ExpoCM: Exposure-Aware One-Step Generative Single-Image HDR Reconstruction

Aoyu Liu[1,*], Zhen Liu[1,*], Ziyi Wang[1], Dian Chen[1], Bing Zeng[1], Shuaicheng Liu[1,‡]
[1]University of Electronic Science and Technology of China
{aoyuliu01@std., liuzhen03@std., eezeng@, liushuaicheng@}uestc.edu.cn

Abstract

Single-image HDR reconstruction aims to recover high dynamic range radiance from a single low dynamic range (LDR) input, but remains highly ill-posed due to detail saturation in over-exposed regions and noise amplification in under-exposed areas. While recent diffusion-based approaches offer powerful generative priors, they often overlook the exposure-dependent nature of the degradation and incur substantial computational costs from iterative sampling. To address these challenges, we propose ExpoCM, a novel one-step generative HDR reconstruction framework that reformulates HDR reconstruction as a Probability Flow ODE (PF-ODE) and constructs exposure-aware consistency trajectories via exposure-dependent perturbations. Specifically, a soft exposure mask is first constructed to separate the LDR image into over-, under-, and well-exposed regions. Based on this partition, region-conditioned consistency trajectories are designed to hallucinate saturated details, suppress noise in dark regions, and preserve reliable structures within a single, distillation-free inference step. To further enhance perceptual quality, we introduce an Exposure-guided Luminance-Chromaticity Loss in the CIE $L^*a^*b^*$ space, which assigns exposure-aware weights to luminance and chromaticity components, effectively mitigating brightness bias and color drift. Extensive experiments on the HDR-REAL, HDR-EYE, and AIM2025 benchmarks demonstrate that ExpoCM achieves state-of-the-art fidelity and perceptual accuracy, while enabling over 400× and 20× faster inference compared to DDPM (1000 steps) and DDIM (50 steps), respectively. Code is available at https://github.com/AoyuLiu01/ExpoCM.

1. Introduction

High dynamic range (HDR) imaging enhances the visual experience of digital content by capturing real-world scene appearances with a wide range of luminance, contrast, and fine details. However, consumer-grade cameras are limited to a narrow dynamic range due to inherent sensor constraints. Substantial research efforts have therefore been devoted to reconstructing HDR images from low dynamic range (LDR) inputs. The most common approach involves fusing multiple LDR images of the same scene at different exposure times [17, 25, 26, 35, 52, 54]. While effective in static scenes, these methods are susceptible to misalignment and ghosting artifacts in dynamic scenarios due to camera motion or moving objects. To address these issues, another line of research seeks to generate HDR content from a single-exposure image.

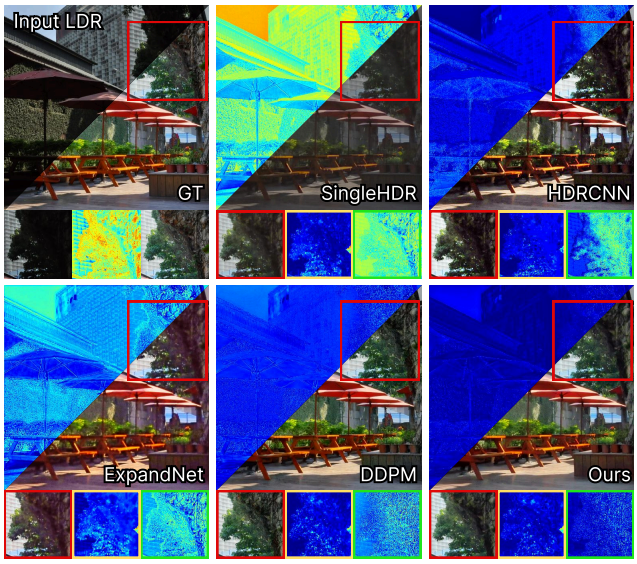

Figure 1. Visual comparisons with previous state-of-the-art methods. For each method, we show the reconstructed result (bottom-right) and the corresponding reconstruction error map (top-left). The highlighted patch (red box) and its associated chrominance error map (yellow box) and luminance error map (green box) are placed below (darker regions indicate smaller errors). The proposed ExpoCM yields results with higher-fidelity global luminance and color information.

*Equal contribution
‡Corresponding author.

Traditional single-image HDR reconstruction methods incorporate hand-crafted priors, such as illumination estimation [3, 4] or camera response modeling [2, 15], to expand the dynamic range of LDR inputs. In contrast, learning-based solutions [6, 9, 24, 34] primarily reconstruct HDR images in a regression manner using convolutional neural networks (CNNs). However, learning a direct LDR-to-HDR mapping via simple regression is fundamentally ill-posed, as the input LDR image often suffers from severe noise in under-exposed regions and detail saturation in over-exposed areas, leading to suboptimal reconstruction quality.

Recently, generative models, particularly diffusion families, have demonstrated remarkable capability in modeling complex data distributions and achieved impressive success across various low-level vision tasks. However, directly applying diffusion models to single-image HDR reconstruction poses unique challenges. First, the degradation is spatially heterogeneous, as over- and under-exposed areas exhibit distinct visual characteristics. A unified diffusion process struggles to simultaneously recover missing details in saturated regions and suppress noise in dark areas. Second, diffusion models are computationally demanding, typically requiring hundreds of iterative sampling steps to produce high-quality results, leading to high computational costs and limited potential for practical HDR applications.

To address the aforementioned challenges, we propose ExpoCM, an exposure-aware one-step generative framework built upon Consistency Models (CMs), for single-image HDR reconstruction. The core idea is to construct an exposure-aware consistency trajectory that explicitly tailors the generative process to the heterogeneous degradation characteristics within the input LDR image. Specifically, we first introduce an exposure mask generation mechanism to partition the input into well-exposed, under-exposed, and over-exposed regions, each exhibiting distinct information loss patterns. Inspired by recent advances in consistency training, we then formulate HDR reconstruction as a trajectory on the Probability Flow Ordinary Differential Equation (PF-ODE) and further develop an Exposure-Aware Consistency Trajectory, where region-specific perturbations and guidance are injected to tailor the PF-ODE flow according to exposure conditions. This exposure-aware consistency formulation enables high-quality one-step HDR generation without requiring distillation from pre-trained multi-step diffusion models. Moreover, to mitigate the inherent luminance and chromaticity bias of generative models [37], we introduce an Exposure-guided Luminance-Chromaticity (ELC) loss in the perceptually uniform CIE L*a*b* space. By assigning exposure-dependent weights to luminance and chromaticity components, the ELC loss adaptively enforces brightness consistency in under-exposed regions and suppresses color drifting in over-exposed areas, leading to more faithful and perceptually accurate HDR reconstruction. To summarize, the main contributions are as follows:

- We propose ExpoCM, a novel one-step generative framework for single-image HDR reconstruction that achieves high-fidelity results within a single inference step, eliminating the need for iterative sampling.
- We design an Exposure-Aware Consistency Trajectory that tailors the generative process of the Probability Flow ODE (PF-ODE) to the spatially heterogeneous degradations in LDR inputs, which is trained from scratch to avoid the costly distillation process.
- We develop an Exposure-guided Luminance-Chromaticity (ELC) loss defined in the perceptually uniform CIE $L^*a^*b^*$ space. This loss adaptively assigns weights to luminance and chromaticity components based on exposure conditions, effectively mitigating brightness imbalances and color distortion for more perceptually faithful reconstructions.

