); Gu et al. (2025); Chen et al. (2018); Jiang & Zheng (2019); Wang et al. (2021b); Xu et al. (2022). Moreover, collecting aligned low/normal-light image pairs is costly, which limits the scale and diversity of available training data. Consequently, models trained on a single paired dataset often generalize poorly to unseen low-light environments.

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.

To address these challenges, we propose RePriFusion, a low-light image enhancement framework that integrates learned retinex-based prior features and visual foundation model features through selected restoration stages and prior-specific fusion modules. As illustrated in Fig. 1, RePriFusion comprises a Cascaded Retinex Prior Estimator(CRPE), a frozen DINO encoder, and a U-Net restoration backbone with different prior-specific fusion modules. CRPE first derives illumination and then uses it to initialize and condition reflectance extraction. It further recombines the two components into an intermediate Retinex reconstruction, which is fed into the U-Net for refinement. This design allows the Retinex relation to affect restoration directly in image space, while leaving the backbone to correct the remaining degradations.

In parallel, we use a the frozen DINOv3 Siméoni et al. (2025) model as vision foundation model, which extracts multi-level contextual features learned from large-scale visual data. Together, the three priors provide complementary but non-interchangeable guidance. The illumination prior describes spatial exposure variation, the reflectance prior retains input-dependent appearance cues, and the DINO prior provides comparatively stable contextual information. RePriFusion therefore designs prior-specific fusion modules and introduces them into selected stages of the U-Net. Finally, mixed-dataset training and scheduled noise augmentation are employed to improve robustness to unseen low-light distributions.

We evaluate RePriFusion on five paired LLIE benchmarks and eight unseen datasets under the zero-shot setting. RePriFusion achieves state-of-the-art enhancement performance on most benchmarks.

In summary, our contributions are four-fold:

- We propose RePriFusion, a low-light image enhancement framework that integrates learned retinex-based prior features and frozen DINO features through selected restoration stages and prior-specific fusion modules.
- We design a Cascaded Retinex Prior Estimator CRPE that progressively extracts illumination and reflectance prior features and constructs an intermediate reconstruction, making their Retinex relation explicit in image space.
- We design prior-specific fusion modules that introduce illumination, reflectance, and DINO priors into selected U-Net stages using dedicated fusion operators tailored to their restoration roles.
- RePriFusion achieves state-of-the-art performance on four real-image paired LLIE benchmarks and eight unseen datasets under zero-shot evaluation.

2 Related Work

2.1 Deep Learning-Based LLIE

Deep learning-based LLIE has evolved from direct image-to-image mapping to model-guided and generative restoration. Retinex-based methods decompose images into illumination and reflectance components and learn their restoration under paired supervision Wei et al. (2018); Xu et al. (2020). Zero-reference, semi-supervised, and self-supervised methods reduce reliance on paired data through curve estimation, unpaired learning, or illumination consistency Guo et al. (2020); Yang et al. (2020); Ma et al. (2022), while deep unfolding combines Retinex optimization with learnable modules Liu et al. (2021b).

Recent methods improve degradation modeling and restoration architectures. SNR-Aware adapts convolutional and Transformer processing to spatial signal reliability Xu et al. (2022), while Retinexformer introduces illumination-guided attention into a one-stage Retinex framework Cai et al. (2023). Other directions explore structure-aware restoration Xu et al. (2023), alternative color representations Yan et al. (2025), and diffusion-based enhancement Lan et al. (2025).

2.2 Priors-Guided LLIE

Prior-guided LLIE incorporates physical, learned, or semantic knowledge into image enhancement. Retinex and SNR priors model illumination, reflectance, and spatial signal reliability Wei et al. (2018); Liu et al. (2021b); Xu et al. (2022); Cai et al. (2023). Physical quadruple priors enable

zero-reference enhancement using normal-light images Wang et al. (2024), while CUE learns illumination and noise priors from masked-autoencoder representations Zheng et al. (2023). Reti-Diff incorporates Retinex priors into a latent diffusion framework, whereas Multinex combines multiple Retinex priors in a lightweight architecture He et al. (2025); Brateanu et al. (2026).

Pretrained high-level guidance has also been explored in LLIE. SKF transfers knowledge from a pretrained segmentation network Wu et al. (2023), NeRCo employs vision-language guidance for unsupervised enhancement Yang et al. (2023), and CLIP-based prompt learning provides semantic and quality supervision without paired references Morawski et al. (2024). FoCo coordinates multiple vision foundation models for self-supervised enhancement and downstream perception Gu et al. (2025). However, how to adaptively incorporate heterogeneous priors into different restoration stages according to their characteristics remains largely unexplored.

3 METHOD

3.1 OVERVIEW

As illustrated in Fig. 1(a), RePriFusion consists of a cascaded Retinex prior estimator, a frozen DINOv3 encoder, and a U-shaped restoration network. Given an low-light image I, the Retinex estimator sequentially extracts illumination and reflectance prior features and constructs an intermediate reconstruction I'. In parallel, DINOv3 extracts multi-level pretrained visual features. These three priors are introduced into selected restoration stages through prior-specific fusion modules.

We use I' rather than the raw low-light image as the image-level input to the restoration network. This allows the Retinex relation to participate directly in image-space restoration instead of remaining only as feature-level guidance. The U-shaped network then predicts the remaining correction $\Delta I'$:

$$\hat{I} = I' + \Delta I' \tag{1}$$

3.2 CASCADED RETINEX PRIOR ESTIMATOR(CRPE)

We design a Cascaded Retinex Prior Estimator(CRPE) $\mathcal{E}_{\mathrm{ret}}$ to sequentially extract illumination and reflectance guidance from I. The two sub-branches are deliberately asymmetric because the Retinex components exhibit different image properties. Illumination mainly describes smooth spatial exposure variation and can be derived from local intensity cues. Reflectance contains richer color, texture, and high-frequency appearance information, but it is also more sensitive to noise in severely dark regions. We therefore first derive the illumination feature F_{ill} and map L, and then use L to initialize and condition reflectance extraction. This dependency defines the cascaded process shown in Fig. 1(b).

Illumination sub-branch. We first compute the channel-wise mean of the input:

$$\bar{I} = \mathrm{Mean}_c(I) \tag{2}$$

and concatenate it with the RGB image. The mean map provides an explicit intensity cue, while the RGB channels retain spatial and chromatic information. Their combination allows the branch to model local exposure variation without discarding scene content.

The concatenated representation is processed by a convolution followed by a depth-wise convolution:

$$F_{\mathrm{ill}} = \mathrm{DWConv}\left(\mathrm{Conv}\left([I, \bar{I}]\right)\right) \tag{3}$$

The convolution mixes the intensity and RGB cues, while the depth-wise operation captures local spatial context with limited overhead. A zero-initialized prediction head maps F_{ill} to a positive illumination map:

$$L = \mathrm{Softplus}\left(\mathrm{Conv}^0_{\mathrm{ill}}\left(F_{\mathrm{ill}}\right)\right) + \epsilon \tag{4}$$

where Conv^0 denotes a zero-initialized convolution and ϵ ensures numerical stability. Positivity makes the subsequent Retinex division well-defined.

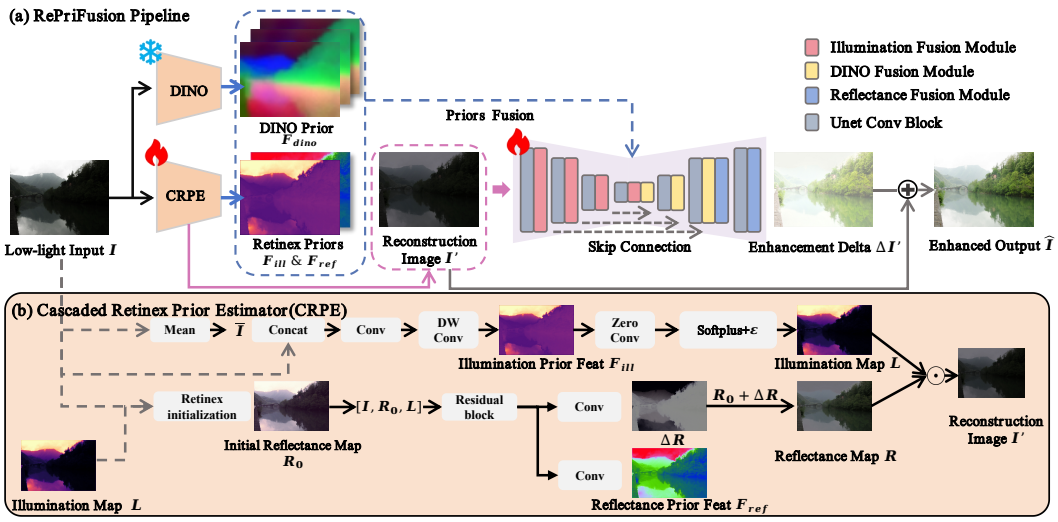


Figure 1: Overview of RePriFusion. (a) Overall pipeline. A frozen DINOv3 encoder and a trainable Retinex prior estimator extract DINO, illumination, and reflectance features from the low-light input I. These priors are incorporated into selected stages of a U-shaped restoration network, which predicts a residual $\Delta I'$ over the reconstruction image I' to produce $\hat{I}$. (b) Cascaded Retinex prior estimator(CRPE). The estimator sequentially predicts F_{ill} and L, constructs R_0, and estimates F_{ref} and ΔR, yielding $R = R_0 + \Delta R$ and $I' = R \odot L$. Snowflake and flame symbols denote frozen and trainable components, respectively.

