GICE: GENERALIZED IMAGE CONTRAST ENHANCEMENT WITH DIFFUSION PRIORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Real-world photographs often exhibit mixed contrast degradations induced by challenging illumination conditions (*e.g.*, low light, backlighting, and over-exposure), together with local corruptions such as noise, blur, and compression artifacts, which reduce visibility and obscure image fine details. Most existing methods target a single enhancement setting, which however often lead to brittle behavior when multiple degradations coexist within an image. In this work, we present **G**eneralized **I**mage **C**ontrast **E**nhancement (**GICE**), a unified framework that addresses diverse contrast degradations while recovering realistic structures and details. Specifically, we cast generalized contrast enhancement as a conditional image generation problem and introduce an exposure-aware decomposition module that separates the input into complementary exposure-variant and exposure-invariant components. The decomposition is learned through multi-exposure supervision and real-world degradation augmentation, enabling robust separation of information requiring contrast correction from structures and textures requiring preservation. We further develop a layer-adaptive conditioning mechanism that adjusts the contributions of the two components across diffusion layers, effectively balancing global contrast correction and local detail reconstruction. Finetuned on a pretrained SD-Turbo model, GICE performs efficient one-step refinement and employs trainable VAE skip modules to compensate for fine-detail loss during latent compression. Extensive experiments demonstrate that GICE generalizes well across diverse lighting conditions and consistently outperforms previous task-specific models. The code and models of GICE will be publicly available.

1 INTRODUCTION

Modern digital imaging devices have achieved great progress in sensors and lenses (Wu et al., 2024). However, real-world photographs still frequently suffer from challenging illumination conditions, such as low-light (Hai et al., 2023), backlit (Liang et al., 2023) and over-exposure (Cao et al., 2020), resulting in insufficient dynamic range and poor contrast. Existing contrast enhancement methods are often designed to address each individual task, such as low-light image enhancement (LLIE) (Hai et al., 2023; Jiang et al., 2024a; Li et al., 2023; Wang et al., 2023c; Feng et al., 2024; Chobola et al., 2024), backlit image enhancement (BIE) (Liang et al., 2023; Lv et al., 2022; Akai et al., 2023) and exposure correction (EC) (Li et al., 2024a; Cui et al., 2024; Huang et al., 2022b; Eyiokur et al., 2022; Wang et al., 2022), without considering that in practice an image can be degraded by multiple factors; for example, an image can have both low-light regions and over-exposure regions to be enhanced. In addition, low-contrast images can be corrupted by noise, blur, and JPEG compression simultaneously, making image contrast enhancement more challenging in practice.

Fig. 1 presents an example scenario where the foreground is dark and affected by noise and JPEG compression, while the background remains bright. Existing methods fail to address these issues effectively. For instance, Reti-Diff (He et al., 2025) enhances visibility but amplifies structural artifacts, whereas other approaches achieve only limited improvement in visibility and details.

In addition to the above-mentioned problem, existing contrast enhancement methods (Hai et al., 2023; Liang et al., 2023) are mostly trained on task-specific and small-scale data, hence having very limited generalization performance. For example, an LLIE model trained on the LSRW dataset (Hai et al., 2023) can perform poorly on other LLIE datasets (*e.g.*, UHDLL (Li et al., 2023)), not to mention

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.

on other tasks such as BIE and EC. One of the reasons for the lack of generality lies in the fact that building training pairs in real-world contrast enhancement tasks is very difficult and costly, making large-scale training data unavailable.

In this work, we address the above issues by presenting a **G**eneralized **I**mage **C**ontrast **E**nhancement model, namely **GICE**. Generalized enhancement requires balancing several potentially conflicting objectives. Enhancing under- and over-exposed regions involves different corrections, while improving contrast may amplify noise and suppressing noise may remove fine structures or introduce stair-step and blocking artifacts. Directly handling these objectives within a shared representation therefore makes it difficult to balance contrast correction, degradation removal, and structure preservation.

Our key idea is to cast generalized contrast enhancement as a conditional image generation problem by leveraging a pre-trained diffusion model, more specifically Stable Diffusion (SD) (Rombach et al., 2022). Previous studies show that different layers and pathways of diffusion U-Nets encode complementary spatial, frequency, and semantic information (Si et al., 2024; Tang et al., 2023; Luo et al., 2023). However, this layer-wise property remains underexplored for contrast enhancement, where appearance correction and structure reconstruction require different information. Moreover, the diffusion prior is degradation-agnostic, so naïve conditioning on degraded inputs may produce unstable corrections under mixed degradations. We therefore learn an additive exposure-aware decomposition from multi-exposure observations without imposing a predefined image formation model. The exposure-variant component captures exposure-related appearance and degradation patterns, while the exposure-invariant residual preserves stable scene structures. Both components are encoded at multiple resolutions and adaptively injected into different U-Net layers, enabling efficient single-step refinement of image contrast and fine textures while preserving structural fidelity.

To train our model, we collect a large scale of multi-exposure sequence (MES) data from multiple high dynamic range (HDR) sources, including AODRAW (Li et al., 2024b), FiveK (Bychkovsky et al., 2011), HDRP (Hasinoff et al., 2016), PASCAL-RAW (Omid-Zohoor et al., 2014-2015), PPR10K (Liang et al., 2021a), RAISE (Dang-Nguyen et al., 2015), and RAW-NOD (Morawski et al., 2022). Since these data vary mainly in exposure but lack local structure degradations such as noise or blur, we augment them with real-world image degradation pipelines (Wang et al., 2021; Zhu et al., 2023; Vo & Bui, 2023) to simulate these effects, enabling robust learning for real-world scenarios where multiple degradations coexist. For evaluation, we test GICE across multiple real-world datasets to verify its effectiveness and generalization performance, demonstrating its superiority to existing state-of-the-art contrast enhancement methods.

Our contributions are summarized as follows. First, we introduce an exposure-aware decomposition framework that separates an image into an exposure-variant component and an exposure-invariant residual using multi-exposure supervision, without relying on a predefined physical decomposition model. Second, we design a layer-adaptive conditioning mechanism that injects the decomposed components into different layers of a pretrained one-step diffusion network, allowing each layer to select the information appropriate for contrast correction or structure reconstruction. Extensive experiments demonstrate that our proposed GICE achieves significantly higher perceptual quality and generalization performance over existing methods.

2 Related Work

Image Contrast Enhancement. Image contrast enhancement has been extensively studied under various scenarios, including low-light conditions (Hai et al., 2023; Jiang et al., 2024a; Li et al., 2023; 2021; Wang et al., 2023c; Feng et al., 2024; Chobola et al., 2024; Wang et al., 2019; Xu et al., 2023; Wu et al., 2023; Yang et al., 2023; Wang et al., 2023a), improper exposure (Li et al., 2024a; Cui et al., 2024; Huang et al., 2022b; Eyiokur et al., 2022; Afifi et al., 2021; Huang et al., 2023; Wang et al., 2022; Huang et al., 2022a), and backlit scenes (Liang et al., 2023; Lv et al., 2022; Ueda et al., 2020; Akai et al., 2023; Zhao et al., 2021). These approaches typically focus on specific tasks and rely on small-scale datasets, limiting their generalization ability across various real-world scenarios. In practice, images often suffer from multiple coexisting degradations, including uneven illumination, noise, blur, and compression artifacts, for which single-task models struggle to address. Recent studies have explored multi-task and unified frameworks, and several image restoration methods (Conde et al., 2024; Chen et al., 2024; Jiang et al., 2024b; Liang et al., 2021b; Mou et al., 2022; Hua et al., 2026) have been proposed to jointly handle degradations such as low-light and blur.

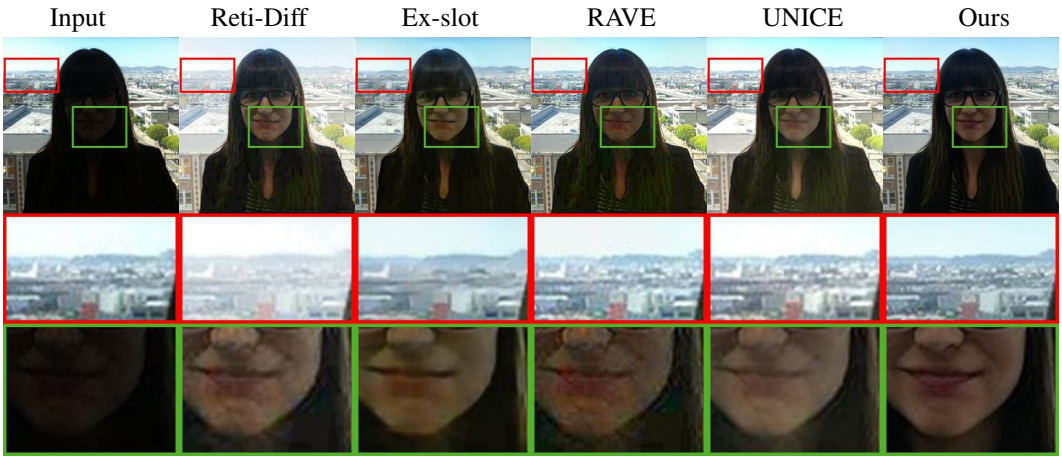


Figure 1: Challenges of image contrast enhancement in real-world. The example is from the backlit dataset (Liang et al., 2023), which does not have GT. The image contains both dark (in green box) and over-exposed regions (in red box). All methods are retrained on the same dataset and setting for fair comparison. We see that existing methods either amplify structural artifacts (Reti-Diff (He et al., 2025)) or achieve limited improvement in visibility and details (Ex-slot (Jung et al., 2025), RAVE (Gaintseva et al., 2024), UNICE (Cui & Zhang, 2025)), while our GICE improves both dark and bright regions with higher structural fidelity and richer details.

However, these methods remain limited in scope, as they often fail to model challenging conditions such as backlit scenes or severe over-exposure. Moreover, most existing approaches are inadequate to bridge domain gaps between different contrast enhancement tasks. This limitation largely stems from training on mixed collections of task-specific, small-scale datasets, which restricts generalization to complex, real-world scenarios.

Diffusion-based Image Contrast Enhancement. With the growing success of diffusion models in image synthesis, researchers have recently extended these models to image contrast enhancement, particularly for LLIE. Most existing approaches condition the diffusion process on structural cues extracted from low-light inputs. For example, PyDiff (Zhou et al., 2023) introduces a pyramid-based sampling strategy to accelerate inference and mitigate global color shifts by incorporating a global corrector. diffLL (Jiang et al., 2023) leverages wavelet-domain diffusion to restore high-frequency details efficiently. GSAD (Hou et al., 2024) enforces global structure-aware regularization, enabling smoother noise schedules during sampling. Diff-Retinex (Yi et al., 2023) integrates Retinex decomposition into the conditioning mechanism, combining physical priors with generative diffusion for improved restoration. ReDDiT (Lan et al., 2025) further proposes a reflectance-aware distillation framework to reduce sampling steps.

3 METHODOLOGY

Figure 2 presents an overview of the proposed **G**eneralized **I**mage **C**ontrast **E**nhancement framework, namely **GICE**, which consists of two modules: an exposure-aware decomposition network $\mathcal{D}$ and a diffusion-based enhancement network $\mathcal{G}$. We learn a task-oriented representation directly from multi-exposure observations without imposing a predefined image formation model, such as Retinex decomposition. Given an input image I, the decomposition network predicts an exposure-variant component $E = \mathcal{D}(I)$, and the exposure-invariant residual is computed as $C = I - E$. The component E captures spatially varying intensity, contrast, and degradation patterns associated with exposure changes, while C preserves scene content that remains stable across exposure levels. The model encodes both components at multiple resolutions and adaptively injects them into different layers of a pretrained diffusion network.