2. Related Work

2.1. HDR Reconstruction

High dynamic range (HDR) image reconstruction can be broadly categorized into multi-image and single-image paradigms according to the number of input LDR images.

Multi-image HDR reconstruction Early HDR reconstruction methods primarily fuse multiple LDR captures of the same scene, either with bracketed exposures [31, 35] or burst sequences [10]. While this approach produces high-quality results in static scenes, it suffers from misalignment and ghosting artifacts when camera motion or dynamic objects are present. Consequently, a large body of research [13, 20, 39, 44] has been devoted to addressing this challenge, commonly referred to as HDR deghosting. Seminal deep learning pipelines [17, 52] established the "alignment-fusion" paradigm, where input LDR images are first aligned (e.g., via optical flow [30] or homography [23] algorithms) and then fused. Recent efforts extend this by integrating implicit alignment modules [25, 54] into end-to-end architectures, or by leveraging advanced designs such as hybrid CNN-ViT networks [5, 26, 48] and novel learning strategies [38, 40]. Despite these advances, multi-image methods still struggle under severe camera or foreground movements, making them primarily suited for static scenes.

Single-image HDR reconstruction To overcome the limitations of multi-exposure methods, several studies have explored generating HDR images from a single LDR input. This task is inherently more challenging due to the limited dynamic range and the severe loss of information in over- or under-exposed regions. Early methods relied on hand-crafted priors, such as illumination estimation [3, 4] or camera response modeling [2, 15]. With the advent of deep learning, CNN-based approaches [6, 9, 21, 24, 34, 43, 57] have significantly advanced the field by learning end-to-

end LDR-to-HDR mappings. Representative methods like HDRCNN [9] and ExpandNet [34] utilize encoder-decoder architectures, while others like HDRUNet [6] and SingleHDR [24] incorporate imaging priors or multi-branch structures to better recover missing details. Some works [21] first synthesize pseudo multi-exposures from the single input before fusion. Despite their success, these regression-based approaches remain fundamentally limited in addressing the highly ill-posed problem, particularly in recovering details from noise and hallucinating content in saturated areas.

2.2. Diffusion Models in Low-level Vision

The remarkable generative capability of diffusion models (DMs) [12, 45] has led to their successful application in various low-level vision tasks, such as image restoration [7, 14, 16, 18, 22, 28, 41], inpainting [29, 53], and editing [19, 56, 58]. However, applying standard DMs to single-image HDR reconstruction presents two major challenges. First, they are computationally expensive, requiring hundreds of iterative sampling steps. Second, the standard diffusion process is spatially-agnostic, treating all pixels uniformly, which is suboptimal for the spatially heterogeneous degradation found in HDR inputs. To address the first challenge, Consistency Models (CMs) [47] have recently been proposed to learn a direct mapping from any point on the PF-ODE trajectory to the clean image, enabling high-quality, one-step generation. Nevertheless, these CMs typically inherit the second challenge from their diffusion counterparts. Furthermore, they often introduce a new requirement of being trained via costly distillation from a pre-trained DM. In this work, we bridge these gaps by designing a distillation-free and exposure-aware consistency framework specifically for the HDR reconstruction task.

3. Method

3.1. Preliminaries

Given a single LDR image $\mathbf{I}_L$, our objective is to reconstruct the corresponding HDR image $\mathbf{I}_H$. Unlike previous regression-based approaches, we formulate this task as a conditional generation task built upon Consistency Models (CMs) [47] and propose an exposure-aware one-step generative framework, termed ExpoCM, to achieve high-quality HDR reconstruction.

Probability Flow ODE Let $\mathbf{y}_0$ be the observed LDR image and $\mathbf{x}_0$ be the target HDR image to be reconstructed, diffusion models synthesize data by reversing a forward noising process that perturbs the target HDR image $\mathbf{x}_0$ into a noisy latent state $\mathbf{x}_t$ at time $t \in [0, T]$. This process can be described by a Stochastic Differential Equation (SDE) [46]:

$$d\mathbf{x}_t = f(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}_t, \tag{1}$$

where $f(\cdot, t)$ and $g(t)$ denote the drift and diffusion coefficients, and $\mathbf{w}_t$ is a standard Wiener process. Correspondingly, there exists a deterministic ordinary differential equation, known as the Probability Flow ODE (PF-ODE), sharing the same marginal probability densities as the SDE:

$$d\mathbf{x}_t = \left[f(\mathbf{x}_t, t) - \frac{1}{2}g(t)^2\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)\right]dt. \tag{2}$$

Solving this ODE from $t{=}T$ to $t{=}0$ recovers $\mathbf{x}_0$, but numerical integration requires tens or even hundreds of steps.

Conditional Consistency Trajectory To overcome this limitation, CMs learn a direct mapping from any point $(\mathbf{x}_t, t)$ on the PF-ODE trajectory to its origin $\mathbf{x}_0$. For conditional tasks like HDR reconstruction, the generative trajectory must be guided by the LDR input $\mathbf{y}_0$. A standard approach [47] defines the intermediate state $\mathbf{x}_t$ as:

$$\mathbf{x}_t = (1 - \alpha(t))\mathbf{x}_0 + \alpha(t)\mathbf{y}_0 + \sigma(t)\boldsymbol{\epsilon}, \tag{3}$$

where $\alpha(t)$ and $\sigma(t)$ are time-dependent coefficients and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$.

