CONTROLLIGHT: TOWARDS CONTROLLABLE, CONSISTENT, AND GENERALIZABLE LOW-LIGHT ENHANCEMENT

Anonymous authors
Paper under double-blind review

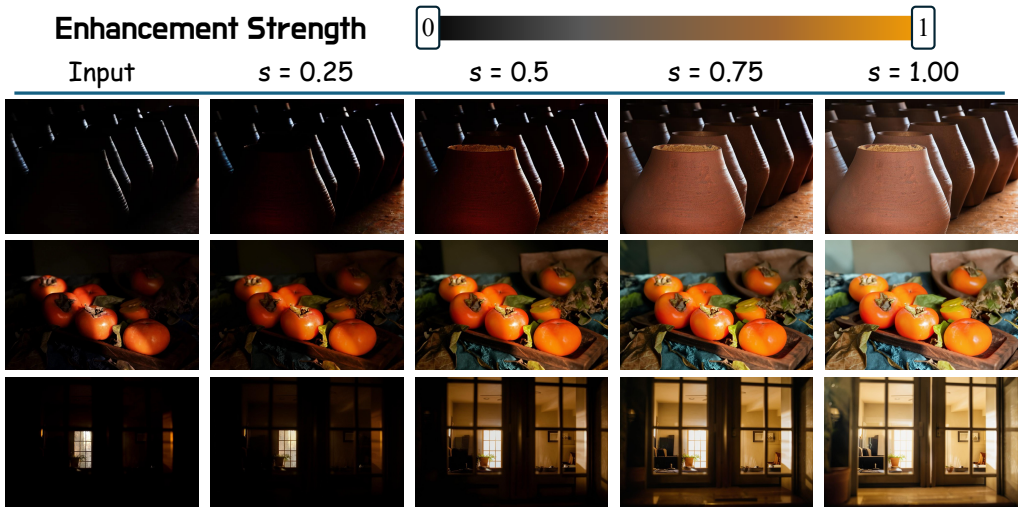

Figure 1: Given a low-light input, **ControlLight** supports continuous adjustment of the enhancement strength from $s = 0$ to $s = 1$, producing smooth and controllable restoration consistent results across real-world scenes.

ABSTRACT

Existing deep learning-based low-light enhancement methods are typically trained on limited datasets with single enhancement targets, which restricts their generalization ability and controllability in real-world applications. To overcome these limitations, we propose ControlLight, a controllable, consistent, and generalizable framework for low-light enhancement. We first construct a large-scale dataset of real-world degraded images with continuous illumination-strength supervision. To further ensure consistent outputs under different control strengths, we introduce a misalignment-aware weighted flow matching loss that preserves image structure across continuous enhancement strengths. ControlLight allows users to edit real-world degraded low-light images toward satisfactory enhancement results by flexibly controlling the strength while preserving visual consistency and realism. Extensive experiments show that ControlLight achieves state-of-the-art performance against existing low-light enhancement approaches while demonstrating strong continuous controllability and generalization to real-world scenarios.

1 INTRODUCTION

Low-light enhancement aims to recover degraded images captured under low-light conditions by restoring details in dark regions while suppressing noise. With the development of deep learning, many methods (Cai et al., 2023; Chen et al., 2018; Pizer, 1990; Weng et al., 2024; Zhang et al., 2019; Wang et al., 2022; Zhou et al., 2023; Wang et al., 2023b) have demonstrated strong capability in low-light image restoration. However, most existing datasets typically provide only a single supervision target for each low-light image, forcing the model to learn a fixed enhancement strength

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.

without controllability. This limitation is critical in practical applications, where users often need to freely adjust the enhancement strength according to different images and personal preferences.

Meanwhile, large-scale image editing models (Esser et al., 2024; Liu et al., 2025; Wu et al., 2025; Team et al., 2025), such as Nano Banana Pro (Team et al., 2023) and FLUX.2-klein (Labs et al., 2025), have demonstrated strong generalization ability across both high-level and low-level vision tasks. Trained on massive image–text paired data, these models possess powerful generative priors and can recover visually plausible details while largely preserving the overall scene structure, making them promising for low-light enhancement. However, most large image editing models provide a single enhancement strength by giving instructions and may introduce hallucinated textures or structural distortions due to their generative nature, which limits their fine-grained controllability and reliability in practical low-light enhancement scenarios.

To address these issues, we construct **Light100K**, a continuous low-light enhancement dataset containing real degraded low-light images and structure-consistent pseudo-enhanced targets with different illumination strengths. This dataset provides fine-grained supervision for controllable enhancement.

We further observe that diffusion-generated pseudo targets, despite offering strong appearance supervision, may contain subtle edge misalignment with the input images. Directly applying flow matching to such targets can cause the model to inherit and amplify these offsets, leading to structural artifacts. To mitigate this issue, we propose a Misalignment-Aware Weighted Flow Matching Loss, which down-weights unreliable target-edge regions and encourages structure preservation from the input image.

Based on Light100K and the proposed Misalignment-Aware Weighted Flow Matching Loss, we train **ControlLight** on FLUX.2-klein-9B with LoRA (Hu et al., 2022). By conditioning on the LoRA strength, ControlLight enables continuous and fine-grained low-light enhancement, producing smooth illumination changes while preserving scene structure.

In summary, our contributions are threefold:

- We construct **Light100K**, a continuous low-light enhancement dataset containing training groups, providing fine-grained supervision for controllable low-light enhancement.
- We reveal that visually plausible diffusion-generated pseudo pairs can still contain subtle edge misalignment, and propose a **Misalignment-Aware Weighted Flow Matching Loss** that anchors the enhanced output edges to the input image structure while down-weighting unreliable target-edge regions.
- We develop **ControlLight**, a continuous low-light enhancement model that enables smooth control over enhancement strength and achieves state-of-the-art performance compared with both continuous and non-continuous low-light enhancement methods.

2 Related Work

2.1 Low-light Enhancement Methods

Many deep learning-based low-light enhancement methods (Wang et al., 2023b; 2024b; Feijoo et al., 2025) incorporate classical imaging priors, especially Retinex theory (Land, 1977), to restore image brightness. With paired datasets such as LOL (Yang et al., 2021) and LSRW (Hai et al., 2023), these methods learn mappings from low-light inputs to normal-light outputs. EnlightenGAN (Jiang et al., 2021b) learns enhancement from unpaired normal-light images with a GAN-based framework (Zhu et al., 2017), while Retinexformer (Cai et al., 2023) uses illumination information to guide a Transformer (Vaswani et al., 2017). CIDNet (Yan et al., 2025) further revisits brightness restoration from the HSV color space.

These methods are generally trained under fixed supervision and therefore tend to produce results with a single enhancement strength. This limitation makes them less suitable for scenarios where flexible brightness control is required.

To address this issue, several works have investigated controllable low-light enhancement. ReCoRo (Xu et al., 2022) adopts GANs to learn enhancement from images with different brightness

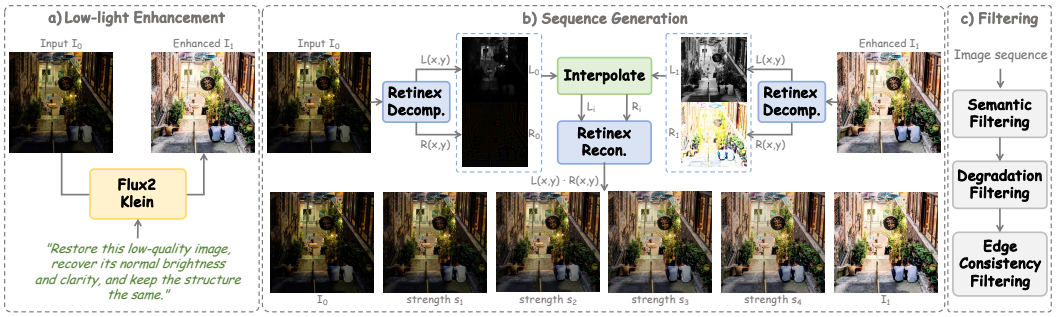

Figure 2: Main data construction pipeline of Light100K. FLUX.2-klein-9B is used to generate normal-light references from real low-light images. We then apply Retinex-inspired decomposition and selective interpolation to construct highly consistent and continuous pseudo-paired data from each pair (I_0, I_1).

levels. CLE Diffusion (Yin et al., 2023) employs a conditional diffusion model that uses brightness alpha blending target images as guidance, enabling controllable enhancement to some extent. Nevertheless, limited by model capacity and the relatively simple interpretation of the training data construction, CLE Diffusion often struggles to generalize to real-world continuous low-light enhancement and may produce noticeable artifacts.