The enhancement module is defined as $\mathcal{G} = \{\Phi(\cdot), \mathbf{p}, \widetilde{\text{VAE}}, \text{U-Net}\}$, where $\Phi(\cdot)$ denotes the proposed layer-adaptive feature injection module, $\mathbf{p}$ is a learnable prompt embedding, $\widetilde{\text{VAE}}$ is a frozen pretrained VAE augmented with trainable skip convolution modules, and the U-Net is adapted for

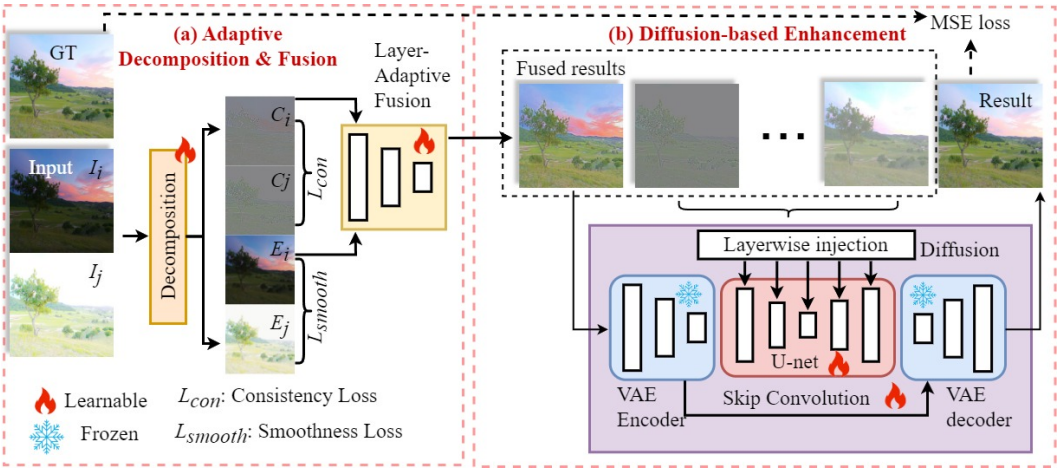


Figure 2: Overview of the GICE framework. Notations are defined in Sec. 3. (a) Exposure-aware Decomposition separates the input into an exposure-variant component and a complementary exposure-invariant residual, disentangling appearance variations from stable scene structures. (b) Diffusion-based Enhancement adaptively injects multi-scale features from both components into different U-Net layers, and performs efficient one-step refinement with learnable prompts and VAE skip convolutions for contrast correction and detail preservation.

efficient one-step enhancement. The decomposition and enhancement modules are jointly optimized in an end to end manner.

3.1 Exposure-Aware Decomposition

Images of the same scene captured at different exposure levels share the underlying scene content but differ in brightness, local contrast, visible details, noise levels, and saturation. Rather than explicitly assigning these variations to illumination or reflectance, we predict an exposure-variant component and derive a complementary exposure-invariant residual from the input.

For each observation I_k, $\mathcal{D}$ predicts an exposure-variant component E_k that captures spatially coherent variations in intensity, contrast, and exposure-related degradations. The corresponding residual C_k preserves structures that remain stable across exposures, including object boundaries, textures, and semantic content. Unlike a conventional degradation residual, C_k is neither constrained by sparsity or noise-like assumptions nor interpreted as an explicit noise or degradation map. This flexibility allows the decomposition to model real exposure-related effects, such as noise, clipping, and compression artifacts, without requiring a single physical image formation model.

Training Data Construction. We use multi-exposure sequences (MES) to train the decomposition network. Given an HDR raw image, we synthesize an MES $\{I_1, I_2, \ldots\}$ by varying the exposure settings. We then apply multi-exposure fusion (MEF) to obtain a pseudo ground-truth image I_{gt}, following (Afifi et al., 2021; Cui & Zhang, 2025). **Appendix A** describes the MES synthesis process and pseudo-GT generation.

During training, we construct triplets $(I_i, I_j, I_{\text{gt}})$, where I_i and I_j depict the same scene at different exposure levels, and I_{gt} is the corresponding fused pseudo ground truth. The paired observations may contain exposure-dependent degradations, including sensor noise, quantization or JPEG artifacts, blur, under-exposed regions, and saturated or clipped highlights.

The paired observations (I_i, I_j) supervise the learning of exposure-variant and exposure-invariant representations, whereas I_{gt} supervises only the final enhancement output. This separation prevents the decomposition network from directly approximating the fused target and encourages it to identify exposure-stable information shared across observations. The pseudo ground truth aggregates complementary information from the MES for reconstructing contrast, color, and fine details.

Network Architecture. As shown in Fig. 2, $\mathcal{D}$ uses a Siamese architecture during training. Its two branches process I_i and I_j using the same fully shared four-layer convolutional network. Each layer uses a 3×3 convolution with a stride of one and is preceded by reflection padding of one pixel, thereby preserving the spatial resolution throughout the network. The first layer maps the three-channel input to N feature channels, the next two layers maintain N channels, and the final layer projects the features back to three channels. The first three convolutional layers are followed by ReLU activations, whereas the output layer uses a sigmoid function to predict the exposure-variant component $E_k \in [0,1]^{H\times W\times 3}$. The network does not independently predict C_k; instead, $C_k = I_k - E_k$ is computed deterministically, avoiding the ambiguity of estimating two unconstrained components while preserving all input information. At inference time, only one branch is required, and $\mathcal{D}$ processes a single input image without requiring a paired observation or exposure metadata. Detailed architectural specifications are provided in **Appendix B**.

Decomposition Objectives. Given a training pair (I_i, I_j), the shared decomposition network predicts E_i and E_j, and the corresponding residuals C_i and C_j are derived. Because the two components preserve the complete input by construction, no reconstruction loss is required. We optimize the decomposition network using smoothness and cross-exposure consistency objectives. Exposure changes typically vary smoothly within locally homogeneous regions. We impose an edge-aware smoothness constraint on E:

$$\mathcal{L}_{\text{smooth}} = \|W_i \odot \nabla E_i\|_1 + \|W_j \odot \nabla E_j\|_1, \qquad W_k = \exp(-\kappa \|\nabla I_k\|_1). \tag{1}$$

Here, ∇ denotes the spatial gradient operator and $\odot$ denotes element-wise multiplication. The edge-aware weights reduce smoothing near strong image boundaries. This constraint maintains the spatial coherence of E while preserving exposure-dependent transitions around saturated regions and high-contrast boundaries. Because I_i and I_j depict the same scene, their exposure-invariant residuals should remain consistent. We define the consistency loss as $\mathcal{L}_{\text{con}} = \|C_i - C_j\|_1$. This objective encourages C_i and C_j to retain exposure-stable structural information. The overall decomposition objective is

$$\mathcal{L}_{\text{decomp}} = \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}} + \lambda_{\text{con}}\mathcal{L}_{\text{con}}. \tag{2}$$

We empirically set $\lambda_{\text{smooth}} = 0.1$ and $\lambda_{\text{con}} = 1$.

3.2 Diffusion-Based Enhancement

The decomposed components condition the enhancement module $\mathcal{G} = \{\Phi(\cdot), \mathbf{p}, \widetilde{\text{VAE}}, \text{U-Net}\}$. Because diffusion layers at different resolutions capture different types of information, we use layer-specific fusion to inject E and C into multiple U-Net layers.

Multi-Scale Conditional Feature Extraction. Two lightweight condition encoders, $\mathcal{F}_E$ and $\mathcal{F}_C$, extract hierarchical features from the exposure-variant component and exposure-invariant residual. Specifically, $\{F_E^\ell\}_{\ell=1}^L = \mathcal{F}_E(E)$ and $\{F_C^\ell\}_{\ell=1}^L = \mathcal{F}_C(C)$, where ℓ indexes the injection layer or feature resolution.

At each layer, learnable convolutional projections map the conditional features to the channel dimensionality of the corresponding diffusion feature: $\bar{F}_E^\ell = P_E^\ell(F_E^\ell)$ and $\bar{F}_C^\ell = P_C^\ell(F_C^\ell)$. The projected features are resized when necessary to match the spatial resolution of the target U-Net layer.

Layer-Adaptive Fusion and Injection. Let H^ℓ denote the intermediate diffusion feature at layer ℓ. A layer-specific gating module predicts the fusion logits: $[a_E^\ell, a_C^\ell] = \mathcal{A}_\ell([H^\ell, \bar{F}_E^\ell, \bar{F}_C^\ell])$, where $[\cdot]$ denotes channel-wise concatenation. A softmax produces the normalized coefficients: $[\alpha_E^\ell, \alpha_C^\ell] = \text{Softmax}([a_E^\ell, a_C^\ell])$, with $\alpha_E^\ell + \alpha_C^\ell = 1$.

The fused condition at layer ℓ is $Q^\ell = \alpha_E^\ell \odot \bar{F}_E^\ell + \alpha_C^\ell \odot \bar{F}_C^\ell$. A residual adapter injects the fused condition into the diffusion feature as $\widetilde{H}^\ell = H^\ell + \eta_\ell Z_\ell(Q^\ell)$, where Z_ℓ is a zero-initialized convolutional layer and η_ℓ is a learnable layer-wise injection strength. The zero initialization preserves the pretrained diffusion prior at the beginning of fine-tuning.

The layer-adaptive mechanism allows different diffusion layers to select appropriate conditioning information. Layers responsible for global tone and contrast can emphasize the exposure component, while high-resolution layers can rely more on the invariant residual to preserve structures and textures. The allocation is learned from enhancement supervision rather than manually specified. We denote the complete set of multi-layer conditions by $\Phi(E, C) = \{Q^1, Q^2, \ldots, Q^L\}$.

One-Step Diffusion. We adopt SD-Turbo (Sauer et al., 2023) and fix the timestep to $t = 1$ for efficient one-step enhancement. A shared learnable prompt embedding $\mathbf{p}$ accesses the pretrained generative prior and maintains compatibility with the cross-attention layers. To compensate for fine-grained details lost during latent compression, we augment the frozen pretrained VAE with trainable skip convolution modules, denoted by $\widetilde{\text{VAE}}$. The input is first encoded as $\mathbf{z} = \widetilde{\text{VAE}}_{\text{enc}}(I)$. Conditioned on the prompt embedding and decomposition features, the U-Net predicts $\widehat{\boldsymbol{\epsilon}} = \epsilon_\theta(\mathbf{z}, t, \mathbf{p}, \Phi(E, C))$. We then update the latent using $\mathbf{z}' = \mathbf{z} - \widehat{\boldsymbol{\epsilon}}$ and reconstruct the enhanced image as $\widetilde{I} = \widetilde{\text{VAE}}_{\text{dec}}(\mathbf{z}')$. We adapt the pretrained U-Net using LoRA (Hu et al., 2022) while keeping its original parameters frozen. Training jointly updates the LoRA parameters, layer-adaptive fusion modules, condition encoders, prompt embedding, VAE skip convolutions, and decomposition network.

Enhancement Objective. The enhanced output is supervised by the fused pseudo ground truth using the pixel-domain mean squared error $\mathcal{L}_{\text{enh}} = \|\widetilde{I} - I_{\text{gt}}\|_2^2$. The complete objective is $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{decomp}} + \lambda_{\text{enh}}\mathcal{L}_{\text{enh}}$, where $\lambda_{\text{enh}} = 1$ in all experiments. During end-to-end training, $\mathcal{L}_{\text{enh}}$ provides task-level supervision to the decomposition network, encouraging E and C to serve as complementary conditions for exposure correction and detail preservation.

4 Experiments

4.1 Experimental Setup

Implementation Details. We train the GICE model on the constructed 48,361 data groups. Each group contains seven images captured at different exposure levels ($\pm$3EV) and one pseudo-GT image produced by MEF. To improve robustness, we apply a degradation pipeline (Wang et al., 2021; Zhu et al., 2023; Vo & Bui, 2023) to the input images in each group to simulate structural degradation, including noise, blur, and JPEG compression. More details of the training data construction are provided in the **Appendix A**. Training is conducted on a server with eight NVIDIA A100 GPUs using mixed precision. We use AdamW with an initial learning rate of 2×10^{-4}, cosine decay scheduling with a 10K-step warm-up, weight decay of 10^{-2}, a batch size of 16, and 500K training iterations.

Evaluation Datasets. To ensure a rigorous evaluation, we select representative datasets for each of the LLIE, BIE and EC sub-tasks using two criteria: images are collected under real-world conditions, and the dataset contains sufficient samples to support statistically reliable comparisons. Following these criteria, we use LSRW (Hai et al., 2023) and UHDLL (Li et al., 2023) for LLIE, and BAID (Lv et al., 2022) and Backlit (Liang et al., 2023) for BIE. For EC, commonly used benchmarks include SICE (Cai et al., 2018), MSEC (Afifi et al., 2021), and LCDP (Wang et al., 2022). Since MSEC (Afifi et al., 2021) and LCDP (Wang et al., 2022) are rendered from the FiveK (Bychkovsky et al., 2011), which overlaps with our training data, we use SICE (Cai et al., 2018) in the experiments.

Compared Methods. For LLIE, we compare with DiffLL (Jiang et al., 2023), DReti++ (Yi et al., 2025), Reti-Diff (He et al., 2025), ReDDiT (Lan et al., 2025), and Multinex (Brateanu et al., 2026). For EC, we compare with Ex-Slot (Jung et al., 2025), CSEC (Li et al., 2024a), CoTF (Li et al., 2024c), LCDPNet (Wang et al., 2022), and FECNet (Huang et al., 2022b). For BIE, we compare with RAVE (Gaintseva et al., 2024) and CLIP-LIT (Liang et al., 2023). Since dedicated BIE methods are relatively few, we also compare with several LLIE methods on this sub-task. In addition, we include recently developed unified image restoration/enhancement methods, namely Bi-Bridge (Hua et al., 2026), AutoDIR (Jiang et al., 2024b), and UNICE (Cui & Zhang, 2025), for comparison.