Reflectance sub-branch. Given a illumination Map L, we initialize reflectance analytically according to the Retinex formulation:

$$R_0 = I/L \tag{5}$$

where the division is performed element-wise. This initialization provides an explicit appearance estimate and reduces the burden of recovering reflectance directly from a severely under-exposed input.

However, division by small illumination values can amplify noise and introduce decomposition artifacts in R_0. We therefore refine the initial estimate using the original input, illumination map, and reflectance initialization:

$$H_{\text{ref}} = \mathcal{B}_{\text{ref}}\left([I, R_0, L]\right) \tag{6}$$

where $\mathcal{B}_{\text{ref}}$ denotes the residual feature extraction block in Fig. 1(b). This combination preserves the original image evidence while providing the branch with both the analytical reflectance estimate and its corresponding illumination condition.

Two convolutional heads map the shared representation H_{ref} to a reflectance prior feature and an image-space correction:

$$\begin{aligned} F_{\text{ref}} &= \text{Conv}_{\text{ref}}\left(H_{\text{ref}}\right), \\ \Delta R &= \text{Conv}^{0}_{\Delta R}\left(H_{\text{ref}}\right), \\ R &= R_0 + \Delta R. \end{aligned} \tag{7}$$

The two heads serve different purposes: F_{ref} provides feature-level appearance guidance, whereas ΔR corrects the analytical reflectance in image space. Predicting a correction instead of a free reflectance map retains R_0 as a reference and focuses the learned component on errors caused by noise amplification and imperfect decomposition.

Image-level Retinex reconstruction. The corrected reflectance is recombined with illumination:

$$\begin{aligned} I' &= R \odot L \\ &= (R_0 + \Delta R) \odot L \\ &= I + \Delta R \odot L \end{aligned} \tag{8}$$

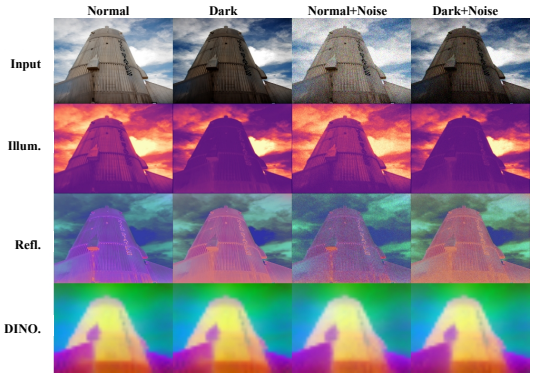


Figure 2: Qualitative analysis of prior features. Illumination mainly responds to exposure variation, reflectance retains input-dependent appearance while being more degradation-sensitive, and frozen DINO features remain comparatively stable.

where $\odot$ denotes element-wise multiplication. The last equality shows that the reconstruction image I' is an input-anchored Retinex correction rather than an independently generated image. Because the correction head is zero-initialized, $\Delta R = 0$ at the beginning of training and $I' = I$ up to numerical precision.

Feeding I' into the U-shaped network places the estimator's image-space output directly on the restoration path. If the raw image I were used instead, the Retinex outputs would affect restoration only through feature modulation. By constructing $I' = R \odot L$, the final enhancement objective also optimizes the coupled image-space output of the illumination and reflectance branches. Although this formulation does not uniquely identify a physical Retinex decomposition, it introduces an additional image-space coupling between the two estimator branches. As shown in Tab. 6, removing this reconstruction path consistently degrades restoration performance. The representation analysis in Appendix C further shows that the learned branches exhibit distinct exposure- and appearance-related responses.

3.2.1 DINO Prior Extraction

We employ a frozen DINOv3 ViT-S/16 encoder $\mathcal{E}_{\text{dino}}$ Siméoni et al. (2025) to extract multi-level pretrained visual features from the low-light input. Let $\mathcal{S} = \{s_1, s_2, \ldots, s_K\}$ denote the selected transformer blocks. The resulting DINO prior features are

$$\mathcal{F}_{\text{dino}} = \left\{ F_{\text{dino}}^{(s)} = \mathcal{E}_{\text{dino}}^{(s)}(I) \mid s \in \mathcal{S} \right\} \tag{9}$$

Features from different transformer depths provide complementary levels of visual abstraction. Intermediate blocks retain richer spatial and local cues, whereas deeper blocks encode more contextual information. RePriFusion therefore uses multi-level DINO features at selected stages of the restoration network. In practice, we extract features from the 6th, 9th, and 12th transformer blocks, and reuse the 12th-block feature for two restoration stages.

3.3 Prior-Specific Fusion

Before introducing the fusion modules, we qualitatively inspect the three prior features. As shown in Fig. 2, the illumination feature responds primarily to exposure variation, the reflectance feature retains richer input-dependent appearance cues but is more sensitive to degradation, and the frozen DINO feature remains comparatively stable. These complementary characteristics motivate RePriFusion to coordinate the three priors through prior-specific fusion operators at selected stages of the restoration network. We provide more analysis in the Appendix C.

Illumination Fusion Module. As shown in Fig. 3(a), the illumination prior feature F_{ill} is fused with the encoder stages and bottleneck to condition feature extraction on spatial exposure variation.

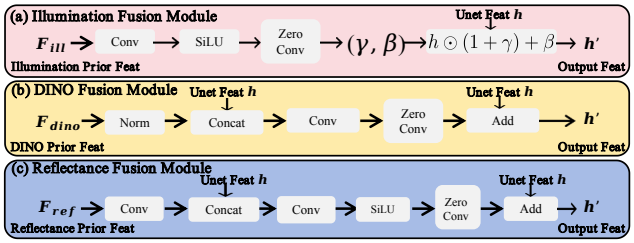


Figure 3: Prior-specific fusion modules in RePriFusion. **(a)** Illumination Fusion Module. **(b)** DINO Fusion Module. **(c)** Reflectance Fusion Module.

For a restoration feature h, F_{ill} is resized to the corresponding resolution and projected into spatial affine parameters:

$$\begin{aligned}(\gamma, \beta) &= \mathcal{Z}_{\text{ill}}\left(\sigma\left(\text{Conv}(F_{\text{ill}})\right)\right) \\ h' &= h \odot (1+\gamma) + \beta\end{aligned} \tag{10}$$

where σ denotes the SiLU activation and $\mathcal{Z}_{\text{ill}}$ is a zero-initialized projection. The multiplicative and additive terms adapt the feature response to local illumination conditions, while zero initialization preserves the original restoration path at the beginning of training. The modulated encoder features are further propagated to the decoder through skip connections.

DINO Fusion Module. As shown in Fig. 3(b), multi-level DINO prior features are fused with the bottleneck and selected low- to mid-resolution decoder stages. For each restoration stage, the corresponding DINO feature $F_{\text{dino}}^{(s)}$ is resized to the target resolution and normalized along the channel dimension:

$$\widehat{F}_{\text{dino}}^{(s)} = \ell_2\text{-norm}\left(F_{\text{dino}}^{(s)}\right) \tag{11}$$

The normalized prior is then incorporated through a residual fusion branch:

$$h' = h + \mathcal{Z}_{\text{dino}}\left(\text{Conv}\left([h, \widehat{F}_{\text{dino}}^{(s)}]\right)\right) \tag{12}$$

where $[\cdot, \cdot]$ denotes channel-wise concatenation and $\mathcal{Z}_{\text{dino}}$ is a zero-initialized projection. Normalization reduces the scale discrepancy between the frozen DINO representation and the trainable restoration feature, while residual fusion preserves the original feature path and introduces complementary pretrained visual information.