Consistency Training (CT) As illustrated in Fig. 2 (a), the CM network f_θ is trained to predict the trajectory's origin $\mathbf{x}_0$, conditioned on the LDR input $\mathbf{y}_0$. The consistency training objective [47] adapted for this conditional task is:

$$\mathcal{L}_{\text{CT}}(\theta, \theta^-) = \mathbb{E}_{\mathbf{x}_0, \mathbf{y}_0, t, t', \boldsymbol{\epsilon}}\left[\left|\left|f_\theta(\mathbf{x}_t, t, \mathbf{y}_0) - f_{\theta^-}(\mathbf{x}_{t'}, t', \mathbf{y}_0)\right|\right|_2^2\right], \tag{4}$$

where f_θ is the online network, f_{θ^-} is its exponential moving average (EMA) target, $t' < t$, and $\mathbf{x}_t, \mathbf{x}_{t'}$ are sampled using Eq. (3). This framework enables one-step inference by computing $\hat{\mathbf{x}}_0 = f_\theta(\mathbf{x}_T, T, \mathbf{y}_0)$, where $\mathbf{x}_T$ can be pure noise or a combination of $\mathbf{y}_0$ and noise.

3.2. ExpoCM: Exposure-Aware Consistency Framework

While the baseline conditional trajectory in Eq. (3) enables one-step generation, its formulation implicitly assumes a spatially uniform degradation. It treats all pixels in the LDR input $\mathbf{y}_0$ identically, regardless of their exposure conditions. This uniform treatment hinders its applicability to single-image HDR reconstruction, which is an inherently spatially heterogeneous problem. Specifically, over-exposed regions contain saturated pixels where information is lost and must be hallucinated. Under-exposed regions suffer from severe noise amplification, requiring careful denoising and detail reconstruction. In contrast, well-exposed regions preserve reliable content that should be preserved and enhanced. To address this, we propose an exposure-aware consistency framework, illustrated in Fig. 2. The core idea is to replace the single, uniform trajectory (Eq. (3)) with a spatially-varying one that adapts to the local exposure condition. Specifically, as shown in Fig. 2(b), we first generate a soft exposure partition of the LDR image through the exposure mask generation module. Then, as depicted

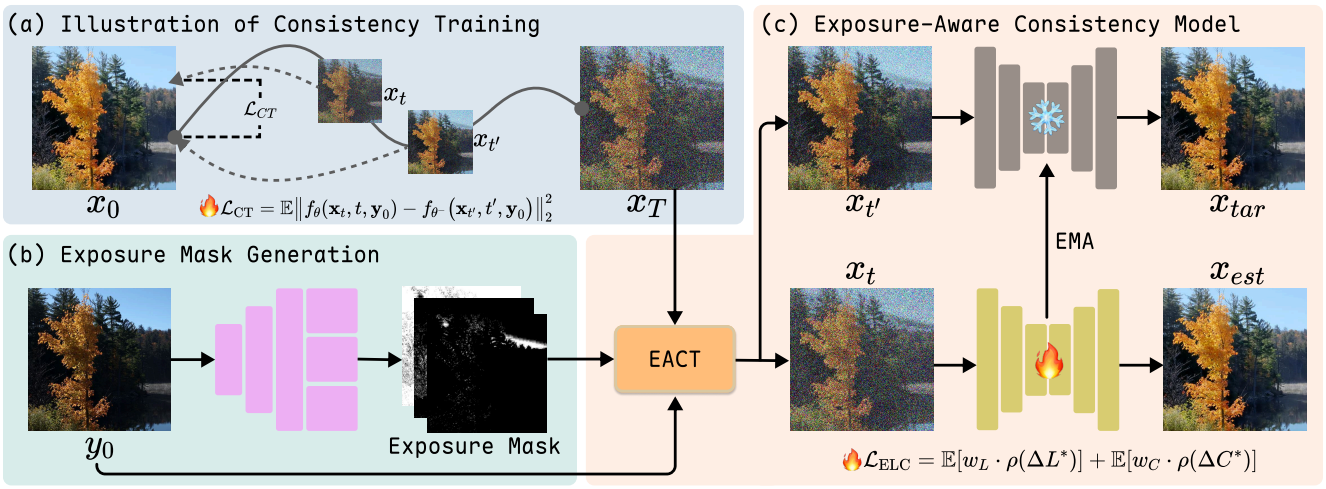

Figure 2. The overall pipeline of our proposed ExpoCM framework. The exposure mask generation module first partitions the input LDR $\mathbf{y}_0$ into over-, under-, and well-exposed regions (Fig. 2(b)). Based on these masks, we construct the exposure-aware consistency trajectory (EACT) by formulating and blending three distinct, region-specific generative flows, and the consistency network f_θ is optimized using the consistency training (CT) loss (Fig. 2(a)) and the proposed exposure-guided luminance-chromaticity (ELC) loss (Fig. 2(c)).

in Fig. 2(c), we construct a blended exposure-aware consistency trajectory (EACT) tailored to each region's characteristics, which is subsequently used for our exposure-aware consistency training.

Exposure Mask Generation To distinguish regions with different exposure characteristics in the input LDR image, we construct a soft exposure mask based on luminance statistics. Given an input $\mathbf{I}$, we first compute its luminance channel as $Y = 0.2126\mathbf{I}_R + 0.7152\mathbf{I}_G + 0.0722\mathbf{I}_B$. Rather than using fixed thresholds, which are highly sensitive to scene content, we adopt a robust, percentile-based strategy. Specifically, the 2^{nd} and 98^{th} luminance percentiles $(q_{\text{lo}}, q_{\text{hi}})$ are extracted, and a narrow transition band is defined using a margin τ=0.02:

$$l_{\text{core}} = q_{\text{lo}} + \tau(q_{\text{hi}} - q_{\text{lo}}), \quad h_{\text{core}} = q_{\text{hi}} - \tau(q_{\text{hi}} - q_{\text{lo}}). \tag{5}$$

Pixels darker than l_{core} are likely under-exposed, while those brighter than h_{core} are prone to saturation. We thus form continuous low- and high-exposure confidence maps by clipping the normalized distance to this core range:

$$\begin{aligned} m_{\text{low}} &= \text{clip}\left(\frac{l_{\text{core}} - Y}{\tau(q_{\text{hi}} - q_{\text{lo}})}, 0, 1\right), \\ m_{\text{high}} &= \text{clip}\left(\frac{Y - h_{\text{core}}}{\tau(q_{\text{hi}} - q_{\text{lo}})}, 0, 1\right). \end{aligned} \tag{6}$$

Finally, three exposure-aware weighting maps are obtained:

$$\begin{aligned} w_{\text{over}} &= m_{\text{high}}(1 - m_{\text{low}}), \\ w_{\text{under}} &= m_{\text{low}}(1 - m_{\text{high}}), \\ w_{\text{good}} &= 1 - \max(w_{\text{over}}, w_{\text{under}}). \end{aligned} \tag{7}$$

These weights softly partition the image into over-exposed, under-exposed, and well-exposed regions.