2.2 Image Editing Methods and Continuous Control

Large-scale image editing models (Labs et al., 2025; Liu et al., 2025; Wu et al., 2025; Team et al., 2025; Huang et al., 2025; Gao et al., 2025; Seedream et al., 2025; Wang et al., 2025; Seedance et al., 2026) have shown strong potential for restoration by leveraging semantic priors learned from massive image–text pairs. However, their generative nature can introduce hallucinations, pixel shifts, and structural deformation, which are undesirable for restoration tasks requiring content consistency. Although high-quality data and consistency reward models (Jiang et al., 2026) can alleviate this issue, existing instruction-based editing methods still lack reliable continuous control.

Recent methods (Baumann et al., 2025; Gandikota et al., 2024; Parihar et al., 2025; Zarei et al., 2025; Sharma et al., 2024; Peng et al., 2026) achieve continuous editing through interpolatable text embeddings, modulation features, or low-rank adaptors, but are limited by scarce continuous supervision. Kontinuous Kontext (Parihar et al., 2025) synthesizes continuous samples via morphing (Cao et al., 2025), which is difficult to keep consistent for global restoration tasks. ConceptSlider (Gandikota et al., 2024) learns controllable LoRA directions, but its control can be unstable without intermediate supervision. To address these limitations, we use Retinex theory to construct continuous pseudo-paired supervision and train a controllable LoRA on FLUX.2-klein-9B. We further propose a Misalignment-Aware Weighted Flow Matching Loss to reduce pixel-level inconsistency during continuous enhancement.

3 Method

3.1 Light100K: Continuous Pseudo-Paired Data Construction

To address the limited availability of real paired training data, we construct paired data from real-world low-light images rather than relying solely on synthetic degradation generated from traditional single-degradation models.

Specifically, we collect high-quality images from open-source image websites, including Pexels and Pinterest, using low-light-related keywords. To build a high-quality real-world degradation dataset, we conduct low-light semantic and degradation filtering. After filtering, we obtain approximately 30K high-quality low-light images.

Given a low-light image I_0, we use a fixed enhancement prompt and the pretrained FLUX.2-klein-9B model to generate its enhanced counterpart I_1. To avoid supervision from structurally inconsistent pseudo pairs, we remove severely mismatched samples using an edge-consistency filtering strategy, leaving approximately 20K high-quality paired samples. For each retained pair (I_0, I_1), we further



Figure 3: Visual comparison of intermediate pseudo-GT construction at $s = 0.5$. Retinex-based interpolation yields more natural illumination transitions and better local contrast than direct alpha blending, making it better suited for continuous low-light enhancement.

construct a continuous pseudo-paired training group:

$$\mathcal{G} = \{I_0, I_{0.2}, I_{0.4}, I_{0.6}, I_{0.8}, I_1\},$$

where I_s denotes the pseudo ground-truth image at enhancement strength $s \in \{0.2, 0.4, 0.6, 0.8\}$.

Additional dataset statistics, including source platforms, scene composition, low-light severity, and filtering counts, are provided in Appendix A.1 (Figure 7).

A straightforward strategy is alpha blending (Yin et al., 2023), i.e., $I_s^{\text{alpha}} = (1-s)I_0 + sI_1$. However, direct RGB-space averaging mixes illumination, reflectance, color, and local contrast, making it suboptimal for continuous low-light enhancement where the target should mainly follow a gradual illumination transition.

To construct a more illumination-consistent trajectory, we propose a **Retinex-inspired interpolation** strategy as shown in Figure 2. The key idea is to use the Retinex (Land, 1977) image formation model $I = R \odot L$, where R represents reflectance-related scene content and L represents illumination. Under this model, continuous enhancement is primarily modeled as a transition in the illumination component rather than as a direct interpolation of the whole image appearance.

Specifically, we first convert I_0 and I_1 from sRGB to linear RGB, and use the same notation for simplicity. We compute their luminance maps by $Y = 0.2126R + 0.7152G + 0.0722B$, and estimate illumination maps using edge-preserving smoothing (bilateral filter): $L_0 = \text{Smooth}(Y_0)$ and $L_1 = \text{Smooth}(Y_1)$, since illumination is assumed to be spatially smooth. The reflectance maps are then estimated according to the Retinex model as $R_0 = I_0/L_0$ and $R_1 = I_1/L_1$.

We interpolate only the illumination maps in the log domain rather than I_0 and I_1 in RGB space:

$$L_s = \exp\left((1-s)\log(L_0) + s\log(L_1)\right). \tag{1}$$

This is equivalent to a multiplicative interpolation $L_s = L_0^{1-s} L_1^s$, which is more consistent with the Retinex assumption than additive image-space averaging. In parallel, we conservatively interpolate the reflectance as $R_s = (1-\beta_s)R_0 + \beta_s R_1$, where $\beta_s = 0.5s$. This design avoids relying only on R_0, which may contain amplified low-light noise, while also avoiding excessive dependence on R_1, which may inherit artifacts or subtle structural deviations from the diffusion-generated target. The intermediate pseudo-GT is finally reconstructed as:

$$I_s = \text{clip}(R_s \odot L_s, 0, 1). \tag{2}$$

The reconstructed image is then converted back to sRGB space for training. More details about the data contrsuction pipeline and the Light100K is provided in the Appendix A.

In Figure 3, the direct RGB-space averaging of Alpha blending flattens shadows and textures by weakening local contrast, while the nonlinear illumination transition of Retinex interpolation preserves local shading, scene depth, and contrast variations. Thus, our use of Retinex interpolation provides a more illumination-aware pseudo-GT trajectory for continuously controllable low-light enhancement.

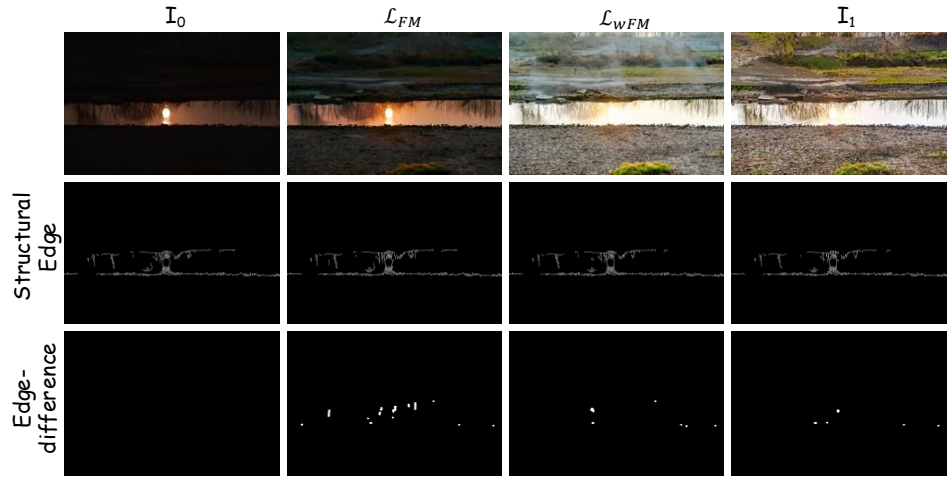

Figure 4: Visualization of edge misalignment and the effect of weighted flow matching. Compared with standard flow matching, the proposed L_{wFM} produces enhanced results with weaker edge-difference responses and better structural alignment to the input.