Reflectance Fusion Module. As shown in Fig. 3(c), the reflectance prior feature F_{ref} is fused with selected high-resolution decoder stages to support appearance reconstruction. After being resized to the target resolution, it is incorporated through a residual branch:

$$h' = h + \mathcal{Z}_{\text{ref}}\left(\text{Conv}\left([h, F_{\text{ref}}]\right)\right) \tag{13}$$

where $\mathcal{Z}_{\text{ref}}$ denotes a zero-initialized projection. Unlike the frozen DINO feature, F_{ref} is learned jointly with the restoration network and is therefore fused without explicit channel normalization. The residual formulation preserves the original decoder feature path while introducing input-dependent appearance cues for local color and detail refinement.

3.4 Training Objective and Strategy

Training objective. Given a low-light input I and its normal-light reference Y, we supervise the final enhanced image $\hat{I}$ using a combination of pixel, structural, and perceptual losses:

$$\begin{aligned}\mathcal{L} = \lambda_1 \|\hat{I} - Y\|_1 + \lambda_2 \left(1 - \text{MS-SSIM}(\hat{I}, Y)\right) \\ + \lambda_3 \,\text{LPIPS}(\hat{I}, Y)\end{aligned} \tag{14}$$

where $\lambda_1 = \lambda_2 = 1.0$ and $\lambda_3 = 0.4$. All trainable components, including the Retinex prior estimator, restoration backbone, projection layers, and fusion modules, are optimized jointly, while the DINOv3 encoder remains frozen.

Table 1: **Full-reference quantitative comparison on the paired low-light image enhancement benchmarks.** Higher PSNR (dB) and SSIM are better. FLOPs are measured on 480×640 inputs. * denotes baselines retrained with our mixed-dataset protocol and noise augmentation. Best and second-best results are highlighted in red and blue, respectively.

Method	Complexity		LOLv1		LOLv2-Real		LOLv2-Syn		SDSD		SMID	
	FLOPs (G)	Params. (M)	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RetinexNet Wei et al. (2018)	5871.47	0.84	16.77	0.560	15.47	0.567	17.13	0.798	20.90	0.623	22.83	0.684
FIDE Xu et al. (2020)	280.51	8.62	18.27	0.665	16.85	0.678	15.20	0.612	22.31	0.644	24.42	0.692
KinD Zhang et al. (2019)	342.01	81.02	20.86	0.790	14.74	0.641	13.29	0.578	21.96	0.663	22.18	0.634
EnGAN Jiang et al. (2021)	620.11	114.35	17.48	0.650	18.23	0.617	16.57	0.734	20.06	0.610	22.62	0.674
Restormer Zamir et al. (2022)	1441.25	26.13	22.43	0.823	19.94	0.827	21.41	0.830	24.73	0.815	25.17	0.758
Retinexformer Cai et al. (2023)	159.54	1.61	27.17	0.850	22.79	0.840	**29.02**	**0.939**	25.28	**0.856**	18.58	0.719
CIDNet Yan et al. (2025)	38.12	1.97	26.32	0.850	27.51	0.867	27.11	0.913	24.09	0.811	22.40	0.739
Multinex Brateanu et al. (2026)	11.90	0.15	26.13	0.841	27.36	0.854	26.96	0.915	23.59	0.830	22.91	0.746
ReDDiT Lan et al. (2025)	1256.61	17.40	27.41	0.862	29.84	0.888	27.76	0.918	23.20	0.808	23.11	0.756
Reti-Diff He et al. (2025)	1049.6	26.20	21.39	0.738	30.31	0.888	**28.27**	**0.934**	25.06	0.841	23.27	0.754
ReDDiT* Lan et al. (2025)	1256.61	17.40	27.52	0.863	29.99	0.885	26.36	0.901	25.44	0.842	24.22	0.749
Reti-Diff* He et al. (2025)	1049.6	26.20	24.81	0.832	30.20	0.881	27.04	0.919	25.54	0.847	23.73	0.729
RePriFusion-Base	810.04	85.96	**28.12**	**0.871**	**31.55**	**0.891**	26.53	0.911	**28.87**	**0.874**	**26.85**	**0.783**
RePriFusion-Small	287.03	39.09	**27.81**	**0.870**	**31.35**	**0.890**	25.08	0.890	**26.69**	0.855	**26.12**	**0.775**

Table 2: **Zero-shot quantitative comparison on unseen low-light image enhancement benchmarks.** We report PSNR/SSIM on paired datasets and NIMA/MUSIQ on unpaired datasets; all metrics are higher-is-better. * denotes baselines retrained with our mixed-dataset protocol and noise augmentation. Best and second-best results are highlighted in red and blue, respectively.

Methods	NT2024		FiveK		SID		DICM		LIME		MEF		NPE		VV	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	NIMA ↑	MUSIQ ↑	NIMA ↑	MUSIQ ↑	NIMA ↑	MUSIQ ↑	NIMA ↑	MUSIQ ↑	NIMA ↑	MUSIQ ↑
CIDNet Yan et al. (2025)	18.491	0.708	20.240	0.792	14.514	0.245	**4.677**	59.027	4.334	58.106	4.372	62.815	4.221	60.146	4.346	51.173
Multinex Brateanu et al. (2026)	18.436	0.714	17.633	0.718	13.688	0.242	4.607	58.213	4.398	58.046	4.494	61.949	4.376	60.908	4.353	50.141
ReDDiT Lan et al. (2025)	18.784	0.712	19.797	0.779	15.647	0.265	4.573	56.412	4.307	56.278	4.222	58.809	4.357	**61.415**	4.203	52.335
Reti-Diff He et al. (2025)	17.467	0.701	17.748	0.727	14.583	0.256	4.539	55.940	**4.544**	57.459	4.201	57.702	4.325	59.121	4.367	53.226
ReDDiT* Lan et al. (2025)	18.961	0.720	20.330	0.789	19.150	**0.625**	4.580	57.890	4.410	58.227	4.415	60.300	4.276	61.348	4.340	53.450
Reti-Diff* He et al. (2025)	17.870	0.722	19.920	0.780	19.020	0.603	4.588	58.450	4.532	57.950	4.489	61.122	4.377	61.043	4.406	54.087
RePriFusion-Base	**19.356**	**0.744**	**20.870**	**0.803**	**19.674**	**0.777**	**4.649**	**62.116**	**4.537**	**60.016**	**4.669**	**64.514**	**4.479**	**61.657**	**4.536**	**56.719**
RePriFusion-Small	**19.513**	**0.740**	**20.448**	**0.796**	**20.354**	0.608	4.618	**61.546**	4.403	58.278	4.505	**63.626**	4.360	61.088	**4.411**	**54.872**

4 Experiments and Results

4.1 Datasets.

RePriFusion is trained once on the combined training splits of five paired LLIE benchmarks: LOLv1 Wei et al. (2018), LOLv2-Real and LOLv2-Synthetic Liu et al. (2021a), SDSD Wang et al. (2021b), and SMID Jiang & Zheng (2019). Their official test splits are used for paired benchmark evaluation. To assess cross-dataset generalization, we further evaluate the trained model on three unseen paired datasets: NT2024 Liu et al. (2024), FiveK Bychkovsky et al. (2011), and SID Chen et al. (2018), and five unseen unpaired datasets: DICM Lee et al. (2012), LIME Guo et al. (2017), MEF Mertens et al. (2007), NPE Wang et al. (2013), and VV Vonikakis et al. (2018). None of these eight zero-shot datasets is used for training, and no fine-tuning or dataset-specific adaptation is performed.

4.2 Low-Light Image Enhancement

4.2.1 Full-Reference Quantitative Comparison

Tab.1 compares RePriFusion with representative LLIE methods on five paired benchmarks. RePriFusion is trained once on the mixed training set and evaluated on the official test split of each dataset.

RePriFusion-Base achieves the best PSNR and SSIM on all four real-captured benchmarks. Compared with the strongest competing result, it improves PSNR by 0.60 dB on LOLv1, 1.24 dB on LOLv2-Real, 3.33 dB on SDSD and 1.68 dB on SMID. RePriFusion-Small retains strong performance while reducing FLOPs by 64.6% relative to RePriFusion-Base. On LOLv2-Synthetic, exist-



Figure 4: Qualitative comparison of state-of-the-art LLIE methods and RePriFusion on different benchmarks.

ing methods remain superior, indicating that the advantage of RePriFusion is more pronounced on real low-light images. Qualitative comparisons are shown in Fig. 4.

4.2.2 Zero-Shot Quantitative Comparison

Tab.2 evaluates RePriFusion on eight unseen datasets without fine-tuning or dataset-specific adaptation. On the three paired datasets, RePriFusion-Base achieves the best SSIM on NT2024, FiveK, and SID, as well as the best PSNR on FiveK. RePriFusion-Small achieves the best PSNR on NT2024 and SID while remaining competitive in structural similarity.

On the five unpaired datasets, RePriFusion-Base achieves the highest MUSIQ score in every case and ranks first or second in NIMA across all benchmarks. RePriFusion-Small also maintains competitive no-reference quality. These results show that RePriFusion transfers effectively to unseen low-light distributions beyond paired evaluation settings. We provide more quantitative visualizations in the Appendix F.