Exposure-Aware Consistency Trajectory After obtaining the exposure masks $w_{\text{over}}, w_{\text{under}}, w_{\text{good}}$, we construct region-specific consistency trajectories to align the probability flow with different exposure characteristics. Specifically, in saturated areas, structural information is completely missing in the LDR observation and cannot be reliably recovered from $\mathbf{y}_0$. Therefore, instead of relying on corrupted inputs, we encourage the model to synthesize plausible details purely from noise. The trajectory for over-exposed regions is defined as:

$$\mathbf{x}_t^o = (1 - \alpha(t))\mathbf{x}_0 + \sigma_o(t)\boldsymbol{\epsilon}, \tag{8}$$

where $\sigma_o(t)$ controls the generation strength.

In dark areas, the signal is not completely lost but heavily buried in noise. Directly using $\mathbf{y}_0$ introduces amplified noise and blurry details. To provide a reliable yet informative guidance, we apply a low-pass filter to extract coarse structural priors from $\mathbf{y}_0$ and inject them into the trajectory:

$$\mathbf{x}_t^u = (1 - \alpha(t))\mathbf{x}_0 + \alpha(t)\lambda_u \mathcal{F}_{\text{low}}(\mathbf{y}_0) + \sigma_u(t)\boldsymbol{\epsilon}, \tag{9}$$

where $\mathcal{F}_{\text{low}}(\cdot)$ denotes a low-frequency operator (e.g., Gaussian blur), and λ_u adjusts its contribution. For pixels that are neither saturated nor heavily underexposed, the LDR observation remains reliable. Hence, we follow a trajectory similar to the baseline:

$$\mathbf{x}_t^g = (1 - \alpha(t))\mathbf{x}_0 + \alpha(t)\mathbf{y}_0 + \sigma_g(t)\boldsymbol{\epsilon}. \tag{10}$$

The full exposure-aware consistency trajectory is obtained by spatially blending the three region-specific trajectories:

$$\mathbf{x}_t = w_{\text{over}} \odot \mathbf{x}_t^o + w_{\text{under}} \odot \mathbf{x}_t^u + w_{\text{good}} \odot \mathbf{x}_t^g, \tag{11}$$

where $\odot$ denotes element-wise multiplication. Note that unlike existing exposure-aware generative approaches [22, 27] that rely on a decoupled two-stage pipeline, which is often slow and artifact-prone, our ExpoCM is a unified one-step framework that solves restoration and generation simultaneously by mathematically blending ODE trajectories.

3.3. Exposure-guided Luminance-Chromaticity Loss

While our Exposure-Aware Consistency Trajectory provides a robust generative prior, the framework's perceptual fidelity can be further enhanced. Reconstructed images may still exhibit subtle luminance imbalance or color deviation, especially in severely over- or under-exposed regions. To further enhance perceptual fidelity, we introduce an Exposure-guided Luminance-Chromaticity (ELC) loss. This loss operates in the perceptually uniform CIE L*a*b* space, which explicitly decouples luminance (L*) from chromaticity (a*, b*), allowing us to apply adaptive, exposure-aware supervision. Given the predicted HDR image $\hat{I}_H$ and ground truth I_H, we first convert them into CIE L*a*b* space:

$$I_H \to (L^*, a^*, b^*), \quad \hat{I}_H \to (\hat{L}^*, \hat{a}^*, \hat{b}^*). \tag{12}$$

The luminance residual is computed as $\Delta L^* = \hat{L}^* - L^*$, and the chromaticity residual as $\Delta C^* = \sqrt{(\hat{a}^* - a^*)^2 + (\hat{b}^* - b^*)^2}$.

Exposure-aware weighting strategy. Different exposure regions exhibit distinct reliability in luminance and chromaticity. In under-exposed areas, chromaticity information is unreliable due to noise, whereas luminance still preserves structural cues. Thus, the loss should strongly enforce luminance consistency while reducing the penalty on unreliable chromaticity. Conversely, in over-exposed regions, pixels saturate towards white, losing color information. Here, the loss must strongly penalize chromaticity errors (i.e., restore color) while being more tolerant of luminance shifts. In well-exposed regions, both components remain reliable and are supervised in a balanced manner.

To implement this behavior in a continuous and differentiable fashion, we design luminance and chromaticity weights w_L and w_C as:

$$w_L = \lambda_L^{(0)} \left(1 + \kappa_L^{\text{lo}}\, s_Y\, w_{\text{under}}^{\alpha} + \kappa_L^{\text{hi}}\, A_{\text{spec}}\, w_{\text{over}}^{\alpha}\right), \tag{13}$$

$$w_C = \lambda_C^{(0)} \left(\kappa_C^{\text{hi}}\, w_{\text{over}}^{\alpha}(1 - A_{\text{spec}})\, h_Y + \kappa_C^{\text{lo}}\, w_{\text{under}}^{\alpha}(1 - s_Y)\right). \tag{14}$$

Here, w_{under} and w_{over} are the exposure-aware weighting maps defined in Eq. (7). The exponent α controls the sharpness of the mask transitions. $s_Y = \sigma\left(\frac{\tau_s - Y}{\delta_s}\right)$ is a shadow-gating function that emphasizes luminance supervision in dark regions. $A_{\text{spec}} = \frac{1}{1 + C_0^*/c_0}$ measures the near-white tendency of highlights (where $C_0^* = \sqrt{a_0^{*2} + b_0^{*2}}$ is the ground-truth chroma and c_0 is a scaling constant). $h_Y = \sigma\left(\frac{Y - \tau_h}{\delta_h}\right)$ is a highlight visibility factor that modulates chromaticity sensitivity in bright regions. Finally, $\lambda_L^{(0)}$ and $\lambda_C^{(0)}$ are baseline loss weights to ensure stable supervision in well-exposed regions.

Finally, the ELC loss is defined as the weighted sum of robust penalties:

$$\mathcal{L}_{\text{ELC}} = \mathbb{E}\left[w_L \cdot \rho(\Delta L^*)\right] + \mathbb{E}\left[w_C \cdot \rho(\Delta C^*)\right], \tag{15}$$

where $\rho(\cdot)$ is a robust penalty function (we use the Charbonnier loss, $\rho(x) = \sqrt{x^2 + \epsilon^2}$). We empirically set $\kappa_L^{\text{lo}} = 3$, $\kappa_L^{\text{hi}} = 1$, $\kappa_C^{\text{hi}} = 3$, $\kappa_C^{\text{lo}} = 0.5$, $\tau_s = 0.2$, $\tau_h = 0.8$, and $\delta_{\{s,h\}} = 0.1$. Notably, varying α, κ_C/κ_L, or τ_s/τ_h slightly results in negligible performance fluctuations (< 0.1dB).