3.2 Misalignment-Aware Weighted Flow Matching

Although the filtered pseudo pairs $\{I_0, I_1\}$ are visually well aligned, they may still contain subtle pixel-level edge misalignment. Such misalignment is difficult to observe directly in RGB space, as the dominant differences between I_0 and I_1 mainly arise from brightness and color variations. After normalizing illumination, however, the remaining high-frequency residuals reveal local structural edge discrepancies. As shown in Figure 4, even a pair that satisfies our matching criterion can still exhibit non-negligible edge differences. When FLUX.2-klein-9B is fine-tuned with the standard flow matching loss (Lipman et al., 2022; Esser et al., 2024; Peebles & Xie, 2023), these misaligned edges may be inherited and amplified, leading to visible structural drift in the enhanced output I_1^{FM}. To address this issue, we introduce a misalignment-aware weighted flow matching loss that reduces the supervision strength in unreliable target-edge regions across the continuous pseudo-paired sequence generated from the same degraded image.

To visualize edge misalignment, we employ a structural edge-difference map that focuses on illumination-invariant features. Specifically, we first convert the images to the log-luminance domain and remove slow-varying brightness by subtracting a smoothed version (via a bilateral filter) to isolate the high-pass structural component $H(I)$. We then compute a high-frequency edge response defined as $E(I) = \|\nabla H(I)\|_1$, where ∇ denotes the gradient operator. Finally, the edge-difference map between any two images A and B is calculated as $I_{\text{edge-diff}}(A, B) = |E(A) - E(B)|$. This operation effectively suppresses low-frequency illumination and color discrepancies, ensuring the resulting response primarily reflects local structural misalignments rather than brightness variations.

As shown in Figure 4, the columns correspond to the input I_0, the output trained with $\mathcal{L}_{\mathrm{FM}}$ (I_1^{FM}), our output (I_1^{wFM}), and the pseudo target I_1. The second row shows the extracted structural edge maps, while the third row shows the edge-difference maps computed with respect to I_0. Compared with standard flow matching, our weighted loss produces fewer edge-difference responses, indicating better preservation of the input structure.

In standard flow matching, given a target image I_s at enhancement strength s, we encode it into the latent space as z_1, sample a noise latent z_0, and construct an intermediate latent $z_t = (1-t)z_0 + tz_1$, where $t \in [0, 1]$. The model predicts a velocity field $v_\theta(z_t, I_0, s)$, and the standard objective is:

$$\mathcal{L}_{\mathrm{FM}} = \|v_\theta(z_t, I_0, s) - v^*\|_2^2, \tag{3}$$

where $v^* = z_1 - z_0$. This objective treats all spatial regions of the pseudo target equally. Therefore, if I_s contains misaligned edges, the model is still encouraged to reproduce those unreliable structures.

We instead assign lower weights to unreliable target-edge regions. For each pseudo target I_s, we compute binary edge maps B_0 and B_s from I_0 and I_s, respectively. We then compute the distance transform D_0 to the nearest edge pixel in B_0. A target edge pixel is regarded as unreliable if it is far

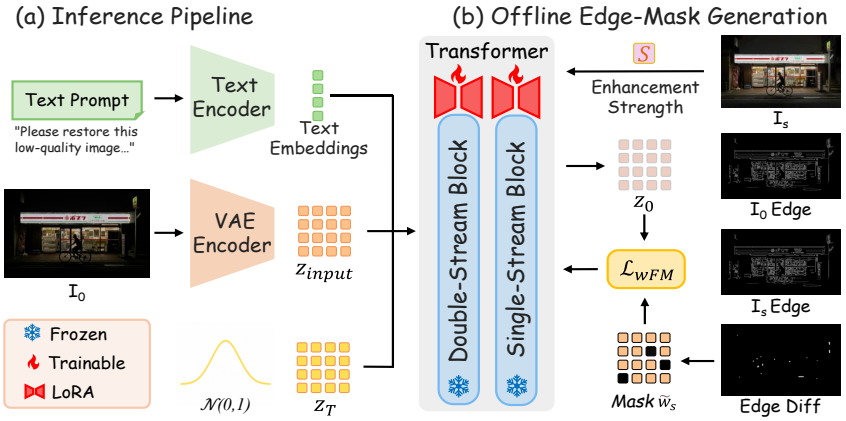

Figure 5: Overview of the proposed framework. (a) During training, the low-light input image and a fixed restoration prompt are encoded and fed into FLUX.2-klein, where LoRA is used for efficient fine-tuning. The enhancement strength s modulates both the LoRA scaling factor and the pseudo ground-truth selection. (b) The edge mask is generated offline from input and target edges, producing the weight map $\widetilde{W}_s$. At inference time, s can be set to any value in $[0, 1]$.

from any input edge:

$$M_s(p) = 1\left[B_s(p) = 1 \text{ and } D_0(p) > d\right], \tag{4}$$

where d is a distance threshold. We dilate M_s slightly to cover the neighborhood around the mismatched edge and obtain a soft weight map:

$$W_s(p) = \text{clip}\left(1 - \alpha M_s(p), w_{\min}, 1\right). \tag{5}$$

Details of the weight map generation and the hyperparameters d, α, and $w_{\min}$ are provided in Appendix B.

The image-space weight map W_s is resized to the latent resolution as $\widetilde{W}_s$, which is then applied over latent spatial locations u, to reweight the flow matching objective:

$$\mathcal{L}_{\text{wFM}} = \frac{\sum_u \widetilde{W}_s(u) \left\| v_\theta(z_t, I_0, s)(u) - v^*(u) \right\|_2^2}{\sum_u \widetilde{W}_s(u)}. \tag{6}$$

Here, $W_s(p)$ remains positive even in unreliable regions, so the model still receives weak appearance supervision but is no longer forced to exactly fit misaligned pseudo-target edges.

3.3 ControlLight

Given the continuous pseudo-paired dataset (Section 3.1) and the misalignment-aware loss (Eq. 6), we now describe how s is incorporated into the model. The Retinex formulation $I = R \odot L$ suggests that continuous enhancement is primarily a smooth transition along the illumination axis, approximately linear in some parameter subspace. This motivates using s directly as the LoRA scaling factor:

$$W' = W + s \cdot AB, \tag{7}$$

where W is frozen and A, B are learnable low-rank matrices.

This formulation resembles Concept Sliders (Gandikota et al., 2024), but the training regimes differ critically. Concept Sliders optimize a LoRA direction via text-guided score matching between opposing prompts, with the scaling factor s applied only at inference time. The linearity of control is assumed but never enforced. In contrast, our s enters the training loop: each $s \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ is paired with a pseudo ground truth I_s, and $\mathcal{L}_{\text{wFM}}$ is computed against that target. The LoRA direction is therefore calibrated against a physically grounded illumination trajectory with per-strength supervision, which is the key reason ControlLight achieves substantially better trajectory smoothness than Concept Sliders (Table 3).

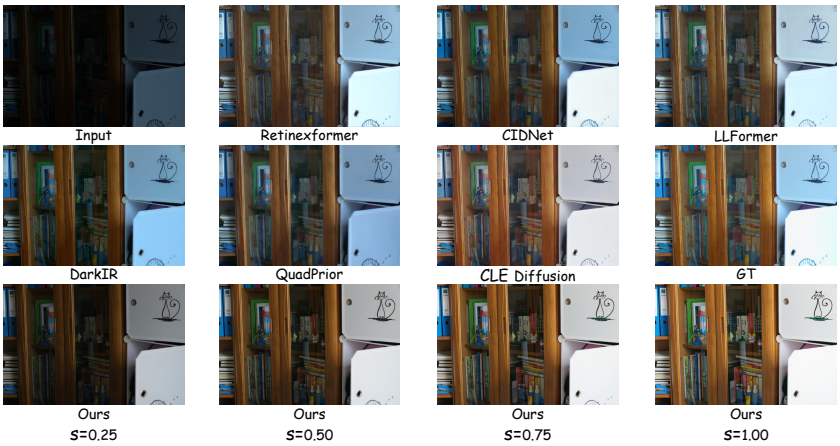

Figure 6: Visual comparison on LOL-v1 benchmark. Our zero-shot results may deviate from the ground truth in color appearance, but they provide natural visual quality with preserved structures and textures. The outputs at different enhancement strengths s show smooth and approximately linear low-light enhancement control.