Experiment Settings. This work focuses on robust and generalizable image enhancement; therefore, we primarily conduct out-of-domain experiments. Specifically, for each sub-task, training is performed on one dataset and testing is performed on a different dataset. Considering that some models compared are small, for a more fair comparison, we adjust their channel configurations to increase the parameter count to approximately 1.2B, comparable to larger models and GICE (1.1B). These enlarged variants are denoted by "*" in the experiments. In addition, for those models that do not include components for repairing structural degradation, we pair them with a diffusion-based image restoration model, PiSA-SR (Sun et al., 2025), which is retrained on our dataset, to improve detail recovery. These variants are denoted by "†" in the experiments.

Evaluation Metrics. We use both full-reference (FR) and no-reference (NR) metrics for evaluation, including PSNR (P.910, 1999), SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018), DISTS (Ding

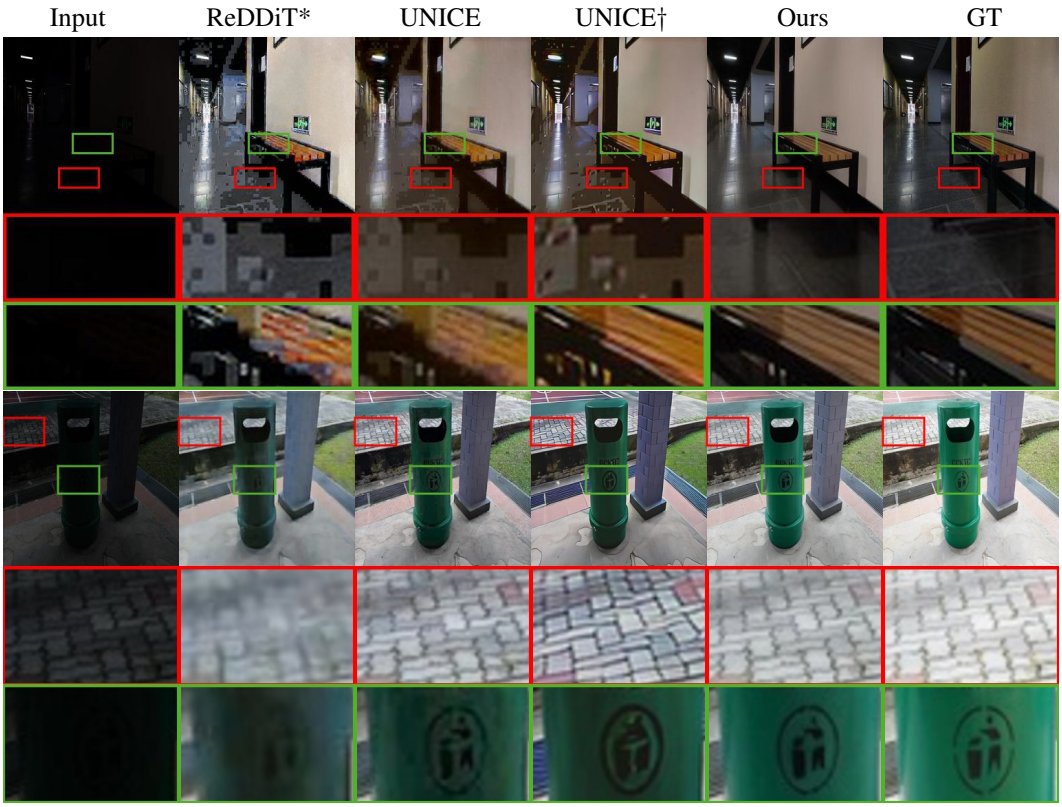


Figure 3: Qualitative comparison of different methods. The top two rows are from the LSRW dataset and the bottom two rows are from UHDLL. Enlarged local regions are highlighted with red and green boxes for better visualization.

et al., 2020), ARNIQA (Agnolucci et al., 2024), MUSIQ (Ke et al., 2021), CLIP-IQA+ (Wang et al., 2023b), and MANIQA (Yang et al., 2022). In addition, a user study is provided in **Appendix C**.

4.2 Experimental Results

Results for LLIE. Table 1 shows the quantitative results for the LLIE task. The upper/lower panels show the results on the LSRW (Hai et al., 2023)/UHDLL (Li et al., 2023) dataset when models are trained on UHDLL (Li et al., 2023)/LSRW (Hai et al., 2023) and our dataset, respectively. We can have several observations. First, those task-specific models, when trained on the small dataset (LSRW or UHDLL), show poor generalization performance. Second, by enlarging the size of these models and retraining them on our dataset, their generalization performance can be greatly improved. Coupling them with PiSA-SR Sun et al. (2025) can further improve their performance. Third, the unified models, AutoDIR, Bi-Bridge and UNICE, overall show better generalization capability than the task-specific models. Finally, our approach consistently achieves the best performance across most metrics, demonstrating its effectiveness.

Fig. 3 provides qualitative comparisons for the LLIE task. In the first example (from the LSRW dataset), both ReDDiT (Lan et al., 2025) and UNICE (Cui & Zhang, 2025) improve overall visibility but simultaneously amplify JPEG compression artifacts. UNICE†, incorporating an image restoration model (Sun et al., 2025), partially recovers structural details (as highlighted in the green box), but it can introduce additional artifacts (as shown in the red box). In the second example (from UHDLL), ReDDiT and UNICE produce inferior visibility to our GICE, and they lack clarity. UNICE† produces results that are inconsistent with the GT due to the hallucinations from the PiSA-SR module. The shape of the floor tile gaps, originally wavy, is straightened (see red box), and the structure of the logo on the trash bin is distorted (see blue box). In contrast, the results produced by our GICE model are not only rich in details but also highly consistent with the GT. Additional visual comparisons are provided in the **Appendix D**.

Table 1: Quantitative comparisons on the LLIE task. Tr./Te. denote the training/testing datasets. LPI (*resp.* DIS, ARN, MUS, CIQA, MAN) denotes LPIPS (*resp.* DISTS, ARNIQA, MUSIQ, CLIP-IQA+, MANIQA). Symbols ↑ and ↓ indicate that a higher or lower value is better. **Bold** and Underlined numbers indicate the best and second-best results. Symbols "*" and "†" mark larger variants and those variants coupled with PiSA-SR (Sun et al., 2025).

Tr.	Te.	Method	PSNR ↑	SSIM ↑	LPI. ↓	DIS. ↓	ARN. ↑	MUS. ↑	CIQA. ↑	MAN. ↑
LSRW	UHDLL	DiffLL	19.89	0.8900	0.1395	0.1202	0.4975	47.98	0.5022	0.3871
		DReti++	20.58	0.8916	0.1268	0.1223	0.5310	45.38	0.5062	0.4095
		Reti-Diff	20.87	0.8871	0.1368	0.1248	0.5241	46.66	0.5112	0.4134
		ReDDiT	21.31	0.8925	0.1084	0.1149	0.5289	48.44	0.5299	0.3982
		Multinex*	19.76	0.8465	0.1481	0.1200	0.4231	44.38	0.4991	0.3870
Ours		DiffLL*	21.45	0.9093	0.1246	0.1201	0.5346	48.94	0.5842	0.4127
		DiffLL†	21.67	0.9120	0.1117	0.1139	0.5429	49.81	0.5959	0.4358
		DReti++*	22.41	0.9246	0.1130	0.1059	0.6287	52.99	0.5891	0.4148
		DReti++†	22.45	0.9257	0.0995	0.1043	0.6329	53.84	0.5992	0.4280
		Reti-Diff*	22.17	0.9240	0.1003	0.1102	0.6580	56.94	0.6259	0.4360
		Reti-Diff†	22.26	0.9267	0.0865	0.1040	0.6647	55.44	0.6401	0.4561
		ReDDiT*	22.72	0.9174	0.0792	0.0987	0.6584	56.69	0.6295	0.4607
		ReDDiT†	22.84	0.9193	0.0765	0.0866	0.6628	57.35	0.6933	0.4738
		Multinex*	19.96	0.9130	0.1354	0.1170	0.5786	46.93	0.5558	0.4166
		Multinex†	20.44	0.9184	0.1405	0.1183	0.6147	45.53	0.5861	0.4287
		Bi-Bridge*	23.01	0.9179	0.0914	0.1071	0.6675	58.92	0.6306	0.4386
		Bi-Bridge†	23.08	0.9251	0.0845	0.0948	0.6617	59.83	0.6323	0.4489
		AutoDIR	23.04	0.9282	0.0930	0.1082	0.6499	57.32	0.6177	0.4254
		AutoDIR†	23.14	0.9289	0.0853	0.0910	0.6564	58.08	0.6301	0.4380
		UNICE	23.65	0.9248	0.0820	0.0899	0.6829	61.26	0.6793	0.5183
		UNICE†	23.70	0.9296	0.0758	0.0865	0.6925	62.23	0.6728	0.5029
		Ours	**24.01**	**0.9367**	**0.0656**	**0.0798**	**0.7094**	**62.52**	**0.6991**	**0.5276**
UHDLL	LSRW	DiffLL	17.93	0.7483	0.1647	0.1692	0.4828	41.18	0.4726	0.3413
		DReti++	18.47	0.7590	0.1462	0.1490	0.4958	43.71	0.5044	0.3664
		Reti-Diff	18.94	0.7426	0.1558	0.1404	0.5161	44.49	0.4837	0.3675
		ReDDiT	19.01	0.7531	0.1556	0.1400	0.5174	45.68	0.4944	0.3675
		Multinex*	18.53	0.7514	0.1598	0.1513	0.5179	41.95	0.4618	0.3303
Ours		DiffLL*	19.79	0.7621	0.1373	0.1418	0.5313	46.62	0.5167	0.3715
		DiffLL†	19.87	0.7648	0.1216	0.1359	0.5407	48.34	0.5164	0.3780
		DReti++*	20.01	0.7834	0.1365	0.1336	0.6278	50.93	0.6038	0.4172
		DReti++†	20.12	0.7841	0.1269	0.1238	0.6379	51.45	0.6174	0.3820
		Reti-Diff*	20.42	0.7672	0.1327	0.1440	0.5979	51.23	0.5674	0.3863
		Reti-Diff†	20.48	0.7894	0.1501	0.1322	0.6060	48.85	0.5816	0.3985
		ReDDiT*	20.81	0.7982	0.1385	0.1152	0.6405	53.83	0.5992	0.4188
		ReDDiT†	19.61	0.7754	0.1430	0.1273	0.6340	49.60	0.5989	0.4059
		Multinex*	19.71	0.7615	0.1582	0.1374	0.5740	48.04	0.5540	0.3817
		Multinex†	19.51	0.7654	0.1431	0.1304	0.5909	46.19	0.5659	0.3740
		Bi-Bridge*	20.57	0.7786	0.1416	0.1247	0.6264	53.69	0.6203	0.3943
		Bi-Bridge†	20.77	0.7785	0.1313	0.1087	0.6455	50.83	0.6122	0.4136
		AutoDIR	20.64	0.8038	0.1452	0.1262	0.6282	52.05	0.6025	0.4189
		AutoDIR†	20.75	0.7747	0.1391	0.1220	0.6396	52.61	0.6127	0.4186
		UNICE	21.37	0.7871	0.1286	0.1114	0.6647	54.30	0.6307	0.4395
		UNICE†	21.25	**0.8157**	0.0995	0.1143	0.6329	54.08	0.6401	0.4561
		Ours	**21.81**	0.8146	**0.0865**	**0.1040**	**0.6996**	**55.44**	**0.6708**	**0.4628**

Results for EC. For the EC task, we compare with Ex-Slot (Jung et al., 2025), CSEC (Li et al., 2024a), CoTF (Li et al., 2024c), LCDPNet (Wang et al., 2022), and FECNet (Huang et al., 2022b), along with the all-in-one approaches AutoDIR (Jiang et al., 2024b) and Bi-Bridge (Hua et al., 2026). All experiments are conducted on SICE (Cai et al., 2018). The training set combines the original SICE training split with our newly constructed dataset. Because SICE includes multiple MES from the same scene, we split the data at the MES level to prevent information leakage. In practice, we use 90% of the MES for training and the remaining 10% for testing. We report quantitative results in Table 2. The upper panel shows models trained and evaluated on SICE as a reference, while the lower panel shows models trained on our dataset and evaluated on SICE. Models trained on our dataset consistently achieve better performance, suggesting that our data provide useful gains in scale, diversity, and quality. Visual comparisons in **Appendix D** show that GICE consistently improves contrast and restores fine details while maintaining structural fidelity across diverse exposure conditions. Although saturated regions remain particularly susceptible to hallucination, the resulting deviations are largely confined to subtle appearance variations (**Appendix E**).

Table 2: Quantitative comparisons on the EC task. Tr./Te. denote the training/testing datasets. LPI (*resp.* DIS, ARN, MUS, CIQA, MAN) denotes LPIPS (*resp.* DISTS, ARNIQA, MUSIQ, CLIP-IQA+, MANIQA). Symbols ↑ and ↓ indicate that a higher or lower value is better. **Bold** and Underlined numbers indicate the best and second-best results. Symbols "*" and "†" mark larger variants and those variants coupled with PiSA-SR (Sun et al., 2025).