3.4. Model Training

Our consistency network f_θ is based on the U-Net architecture [42], utilizing three downsampling and three upsampling stages, with several residual blocks embedded in each. The network's input is formed by concatenating the noisy state $\mathbf{x}_t$ and the LDR image $\mathbf{y}_0$ along the channel dimension. The time step t is converted into a positional embedding and injected into each residual block. We adopt a two-stage training strategy to ensure both generative stability and high perceptual fidelity. In the first stage, we train the network using the consistency training loss ($\mathcal{L}_{\text{CT}}$), as defined in Eq. (4), to learn the exposure-aware consistency trajectories and enable stable, one-step inference. In the second stage, we finetune the model using our proposed ELC loss ($\mathcal{L}_{\text{ELC}}$) to explicitly mitigate luminance imbalances and color drift, obtaining the final high-fidelity HDR results.

4. Experiments

4.1. Experimental Settings

Implementation Details We implement our framework in PyTorch and train our models on NVIDIA 3090 GPUs. The network is trained for 500,000 iterations with a total batch size of 4, using randomly cropped 256×256 patches. We employ the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$. The initial learning rate is set to 5×10^{-5} and is gradually decayed to a minimum of 1×10^{-7} using a cosine annealing scheduler. For inference, our ExpoCM processes a 512×512 image in 0.33s, which is significantly faster than DDPM (174.10s) and DDIM (7.85s).

Datasets Following [8, 24], we conduct our experiments on three widely used benchmarks: HDR-REAL [24], HDR-EYE [36], and the AIM2025 [49] Challenge dataset. Specifically, HDR-REAL and HDR-EYE contain 1,838 and 46 LDR-HDR pairs of size 512×512, respectively, while the AIM2025 dataset consists of 18,898 paired samples.

Table 1. Quantitative comparisons on the HDR-REAL [24], HDR-EYE [36], and AIM2025 [49] challenge datasets. '-l', '-μ', and '-PU' denote metrics computed on linear, μ-law tonemapped, and perceptually-uniform (PU) encoded domain, respectively. ↑ indicates higher is better, and ↓ indicates lower is better. The best and second-best results are highlighted in **bold** and underlined, respectively.

Dataset	Method	PSNR-μ ↑	SSIM-μ ↑	PSNR-PU ↑	SSIM-PU ↑	PSNR-l ↑	SSIM-l ↑	MS-SSIM ↑	HDR-VDP-2/-3 ↑	LPIPS ↓	ΔE_{2000} ↓
HDR-REAL [24]	HDRCNN [9]	14.99	0.5638	16.25	0.5305	25.81	0.5936	0.7678	31.52 / 5.71	0.3497	21.14
	SingleHDR [24]	17.92	0.5906	19.93	0.5695	31.08	0.7758	0.8081	34.44 / 6.16	0.3156	14.36
	ExpandNet [34]	18.07	0.5999	20.21	0.5953	31.70	0.7878	0.8186	30.35 / 5.91	0.3416	13.77
	HDRUNet [6]	16.22	0.6327	17.79	0.6157	28.35	0.7118	0.7293	24.23 / 4.41	0.3937	15.72
	DDPM [12]	25.45	0.8173	24.78	0.7901	**37.46**	0.9063	0.9041	43.52 / 7.45	0.1921	10.40
	DDIM [45]	20.77	0.6925	22.65	0.6716	34.23	0.8539	0.8307	39.14 / 6.76	0.2365	14.26
	HDR-Trans. [26]	15.71	0.6027	16.83	0.5532	25.08	0.6259	0.7870	31.72 / 5.69	0.3665	19.24
	Reti-Diff [11]	27.64	0.8354	29.15	0.8666	35.93	0.9397	0.9147	42.08 / 7.31	0.2645	4.83
	Ours	**28.66**	**0.8684**	**30.07**	**0.8935**	36.22	**0.9521**	**0.9304**	**44.27 / 7.72**	**0.1919**	**4.02**
HDR-EYE [36]	HDRCNN [9]	15.55	0.5986	16.12	0.5673	22.84	0.7030	0.8049	37.84 / 7.08	0.2811	14.05
	SingleHDR [24]	15.04	0.6535	14.36	0.5536	19.04	0.5612	0.8813	45.23 / 7.66	0.2436	19.28
	ExpandNet [34]	16.09	0.7023	15.15	0.6073	17.05	0.5605	0.8878	27.97 / 7.32	0.3105	17.55
	HDRUNet [6]	14.81	0.6883	13.99	0.6289	17.69	0.6014	0.8149	26.56 / 5.79	0.3054	15.35
	DDPM [12]	17.45	0.7496	16.56	0.6859	**23.38**	0.6191	0.9040	53.12 / **7.99**	**0.2005**	13.81
	DDIM [45]	16.98	0.7647	16.03	0.7062	21.57	0.6270	0.9044	**53.47** / 7.92	0.2007	13.19
	HDR-Trans. [26]	17.23	0.7453	16.47	0.6889	20.85	0.6576	0.8801	44.62 / 7.51	0.2537	12.78
	Reti-Diff [11]	15.36	0.6944	14.97	0.6163	18.77	0.5626	0.8974	46.26 / 7.74	0.2475	17.52
	Ours	**20.75**	**0.8017**	**19.32**	**0.7638**	21.30	**0.7424**	**0.9053**	44.09 / 7.94	0.2353	**9.68**
AIM2025 [49]	HDRCNN [9]	17.67	0.6147	16.73	0.6250	23.96	0.6750	0.8097	42.16 / 6.57	0.3605	10.92
	SingleHDR [24]	20.77	0.7328	21.55	0.7053	28.48	0.7930	0.9027	64.48 / 7.90	0.2460	10.64
	ExpandNet [34]	24.94	0.8281	25.06	0.8402	29.16	0.8765	0.9455	66.94 / 8.21	0.2149	5.93
	HDRUNet [6]	25.88	0.8709	26.38	0.8808	30.98	0.9131	0.9312	57.83 / 7.06	0.2218	4.46
	DDPM [12]	23.03	0.8320	23.36	0.8145	29.71	0.8316	0.9550	**75.57 / 8.78**	**0.1286**	7.91
	DDIM [45]	19.50	0.7733	19.72	0.7277	27.22	0.7602	0.9275	69.96 / 8.36	0.1580	11.24
	HDR-Trans. [26]	17.12	0.7253	16.70	0.6702	20.26	0.6174	0.8978	62.66 / 7.78	0.2604	14.94
	Reti-Diff [11]	23.31	0.8269	23.49	0.8229	28.65	0.8370	0.9363	66.93 / 8.19	0.2332	7.74
	Ours	**29.02**	**0.8922**	**29.06**	**0.9069**	**31.97**	**0.9360**	**0.9654**	74.01 / 8.68	0.1511	**3.90**

Compared Methods To assess the performance of the proposed ExpoCM, we compare it with state-of-the-art methods, including HDRCNN [9], SingleHDR [24], ExpandNet [34], HDRUNet [6], HDR-Transformer [26], and Reti-Diff [11]. In addition, we also include two representative diffusion models, DDPM [12] and DDIM [45], to demonstrate the advantages of our one-step diffusion framework.