4.3 Ablation Studies

4.3.1 Prior Component Ablation

We evaluate the contribution and complementarity of the three prior features using the RePriFusion-Small backbone. All variants share the same Retinex reconstruction, restoration network, and training protocol; only the prior features enabled for fusion are changed. Each active prior retains its adopted fusion stage and operator.

As shown in Tab. 3, each prior improves PSNR over the baseline on all three benchmarks. Illumination provides consistent gains in both PSNR and SSIM, while DINO achieves the strongest single-prior result on the unseen FiveK dataset, improving PSNR/SSIM from $18.72/0.754$ to $19.86/0.779$. Reflectance yields smaller standalone gains and slightly reduces SSIM on LOLv2-Real, suggesting that it is more effective when combined with complementary illumination or pretrained guidance.

Pairwise combinations generally outperform their corresponding single-prior variants. Illumination and DINO form the strongest pair on LOLv1 and LOLv2-Real, reaching $27.45/0.864$ and $30.92/0.884$, respectively. On FiveK, reflectance with DINO achieves the second-best PSNR, while illumination with DINO obtains the second-best SSIM. Using all three priors yields the best result on every metric, improving PSNR over the baseline by 1.01, 2.04, and 1.73 dB on LOLv1, LOLv2-Real, and FiveK, respectively. These results support the complementarity of the three prior representations.

Table 3: **Ablation of prior components.** All experiments use the RePriFusion-Small setting. Only the prior features enabled for fusion are changed, and each active prior uses its adopted fusion stage and operator. "Illum." and "Refl." denote the illumination and reflectance priors, respectively. The best and second-best results are highlighted in red and blue, respectively.

Configuration	LOL-v1		LOL-v2-Real		FiveK	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Baseline U-Net	26.80	0.844	29.31	0.872	18.72	0.754
Illum. only	27.01	0.852	29.94	0.880	19.27	0.765
DINO only	27.00	0.850	29.96	0.876	19.86	0.779
Refl. only	26.95	0.848	29.87	0.870	19.24	0.762
Illum. + DINO	**27.45**	**0.864**	**30.92**	**0.884**	20.16	**0.787**
Illum. + Refl.	27.31	0.861	30.54	0.881	19.62	0.778
Refl. + DINO	27.22	0.856	30.65	0.879	**20.20**	0.786
All priors	**27.81**	**0.870**	**31.35**	**0.890**	**20.45**	**0.796**

Table 4: **Comparison of prior-stage configurations.** F and P denote FLOPs (G) and parameters (M), respectively. All experiments use the RePriFusion-Small setting. "All → Enc./Dec./Full" introduces all three priors into the encoder, decoder, or entire U-Net, respectively. Single-prior relocation changes only the indicated prior while retaining the adopted configuration for the remaining priors. The ↔ symbol denotes an exchange of stage assignments between two priors.

Prior-Stage Configuration	Complexity		LOL-v1		LOLv2-Real	
	F	P	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Uniform stage assignment						
All → Enc.	330	41.0	27.53	0.864	30.99	0.885
All → Dec.	380	38.9	27.44	0.859	30.74	0.887
All → Full U-Net	444	45.7	**27.71**	0.867	**31.26**	**0.889**
Single-prior relocation						
DINO → Enc.	320	39.2	27.38	0.853	30.45	0.873
Refl. → Enc.	289	40.9	27.70	**0.868**	31.25	**0.890**
Illum. → Dec.	280	38.9	27.59	0.863	31.05	0.888
Pairwise stage exchange						
Illum. ↔ DINO	320	39.0	27.45	0.861	30.87	0.886
Illum. ↔ Refl.	280	40.8	27.61	0.864	30.77	0.879
RePriFusion-Small	287	39.0	**27.81**	**0.870**	**31.35**	**0.890**

4.3.2 Prior-Stage Configuration Ablation

We compare different prior-stage configurations using the RePriFusion-Small backbone. The prior definitions, fusion formulations, and training protocol are kept fixed, while the target fusion stages are varied. Moving a prior to a different stage may require additional channel projection or spatial resampling to match the receiving feature. The resulting variants therefore differ in parameter count and FLOPs and are evaluated as performance–complexity trade-offs rather than strictly capacity-matched ablations.

As shown in Tab. 4, assigning all priors uniformly to either the encoder or decoder yields lower performance. Extending all priors throughout the full U-Net achieves comparable accuracy but increases the cost to 444 G FLOPs and 45.7 M parameters. In comparison, the adopted RePriFusion-Small configuration achieves the best overall results with 287 G FLOPs and 39.0 M parameters, reducing computation by 35.5%. It improves PSNR over full-U-Net fusion by 0.10 dB on LOLv1 and 0.09 dB on LOLv2-Real. These results indicate that dense multi-stage fusion is unnecessary and that the adopted configuration provides a more favorable performance–complexity trade-off.

Among the relocation variants, moving illumination to the decoder consistently reduces performance. Relocating reflectance to the encoder remains competitive, particularly on LOLv2-Real, but does not improve the overall results. Moving DINO to the encoder also performs worse on both benchmarks. Pairwise exchanges involving illumination further reduce overall performance. These observations suggest that the three priors exhibit different sensitivities to fusion stages, while the adopted configuration provides the strongest overall balance across the two evaluated benchmarks.

5 Conclusion

We presented RePriFusion, a Retinex-Based Low-Light Image Enhancement via Prior-Specific Fusion. A cascaded Retinex estimator jointly produces illumination and reflectance prior features together with an input-anchored reconstruction, while a frozen DINOv3 encoder provides multi-level pretrained visual features. These heterogeneous priors are incorporated through prior-specific fusion modules at selected stages of a U-shaped restoration network. Extensive experiments on five paired benchmarks and eight unseen datasets demonstrate strong enhancement quality and cross-dataset generalization.

AI USE STATEMENT

(This section is **required** and does not count toward the page limit.)

In this work, we used generative AI tools solely for language polishing and grammar checking. We did not use generative AI tools for research ideation, method or experimental design, code writing, experiment execution, or the analysis and interpretation of experimental results. All AI-assisted edits were manually reviewed by the authors to ensure technical accuracy and consistency with the intended meaning. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

See the ICLR 2027 AI Policy for Authors for more details. This statement should not be more than 1 page.

REPRODUCIBILITY STATEMENT

We facilitate reproducibility by specifying the RePriFusion architecture, the construction of the Retinex priors, the extraction of DINO features, the prior-specific fusion operations, and the training objective in Section 3. Section 4 documents the training and evaluation datasets, zero-shot protocol, evaluation metrics, and settings used in the main ablations. Appendix A reports the optimization schedule, batch size, training duration, hardware, and frozen-encoder setup. Appendices B, C, and D provide the downstream depth-evaluation protocol, prior-representation analysis, and additional controlled ablations supporting the reported results.

REFERENCES

Jinwoo Bae, Sungho Moon, and Sunghoon Im. Deep digging into the generalization of self-supervised monocular depth estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 187–196, 2023. doi: 10.1609/aaai.v37i1.25090.

Alexandru Brateanu, Tingting Mu, Codruta O. Ancuti, and Cosmin Ancuti. Multinex: Lightweight low-light image enhancement via multi-prior retinex, 2026.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 97–104, 2011.

Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11618–11628, 2020.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12470–12479, 2023.

Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3291–3300, 2018.

Yuxuan Gu, Haoxuan Wang, Pengyang Ling, Zhixiang Wei, Huaian Chen, Yi Jin, and Enhong Chen. Improving visual and downstream performance of low-light enhancer with vision foundation models collaboration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16071–16080, 2025.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1777–1786, 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.

Chunming He, Chengyu Fang, Yulun Zhang, Longxiang Tang, Jinfa Huang, Kai Li, Zhenhua Guo, Xiu Li, and Sina Farsiu. Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. In *International Conference on Learning Representations (ICLR)*, 2025.

Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):10579–10596, 2024. doi: 10.1109/TPAMI.2024.3444912. URL https://doi.org/10.1109/TPAMI.2024.3444912.

Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7323–7332, 2019.

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.

Guanzhou Lan, Qianli Ma, Yuqi Yang, Zhigang Wang, Dong Wang, Xuelong Li, and Bin Zhao. Efficient diffusion as low light enhancer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21277–21286, 2025.

Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *IEEE International Conference on Image Processing (ICIP)*, pp. 965–968, 2012.

Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129(4):1153–1184, 2021a.

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 *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pp. 10561–10570. Computer Vision Foundation / IEEE, 2021b.

Xiaoning Liu, Zongwei Wu, Ao Li, et al. NTIRE 2024 challenge on low light image enhancement: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 6571–6594, 2024.

Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 5627–5636. IEEE, 2022.

Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017.

Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Pacific Graphics*, pp. 382–390, 2007.

Igor Morawski, Kai He, Shusil Dangi, and Winston H. Hsu. Unsupervised image prior via prompt learning and CLIP semantic guidance for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 5971–5981, 2024.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3, 2025. URL https://arxiv.org/abs/2508.10104.

Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77(8):9211–9231, 2018.

Kun Wang, Zhenyu Zhang, Zhiqiang Yan, Xiang Li, Baobei Xu, Jun Li, and Jian Yang. Regularizing nighttime weirdness: Efficient self-supervised monocular depth estimation in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16055–16064, 2021a.

Ruicheng Wang, Sicheng Xu, Yue Dong, Yu Deng, Jianfeng Xiang, Zelong Lv, Guangzhong Sun, Xin Tong, and Jiaolong Yang. Moge-2: Accurate monocular geometry with metric scale and sharp details, 2025.

Ruixing Wang, Xiaogang Xu, Chi-Wing Fu, Jiangbo Lu, Bei Yu, and Jiaya Jia. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9680–9689, 2021b.

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.

Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26057–26066, 2024.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference (BMVC)*, pp. 155, 2018.

Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1662–1671, 2023.

Ke Xu, Xin Yang, Baocai Yin, and Rynson W. H. Lau. Learning to restore low-light images via decomposition-and-enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2278–2287, 2020.

Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. SNR-aware low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17693–17703, 2022.

Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9893–9903, 2023.

Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. HVI: A new color space for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5678–5687, 2025.

Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12872–12881, 2023.

Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3060–3069, 2020.

Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with A generative diffusion model. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pp. 12268–12277. IEEE, 2023.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5718–5729, 2022.

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM International Conference on Multimedia (ACM MM)*, pp. 1632–1640, 2019.

Naishan Zheng, Man Zhou, Yanmeng Dong, Xiangyu Rui, Jie Huang, Chongyi Li, and Feng Zhao. Empowering low-light image enhancer through customized learnable priors. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12525–12535, 2023.

A Training Details

RePriFusion uses a frozen DINOv3 ViT-S/16 encoder to extract multi-level visual prior features. The unified Retinex prior estimator and the five-level U-shaped restoration network are jointly optimized together with all projection and fusion modules, while the DINOv3 encoder remains fixed. We use AdamW with an initial learning rate of 1×10^{-4} and a cosine decay schedule. The model is trained for approximately 90,000 iterations with an effective batch size of 32 on a single NVIDIA GeForce RTX 5090 GPU.

To improve cross-dataset generalization, RePriFusion is trained on the combined training splits of five paired LLIE datasets rather than using a separate model for each benchmark. This exposes the network to more diverse cameras, scenes, and illumination conditions during optimization. We further augment the low-light inputs with Poisson noise at a fixed intensity of 0.1 to reduce overfitting to dataset-specific noise patterns and improve robustness to real low-light degradations.

B Downstream Nighttime Depth Estimation

We evaluate whether low-light enhancement benefits downstream monocular depth estimation on nuScenes-Night Caesar et al. (2020) and RobotCar-Night Maddern et al. (2017). Following established nighttime depth evaluation protocols Wang et al. (2021a); Bae et al. (2023), we feed the enhanced images into fixed MoGe v2 Wang et al. (2025) and Metric3D v2 Hu et al. (2024) models without task-specific fine-tuning or adaptation.

As shown in Tab.5, RePriFusion achieves the best performance under both fixed depth estimators. With MoGe v2, it improves AbsRel/δ_1 from $0.284/0.644$ to $0.204/0.740$ on nuScenes and from $0.162/0.788$ to $0.136/0.834$ on RobotCar. Compared with the strongest enhancement baseline, RePriFusion reduces AbsRel by 15.0% and 12.3%, respectively. With Metric3D v2, it achieves AbsRel values of 0.178 and 0.122, corresponding to reductions of 28.5% and 18.7% over the strongest enhancement baselines.

Table 5: **Downstream nighttime monocular depth estimation.** Enhanced nuScenes-Night Caesar et al. (2020) and RobotCar-Night Maddern et al. (2017) images are evaluated using fixed MoGe v2 Wang et al. (2025) and Metric3D v2 Hu et al. (2024) models without task-specific fine-tuning. RNW Wang et al. (2021a) and MonoFormer Bae et al. (2023) are reported as nighttime-specific depth methods for reference. The best and second-best results within each fixed depth estimator are highlighted in red and blue, respectively.

Enhancement	nuScenes		RobotCar	
	AbsRel ↓	δ_1 ↑	AbsRel ↓	δ_1 ↑
Nighttime-Specific Methods				
RNW	0.376	0.451	0.237	0.683
MonoFormer	0.307	0.521	0.289	0.543
Depth estimator: MoGe v2				
No enhancement	0.284	0.644	0.162	0.788
CIDNet	0.259	0.638	0.186	0.738
ReDDiT	**0.240**	**0.726**	**0.155**	**0.810**
Reti-Diff	0.277	0.625	0.174	0.772
Multinex	0.289	0.546	0.191	0.761
RePriFusion	**0.204**	**0.740**	**0.136**	**0.834**
Depth estimator: Metric3D v2				
No enhancement	0.335	0.791	0.165	0.852
CIDNet	0.264	0.799	0.151	0.873
ReDDiT	0.261	**0.802**	0.180	0.814
Reti-Diff	**0.249**	0.789	**0.150**	0.874
Multinex	0.253	0.801	0.154	**0.876**
RePriFusion	**0.178**	**0.806**	**0.122**	**0.880**

These results indicate that RePriFusion preserves visual structures that are more compatible with existing monocular depth estimators. The consistent gains across two datasets and two fixed depth models further suggest that the enhanced images retain task-relevant cues for nighttime geometric perception.

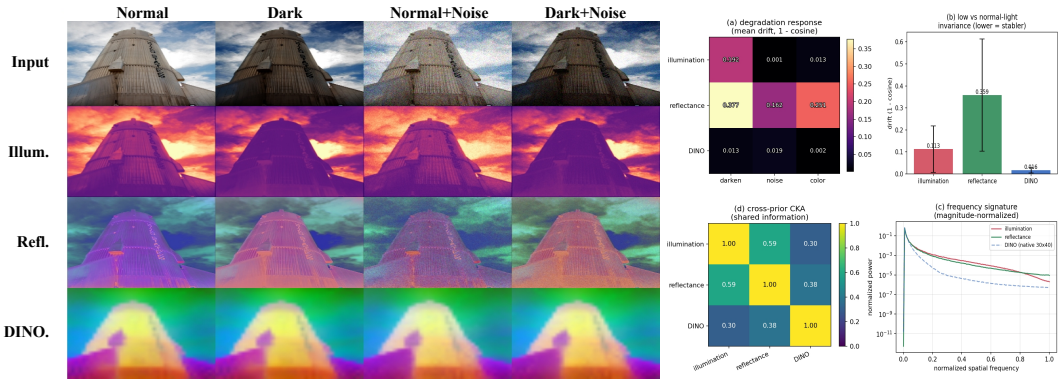


Figure 5: Analysis of prior representations learned by RePriFusion. *Left:* Visualizations of a representative scene under normal, darkened, noisy, and jointly darkened-and-noisy conditions. The illumination feature is visualized using its activation magnitude, while the reflectance and DINO features are projected onto their top three principal components. *Right:* (a) mean feature drift, measured by $1 - \text{cosine}$ similarity, under darkening, noise, and color shift; (b) feature drift between paired low- and normal-light images of the same scene; (c) magnitude-normalized radial power spectrum; and (d) linear CKA similarity.

C Analysis of Learned Prior Representations

We analyze the prior representations learned by RePriFusion on 10K images sampled from the training and evaluation datasets. As shown in Fig.5, we examine their responses to image degradations, variation across illumination conditions, spatial-frequency characteristics, and representational similarity. This analysis characterizes the features learned by the instantiated RePriFusion model rather than universal properties of arbitrary prior extractors.

The illumination prior is most responsive to exposure variation. Its feature drift, measured by $1 - \text{cosine}$ similarity, is 0.192 under darkening, but only 0.001 under noise and 0.013 under color shift. The drift between paired low- and normal-light images is 0.113. This selective response is consistent with the use of illumination features for spatial exposure conditioning.

The reflectance prior is more sensitive to input degradations, with feature drifts of 0.377, 0.162, and 0.251 under darkening, noise, and color shift, respectively. It also exhibits the largest drift between paired low- and normal-light images, reaching 0.359, and retains stronger high-frequency responses. These results indicate that the reflectance representation preserves rich input-dependent appearance information, while also carrying degradation-related variations.