Tr.	Method	PSNR ↑	SSIM ↑	LPI. ↓	DIS. ↓	ARN. ↑	MUS. ↑	CIQA. ↑	MAN. ↑
SICE	LCDPNet	14.64	0.6501	0.2636	0.2048	0.6273	43.41	0.5441	0.3288
	LCDPNet†	14.64	0.6707	0.2675	0.2024	0.6286	43.35	0.5878	0.3390
	FECNet	14.86	0.6607	0.2692	0.1913	0.6232	43.30	0.5551	0.3345
	FECNet†	15.05	0.6675	0.2360	0.2077	0.6267	43.35	0.5770	0.3357
	CoTF	15.43	0.6557	0.2436	0.1832	0.6218	43.84	0.6006	0.3256
	CoTF†	15.69	0.7086	0.2665	0.1855	0.6359	43.43	0.5957	0.3475
	CSEC	15.91	0.6879	0.2432	0.1861	0.6263	43.36	0.5776	0.3359
	CSEC†	16.17	0.6914	0.2245	0.1890	0.6333	43.94	0.5871	0.3417
	Ex-Slot	16.01	0.6968	0.2337	0.1845	0.6394	43.24	0.5843	0.3433
	Ex-Slot†	16.29	0.7125	0.2202	0.1762	0.6473	44.29	0.6105	0.3464
Ours	LCDPNet*	18.83	0.8201	0.2068	0.1356	0.6743	48.44	0.6443	0.2969
	LCDPNet†	18.86	0.8154	0.1932	0.1354	0.6880	49.38	0.6654	0.3220
	FECNet*	18.87	0.8207	0.1876	0.1378	0.6606	48.08	0.6474	0.3154
	FECNet†	18.94	0.8331	0.1792	0.1374	0.6627	47.57	0.6317	0.3236
	CSEC*	19.35	0.8253	0.1742	0.1351	0.6826	49.35	0.6631	0.3261
	CSEC†	19.54	0.7998	0.1701	0.1990	0.6490	39.56	0.5405	0.3333
	CoTF*	19.59	0.8272	0.1632	0.1448	0.6895	48.86	0.6428	0.3309
	CoTF†	19.75	0.8411	0.1548	0.1486	0.6659	44.05	0.6139	0.3592
	Bi-Bridge*	20.25	0.8363	0.1503	0.1505	0.6534	47.33	0.5998	0.3643
	Bi-Bridge†	20.49	0.8318	0.1495	0.1849	0.6549	42.16	0.5684	0.3730
	AutoDIR*	20.67	0.8261	0.1486	0.1871	0.6581	41.94	0.5760	0.3833
	AutoDIR†	20.73	0.8573	0.1484	0.1481	0.6798	45.28	0.6272	0.3844
	Ex-Slot*	20.78	0.8435	0.1463	0.1666	0.6713	44.75	0.5993	0.3866
	Ex-Slot†	21.13	0.8501	0.1444	0.1600	0.6539	44.11	0.5835	0.3885
	UNICE	21.26	0.8678	0.1421	0.1388	0.6977	49.39	0.6855	0.3962
	UNICE†	21.40	0.8704	0.1389	0.1316	0.6657	48.46	0.6345	0.3976
	Ours	**21.80**	**0.8801**	**0.1325**	**0.1305**	**0.6986**	**49.85**	**0.6910**	**0.4016**

Results for BIE. For the BIE task, we evaluate on BAID (Lv et al., 2022) and Backlit (Liang et al., 2023) datasets, comparing against RAVE (Gaintseva et al., 2024), CLIP-LIT (Liang et al., 2023), and several adapted LLIE methods. Detailed results are provided in the **Appendix F**.

Additional Experiments. **Appendix G** visualizes the learned exposure-aware decomposition, illustrating that the exposure-variant component captures large-scale appearance variations, while the exposure-invariant residual preserves scene structures and fine textures across exposure levels. **Appendix H** provides comprehensive ablation studies of the major design components of GICE, and **Appendix I** compares its computational complexity with competing diffusion-based methods. These experiments further demonstrate the effectiveness and efficiency of the proposed framework.

5 Conclusion

We presented GICE, a unified and generalized framework for image contrast enhancement, to robustly handle multiple types of contrast degradation within a single model. By integrating exposure-aware decomposition with diffusion priors, our model effectively separates exposure-variant appearance from exposure-stable scene content and leverages generative priors for high-fidelity detail restoration. The proposed layer-adaptive conditioning mechanism further balances global contrast correction and local structure preservation across diffusion layers. Extensive experiments across low-light enhancement, backlit enhancement, and exposure correction tasks demonstrated that GICE achieved superior perceptual quality, structural fidelity, and generalization performance compared with state-of-the-art task-specific and unified models.

Limitations. While effective, GICE has some limitations. First, it takes SD-Turbo as the backbone and employs a one-step diffusion strategy for efficiency, which sacrifices a little visual quality. Second, the model size of GICE can be further reduced by pruning and quantization techniques. In addition, broader downstream enhancement tasks can be explored in the future.

LLM USAGE STATEMENT

Large language models were used exclusively to assist with the linguistic refinement of this manuscript. Their use was limited to improving grammar, wording, sentence structure, readability, and clarity. The authors carefully reviewed and, where appropriate, revised all suggested edits and take full responsibility for the final manuscript.

REFERENCES

Mahmoud Afifi, Konstantinos G. Derpanis, Bjorn Ommer, and Michael S. Brown. Learning multi-scale photo exposure correction. In *CVPR*, pp. 9157–9167, 2021.

Lorenzo Agnolucci, Leonardo Galteri, Marco Bertini, and Alberto Del Bimbo. Arniqa: Learning distortion manifold for image quality assessment. In *WACV*, 2024.

Masato Akai, Yoshiaki Ueda, Takanori Koga, and Noriaki Suetake. Low-artifact and fast backlit image enhancement method based on suppression of lightness order error. *IEEE Access*, 2023.

Alexandru Brateanu, Tingting Mu, Codruta O. Ancuti, and Cosmin Ancuti. Multinex: Lightweight low-light image enhancement via multi-prior retinex. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, pp. 97–104, 2011.

Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *TIP*, 27(4):2049–2062, 2018.

Yuhui Cao, Yurui Ren, Thomas H. Li, and Ge Li. Over-exposure correction via exposure and scene information disentanglement. In *ACCV*, 2020.

Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Sixiang Chen, Tian Ye, Renjing Pei, Kaiwen Zhou, Fenglong Song, and Lei Zhu. Restoreagent: Autonomous image restoration agent via multimodal large language models. In *NeurIPS*, 2024.

Tomáš Chobola, Yu Liu, Hanyi Zhang, Julia A. Schnabel, and Tingying Peng. Fast context-based low-light image enhancement via neural implicit representations. In *ECCV*, 2024.

Marcos V. Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *ECCV*, 2024.

Ruodai Cui and Lei Zhang. Unice: Training a universal image contrast enhancer, 2025.

Ruodai Cui, Niu Li, and Guosheng Hu. Unsupervised exposure correction. In *ECCV*, 2024.

Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato. Raise: A raw images dataset for digital image forensics. In *MMSys*, pp. 219–224, 2015.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE TPAMI*, 2020.

F. Eyiokur, Dogucan Yaman, Hazim Kemal Ekenel, and Alexander Waibel. Exposure correction model to enhance image quality. In *CVPR*, pp. 676–686, 2022.

Yixu Feng, Shuo Hou, Haotian Lin, Yu Zhu, Peng Wu, Wei Dong, Jinqiu Sun, Qingsen Yan, and Yanning Zhang. Difflight: Integrating content and detail for low-light image enhancement. In *CVPR*, pp. 6143–6152, 2024.

Tatiana Gaintseva, Michael Benning, and Greg Slabaugh. Rave: Residual vector embedding for clip-guided backlit image enhancement. In *ECCV*, pp. 412–428, 2024.

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

Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *TOG*, 35(6), 2016.

Chunming He, Chengyu Fang, Yulun Zhang, Kai Li, Longxiang Tang, Chenyu You, Fengyang Xiao, Zhenhua Guo, and Xiu Li. Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. *ICLR*, 2025.

Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. In *NeurIPS*, 2024.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *ICLR*, 2022.

Zeyu Hua, Hui Li, Yu Wang, Song Wang, Congchao Zhu, and Caixia Zheng. Bi-bridge: Bidirectional diffusion bridges for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 37455–37464, 2026.

Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *CVPR*, pp. 6043–6052, 2022a.

Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *ECCV*, pp. 163–180, 2022b.

Jie Huang, Feng Zhao, Man Zhou, Jie Xiao, Naishan Zheng, Kaiwen Zheng, and Zhiwei Xiong. Learning sample relationship for exposure correction. In *CVPR*, pp. 9904–9913, 2023.

Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *TOG*, 42(6):1–14, 2023.

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

Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *ECCV*, 2024b.

Donggoo Jung, Daehyun Kim, Guanghui Wang, and Tae Hyun Kim. Exposure-slot: Exposure-centric representations learning with slot-in-slot attention for region-aware exposure correction. In *CVPR*, 2025.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *ICCV*, pp. 5148–5157, 2021.

Guanzhou Lan, Qianli Ma, Yuqi Yang, Zhigang Wang, Dong Wang, Xuelong Li, and Bin Zhao. Efficient diffusion as low light enhancer. In *CVPR*, pp. 21277–21286, 2025.

Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *TPAMI*, 2021. doi: 10.1109/TPAMI.2021.3063604.

Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. In *ICLR*, 2023.

Yiyu Li, Ke Xu, Gerhard Petrus Hancke, and Rynson W.H. Lau. Color shift estimation-and-correction for image enhancement. In *CVPR*, 2024a.

Zhong-Yu Li, Xin Jin, Boyuan Sun, Chun-Le Guo, and Ming-Ming Cheng. Towards raw object detection in diverse conditions. *arXiv preprint arXiv:2411.15678*, 2024b.

Ziwen Li, Feng Zhang, Meng Cao, Jinpu Zhang, Yuanjie Shao, Yuehuan Wang, and Nong Sang. Real-time exposure correction via collaborative transformations and adaptive sampling. In *CVPR*, 2024c.

Jie Liang, Hui Zeng, Miaomiao Cui, Xuansong Xie, and Lei Zhang. Ppr10k: A large-scale portrait photo retouching dataset with human-region mask and group-level consistency. In *CVPR*, 2021a.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, 2021b.

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

Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, and Trevor Darrell. Diffusion hyperfeatures: Searching through time and space for semantic correspondence. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 47500–47510. Curran Associates, Inc., 2023. doi: 10.52202/075280-2057. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/942032b61720a3fd64897efe46237c81-Abstract-Conference.html.

Xiaoqian Lv, Shengping Zhang, Qinglin Liu, Haozhe Xie, Bineng Zhong, and Huiyu Zhou. Backlitnet: A dataset and network for backlit image enhancement. *CVIU*, 218:103403, 2022.

Igor Morawski, Yu-An Chen, Yu-Sheng Lin, Shusil Dangi, Kai He, and Winston H. Hsu. Genisp: Neural isp for low-light machine cognition. In *CVPR*, pp. 630–639, 2022.

Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *CVPR*, 2022.

Alex Omid-Zohoor, David Ta, and Boris Murmann. Pascalraw: Raw image database for object detection, 2014-2015.

ITU-T Recommendation P.910. Subjective video quality assessment methods for multimedia applications, 1999.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pp. 10684–10695, 2022.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. *arXiv preprint arXiv:2311.17042*, 2023.

Chenyang Si, Ziqi Huang, Yuming Jiang, and Ziwei Liu. FreeU: Free lunch in diffusion U-Net. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4733–4743, June 2024.

Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, and Lei Zhang. Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. *CVPR*, 2025.

Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, and Bharath Hariharan. Emergent correspondence from image diffusion. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 1363–1389. Curran Associates, Inc., 2023. doi: 10.52202/075280-0068. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/0503f5dce343a1d06d16ba103dd52db1-Abstract-Conference.html.

Yoshiaki Ueda, Daiki Moriyama, Takanori Koga, and Noriaki Suetake. Histogram specification-based image enhancement for backlit image. In *ICIP*, pp. 958–962, 2020.

Khoa D. Vo and Len T. Bui. Starsrgan: Improving real-world blind super-resolution. *arXiv*, 2023.

Chenxi Wang, Hongjun Wu, and Zhi Jin. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *ACMMM*, pp. 7459–7469, 2023a.

Haoyuan Wang, Ke Xu, and Rynson W.H. Lau. Local color distributions prior for image enhancement. In *ECCV*, pp. 343–359, 2022.

Jianyi Wang, Kelvin C.K. Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, 2023b.

Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, pp. 6849–6857, 2019.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV*, pp. 1905–1914, 2021.

Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C. Kot, and Bihan Wen. Exposurediffusion: Learning to expose for low-light image enhancement. In *ICCV*, pp. 12438–12448, 2023c.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.

Wen-Qiang Wu, Chun-Feng Wang, Su-Ting Han, and Cao-Feng Pan. Recent advances in imaging devices: image sensors and neuromorphic vision sensors. *Rare Metals*, 43(11):5487–5515, 2024.

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 *CVPR*, pp. 1662–1671, 2023.

Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *CVPR*, pp. 9893–9903, 2023.

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

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. *arXiv preprint arXiv:2204.08958*, 2022.

Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *ICCV*, pp. 12302–12311, 2023.

Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex++: Retinex-driven reinforced diffusion model for low-light image enhancement. *IEEE TPAMI*, 2025.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pp. 586–595, 2018.

Minghua Zhao, Danni Cheng, and Li Wang. Backlit image enhancement based on foreground extraction. In *ICGIP*, pp. 366–373, 2021.

Dewei Zhou, Zongxin Yang, and Yi Yang. Pyramid diffusion models for low-light image enhancement. *arXiv preprint arXiv:2305.10028*, 2023.

Zhengwei Zhu, Yushi Lei, Yilin Qin, Chenyang Zhu, and Yanping Zhu. Ire: Improved image super-resolution based on real-esrgan. *IEEE Access*, 11:45334–45348, 2023. doi: 10.1109/ACCESS.2023.3256086.

Appendix

In this Appendix, we provide the following materials:

1. Appendix. A: A detailed description of the training data construction pipeline (referring to Sec. 3.1 and Sec. 4.1 in the main paper);
2. Appendix. B: Architectural details of the exposure-aware decomposition network $\mathcal{D}$ (referring to Sec. 3.1 in the main paper);
3. Appendix. C: A user study comparing the competing methods (referring to Sec. 4.1 in the main paper);
4. Appendix. D: Additional visualization results for LLIE, EC, and BIE (referring to Sec. 4.2 in the main paper);
5. Appendix. E: Analysis of hallucinations induced by the diffusion prior and their influence on structural fidelity and aggregate metrics;
6. Appendix. F: Quantitative results and analysis for the BIE task (referring to Sec. 4.2 in the main paper);
7. Appendix. G: Visual analysis of the exposure-aware decomposition results (referring to Sec. 4.2 in the main paper);
8. Appendix. H: Ablation study on the design components of GICE (referring to Sec. 4.2 in the main paper);
9. Appendix. I: Complexity analysis comparing GICE with competing methods (referring to Sec. 4.2 in the main paper).

A Training Data Construction

Based on previous works (Cui & Zhang, 2025; Cui et al., 2024; Afifi et al., 2021), we build a pipeline to construct our training data, as illustrated in Fig. 4. We first collect high-bit RAW images from publicly available datasets, including AODRAW (Li et al., 2024b), FiveK (Bychkovsky et al., 2011), HDRP (Hasinoff et al., 2016), PASCAL-RAW (Omid-Zohoor et al., 2014-2015), PPR10K (Liang et al., 2021a), RAISE (Dang-Nguyen et al., 2015), and RAW-NOD (Morawski et al., 2022). We then process each RAW image using the Adobe Camera RAW SDK, where the exposure value (EV) is adjusted to $\{+3, +2, +1, 0, -1, -2, -3\}$ EV, resulting in a multi-exposure sequence (MES). The pseudo ground-truth (GT) image is generated by fusing the MES images using multi-exposure fusion (MEF) techniques, following the strategy introduced in UNICE (Cui & Zhang, 2025). To enhance diversity and quality, we apply multiple MEF algorithms to each MSE, producing different MEF results, which are randomly mixed to create additional MEF variations. Finally, we assess the synthesized images using no-reference image quality assessment (NR-IQA) models, including ARNIQA (Agnolucci et al., 2024), MUSIQ (Ke et al., 2021), CLIP-IQA+ (Wang et al., 2023b), and MANIQA (Yang et al., 2022). The best-performing image, which achieves the highest average score over these NR-IQA metrics, is selected as the pseudo-GT.

The above process can result in high-quality pseudo-GT images; however, we also need low-quality input images to build the triplet for GICE network training. The MES images synthesized in the above process already contain various degradations in contrast and exposure, but lack structural degradations. One direct solution is to apply the commonly used ESRGAN degradation pipeline Wang et al. (2021) to the MES images. However, this method generally imposes global structural distortions, which are not well suited to our task. Empirical observations indicate that darker regions are more susceptible to degradation than brighter areas due to their lower signal-to-noise ratio (SNR), making them more vulnerable to noise and artifacts. Motivated by this, we introduce a luminance-aware structural degradation strategy. Specifically, we first apply a global degradation to the entire image and then select a luminance threshold. Pixels below this threshold undergo additional high-order degradations, thereby simulating the spatially variant degradation patterns commonly observed in real-world conditions.

Finally, after manually filtering out unsuitable images, such as those of poor quality or calibration targets, we construct a dataset comprising 47,783 groups, each group containing seven degraded

Table 3: Statistics of our constructed training dataset.

Subdataset	Total	Filtered	Retained
AODRAW	7,785	457	7,328
fivek	5,000	57	4,943
HDRP	3,640	15	3,625
PASCAL-RAW	4,259	0	4,259
PPR10K	11,161	13	11,148
RAISE	8,155	34	8,121
RAW-NOD	8,361	2	8,359
OVERALL	48,361	578	47,783

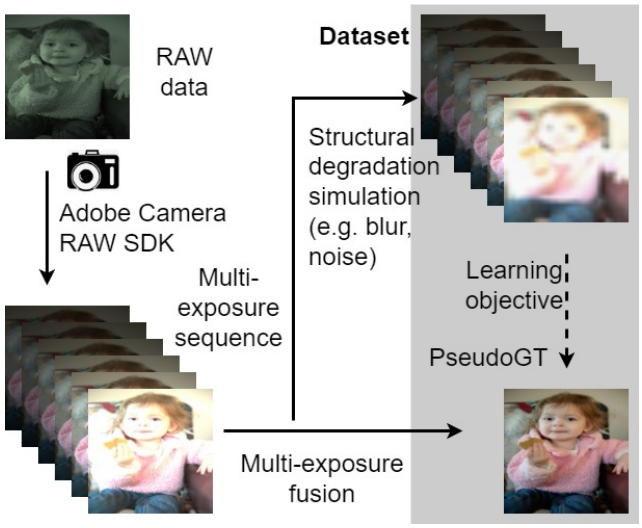


Figure 4: Illustration of our training data construction pipeline.

images at different exposure levels ($\pm$3EV) and one pseudo-GT image obtained through multi-exposure fusion (MEF). The statistics of the dataset are summarized in Tab. 3. Fig. 5 presents some examples of our dataset.

B Architecture of the Decomposition Network

This section provides the architectural details of the exposure-aware decomposition network $\mathcal{D}$. The network is designed as a lightweight fully convolutional module that preserves the spatial resolution of the input image. Given an RGB image $I \in [0,1]^{H\times W\times 3}$, $\mathcal{D}$ predicts a three-channel exposure-variant component

$$E = \mathcal{D}(I), \tag{3}$$

and the corresponding exposure-invariant residual is obtained deterministically as

$$C = I - E. \tag{4}$$

Thus, the two components preserve the complete input information by construction.

Specifically, $\mathcal{D}$ consists of four 3×3 convolutional layers. Each convolution is preceded by reflection padding of one pixel. Reflection padding avoids introducing artificial constant values near image boundaries and ensures that all intermediate features have the same spatial resolution as the input. The first convolution maps the three-channel input to N feature channels, where N denotes the base channel number. The following two convolutions maintain N channels, and the final convolution projects the features back to three channels. The first three convolutional layers are followed by



Figure 5: Examples of our dataset. Each group consists of seven degraded multi-exposure images and one pseudo-GT image.

ReLU activation functions, while a sigmoid function is applied after the final convolution to constrain the predicted exposure-variant component to the range $[0, 1]$.

The decomposition process can be written as

$$X_1 = \mathrm{ReLU}(\mathrm{Conv}_{3\to N}(\mathrm{RP}(I)))\,, \tag{5}$$
$$X_2 = \mathrm{ReLU}(\mathrm{Conv}_{N\to N}(\mathrm{RP}(X_1)))\,, \tag{6}$$
$$X_3 = \mathrm{ReLU}(\mathrm{Conv}_{N\to N}(\mathrm{RP}(X_2)))\,, \tag{7}$$
$$E = \mathrm{Sigmoid}(\mathrm{Conv}_{N\to 3}(\mathrm{RP}(X_3)))\,, \tag{8}$$

where $\mathrm{RP}(\cdot)$ denotes reflection padding with a padding size of one pixel. All convolutional layers use a stride of one and no internal zero padding. Consequently, both E and C have the same spatial resolution as the input image.

During training, the two branches of the Siamese decomposition network process observations of the same scene captured at different exposure levels. The two branches share the complete set of parameters of $\mathcal{D}$. At inference time, only a single branch is required, and the decomposition network operates on one input image without requiring a paired observation or exposure metadata. The decomposition network does not independently predict C, which removes the ambiguity of estimating two unconstrained components and guarantees the additive relation $I = E + C$.

C User Study

To more comprehensively evaluate the competing image contrast enhancement methods, we conduct a user study on the LLIE, EC and BIE tasks. For each task, we compared both task-specific models and general-purpose approaches. The task-specific models include DReti++ (Yi et al., 2025), Reti-Diff (He et al., 2025), and ReDDiT (Lan et al., 2025) for LLIE; Ex-Slot (Jung et al., 2025), CSEC (Li et al., 2024a), and CoTF (Li et al., 2024c) for EC; and CLIP-LIT (Liang et al., 2023), RAVE (Gaintseva et al., 2024), and ReDDiT (Lan et al., 2025) for BIE. The general-purpose methods include Bi-Bridge (Hua et al., 2026), AutoDIR (Jiang et al., 2024b), and UNICE (Cui & Zhang, 2025). In addition, we couple these methods with PiSA-SR (Sun et al., 2025) to further enhance their perceptual quality, with results denoted by "†". All models are trained on our constructed dataset to ensure a fair and consistent comparison.

For each task, 200 images are randomly sampled from the corresponding evaluation dataset. Each image is processed by all competing methods and the outputs are anonymized and randomly shuffled. Ten volunteers participated in the study. During evaluation, each volunteer was presented with the enhanced images for each sampled image, displayed side-by-side in random order. Each enhanced image was rated on a five-point Likert scale, considering three aspects: contrast improvement, detail restoration, and overall visual quality. All ratings were aggregated and averaged to obtain the final score for each method. Statistical significance was assessed using paired t-tests between methods. This user study complements our quantitative evaluations and offers insights into the real-world perceptual performance of each approach.

The results are summarized in Tab. 4. Our method achieves a substantial improvement in perceptual quality, consistently outperforming competing approaches in contrast enhancement, detail restoration, and overall visual appeal. This advantage highlights the effectiveness of a unified training strategy that harmonizes global contrast adjustment with fine-grained detail recovery, enabling robust performance across low-light, exposure correction, and backlit scenarios.

D More Visual Results

In this section, we provide additional visual results for low-light image enhancement (LLIE), exposure correction (EC), and backlit image enhancement (BIE).

D.1 More Visual Results for LLIE

Fig. 6 presents additional qualitative results for LLIE. The first two examples are sampled from LSRW (Hai et al., 2023), while the last example comes from UHDLL (Li et al., 2023). In the first example, competing methods produce overly blurry outputs (*e.g.*, ReDDiT (Lan et al., 2025)) or amplify noise (*e.g.*, UNICE (Cui & Zhang, 2025)). Combining them with PiSA-SR (Sun et al., 2025) can further aggravate noise artifacts. In the second and third examples, ReDDiT (Lan et al., 2025) again produces blurred reconstructions, while UNICE† introduces visible distortions due to diffusion hallucination, such as inconsistent text compared with the ground truth.

D.2 Visual Results for EC

Fig. 7 shows qualitative comparisons for the EC task for models trained on our dataset. The first example contains severe overexposure: our method reconstructs plausible missing content, whereas competing approaches mainly adjust global brightness/contrast without recovering textures. The second example exhibits localized overexposure, where our approach is the only one that consistently

Table 4: User study results on LLIE, EC, and BIE tasks. Higher scores indicate better perceptual quality.