Evaluation Metrics We evaluate the reconstruction fidelity using five full-reference metrics, including PSNR, SSIM [51], MS-SSIM [50], HDR-VDP-2 [32], and HDR-VDP-3 [33]. For PSNR and SSIM, evaluations are conducted in three domains: the linear domain, the tonemapped domain obtained by μ-law, and the PU21 [1] encoded domain, denoted as '-l', '-μ', and '-PU', respectively. For the HDR-VDP-2, we use the standard configuration of 30 pixels per degree (PPD) at a viewing distance of 0.5m. In addition, we report LPIPS [55] and ΔE_{2000} to assess the perceptual quality and color accuracy of the reconstructed results.

4.2. Quantitative Comparison

The quantitative results on the HDR-REAL, HDR-EYE, and the AIM2025 challenge datasets are summarized in Table 1. Several key observations can be made. First, recent advances in learning-based architectures [6, 9, 11, 24, 26, 34] have substantially improved the fidelity of single-image HDR reconstruction, achieving high PSNR and SSIM scores (e.g., HDRUNet [6] and Reti-Diff [11]). Second, diffusion-based method DDPM [12] markedly enhance perceptual quality and obtain the highest LPIPS values, yet often at the expense of pixel-wise fidelity due to their stochastic generation process. Moreover, its efficient variant DDIM [45] exhibit pronounced performance degradation when the number of sampling steps is reduced, revealing their heavy reliance on iterative denoising. Finally, our proposed ExpoCM, empowered by exposure-aware consistency training, attains state-of-the-art fidelity among single-step methods, while maintaining highly competitive perceptual quality across most datasets with a significantly accelerated inference speed. Furthermore, the incorporation of the proposed Exposure-guided Luminance-Chromaticity (ELC) loss effectively mitigates the inherent luminance and color bias of diffusion models, resulting in more faithful color reconstruction (i.e., the lowest ΔE_{2000}).

4.3. Qualitative Comparison

Fig. 3 illustrates the qualitative comparisons with state-of-the-art methods on challenging scenes from the HDR-REAL [24] and AIM2025 [49] datasets. As can be seen, previous CNN-based methods (e.g., HDRCNN, SingleHDR, ExpandNet, and HDRUNet) struggle to reconstruct

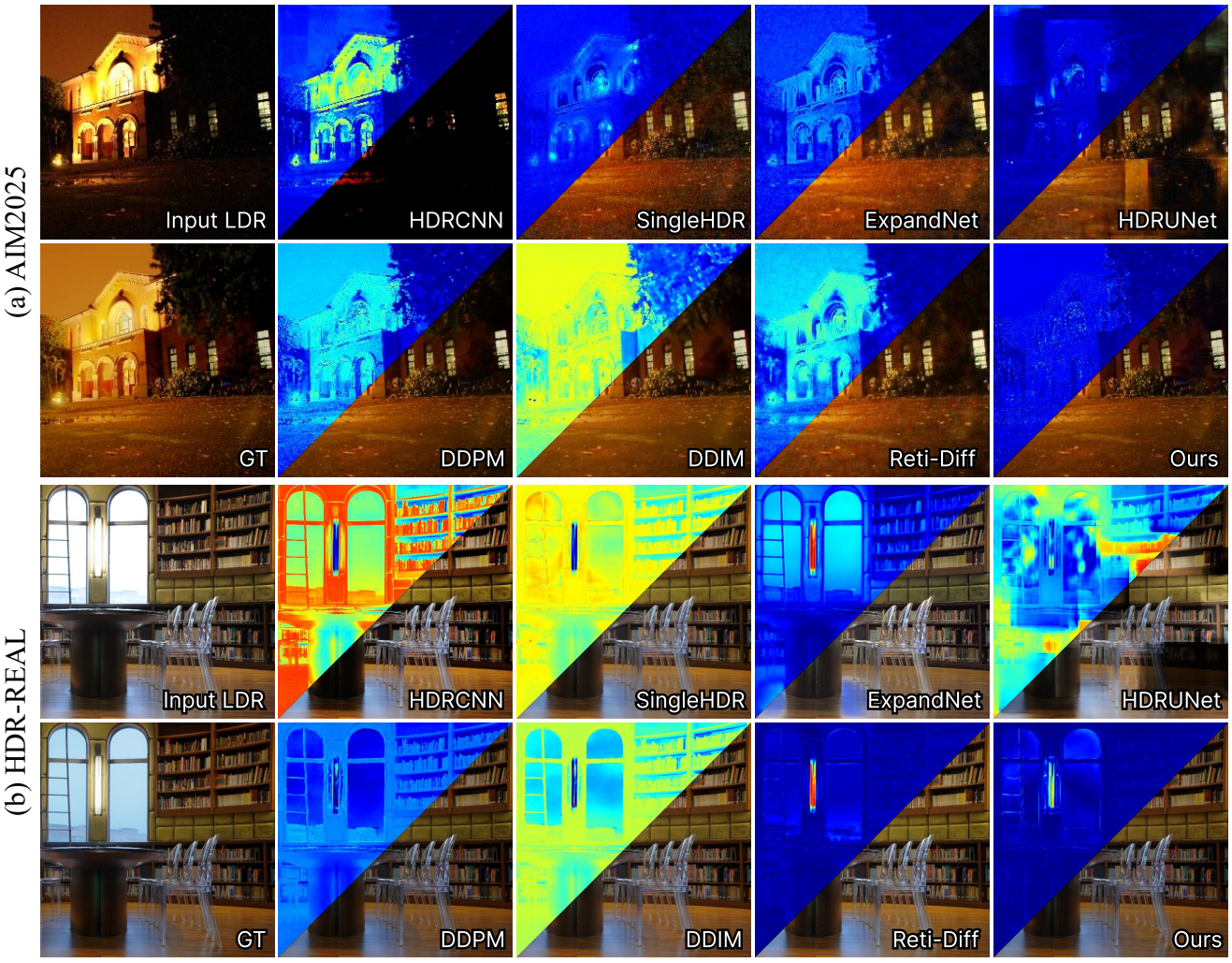

Figure 3. Qualitative comparisons with state-of-the-art single-image HDR reconstruction methods on the AIM2025 and HDR-REAL datasets. For each method, we show the reconstructed HDR image and its corresponding error map, which visualizes the pixel-wise difference from the ground-truth HDR image (darker regions indicate smaller errors).