During training, the input image and fixed text prompt are encoded by Flux2-VAE (Labs et al., 2025) and Qwen3-VL (Team, 2025), respectively. Since the prompt remains fixed, the Qwen3-VL text encoder can be offloaded during inference. The weight maps $\widetilde{W}_s$ are precomputed offline. We train at 1024×1024 resolution with a fixed learning rate of 1×10^{-4} and a global batch size of 16. The LoRA modules contain about 300M trainable parameters. Additional implementation details are provided in Appendix C.

4 Experiments

4.1 Quantitative Metrics and Evaluation Protocol

ControlLight is compared with two baseline groups: low-light enhancement methods and universal continuous image editing methods. For low-light enhancement, we evaluate on five benchmarks: LOL (Yang et al., 2021) and LSRW (Hai et al., 2023) with paired reference, as well as real-world DICM (Lee et al., 2013), LIME (Guo et al., 2016), and RealIR-Bench (Yang et al., 2026) with non-reference. As generative restoration models such as SUPIR (Yu et al., 2024) may synthesize perceptually plausible details that are penalized by reference-based metrics like PSNR and SSIM (Wang et al., 2004), we mainly report non-reference perceptual metrics, including CLIP-IQA (Wang et al., 2023a), MUSIQ (Ke et al., 2021), NIQE (Mittal et al., 2012), and MANIQA (Yang et al., 2022). To further evaluate Linear Control, we compare ControlLight with universal continuous image editing methods on real-world non-reference test sets, as they can potentially perform continuous low-light enhancement. We assess the smoothness and directionality of the enhancement trajectory using δ_{smooth} (Parihar et al., 2025) and CLIP-Dir (Patashnik et al., 2021), respectively.

4.2 Low-light Enhancement Evaluation

We compare ControlLight with representative low-light enhancement methods, including Retinexformer (Cai et al., 2023), HVI-CIDNet (Yan et al., 2025), LLFormer (Wang et al., 2023b), DarkIR (Feijoo et al., 2025), CLE Diffusion (Yin et al., 2023), and QuadPrior (Wang et al., 2024a). Paired evaluation is conducted on LOL-v1 (Yang et al., 2021) (15 test images) and LSRW (Hai et al., 2023) (50 test images, averaged over Huawei and Nikon subsets).

As a continuous enhancement model, ControlLight is evaluated at $s \in 0.25, 0.50, 0.75, 1.00$ and averaged across strengths. CLE Diffusion uses ground-truth-guided result selection on paired benchmarks; on unpaired benchmarks, we evaluate it using the same four-strength protocol for fair comparison.

Table 1: FLUX.2-klein-9B is evaluated without LoRA using the fixed restoration prompt; ControlLight scores are averaged over four strengths. Quantitative comparison on real-world and unpaired datasets: DICM (Lee et al., 2013), LIME (Guo et al., 2016), and RealIR-Bench (Yang et al., 2026).

Method	DICM (Lee et al., 2013)				LIME (Guo et al., 2016)				RealIR-Bench (Yang et al., 2026)			
	NIQE↓	CLIPIQA↑	MANIQA↑	MUSIQ↑	NIQE↓	CLIPIQA↑	MANIQA↑	MUSIQ↑	NIQE↓	CLIPIQA↑	MANIQA↑	MUSIQ↑
Retinexformer (Cai et al., 2023)	3.962	0.377	0.291	54.27	4.300	0.394	0.367	59.41	4.200	0.286	0.277	52.98
CIDNet (Yan et al., 2025)	3.657	0.501	0.384	57.90	4.182	0.439	0.399	60.72	4.129	0.377	0.353	62.41
LLFormer (Wang et al., 2023b)	3.943	0.435	0.274	55.03	4.392	0.382	0.297	57.39	3.866	0.236	0.250	49.14
DarkIR (Feijoo et al., 2025)	3.869	0.463	0.345	57.44	4.523	0.441	0.373	61.73	5.097	0.374	0.358	63.30
QuadPrior (Wang et al., 2024a)	4.797	0.488	0.315	58.21	5.310	0.396	0.292	58.92	4.659	0.305	0.270	51.86
CLE Diffusion (Yin et al., 2023)	4.368	0.390	0.218	47.38	5.317	0.433	0.279	57.76	3.887	0.423	0.347	60.10
FLUX.2-klein-9B ($s = 1$)	3.857	0.593	0.444	62.411	3.879	0.540	0.472	62.494	5.893	0.486	0.271	48.243
ControlLight (Ours)	**3.522**	**0.698**	**0.505**	**68.22**	**3.638**	**0.576**	**0.526**	**67.68**	**3.748**	**0.550**	**0.491**	**67.96**

Table 2: Quantitative comparison on paired enhancement benchmarks: LOL-v1 (Yang et al., 2021) and LSRW (Hai et al., 2023), using full-reference metrics. ControlLight is evaluated with fixed $s = 1$. The best and second-best results are highlighted with yellow and blue backgrounds, respectively.

Method	LOL-v1 (Yang et al., 2021)			LSRW (Hai et al., 2023)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Retinexformer (Cai et al., 2023)	25.397	0.843	0.255	17.652	0.506	**0.412**
CIDNet (Yan et al., 2025)	23.809	**0.857**	**0.188**	16.613	0.465	0.437
LLFormer (Wang et al., 2023b)	23.649	0.816	0.280	18.127	0.519	0.417
DarkIR (Feijoo et al., 2025)	17.086	0.745	0.350	17.697	0.520	0.421
QuadPrior (Wang et al., 2024a)	18.789	0.776	0.307	16.926	0.546	0.443
CLE Diffusion (Yin et al., 2023)	**25.421**	0.813	0.280	**20.746**	**0.550**	0.431
FLUX.2-klein-9B ($s = 1$)	11.411	0.349	0.647	11.039	0.292	0.638
ControlLight (Ours)	15.858	0.596	0.326	13.171	0.387	0.500

As shown in Table 1, ControlLight consistently outperforms all baselines on real-world benchmarks, demonstrating strong generalization to diverse low-light degradations. Figure 6 further shows more natural textures and colors, although perceptually plausible outputs may deviate from paired references in local details (the cat color in Figure 6). In contrast, task-specific methods without large-scale generative priors generalize less effectively to complex real-world degradations, as shown in Figure 9.

We further evaluate reference fidelity on LOL-v1 and LSRW in Table 2. Although ControlLight is not trained on these dataset-specific domains and therefore trails several specialized methods in PSNR and SSIM, it substantially improves all three reference metrics over the unadapted FLUX.2-klein-9B baseline. It also preserves high fidelity in challenging regions such as faces and text (please see Figure 12). More quantitative and qualitative results are provided in Appendix D.

4.3 Linear Control Evaluation

Following the evaluation protocol of KSlider (Parihar et al., 2025), we report δ_{smooth} to measure the smoothness of the continuous enhanment trajectory based on LPIPS feature distances. We also report CLIP-Dir to evaluate whether the enhancement trajectory consistently moves away from dark or under-

Table 3: Quantitative comparison of controllable editing performance across three datasets. We focus on the trajectory smoothness ($\delta_{\text{smooth}} \downarrow$) and semantic directional consistency (CLIP-Dir ↑). To ensure fairness, all the methods are evaluated using aligned four-point control strengths.