The frozen DINO prior is comparatively stable across the evaluated perturbations, with feature drifts of 0.013, 0.019, and 0.002 under darkening, noise, and color shift, respectively. Its drift between paired low- and normal-light images is also low at 0.016. Moreover, its radial spectrum is concentrated more strongly in low-frequency components. These results suggest that DINO provides a stable pretrained representation that is less affected by low-level photometric degradation.

The CKA analysis further shows that the frozen DINO prior has limited representational overlap with the two learned Retinex priors. Its CKA similarities with illumination and reflectance are 0.30 and 0.38, respectively, whereas the similarity between illumination and reflectance reaches 0.59. This indicates that DINO contributes complementary information beyond the two Retinex-derived representations.

Overall, the three priors exhibit complementary representation characteristics: illumination is primarily responsive to exposure variation, reflectance preserves richer but degradation-sensitive appearance cues, and DINO provides comparatively stable pretrained features with limited overlap with the Retinex representations.

Table 6: **Ablation of the cascaded Retinex prior estimator.** All variants use RePriFusion-Small with identical prior fusion modules, loss functions, and training settings. "Parallel branches" independently predict illumination and reflectance from the input. "Direct reflectance regression" predicts R without the analytical initialization R_0, while "w/o Reconstruction image" replaces $\hat{I} = I' + \Delta I'$ with $\hat{I} = I + \Delta I$. The best and second-best results are highlighted in red and blue, respectively.

Variant	Illum.-Cond. Reflectance	Analytical R_0	Residual ΔR	Reconstruction image I'	Zero-Init. ΔR Head	LOL-v1		LOL-v2-Real	
						PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Parallel Retinex branches	–	–	–	✓	–	27.47	0.856	30.77	0.876
Direct reflectance regression	✓	–	–	✓	–	27.03	0.841	29.85	0.862
w/o Reflectance Correction	✓	✓	–	✓	–	27.33	0.850	30.52	0.867
w/o Reconstruction image	✓	✓	✓	–	✓	27.21	0.850	30.05	0.861
w/o Zero Initialization	✓	✓	✓	✓	–	**27.69**	**0.865**	**31.02**	**0.883**
RePriFusion-Small	✓	✓	✓	✓	✓	**27.81**	**0.870**	**31.35**	**0.890**

D Additional Ablation Studies

Note. Unless otherwise specified, all ablations use the RePriFusion-Small configuration.

D.1 Cascaded Retinex Estimator(CRPE) Ablation

We ablate the main components of the cascaded Retinex prior estimator using the RePriFusion-Small configuration. All variants retain the same restoration backbone, prior fusion modules, loss functions, and training protocol. "Parallel Retinex branches" estimates illumination and reflectance independently from the input, removing the illumination-conditioned cascade. "Direct reflectance regression" retains illumination conditioning but directly predicts R without the analytical initialization $R_0 = I/L$. "w/o Reflectance Correction" removes the residual correction by setting $R = R_0$. "w/o Reconstruction image" retains the Retinex priors but replaces the input-anchored formulation $\hat{I} = I' + \Delta I'$ with $\hat{I} = I + \Delta I$. Finally, "w/o Zero Initialization" replaces the zero-initialized ΔR prediction head with a conventionally initialized convolution, while leaving all other components unchanged.

As shown in Tab.6, the complete estimator achieves the best performance on both datasets. Replacing the cascaded estimator with parallel Retinex branches reduces PSNR by 0.34 dB on LOL-v1 and 0.58 dB on LOL-v2-Real, supporting the use of illumination-conditioned reflectance estimation. Directly regressing reflectance causes the largest PSNR degradation, decreasing the results from $27.81/0.870$ to $27.03/0.841$ on LOL-v1 and from $31.35/0.890$ to $29.85/0.862$ on LOL-v2-Real. This comparison support the benefit of constructing reflectance from the analytical initialization R_0 rather than learning it entirely from scratch.

Removing the reflectance correction also consistently degrades performance, showing that the analytical initialization benefits from a learned residual refinement. Moreover, replacing the Retinex reconstruction image I' with the original low-light input I reduces PSNR by 0.60 dB on LOL-v1 and 1.30 dB on LOL-v2-Real, indicating that I' provides a more effective starting point for subsequent restoration. Standard initialization results in smaller but consistent decreases on both datasets. Overall, these results validate the cascaded estimation, analytical reflectance initialization, residual correction, input-anchored reconstruction, and identity-preserving initialization used in the Retinex estimator.

D.2 Fusion Operator Ablation

We examine the effect of the fusion operator used for each prior while keeping the prior components, adopted fusion stages, restoration backbone, and training objective fixed. For each variant, only one operator is replaced: illumination affine modulation is substituted with addition or concatenation, whereas the DINO and reflectance residual modules are replaced with FiLM or direct addition.

As shown in Tab.7, RePriFusion-Small achieves the best overall performance on both datasets. Replacing illumination affine modulation with concatenation reduces PSNR by 0.15 and 0.21 dB on

Table 7: **Ablation of fusion operators.** For each variant, only the indicated fusion operator is replaced, while the prior components, adopted fusion stages, backbone, and training objective remain unchanged. The best and second-best results are highlighted in red and blue, respectively. Ties are highlighted identically.

Fusion Variant	LOL-v1		LOL-v2-Real	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Illum. → Addition	27.43	0.859	30.75	0.869
Illum. → Concatenation	27.66	**0.863**	**31.14**	**0.885**
Refl. → FiLM	**27.70**	**0.870**	30.70	0.868
DINO → FiLM	27.60	0.860	31.04	0.880
Refl. → Addition	27.44	0.854	30.35	0.862
DINO → Addition	27.50	0.861	30.57	0.862
RePriFusion-Small	**27.81**	**0.870**	**31.35**	**0.890**

Table 8: **Ablation study of the training objective.** The network architecture and training settings are fixed, while the loss terms are progressively added. Higher values are better. The best and second-best results are highlighted in red and blue, respectively.

Training Objective	LOL-v1		LOL-v2-Real	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
$\mathcal{L}_1$	27.01	0.843	30.05	0.864
$\mathcal{L}_1 + \mathcal{L}_{\text{MS-SSIM}}$	**27.45**	**0.859**	**30.64**	**0.879**
$\mathcal{L}_1 + \mathcal{L}_{\text{MS-SSIM}} + \mathcal{L}_{\text{LPIPS}}$	**27.81**	**0.870**	**31.35**	**0.890**

LOL-v1 and LOL-v2-Real, respectively, while direct addition causes larger degradation. Replacing the DINO or reflectance residual module with FiLM or direct addition also lowers most metrics. Although reflectance FiLM matches the full model in SSIM on LOL-v1, it degrades to $30.70/0.868$ on LOL-v2-Real. Overall, the results support using prior-specific fusion operators rather than applying a uniform interaction pattern to all three representations.

D.3 Loss Ablation

We evaluate the contribution of each loss term by progressively adding MS-SSIM and LPIPS to the ℓ_1 reconstruction objective. The network architecture, prior components, fusion stages, and fusion operators are kept unchanged across all variants.

As shown in Tab.8, the ℓ_1 objective yields $27.01/0.843$ on LOL-v1 and $30.05/0.864$ on LOL-v2-Real. Adding MS-SSIM improves the results to $27.45/0.859$ and $30.64/0.879$, respectively. Incorporating LPIPS further achieves the best performance, reaching $27.81/0.870$ on LOL-v1 and $31.35/0.890$ on LOL-v2-Real. The progressive improvements indicate that pixel fidelity, structural consistency, and perceptual similarity provide complementary supervision for enhancement.

D.4 Training Data and Noise Augmentation

We examine the effects of mixed-dataset training and noise augmentation while keeping the architecture, loss function, and other training settings unchanged. The "w/o Noise Augmentation" variant removes noise augmentation from the mixed-dataset training protocol. For "w/o Mixed Training," separate models are trained on individual paired datasets: the LOL-v1 and LOL-v2-Real results are obtained from their corresponding dataset-specific models, while SID is evaluated zero-shot using the model trained on LOL-v2-Real.

As shown in Tab.9, removing noise augmentation slightly improves performance on LOL-v1 and LOL-v2-Real, but substantially reduces zero-shot SID performance from $20.354/0.608$ to $14.450/0.289$. This result reveals a clear trade-off: noise augmentation introduces a small performance cost on the paired benchmarks while markedly improving robustness to noise-corrupted low-light images. We therefore retain it in the final training protocol.

Table 9: **Ablation study of training data and noise augmentation.** "w/o Noise Aug." removes the scheduled noise augmentation, while "w/o Mixed Training" trains a separate model for each paired source benchmark: LOL-v1 and LOL-v2-Real are evaluated using models trained on their respective training sets, while SID is evaluated zero-shot using the model trained on LOL-v2-Real. All other variants use a single model trained on the combined training splits. Higher values are better. The best and second-best results are highlighted in red and blue, respectively. Ties are highlighted identically.