missing details in saturated, over-exposed regions. Furthermore, they tend to produce noticeable blurriness or artifacts in under-exposed areas after denoising. While diffusion-based methods (e.g., DDPM) can alleviate this issue with their strong generative priors, they often suffer from a global brightness bias. Moreover, accelerating these models via reduced sampling steps (e.g., DDIM) significantly degrades reconstruction quality. In contrast, our proposed ExpoCM achieves high-quality HDR reconstruction within a single inference step and can effectively mitigate both global brightness and local color biases, faithfully recovering details in both over- and under-exposed regions.

4.4. Ablation Study

In this section, we conduct comprehensive ablation studies to validate the effectiveness of each core component. Unless otherwise specified, all variants are trained following the implementation details in Sec 4.1. The quantitative results on the HDR-REAL and AIM2025 datasets are reported in Table 2 and Table 3.

Exposure-aware Consistency Trajectory We first analyze the effectiveness of our core design, the Exposure-Aware Consistency Trajectory. Our central hypothesis is that a spatially heterogeneous trajectory, tailored to regional degradation, is superior to a spatially-agnostic one. To validate this, we compare the three variants detailed in Table 2: (1) a Baseline model using the uniform trajectory from Eq. (3), (2) a Two-Mask variant that only distinguishes between well-exposed and ill-posed regions, and (3) our full Three-Mask framework which uses three distinct, region-specific trajectories. As shown in Table 2, the baseline model yields the poorest results. The Two-Mask variant significantly

Table 2. Quantitative results of the ablation study on our Exposure-Aware Consistency Trajectory (EACT). We compare: (1) **Baseline**, which uses a uniform, spatially-agnostic trajectory. (2) **Two-Mask**, a simplified variant that distinguishes only between well-exposed and ill-posed (over- and under-exposed combined) regions. (3) **Three-Mask**, our full framework using three distinct trajectories for over-, under-, and well-exposed regions.

Method	HDR-REAL [24]			AIM2025 [49]		
	PSNR-$\mu\uparrow$	SSIM-$\mu\uparrow$	LPIPS $\downarrow$	PSNR-$\mu\uparrow$	SSIM-$\mu\uparrow$	LPIPS $\downarrow$
Baseline	21.09	0.6917	0.3041	27.90	0.8842	0.1920
Two-Mask	25.75	0.8076	0.2785	28.48	**0.8936**	0.1543
Three-Mask	**25.84**	**0.8282**	**0.2754**	**28.89**	0.8921	**0.1504**

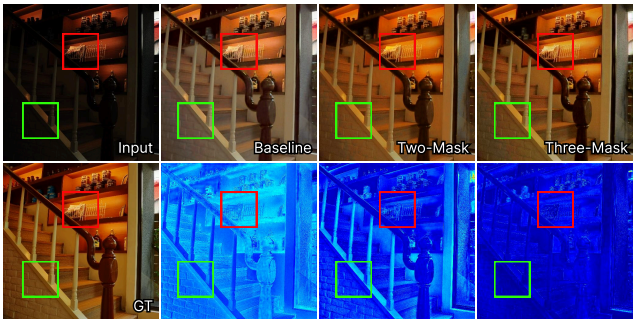

Figure 4. Visual results of our ablation studies on the proposed exposure-aware consistency trajectory. Our full method exhibits the lowest reconstruction error in both the over-exposed (green box) and highlight (red box) regions.

improves all metrics (e.g., +4.66 PSNR on HDR-REAL), demonstrating the clear benefit of separating reliable from unreliable regions. Our full Three-Mask model achieves the best performance across both datasets. The qualitative results presented in Fig. 4 further reinforce these findings, visually validating our hypothesis that explicitly distinguishing between over-exposure and under-exposure is critical for high-fidelity reconstruction.

Effectiveness of ELC Loss We next validate the contribution of our Exposure-guided Luminance-Chromaticity (ELC) loss. The quantitative results are presented in Table 3. We conduct two key comparisons. First, to demonstrate the general effectiveness of the ELC loss, we compare the baseline model with the 'w/o EACT' variant. The 'w/o EACT' model simply adds our ELC loss onto the baseline trajectory. As shown in the first two rows of Table 3, this addition provides a consistent improvement across all metrics, most notably reducing the color error ΔE_{2000} from 12.23 to 12.04 on HDR-REAL. This confirms that the ELC loss enhances perceptual fidelity even on a suboptimal backbone. Second, and more importantly, we isolate the benefit of our exposure-guided weighting strategy. We achieve this by comparing the 'w/o weighting' model against our default (i.e., full) framework. Both models utilize our powerful EACT (Three-Mask) trajectory, but the 'w/o weighting' variant applies a spatially uniform CIE L*a*b* loss, (i.e., w_L and w_C are constant). The results clearly show that our default model, with its adaptive weights, outperforms this variant on both datasets, achieving the lowest ΔE_{2000}. The visual comparisons in Fig. 5 further validate the effectiveness of our ELC loss in enforcing decoupled constraints on luminance and chrominance, leading to visibly superior color and brightness fidelity.

Table 3. Quantitative results of ablation studies about the proposed ELC loss. **Baseline**: Uniform Trajectory. **'w/o' EACT**: Uniform Trajectory + Our full ELC loss. **'w/o' weighting**: EACT Trajectory + Uniform CIE L*a*b* loss. **Default** (Ours): EACT Trajectory + Our full ELC loss. Best results are in bold.