Task	Method	Contrast	Detail	Overall
LLIE	DReti++*	2.8	2.7	2.8
	DReti++*†	3.0	2.9	3.0
	Reti-Diff*	2.9	2.8	2.9
	Reti-Diff*†	3.1	3.0	3.1
	ReDDiT*	3.0	2.9	3.0
	ReDDiT*†	3.2	3.1	3.2
	Bi-Bridge*	3.1	3.2	3.3
	Bi-Bridge*†	3.4	3.5	3.4
	AutoDIR*	3.1	3.0	3.1
	AutoDIR*†	3.3	3.2	3.3
	UNICE	3.6	3.5	3.6
	UNICE†	3.8	3.7	3.8
	GICE (Ours)	**4.2**	**3.9**	**4.3**
EC	Ex-Slot*	2.9	2.8	2.9
	Ex-Slot*†	3.1	3.0	3.1
	CSEC*	3.0	2.9	3.0
	CSEC*†	3.2	3.1	3.2
	CoTF*	2.8	2.7	2.8
	CoTF*†	3.0	2.9	3.0
	Bi-Bridge*	3.1	3.0	3.3
	Bi-Bridge*†	3.5	3.4	3.5
	AutoDIR*	3.0	2.9	3.0
	AutoDIR*†	3.2	3.1	3.2
	UNICE	3.7	3.6	3.7
	UNICE†	3.9	3.8	3.9
	GICE (Ours)	**4.2**	**4.3**	**4.3**
BIE	CLIP-LIT*	2.7	2.6	2.7
	CLIP-LIT*†	2.9	2.8	2.9
	RAVE*	2.8	2.7	2.8
	RAVE*†	3.0	2.9	3.0
	ReDDiT*	2.9	2.8	2.9
	ReDDiT*†	3.1	3.0	3.1
	Bi-Bridge*	3.2	3.1	3.2
	Bi-Bridge*†	3.4	3.1	3.2
	AutoDIR*	3.0	2.9	3.0
	AutoDIR*†	3.2	3.1	3.2
	UNICE	3.5	3.4	3.5
	UNICE†	3.7	3.6	3.7
	GICE (Ours)	**4.0**	**4.1**	**4.1**

restores missing details. The third example is a low-light case, where our method achieves better visibility while preserving structures.

Fig. 8 further compares models trained on SICE (Cai et al., 2018) with our model. Even when competitors are trained in-domain, they tend to under-perform in restoring structurally degraded regions (such as saturated highlights or noisy dark areas), while our method maintains higher structural fidelity with more natural appearance.

D.3 Visual Results for BIE

Fig. 10 presents visual comparisons on the Backlit dataset (Liang et al., 2023), which does not provide ground-truth. In the first example, our approach produces the most balanced local contrast and avoids

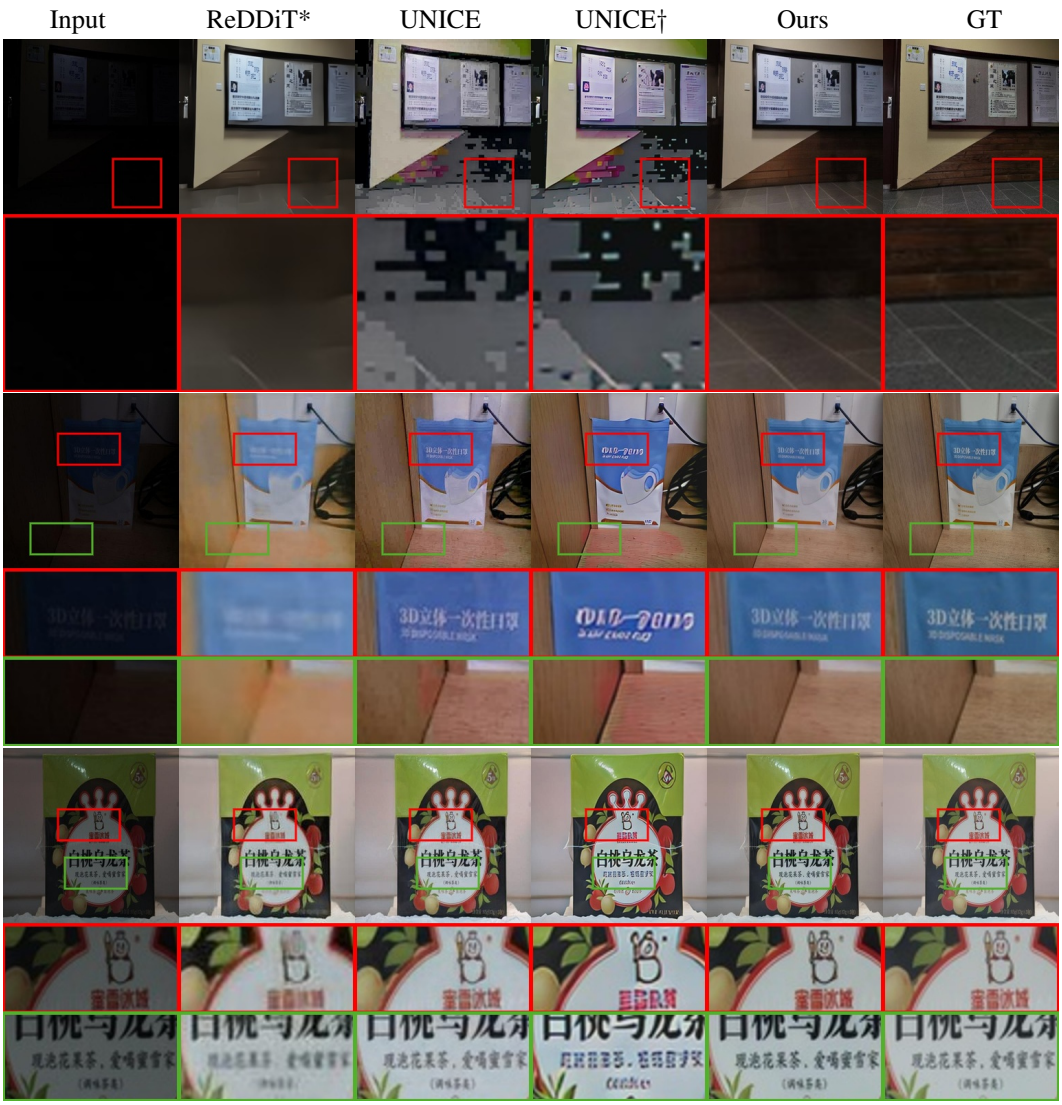


Figure 6: Qualitative comparison of different methods for the LLIE task. Enlarged local regions are highlighted with red and green boxes for better visualization.

over-amplification. In the second example, several baselines (*e.g.*, RAVE* (Gaintseva et al., 2024), ReDDiT* (Lan et al., 2025), UNICE (Cui & Zhang, 2025)) noticeably amplify JPEG compression artifacts, while UNICE† only partially recovers fine structures. In contrast, our method yields cleaner textures with stronger detail restoration. In the third example, RAVE* (Gaintseva et al., 2024) and ReDDiT* (Lan et al., 2025) fail to recover fine details, and UNICE (with or without †) suffers from limited visibility, whereas GICE delivers the best perceptual quality.

E Effect of Diffusion-Prior Hallucination

Diffusion priors can potentially introduce content that is not supported by the input. We therefore examine this issue using representative exposure-correction examples in Fig. 11. From left to right, each group shows the input, Ex-Slot, ReDDiT, UNICE, our result, and the ground truth; the lower half of each group enlarges the red-box region.

The examples show that the proposed generation process remains strongly constrained by the observed scene. In the first example, GICE restores a clean sky and preserves the skyline and tower geometry,

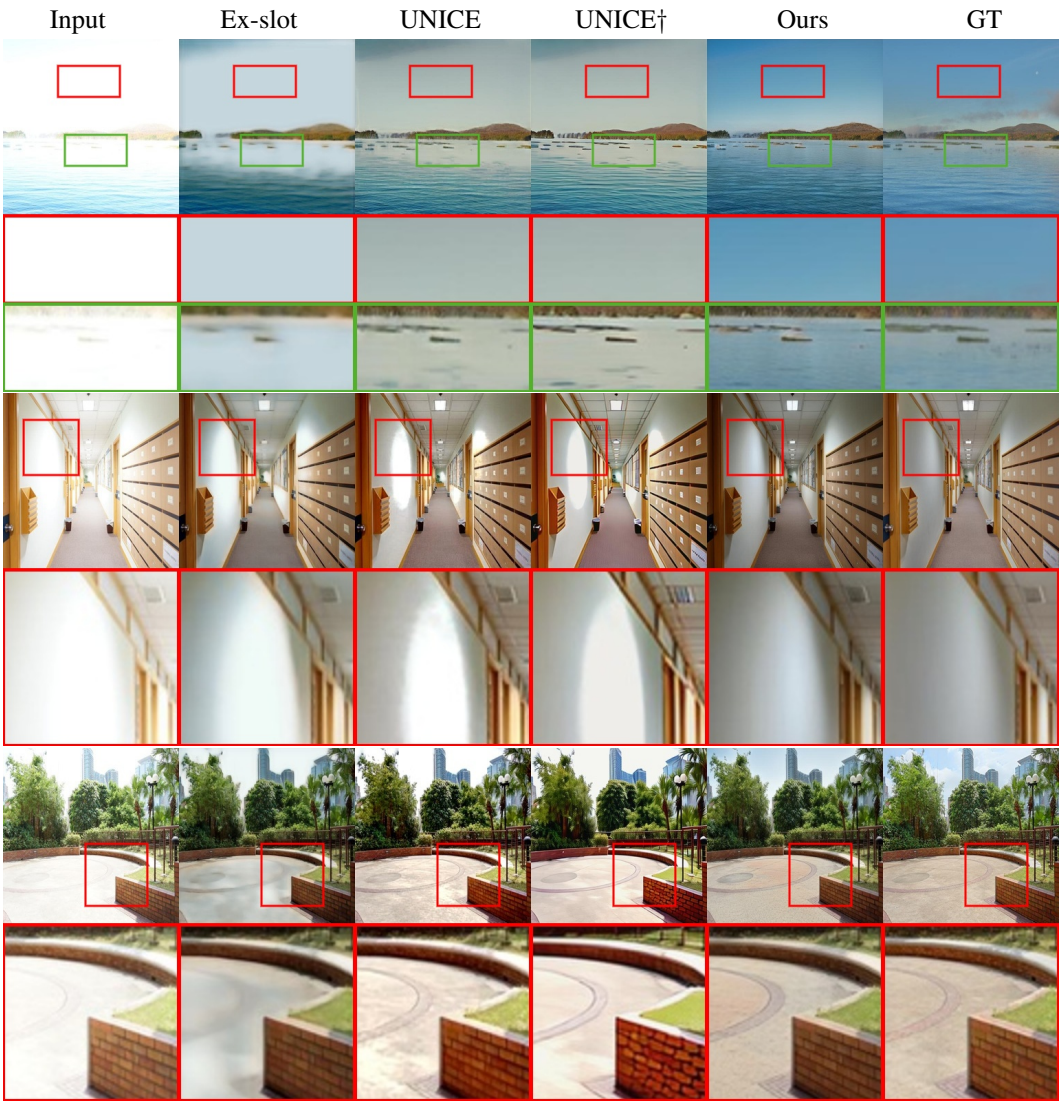


Figure 7: Qualitative comparison of different methods for the EC task. Enlarged local regions are highlighted with red and green boxes for better visualization.

without introducing the conspicuous cloud structures visible in the ground truth but absent from the saturated input. In the second example, GICE recovers the pavement, tree trunk, railing, and the person-like foreground structure at locations consistent with the input. Although the ground truth contains an additional signboard that is largely unobservable in the over-exposed input, GICE does not fabricate this missing object. In the third example, GICE reconstructs the river region with spatially coherent shading while maintaining the shoreline and surrounding vegetation. The recovered water texture is smoother than the ground truth, but no salient new object or geometric structure is generated. In contrast, the other compared methods fail to recover these clipped regions, often producing visually unnatural content and noticeable artifacts. These observations indicate that hallucination, when present, is limited mainly to low-amplitude appearance differences in severely clipped regions, rather than changes to scene layout or object geometry.