Method	RealIR-Bench		DICM		LIME	
	$\delta_{\text{smooth}} \downarrow$	CLIP-Dir ↑	$\delta_{\text{smooth}} \downarrow$	CLIP-Dir ↑	$\delta_{\text{smooth}} \downarrow$	CLIP-Dir ↑
ConceptSlider (Gandikota et al., 2024)	0.9237	-0.0530	0.8700	0.3872	0.8589	0.0256
AttributeControl (Baumann et al., 2025)	0.7262	0.3520	0.7928	0.3593	0.8176	0.3605
KSlider (Parihar et al., 2025)	**0.1956**	0.0901	0.3570	0.4488	**0.0485**	0.0434
SliderEdit (Zarei et al., 2025)	0.3840	-0.3125	0.4818	0.1768	0.3741	-0.1061
CLE Diffusion (Yin et al., 2023)	0.7503	-0.2624	0.7063	-0.2946	0.6643	0.1830
ControlLight (Ours)	0.2195	**0.9138**	**0.2382**	**0.9012**	0.1786	**0.9159**

exposed semantics. We compare with several universal continuous image editing methods, including ConceptSlider (Gandikota et al., 2024), AttributeControl (Baumann et al., 2025), KSlider (Parihar et al., 2025), SliderEdit (Zarei et al., 2025), and CLE Diffusion (Yin et al., 2023). For a fair comparison, all methods are evaluated at the same four control strengths, $s \in \{0.25, 0.50, 0.75, 1.00\}$, by mapping each method's control variable linearly to this range. As shown in Table 3, our method achieves the highest CLIP-Dir score, demonstrating that its enhancement trajectory is more semantically aligned with the increasing enhancement strength and exhibits stronger linear controllability. More Qualitative Results are provide in the Appendix D.

4.4 Ablation Study

We examine the effects of the training objective and intermediate supervision.

Misalignment-Aware Weighted Flow Matching Loss. We adopt LI-LPIPS (Yin et al., 2023) from CLE Diffusion, an edge-aware and color-normalized perceptual distance designed to provide a more stable measure of consistency across continuous enhancement outputs than the original LPIPS. As shown in Table 4, introducing $\mathcal{L}_{\mathrm{wFM}}$ consistently improves LI-LPIPS under different interpolation settings. The consistent improvement in LI-LPIPS therefore indicates that $\mathcal{L}_{\mathrm{wFM}}$ enhances structural consistency across different enhancement strengths. We further examine its effect on standard reference-fidelity metrics in Table 5. The improvements in LPIPS, PSNR, and SSIM further support the effectiveness of $\mathcal{L}_{\mathrm{wFM}}$ in preserving image fidelity.

Data Interpolation Methods. Table 4 further shows that our interpolation strategy improves CLIP-Dir, indicating better directional consistency when continuously varying the enhancement strength. This suggests that the proposed interpolation provides a more coherent supervision signal for learning the transition from low-light to normally exposed images. We further analyze the quality of the interpolation for the effectiveness of the proposed interpolation strategy in Appendix D.2.

Table 4: Factorial ablation on LSRW across $s \in \{0.25, 0.50, 0.75, 1.00\}$. LI-LPIPS is averaged across the four strengths.

Data Interpolation Methods	Objective	CLIP-Dir↑	LI-LPIPS↓
Alpha blending	$\mathcal{L}_{\mathrm{FM}}$	0.8854	0.1616
Alpha blending	$\mathcal{L}_{\mathrm{wFM}}$	0.8611	**0.1572**
Retinex interpolation	$\mathcal{L}_{\mathrm{FM}}$	0.9244	0.1618
Retinex interpolation	$\mathcal{L}_{\mathrm{wFM}}$	**0.9260**	**0.1572**

Table 5: Loss ablation on LSRW-Huawei with fixed Retinex supervision. All metrics are averaged over four evaluation strengths.

Objective	PSNR↑	SSIM↑	LPIPS↓
$\mathcal{L}_{\mathrm{FM}}$	13.290	0.390	0.540
$\mathcal{L}_{\mathrm{wFM}}$	**13.468**	**0.402**	**0.520**

5 Conclusions

We introduced a Retinex-inspired interpolation strategy and a high-quality dataset, Light100K, to facilitate real-world low-light enhancement. To tackle hallucinations and inconsistencies in outputs, we also developed the Misalignment-Aware Weighted Flow Matching Loss with Offline Edge-Mask Generation, which suppresses the effects of edge shifts during training. By fine-tuning the FLUX.2-klein-9B model with LoRA using our proposed $\mathcal{L}_{\mathrm{wFM}}$, ControlLight establishes new state-of-the-art performance. It outperforms existing enhancement and continuous editing methods, delivering superior consistency, strength controllability, and generalization in real-world scenarios.

AI USE STATEMENT

In this work, we used generative AI tools to design or provide feedback on research methodology or experiments and to assist with translation. We have not used generative AI tools to generate synthetic datasets, develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, or assist in the writing of proofs, and the remaining tasks with required disclosure are not applicable to this work. Additionally, we used generative AI tools to create or edit software code and to edit the research paper to improve readability. We have reviewed all AI-assisted work. All AI-assisted content, including experimental suggestions, translations, code, and manuscript edits, was manually reviewed by the authors to ensure correctness and consistency with the intended methodology and technical meaning. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Stefan Andreas Baumann, Felix Krause, Michael Neumayr, Nick Stracke, Melvin Sevi, Vincent Tao Hu, and Björn Ommer. Continuous, subject-specific attribute control in t2i models by identifying semantic directions. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 13231–13241, 2025.

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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12504–12513, 2023.

Yukang Cao, Chenyang Si, Jinghao Wang, and Ziwei Liu. Freemorph: Tuning-free generalized image morphing with diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 18111–18120, 2025.

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

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*, 2024.

Daniel Feijoo, Juan C. Benito, Alvaro Garcia, and Marcos V. Conde. Darkir: Robust low-light image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 10879–10889, June 2025.

Rohit Gandikota, Joanna Materzyńska, Tingrui Zhou, Antonio Torralba, and David Bau. Concept sliders: Lora adaptors for precise control in diffusion models. In *European Conference on Computer Vision*, pp. 172–188. Springer, 2024.

Yu Gao, Lixue Gong, Qiushan Guo, Xiaoxia Hou, Zhichao Lai, Fanshi Li, Liang Li, Xiaochen Lian, Chao Liao, Liyang Liu, et al. Seedream 3.0 technical report. *arXiv preprint arXiv:2504.11346*, 2025.

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, 2016.

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

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.

Yi Huang, Jiancheng Huang, Yifan Liu, Mingfu Yan, Jiaxi Lv, Jianzhuang Liu, Wei Xiong, He Zhang, Liangliang Cao, and Shifeng Chen. Diffusion model-based image editing: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Bernd Jähne. *Digital image processing*. Springer, 2005.

Kui Jiang, Zhongyuan Wang, Zheng Wang, Peng Yi, Xiao Wang, Yansheng Qiu, Chen Chen, and Chia-Wen Lin. Degrade is upgrade: Learning degradation for low-light image enhancement. In *arXiv*, March 2021a.

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, 2021b.

Zhangqi Jiang, Zheng Sun, Xianfang Zeng, Yufeng Yang, Xuanyang Zhang, Yongliang Wu, Wei Cheng, Gang Yu, Xu Yang, and Bihan Wen. Geditbench v2: A human-aligned benchmark for general image editing. *arXiv preprint arXiv:2603.28547*, 2026.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5148–5157, 2021.

Black Forest Labs, Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, et al. Flux. 1 kontext: Flow matching for in-context image generation and editing in latent space. *arXiv preprint arXiv:2506.15742*, 2025.

Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

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

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Shiyu Liu, Yucheng Han, Peng Xing, Fukun Yin, Rui Wang, Wei Cheng, Jiaqi Liao, Yingming Wang, Honghao Fu, Chunrui Han, et al. Step1x-edit: A practical framework for general image editing. *arXiv preprint arXiv:2504.17761*, 2025.

Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

Rishubh Parihar, Or Patashnik, Daniil Ostashev, R Venkatesh Babu, Daniel Cohen-Or, and Kuan-Chieh Wang. Kontinuous kontext: Continuous strength control for instruction-based image editing. *arXiv preprint arXiv:2510.08532*, 2025.

Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2085–2094, October 2021.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

Yuqi Peng, Lingtao Zheng, Yufeng Yang, Yi Huang, Mingfu Yan, Jianzhuang Liu, and Shifeng Chen. Tara: Token-aware lora for composable personalization in diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 8385–8393, 2026.

Stephen M Pizer. Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In *Proceedings of the first conference on visualization in biomedical computing, Atlanta, Georgia*, volume 337, pp. 2, 1990.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pp. 8748–8763. PmLR, 2021.

Sudarshan Rajagopalan, Nithin Gopalakrishnan Nair, Jay N Paranjape, and Vishal M Patel. Gendeg: Diffusion-based degradation synthesis for generalizable all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28144–28154, 2025.

Team Seedance, De Chen, Liyang Chen, Xin Chen, Ying Chen, Zhuo Chen, Zhuowei Chen, Feng Cheng, Tianheng Cheng, Yufeng Cheng, et al. Seedance 2.0: Advancing video generation for world complexity. *arXiv preprint arXiv:2604.14148*, 2026.

Team Seedream, Yunpeng Chen, Yu Gao, Lixue Gong, Meng Guo, Qiushan Guo, Zhiyao Guo, Xiaoxia Hou, Weilin Huang, Yixuan Huang, et al. Seedream 4.0: Toward next-generation multimodal image generation. *arXiv preprint arXiv:2509.20427*, 2025.

Prafull Sharma, Varun Jampani, Yuanzhen Li, Xuhui Jia, Dmitry Lagun, Fredo Durand, Bill Freeman, and Mark Matthews. Alchemist: Parametric control of material properties with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24130–24141, 2024.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

Meituan LongCat Team, Hanghang Ma, Haoxian Tan, Jiale Huang, Junqiang Wu, Jun-Yan He, Lishuai Gao, Songlin Xiao, Xiaoming Wei, Xiaoqi Ma, Xunliang Cai, Yayong Guan, and Jie Hu. Longcat-image technical report. *arXiv preprint arXiv:2512.07584*, 2025.

Qwen Team. Qwen3 technical report, 2025. URL `https://arxiv.org/abs/2505.09388`.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 2555–2563, 2023a.

Peng Wang, Yichun Shi, Xiaochen Lian, Zhonghua Zhai, Xin Xia, Xuefeng Xiao, Weilin Huang, and Jianchao Yang. Seededit 3.0: Fast and high-quality generative image editing. *arXiv preprint arXiv:2506.05083*, 2025.

Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 2654–2662, 2023b.

Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors, 2024a.

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

Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 2604–2612, 2022.

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

Jiangwei Weng, Zhiqiang Yan, Ying Tai, Jianjun Qian, Jian Yang, and Jun Li. Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. *Advances in neural information processing systems*, 37:27440–27462, 2024.

Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhengyi Wang, An Yang, Bowen Yu, Chen Cheng, Dayiheng Liu, Deqing Li, Hang Zhang, Hao Meng, Hu Wei, Jingyuan Ni, Kai Chen, Kuan Cao, Liang Peng, Lin Qu, Minggang Wu, Peng Wang, Shuting Yu, Tingkun Wen, Wensen Feng, Xiaoxiao Xu, Yi Wang, Yichang Zhang, Yongqiang Zhu, Yujia Wu, Yuxuan Cai, and Zenan Liu. Qwen-image technical report, 2025. URL `https://arxiv.org/abs/2508.02324`.

Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Chunyi Li, Wenxiu Sun, Qiong Yan, Guangtao Zhai, and Weisi Lin. Q-bench: A benchmark for general-purpose foundation models on low-level vision. In *ICLR*, 2024.

Dejia Xu, Hayk Poghosyan, Shant Navasardyan, Yifan Jiang, Humphrey Shi, and Zhangyang Wang. Recoro: Re gion-co ntrollable ro bust light enhancement with user-specified imprecise masks. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 1376–1386, 2022.

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 *Proceedings of the computer vision and pattern recognition conference*, pp. 5678–5687, 2025.

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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1191–1200, 2022.

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

Yufeng Yang, Xianfang Zeng, Zhangqi Jiang, Fukun Yin, Jianzhuang Liu, Wei Cheng, Shiyu Liu, Yuqi Peng, Gang YU, Shifeng Chen, et al. Realrestorer: Towards generalizable real-world image restoration with large-scale image editing models. *arXiv preprint arXiv:2603.25502*, 2026.

Yuyang Yin, Dejia Xu, Chuangchuang Tan, Ping Liu, Yao Zhao, and Yunchao Wei. Cle diffusion: Controllable light enhancement diffusion model. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 8145–8156, 2023.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25669–25680, 2024.

Arman Zarei, Samyadeep Basu, Mobina Pournemat, Sayan Nag, Ryan Rossi, and Soheil Feizi. Slideredit: Continuous image editing with fine-grained instruction control. *arXiv preprint arXiv:2511.09715*, 2025.

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

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

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232, 2017.

Appendix

A Continuous Pseudo-Paired Data Construction Details

During the construction of Light100K, we first collect high-resolution low-light images from open-source image websites, including Pexels and Pinterest, using low-light-related keywords. We then use the CLIP text encoder (Radford et al., 2021) to compute the cosine similarity between each image and darkness-related prompts, such as "a dark photo", "underexposed", and "low illumination", in order to filter images with relevant low-light semantics. Next, we employ Qwen3-VL-8B-Instruct (Team, 2025) to assess the degradation level (Wu et al., 2024), ensuring that the retained images contain sufficient degradation cues for model learning (Jiang et al., 2021a; Rajagopalan et al., 2025). After semantic and degradation filtering, we obtain 27,529 high-quality low-light images, all with resolutions higher than 1024×1024.

We then use FLUX.2-klein-9B to generate restored normal-light references for the collected low-light images. To ensure pairwise structural consistency, we apply Sobel edge detection (Jähne, 2005) to the low-light and restored images and filter out pairs with obvious edge shifts or structural misalignment. This process yields 17,809 high-consistency low-/normal-light image pairs.

Finally, we apply Retinex-inspired interpolation with enhancement strengths $s \in \{0.2, 0.4, 0.6, 0.8\}$ to construct intermediate pseudo targets. The resulting Light100K is a high-quality, real-world, continuous pseudo-paired dataset for controllable low-light enhancement.