Training Configuration	LOL-v1		LOL-v2-Real		SID	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
w/o Noise Aug.	**28.03**	**0.874**	**31.49**	**0.894**	14.450	0.289
w/o Mixed Training	**27.87**	0.869	**31.41**	**0.890**	**19.544**	**0.587**
Full	27.81	**0.870**	31.35	**0.890**	**20.354**	**0.608**

Table 10: **Ablation of the pretrained visual encoder and DINOv3 layer selection.** All variants use RePriFusion-Small with a frozen visual encoder and identical Retinex estimator, restoration backbone, fusion modules, loss functions, and training protocol. The block tuple lists the features assigned to the bottleneck and three decoder stages. Parameters and FLOPs include the frozen encoder and its forward pass, respectively. The best and second-best results within each group are highlighted in red and blue, respectively.

Visual Encoder	Selected Blocks	Params. (M)	FLOPs (G)	LOL-v1		LOL-v2-Real	
				PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
(a) Pretrained visual encoder comparison							
ImageNet DeiT-S/16	$\{6, 9, 12, 12\}$	39.2	286.9	27.22	0.834	30.21	0.860
CLIP ViT-B/16	$\{6, 9, 12, 12\}$	104.0	377.0	**27.69**	**0.866**	**31.08**	0.876
DINOv2 ViT-S/14	$\{6, 9, 12, 12\}$	39.6	302.2	27.42	0.843	30.89	**0.879**
DINOv3 ViT-S/16	$\{6, 9, 12, 12\}$	39.09	287.03	**27.81**	**0.870**	**31.35**	**0.890**
(b) DINOv3 layer-selection ablation							
DINOv3 ViT-S/16	$\{12, 12, 12, 12\}$	39.09	287.03	27.64	0.859	31.19	0.881
DINOv3 ViT-S/16	$\{3, 6, 9, 12\}$	39.09	287.03	27.50	0.848	31.27	0.876
DINOv3 ViT-S/16	$\{6, 8, 10, 12\}$	39.09	287.03	**27.73**	0.863	**31.32**	**0.888**
DINOv3 ViT-S/16	$\{9, 10, 11, 12\}$	39.09	287.03	27.66	**0.864**	31.17	0.881
DINOv3 ViT-S/16	$\{6, 9, 12, 12\}$	39.09	287.03	**27.81**	**0.870**	**31.35**	**0.890**

Compared with dataset-specific training, mixed-dataset training maintains comparable performance on LOL-v1 and LOL-v2-Real while improving zero-shot SID performance from $19.544/0.587$ to $20.354/0.608$. These results suggest that exposure to more diverse training distributions improves cross-dataset transfer while preserving strong performance on the paired benchmarks.

D.5 Visual Encoder and Layer Selection Ablation

We examine the choice of pretrained visual encoder and the selection of intermediate feature blocks using the RePriFusion-Small configuration. To isolate the visual prior, all encoders remain frozen, while the Retinex estimator, restoration backbone, fusion stages, fusion operators, loss functions, and training protocol are kept unchanged. All configurations are compared on the source-domain LOL-v1 and LOL-v2-Real benchmarks, without using unseen target datasets for model selection.

As shown in Tab.10(a), DINOv3 achieves the best performance on both datasets. Compared with the ImageNet-pretrained DeiT-S/16 encoder, DINOv3 improves PSNR by 0.59 dB on LOL-v1 and 1.14 dB on LOL-v2-Real. It also outperforms DINOv2 by 0.39 and 0.46 dB, respectively. CLIP provides the strongest alternative visual prior, but remains below DINOv3 on all four metrics while requiring substantially more parameters and FLOPs. These results indicate that DINOv3 provides the most favorable balance between restoration accuracy and computational cost among the evaluated encoders. We therefore adopt DINOv3 ViT-S/16 as the frozen visual prior encoder.

Tab.10(b) further compares different DINOv3 block selections. Using only the final-block representation $\{12, 12, 12, 12\}$ performs consistently worse than combining features from different transformer depths, suggesting that intermediate and deep representations provide complementary guidance. The $\{6, 8, 10, 12\}$ configuration achieves competitive results, whereas including earlier fea-

tures through $\{3, 6, 9, 12\}$ or concentrating on the deepest blocks through $\{9, 10, 11, 12\}$ leads to lower overall performance. The adopted configuration $\{6, 9, 12, 12\}$ achieves the best PSNR and SSIM on both datasets. This configuration combines intermediate and deep DINO representations and reuses the final-block feature for the last two restoration stages, providing a small but consistent improvement over using four distinct block depths.

E Limitations and Future Work

RePriFusion has two main limitations. First, it may over-enhance regions containing strong local light sources or highly non-uniform illumination, producing saturated highlights and reduced local contrast. More explicit exposure constraints or region-adaptive illumination regularization may alleviate this issue.

Second, the frozen DINO encoder increases model size and inference cost. Even the smallest variant contributes approximately 18M parameters and remains part of the inference pipeline. Future work will explore feature distillation and shared multi-task architectures in which the same pretrained representation supports enhancement and downstream perception, thereby amortizing the additional cost.

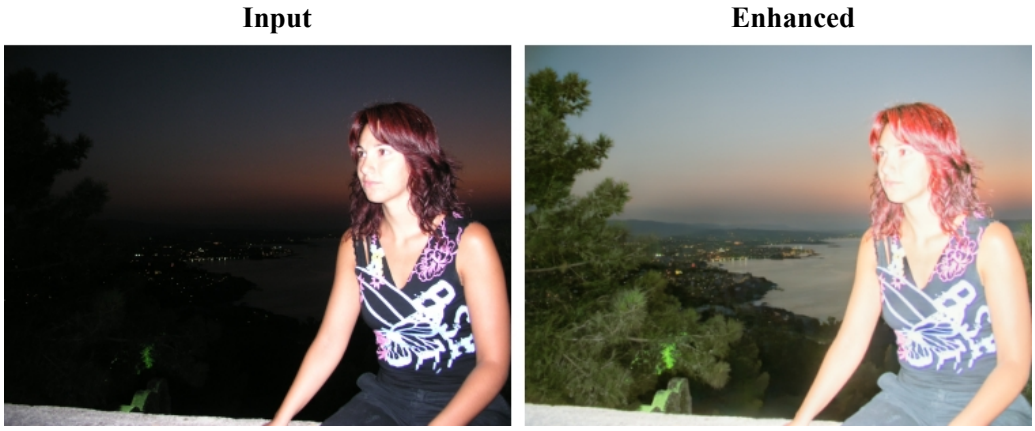


Figure 6: Failure case under strong local illumination.

F Appendix F: Additional Qualitative Results

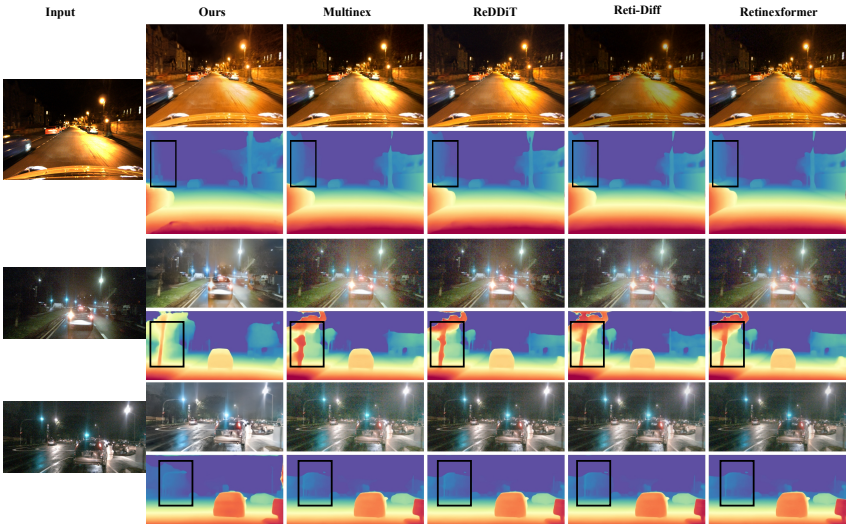


Figure 7: **Qualitative comparison of state-of-the-art LLIE methods and RePriFusion on nighttime depth-estimation benchmarks. Each example shows the enhanced image and the corresponding depth prediction produced by the fixed depth estimator. Boxes highlight regions with notable geometric differences.**