F Quantitative Results for BIE

For the BIE task, in addition to RAVE (Gaintseva et al., 2024) and CLIP-LIT (Liang et al., 2023), we also adapt several LLIE methods, including DReti++ (Yi et al., 2025), Reti-Diff (He et al., 2025),

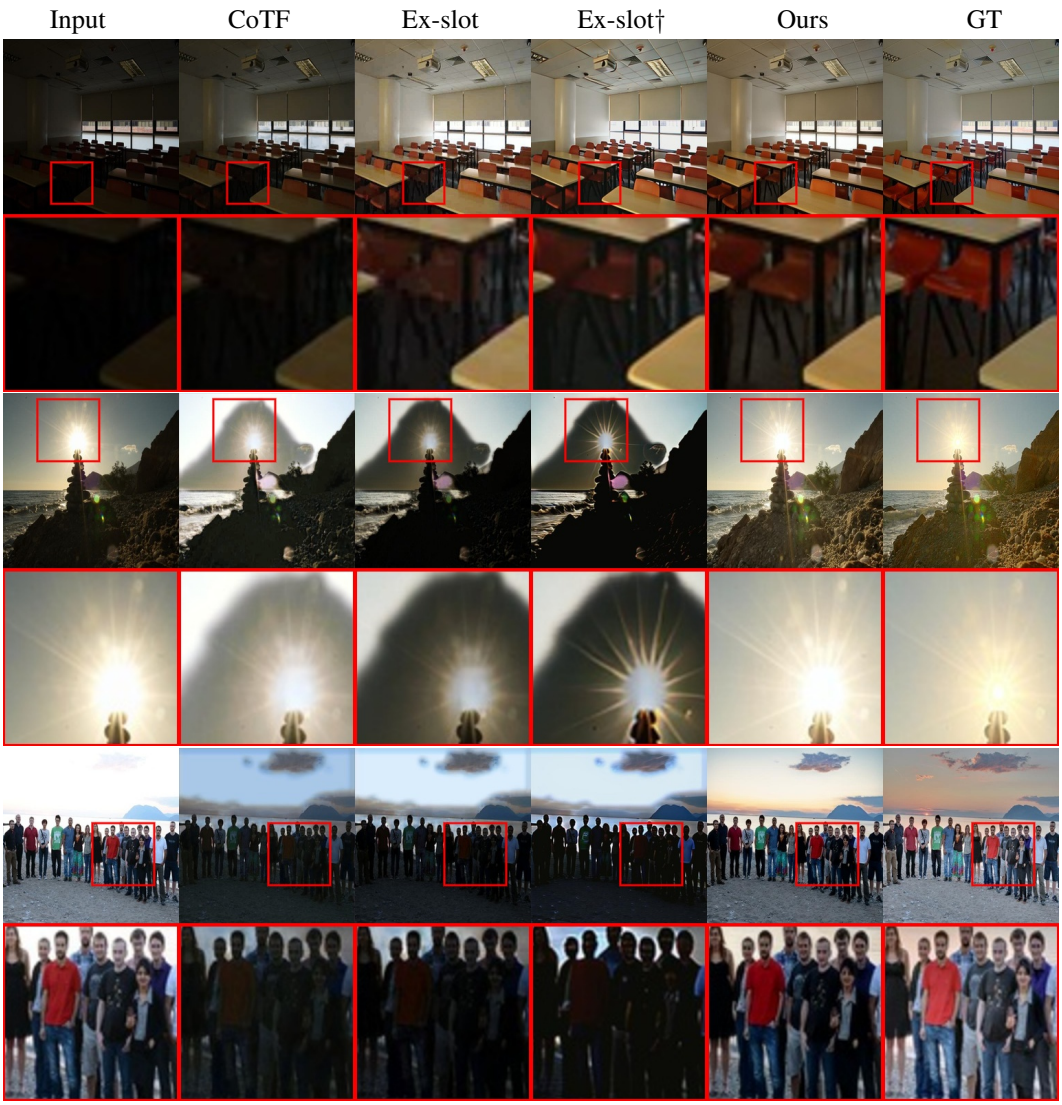


Figure 8: Qualitative comparison of different methods for the EC task. Enlarged local regions are highlighted with red and green boxes for better visualization.

and ReDDiT (Lan et al., 2025), for comparison. For CLIP-LIT, we use the implementation provided in (Gaintseva et al., 2024), which incorporates several improvements to enhance performance. Two datasets are used in the evaluation: Backlit (Liang et al., 2023) and BAID (Lv et al., 2022). The Backlit dataset does not provide GT, whereas BAID includes GT. We train models separately on BAID and our dataset and conduct both in-domain and out-of-domain evaluations on BAID and Backlit, respectively.

The experimental results are shown in Table 5. We have the following observations. First, the in-domain test (train and test on BAID) leads to high PSNR and SSIM indices; however, the perceptual quality measured by no-reference metrics is low, implying that optimization toward pixel-level accuracy does not necessarily improve domain-independent visual quality. Second, when BAID-trained models are evaluated on the Backlit dataset, the degradation in no-reference scores becomes substantial, highlighting the limited generalization capability of existing approaches and underscoring the sensitivity of perceptual quality to distribution shifts. Finally, our proposed method consistently achieves superior performance. It not only attains the highest full-reference scores on BAID but also demonstrates generalization capability on Backlit, outperforming all baselines (RAVE, CLIP-LIT,

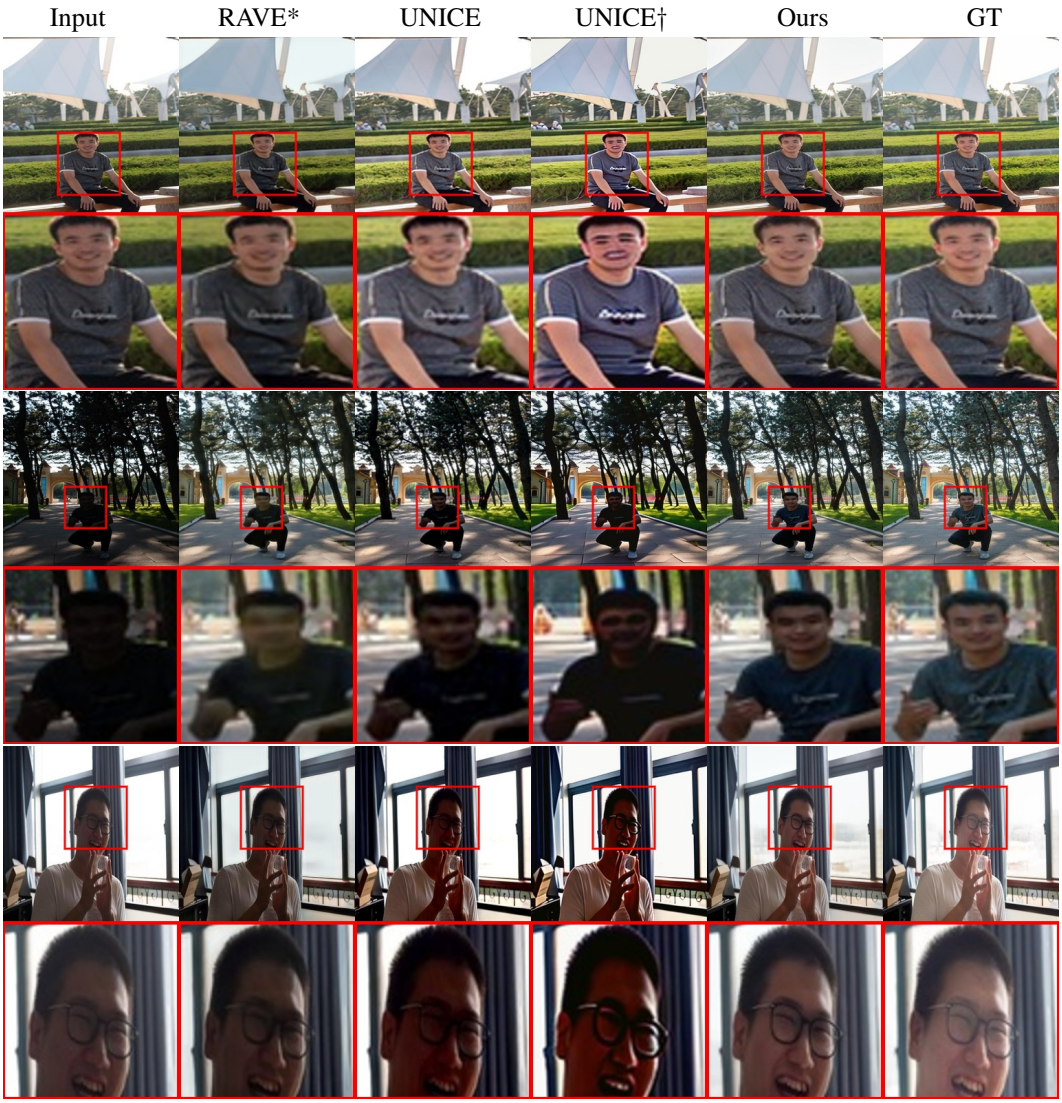


Figure 9: Qualitative comparison of different methods for the BIE task on the BAID (Lv et al., 2022) dataset. Enlarged local regions are highlighted with red and green boxes for better visualization.

DReti++, Reti-Diff, ReDDiT) in perceptual metrics. These findings validate the effectiveness of our approach in harmonizing structural fidelity with perceptual enhancement.

G Visualization of Decomposition Results

Figure 12 visualizes the decomposition of two differently exposed observations of the same scene. In both rows, the exposure-variant component preserves the overall color and coarse spatial layout while suppressing high-frequency details, resulting in a smoother representation of the sky, ground, vegetation, and rocks. The component also changes with the input exposure: the darker observation produces a correspondingly darker appearance component, showing that this branch responds to exposure-dependent intensity and contrast.

The exposure-invariant residual highlights object boundaries and fine structures, including the outlines of the tree branches, rock contours, shrubs, and cloud edges. Despite the clear brightness difference between the two inputs, these structures appear at similar locations in both residual maps. The visual results therefore show a complementary separation: the exposure-variant component mainly carries

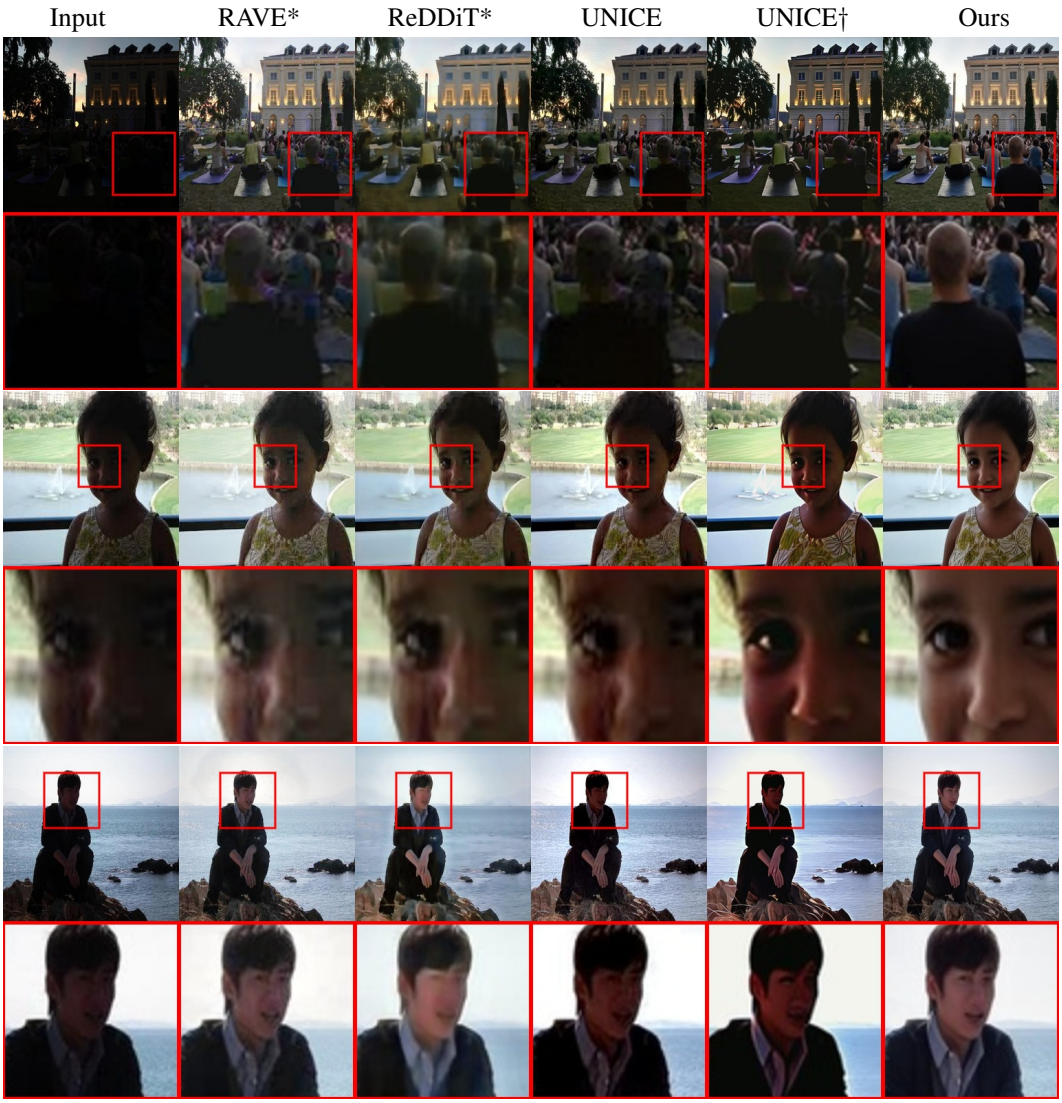


Figure 10: Qualitative comparison of different methods for the BIE task on the Backlit (Akai et al., 2023) dataset. **Note that this dataset does not provide GT, and the last column shows our results.** Enlarged local regions are highlighted with red and green boxes for better visualization.

large-scale appearance variations, whereas the residual retains scene geometry and texture cues that remain comparatively stable across exposures.