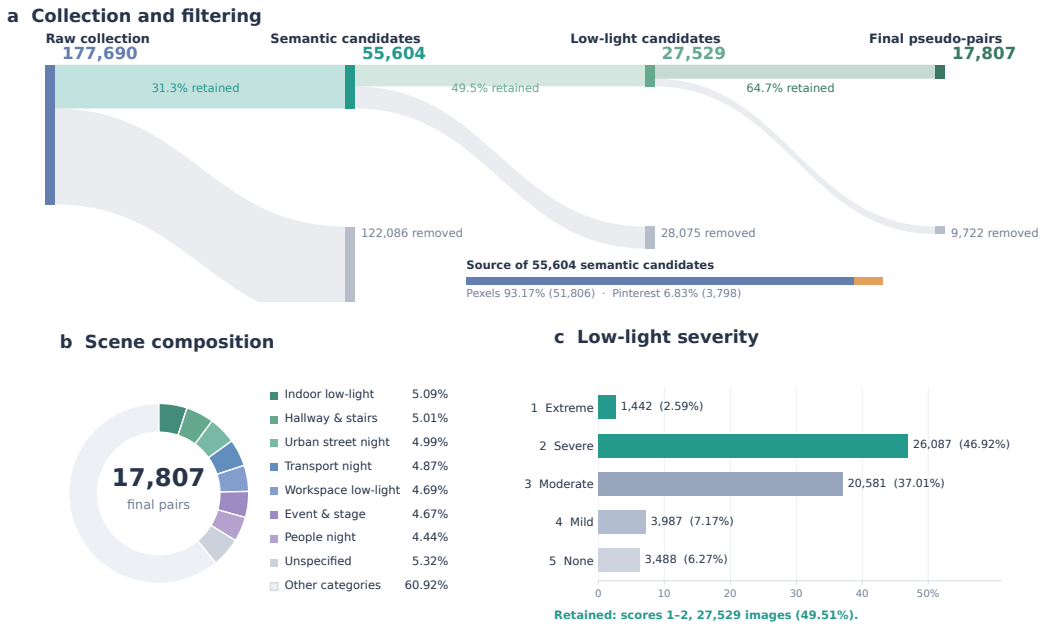

Figure 7: Light100K composition and filtering statistics. (a) Image counts through successive filtering stages and source-platform proportions among the 55,604 semantic candidates. (b) Scene proportions in the final 17,807-pair set; other categories aggregate the remaining share computed from rounded percentages. (c) Low-light severity in the semantic candidate pool, with scores 1–2 retained. The reported severity bins account for 55,585 candidates.

A.1 Dataset Composition and Filtering Statistics

Figure 7 summarizes the collection and filtering process and the composition of Light100K. Starting from 177,690 collected images, semantic filtering retains 55,604 candidates, of which 27,529 satisfy

the low-light criterion (severity score < 3); edge-consistency filtering then yields 17,807 pseudo pairs. The semantic candidate pool contains 51,806 images from Pexels and 3,798 from Pinterest. The final set covers indoor, urban street, transport, workspace, event, and people scenes. These statistics characterize scene coverage and degradation severit.

B Misalignment Analysis and Offline Edge-Mask Generation

During the construction of Light100K, subtle visual misalignment may still remain between low-light inputs and their paired enhanced images, even after edge-consistency filtering. Although such misalignment is below the filtering threshold and is often visually negligible, the generative nature of the base model makes it problematic during training. Under the standard flow matching loss,

$$\mathcal{L}_{\mathrm{FM}} = \|v_\theta(z_t, I_0, s) - v^*\|_2^2 ,$$

The model is encouraged to fit all target regions equally, which may introduce additional randomness when learning from slightly misaligned pseudo targets and lead to inconsistent outputs.

Our key insight is to preserve the structural edges of the input image while learning the desired illumination enhancement. To this end, we compute illumination-normalized log-luminance representations instead of directly comparing RGB values, so that images with different brightness levels can still share similar structural responses. We then apply a gradient operator to extract the main structural edges.

For each pair in Light100K, we compute a structural edge-difference map and use it to generate a spatial mask that guides flow matching with adaptive weights:

$$W_s(p) = \mathrm{clip}\left(1 - \alpha M_s(p), w_{\min}, 1\right).$$

The resulting weight map is resized to the latent resolution as $\widetilde{W}_s$ and used in the weighted flow matching objective:

$$\mathcal{L}_{\mathrm{wFM}} = \frac{\sum_u \widetilde{W}_s(u)\, \|v_\theta(z_t, I_0, s)(u) - v^*(u)\|_2^2}{\sum_u \widetilde{W}_s(u)}.$$

In practice, we set $d = 3$ pixels, $\alpha = 0.8$, and $w_{\min} = 0.2$. To improve training efficiency, all weight maps are generated offline and cached before training.

Table 6: Training hyperparameters for ControlLight fine-tuning.

Hyperparameters	
LoRA Setting	Rank=64, Alpha=64
Trainable Parameters	317M
Learning Rate	1×10^{-4}
Optimizer	AdamW 8-bit
Precision	BFloat16 (BF16)
Scheduler	Flow Matching
Global Batch Size	16
Training Steps	3,000
Resolution	1024×1024

C Implementation Details

During training, we fine-tune only the DiT blocks with LoRA, while freezing the VAE and text encoders. LoRA layers are applied to both the single-stream and double-stream DiT blocks (Peebles & Xie, 2023) with a rank of 64. The bucket resolution is fixed at 1024×1024, and the global batch size is set to 16. Detailed hyperparameters are provided in Table 6. All experiments are conducted on 4 NVIDIA A6000 GPUS.

D More Qualitative Results and Ablation Study Details

D.1 More Qualitative Results

Additional qualitative comparisons with low-light enhancement methods and general continuous image editing methods are presented in Figure 8 and Figure 10.

D.2 Interpolated Target Quality Analysis

We report NIQE and MUSIQ to evaluate the image-quality trajectories of different interpolation methods across four intermediate enhancement strengths using 200 randomly sampled images from Light100K. Since the low-light image I_0 is expected to have lower perceptual quality than the normal-light image I_1, a desirable interpolation method should produce a smooth and monotonic quality transition between them. Table 7 characterizes the quality trajectories of the constructed targets. This is an analysis of supervision images, not an ablation of trained models; no-reference scores alone do not establish physical realism or downstream restoration gains.

Table 7: Image-quality scores of interpolated Light100K targets. Metric rankings do not establish which supervision is preferable for training.

Metric	Strategy	I_0	$I_{0.2}$	$I_{0.4}$	$I_{0.6}$	$I_{0.8}$	I_1
NIQE ↓	Alpha Blending	**4.588**	**3.931**	**3.561**	3.356	3.461	**3.695**
	Ours	**4.588**	4.171	3.649	**3.315**	**3.419**	**3.695**
MUSIQ ↑	Alpha Blending	**55.936**	**62.620**	**66.289**	**70.047**	**68.469**	**70.019**
	Ours	**55.936**	58.780	60.889	67.626	67.716	**70.019**



Figure 8: Qualitative comparison with state-of-the-art low-light enhancement methods on the paired benchmarks LOL-v1 and LSRW.

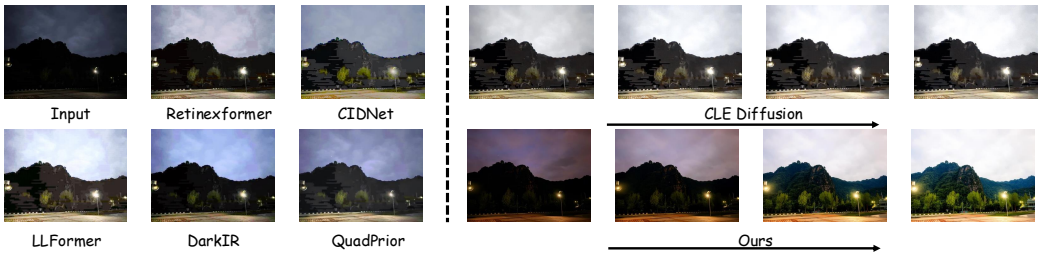

Figure 9: Visual comparison on RealIR-Bench. The arrows indicate increasing enhancement strength from left to right. Compared with traditional methods, our method achieves more natural restoration and better preserves scene structure. The arrows indicate increasing enhancement strength, showing that our model provides continuous and controllable low-light enhancement.