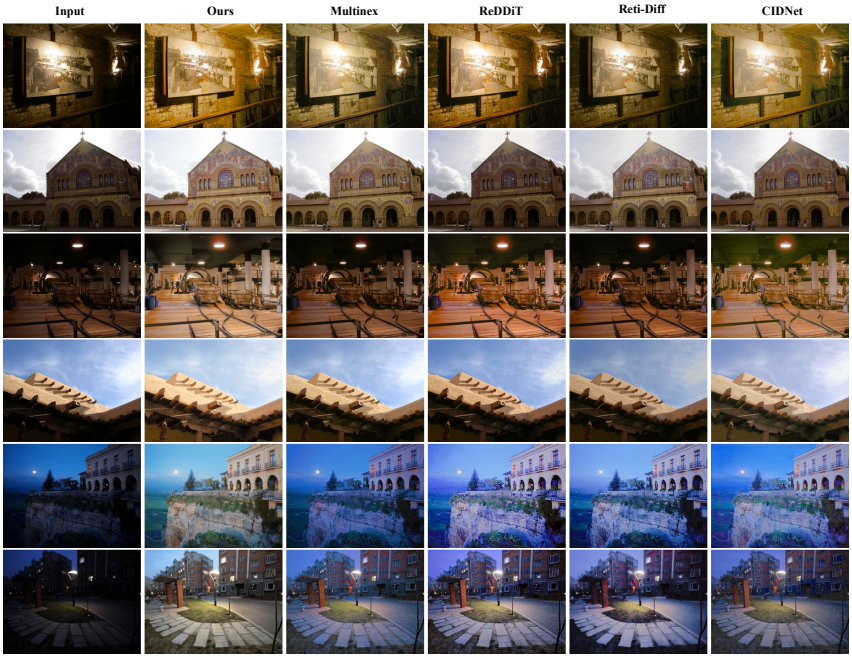


Figure 8: **Qualitative comparison of state-of-the-art LLIE methods and RePriFusion on no-reference benchmarks.**

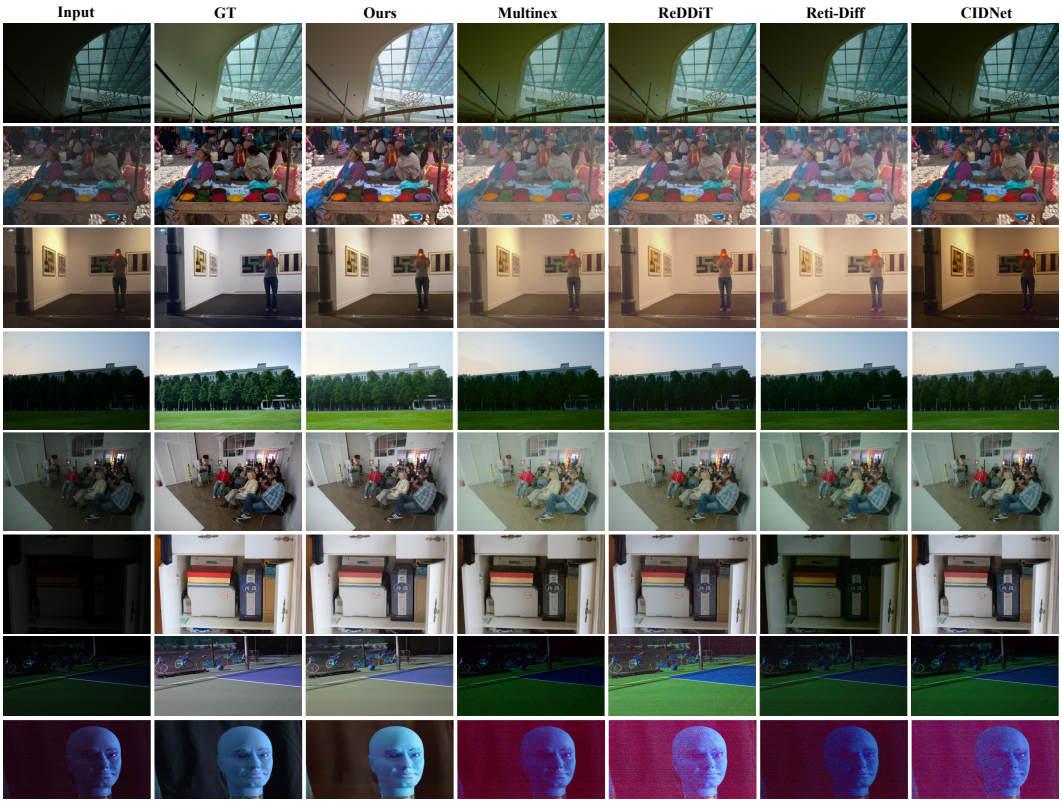


Figure 9: **Qualitative comparison of state-of-the-art LLIE methods and RePriFusion on paired benchmarks.**

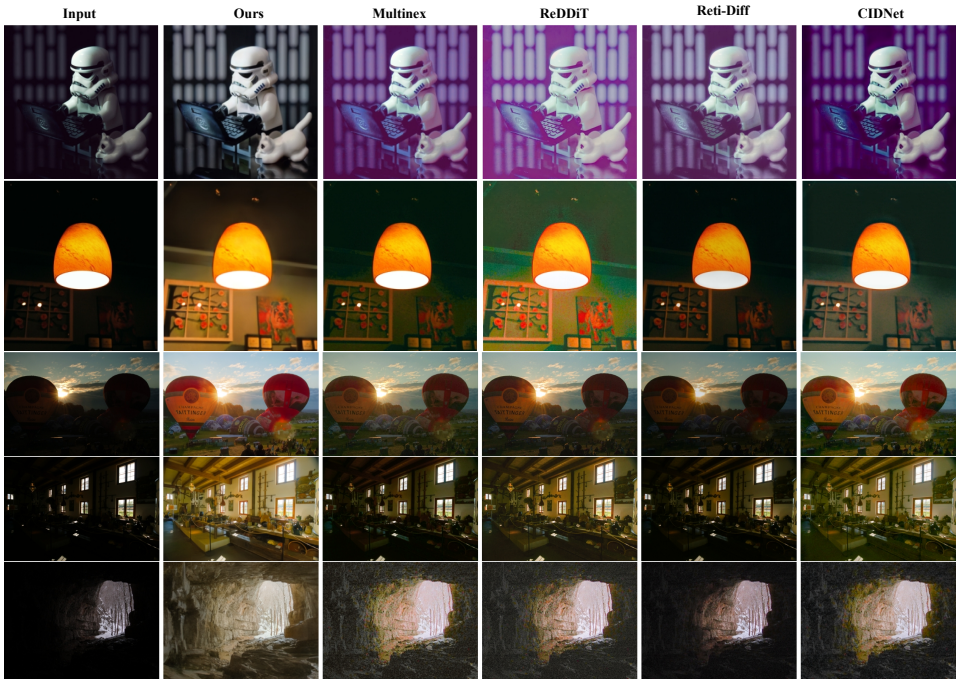


Figure 10: **Additional qualitative comparison on no-reference benchmarks.**

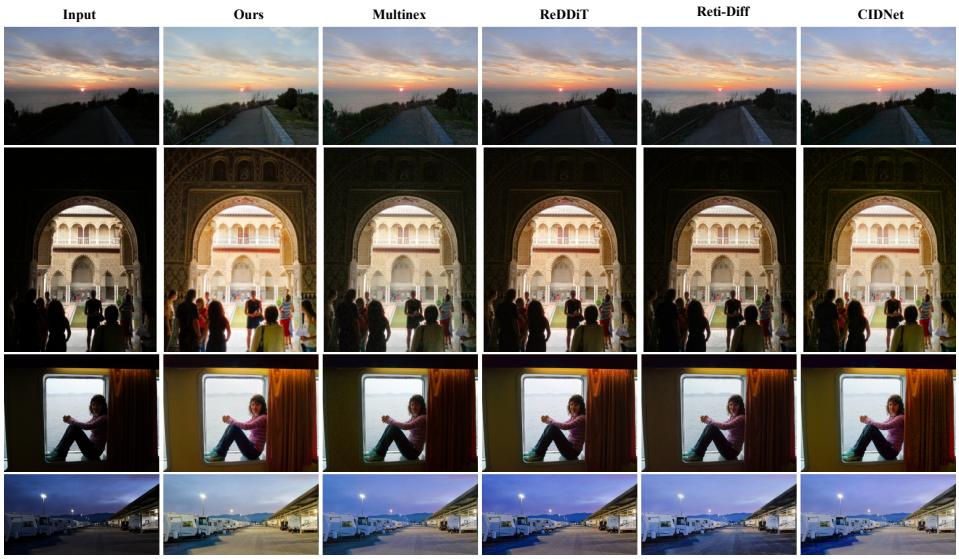


Figure 11: **Additional qualitative comparison on no-reference benchmarks.**