H Ablation Study

We conduct ablation studies to quantify the contribution of each design component in GICE. We evaluate ten variants: (1) w/o Triplet Supervision, where we replace triplet training $(I_i, I_j, I_{\text{gt}})$ with paired training (I_i, I_{gt}); (2) w/o Adaptive Decomposition, where we remove the adaptive decomposition and directly feed the input image into the diffusion model; (3) w/o Skip Convolution, where we remove the residual skip-convolution pathway; (4) w/o Learnable Prompts, where the learnable prompt embeddings are disabled; (5) w/o Degradation Simulation, where we remove the proposed degradation synthesis for constructing structurally degraded inputs; (6) w/o Diffusion Prior, where we do not load pretrained diffusion weights and train the backbone from scratch; (7) w/o Multi-layer Injection, where multi-layer feature injection is disabled; (8) w/o Multi-layer Adaptive



Figure 11: Visual analysis of diffusion-prior hallucination on representative exposure-correction examples. Red boxes indicate the enlarged regions shown beneath each example.

Weights, where the adaptive weights for multi-layer feature injection are removed; (9) w/o E, where the E module is removed; and (10) w/o C, where the C module is removed.

All variants are trained under the same setting as the full model and evaluated on LSRW. The results are summarized in Tab. 6. We can see that removing any component degrades the performance, validating the contribution of each component in our model design.

I Complexity Analysis

We compare the complexity of GICE with those diffusion-based approaches (modified and re-trained by us, marked with "*"), including DiffLL (Jiang et al., 2023), Diff-Reti++ (Lan et al., 2025), Reti-Diff (He et al., 2025), ReDDiT (Lan et al., 2025), Bi-Bridge (Hua et al., 2026), AutoDIR (Jiang et al., 2024b), Multinex (Brateanu et al., 2026), and UNICE (Cui & Zhang, 2025), as well as their variants combined with PiSA-SR (Sun et al., 2025) (marked with †). As shown in Tab. 7, GICE has a reasonable number of parameters (inherent from its SD-Turbo backbone) and memory usage and shows the fastest speed. This is primarily because it adopts a single-step diffusion process, whereas

Table 5: Quantitative comparisons on the BIE task. Tr./Te. denote the training/testing datasets. LPI (*resp.* DIS, ARN, MUS, CIQA, MAN) denotes LPIPS (*resp.* DISTS, ARNIQA, MUSIQ, CLIP-IQA+, MANIQA). Symbols ↑ and ↓ indicate that a higher or lower value is better. **Bold** and Underlined numbers indicate the best and second-best results. Symbols "*" and "†" mark larger variants and those variants coupled with PiSA-SR (Sun et al., 2025). For Backlit dataset, FR-IQA metrics are unavailable due to the lack of ground-truth and are reported as "–".

Tr.	Te.	Method	PSNR ↑	SSIM ↑	LPI. ↓	DIS. ↓	ARN. ↑	MUS. ↑	CIQA. ↑	MAN. ↑
BAID	BAID	CLIP-LIT	22.42	0.9365	0.0889	0.1092	0.6791	47.37	0.5429	0.3370
		RAVE	22.56	0.9325	0.0820	0.0997	0.6765	48.73	0.5471	0.3869
		DReti++	23.13	0.9427	0.0880	0.0996	0.6704	47.80	0.5553	0.3445
		Reti-Diff	24.32	0.9556	0.0762	0.0737	0.6839	49.55	0.5650	0.4009
		ReDDiT	**25.32**	**0.9639**	**0.0735**	**0.0649**	0.6903	49.13	0.5762	0.3890
Ours		CLIP-LIT*	18.64	0.8619	0.1096	0.1114	0.6879	50.84	0.5920	0.4138
		CLIP-LIT†	18.70	0.8719	0.1032	0.1148	0.6838	51.36	0.6016	0.4188
		RAVE*	18.71	0.8781	0.1024	0.1176	0.6859	51.64	0.6072	0.4176
		RAVE†	18.83	0.8766	0.0998	0.1002	0.6907	50.13	0.5956	0.4211
		DReti++*	19.01	0.8881	0.0825	0.1002	0.6943	50.21	0.6045	0.4375
		DReti++†	19.08	0.8977	0.0892	0.1025	0.6962	50.24	0.6013	0.4378
		Reti-Diff*	19.11	0.8981	0.0935	0.1052	0.6817	51.48	0.6043	0.4335
		Reti-Diff†	19.15	0.8914	0.0904	0.1055	0.6885	51.36	0.6166	0.4315
		ReDDiT*	19.20	0.9048	0.0952	0.1095	0.6802	51.27	0.6145	0.4324
		ReDDiT†	19.31	0.8868	0.0906	0.1049	0.6881	52.90	0.6166	0.4369
		Bi-Bridge*	19.44	0.9009	0.0902	0.1059	0.6906	51.15	0.6072	0.4375
		Bi-Bridge†	19.26	0.9066	0.0862	0.1037	0.6925	50.73	0.6096	0.4436
		AutoDIR	19.46	0.8898	0.0865	0.0974	0.6940	52.06	0.6128	0.4451
		AutoDIR†	19.67	0.9110	0.0846	0.0998	0.6935	51.14	0.6125	0.4588
		UNICE	19.80	0.9018	0.0853	0.0984	0.6971	52.20	0.6181	0.4526
		UNICE†	19.86	0.8965	0.0873	0.0977	0.6961	53.49	0.6254	0.4470
		Ours	20.22	0.9044	0.0765	0.0861	**0.6973**	**53.85**	**0.6398**	**0.4620**
BAID	Backlit	CLIP-LIT	–	–	–	–	0.6418	42.93	0.5390	0.2822
		RAVE	–	–	–	–	0.6431	43.34	0.5428	0.3051
		DReti++	–	–	–	–	0.6520	44.93	0.5512	0.3330
		Reti-Diff	–	–	–	–	0.6561	43.60	0.5694	0.3510
		ReDDiT	–	–	–	–	0.6580	44.54	0.5634	0.3674
Ours		CLIP-LIT*	–	–	–	–	0.6443	52.21	0.6441	0.4122
		CLIP-LIT†	–	–	–	–	0.6445	52.99	0.6538	0.4120
		RAVE*	–	–	–	–	0.6495	52.99	0.6366	0.3887
		RAVE†	–	–	–	–	0.6495	52.96	0.6454	0.3805
		DReti++*	–	–	–	–	0.6535	52.01	0.6380	0.3671
		DReti++†	–	–	–	–	0.6547	52.50	0.6354	0.3621
		Reti-Diff*	–	–	–	–	0.6552	49.61	0.6600	0.4195
		Reti-Diff†	–	–	–	–	0.6557	49.43	0.6603	0.4199
		ReDDiT*	–	–	–	–	0.6565	52.31	0.6381	0.3501
		ReDDiT†	–	–	–	–	0.6573	49.43	0.6521	0.4151
		Bi-Bridge*	–	–	–	–	0.6548	49.02	0.6670	0.4240
		Bi-Bridge†	–	–	–	–	0.6611	51.83	0.6396	0.3416
		AutoDIR	–	–	–	–	0.6639	46.61	0.6641	0.4286
		AutoDIR†	–	–	–	–	0.6652	47.83	0.6555	0.3946
		UNICE	–	–	–	–	0.6630	53.30	0.6508	0.4109
		UNICE†	–	–	–	–	0.6640	53.25	0.6674	0.4286
		Ours	–	–	–	–	**0.6712**	**54.13**	**0.6854**	**0.4483**

many competing methods rely on multiple steps or iterative refinement stages, which substantially increase computational cost. When PiSA-SR (Sun et al., 2025) is integrated, these methods become even more resource-intensive. Overall, GICE provides a favorable trade-off between complexity and performance, making it well-suited for practical scenarios.

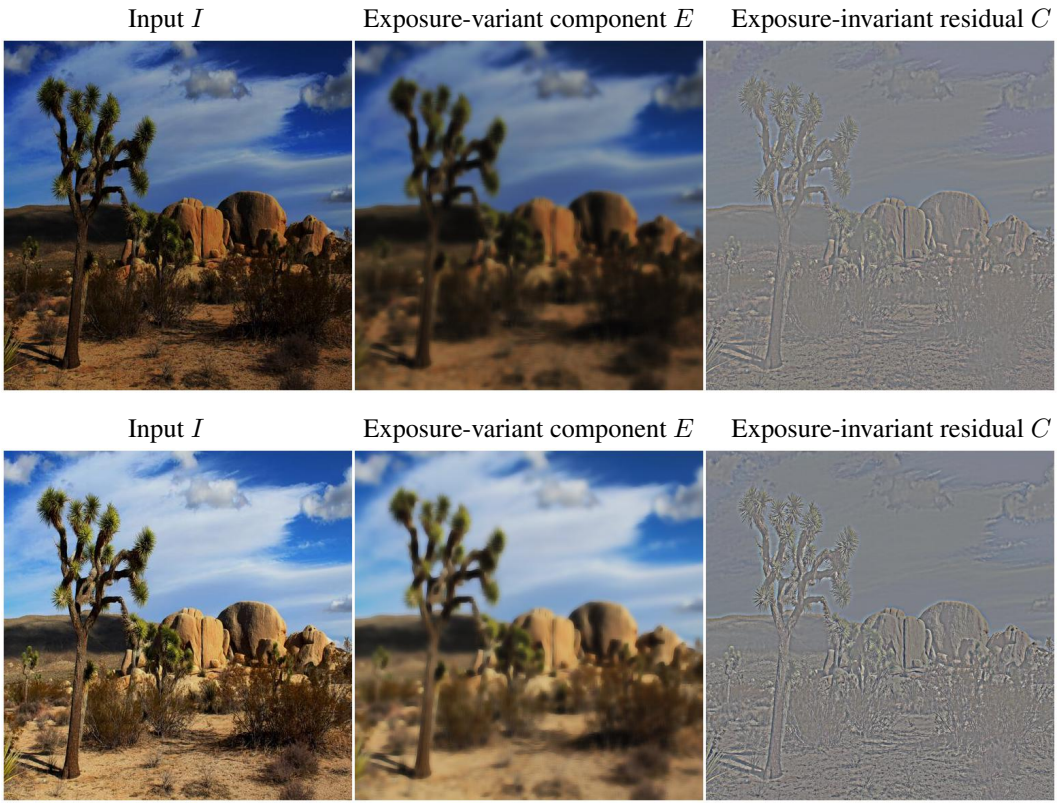


Figure 12: Visualization of the exposure-aware decomposition for two observations of the same scene captured at different exposure levels.

Table 6: Ablation study on the variants of GICE.

Variant	PSNR ↑	SSIM ↑	MUSIQ ↑	MANIQA ↑
w/o Triplet Supervision	20.72	0.7892	51.59	0.4463
w/o Adaptive Decomposition	19.84	0.7706	50.86	0.4274
w/o Skip Convolution	21.28	0.8011	53.08	0.4539
w/o Learnable Prompts	21.35	0.8030	53.12	0.4551
w/o Degradation Simulation	20.36	0.7737	51.24	0.4382
w/o Diffusion Prior	20.78	0.7698	49.60	0.4180
w/o Multi-layer Injection	21.39	0.8042	53.71	0.4546
w/o Multi-layer Adaptive Weights	21.55	0.8087	54.26	0.4573
w/o E	21.08	0.7951	52.96	0.4487
w/o C	20.69	0.7879	52.10	0.4421
Full model	**21.81**	**0.8146**	**55.44**	**0.4628**

Table 7: Complexity comparison between GICE and those diffusion-based methods. For ReDDiT*, FPS is computed using the 2-step variant. For methods Multinex, UNICE and AutoDIR, we average their runtime on a test set to calculate FPS.

Method	Steps	Parameters	GPU memory (GB).	FPS
DiffLL*	10	1.2B	6.8	0.49
DiffLL*†	11	2.3B	6.8	0.46
Diff-Reti++*	5	1.2B	6.8	1.03
Diff-Reti++*†	6	2.3B	6.8	0.89
Reti-Diff*	4	1.2B	6.8	1.25
Reti-Diff*†	5	2.3B	6.8	1.05
ReDDiT*	2–8	1.2B	6.8	2.52
ReDDiT*†	3–9	2.3B	6.8	1.83
Bi-Bridge*	4	1.2B	6.7	1.25
Bi-Bridge*†	5	2.3B	6.9	1.07
AutoDIR	50–350	1.2B	26.2	1.09
AutoDIR†	51–351	2.3B	26.2	0.94
Multinex	1	7.0B	28.4	0.86
Multinex†	2	8.1B	28.4	0.74
UNICE	4	2.2B	8.3	0.67
UNICE†	5	3.3B	8.3	0.61
Ours	1	1.1B	6.9	5.36