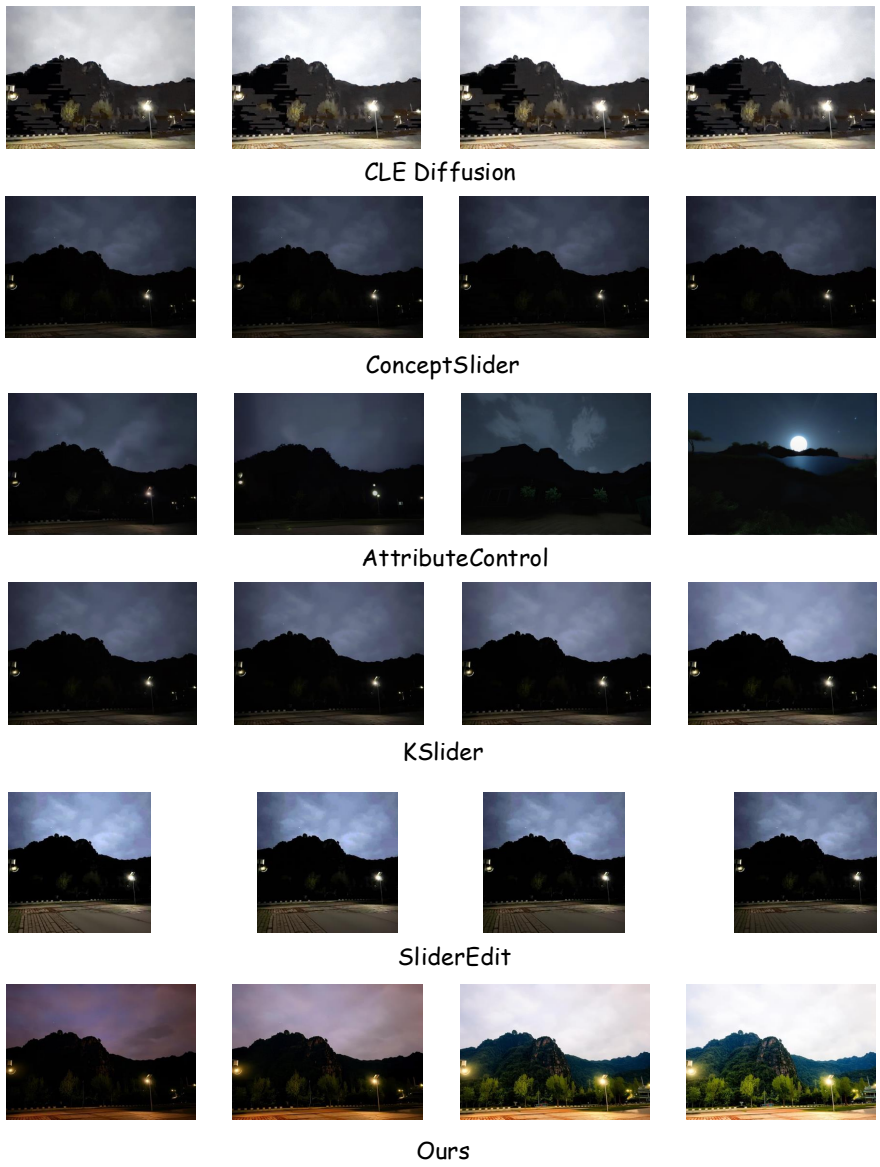

Figure 10: Qualitative comparison with universal continuous image editing methods on RealIR-Bench for low-light enhancement. All methods are evaluated under the same four-point control strengths.

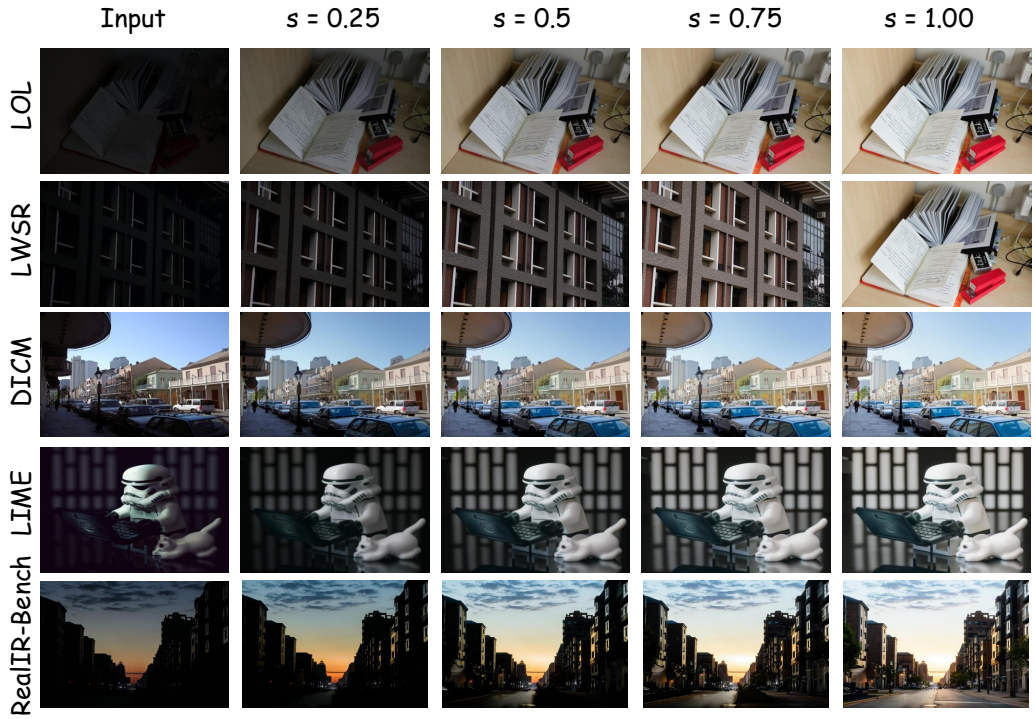

Figure 11: Qualitative results of ControlLight on the reported benchmarks under four continuous enhancement strengths.

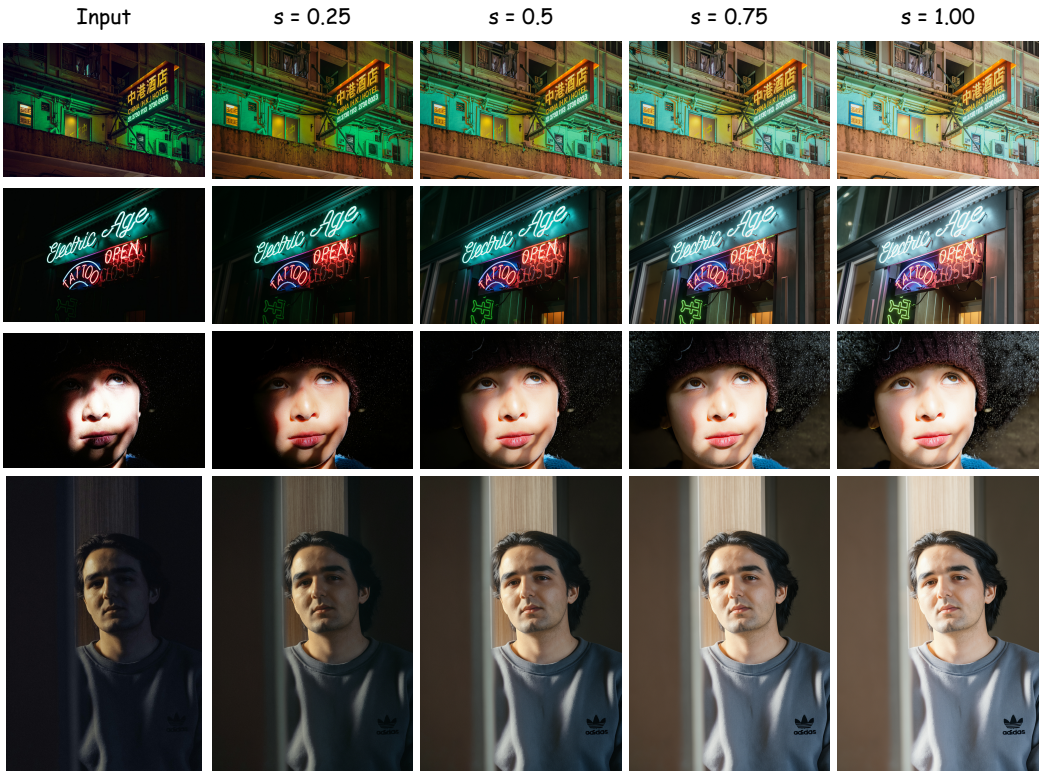

Figure 12: Qualitative results showing that ControlLight preserves high consistency and fidelity of faces and text in real-world low-light enhancement.