Method	HDR-REAL [24]			AIM2025 [49]		
	PSNR-$\mu\uparrow$	SSIM-$\mu\uparrow$	$\Delta E_{2000}\downarrow$	PSNR-$\mu\uparrow$	SSIM-$\mu\uparrow$	$\Delta E_{2000}\downarrow$
Baseline	21.09	0.6917	12.23	27.90	0.8842	4.24
'w/o' EACT	21.26	0.6988	12.04	28.23	0.8833	4.22
'w/o' weighting	28.56	0.8663	4.06	28.91	0.8898	3.95
Default	**28.66**	**0.8684**	**4.02**	**29.02**	**0.8922**	**3.90**

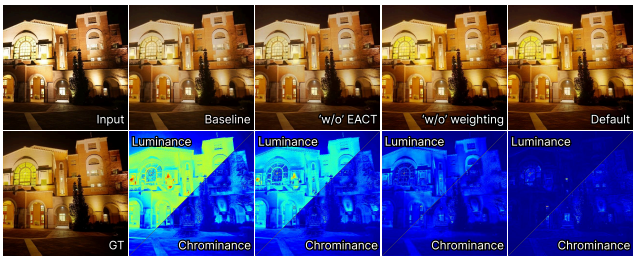

Figure 5. Visual comparisons of our ablation studies on the proposed ELC loss. Compared to all variants, the model optimized with our full ELC loss demonstrates the minimal luminance (top-left) and chrominance (bottom-right) error.

5. Conclusion

In this paper, we have proposed ExpoCM, a novel one-step generative framework for single-image HDR reconstruction. Our method is designed to address the highly ill-posed nature of this task, specifically the exposure-dependent degradation and the high computational cost of recent generative models. We have introduced an Exposure-Aware Consistency Trajectory (EACT) that partitions the LDR input and tailors the generative PF-ODE flow to over-, under-, and well-exposed regions, enabling high-fidelity reconstruction within a single, distillation-free inference step. To further enhance perceptual quality, we have developed an Exposure-guided Luminance-Chromaticity (ELC) loss in the perceptually uniform CIE L*a*b* space, which adaptively mitigates brightness imbalance and color drift. Extensive experiments have demonstrated that ExpoCM achieves state-of-the-art fidelity and perceptual accuracy, while being substantially faster than diffusion-based counterparts.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (NSFC) under grant 62372091, and in part by the Hainan Province Science and Technology Plan Project under Grant ZDYF2024(LALH)001.

References

[1] Maryam Azimi et al. Pu21: A novel perceptually uniform encoding for adapting existing quality metrics for hdr. In *2021 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2021. 6

[2] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Inverse tone mapping. In *CGIT*, pages 349–356, 2006. 2

[3] Francesco Banterle, Patrick Ledda, Kurt Debattista, Alan Chalmers, and Marina Bloj. A framework for inverse tone mapping. *The Visual Computer*, 23(7):467–478, 2007. 2

[4] Francesco Banterle, Kurt Debattista, Alessandro Artusi, Sumanta Pattanaik, Karol Myszkowski, Patrick Ledda, and Alan Chalmers. High dynamic range imaging and low dynamic range expansion for generating hdr content. In *Comput. Graph. Forum*, pages 2343–2367. Wiley Online Library, 2009. 2

[5] Rufeng Chen, Bolun Zheng, Hua Zhang, Quan Chen, Chenggang Yan, Gregory Slabaugh, and Shanxin Yuan. Improving dynamic hdr imaging with fusion transformer. In *AAAI*, pages 340–349, 2023. 2

[6] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In *CVPR*, pages 354–363, 2021. 2, 3, 6

[7] Shen Cheng, Haipeng Li, Haibin Huang, Xiaohong Liu, and Shuaicheng Liu. Blind-spot guided diffusion for self-supervised real-world denoising. *arXiv preprint arXiv:2509.16091*, 2025. 3

[8] Dwip Dalal, Gautam Vashishtha, Prajwal Singh, and Shanmuganathan Raman. Single image ldr to hdr conversion using conditional diffusion. In *ICIP*, pages 3533–3537. IEEE, 2023. 5

[9] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM TOG*, 36(6):1–15, 2017. 2, 3, 6

[10] 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. *ACM TOG*, 35(6):1–12, 2016. 2

[11] 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. In *ICLR*, 2025. 6

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, pages 6840–6851, 2020. 3, 6

[13] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. Hdr deghosting: How to deal with saturation? In *CVPR*, pages 1163–1170, 2013. 2

[14] Yan Huang, Xiaoshan Liao, Jinxiu Liang, Boxin Shi, Yong Xu, and Patrick Le Callet. Detail-preserving diffusion models for low-light image enhancement. *IEEE TCSVT*, 2024. 3

[15] Yongqing Huo, Fan Yang, Le Dong, and Vincent Brost. Physiological inverse tone mapping based on retina response. *The Visual Computer*, 30(5):507–517, 2014. 2

[16] Hai Jiang, Ao Luo, Songchen Han, Haoqiang Fan, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM TOG*, 42(6):1–14, 2023. 3

[17] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM TOG*, 36 (4):144, 2017. 1, 2

[18] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS*, pages 23593–23606, 2022. 3

[19] Bahjat Kawar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. Imagic: Text-based real image editing with diffusion models. In *CVPR*, pages 6007–6017, 2023. 3

[20] Erum Arif Khan, Ahmet Oguz Akyuz, and Erik Reinhard. Ghost removal in high dynamic range images. In *ICIP*, pages 2005–2008, 2006. 2

[21] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *ECCV*, pages 596–611, 2018. 2, 3

[22] Baiang Li, Sizhuo Ma, Yanhong Zeng, Xiaogang Xu, Youqing Fang, Zhao Zhang, Jian Wang, and Kai Chen. Exposure-limited image enhancement with generative diffusion prior. In *2025 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2025. 3, 5

[23] Haipeng Li, Hai Jiang, Ao Luo, Ping Tan, Haoqiang Fan, Bing Zeng, and Shuaicheng Liu. Dmhomo: Learning homography with diffusion models. *ACM TOG*, 43(3):1–16, 2024. 2

[24] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In *CVPR*, pages 1651–1660, 2020. 2, 3, 5, 6, 8

[25] Zhen Liu, Wenjie Lin, Xinpeng Li, Qing Rao, Ting Jiang, Mingyan Han, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Adnet: Attention-guided deformable convolutional network for high dynamic range imaging. In *CVPRW*, pages 463–470, 2021. 1, 2

[26] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *ECCV*, pages 344–360. Springer, 2022. 1, 2, 6

[27] Zhen Liu, Hai Jiang, Haipeng Li, Shuaicheng Liu, and Bing Zeng. Solving ill-posed regions in high dynamic range reconstruction with uncertainty-aware diffusion models. *IEEE TCSVT*, 2025. 5

[28] Zhen Liu, Diedong Feng, Hai Jiang, Liaoyuan Zeng, Hao Wang, Chaoyu Feng, Lei Lei, Bing Zeng, and Shuaicheng Liu. Raw-flow: Advancing rgb-to-raw image reconstruction with deterministic latent flow matching. *AAAI*, 40(9):7431–7439, 2026. 3

[29] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, pages 11461–11471, 2022. 3

[30] Xinglong Luo, Ao Luo, Kunming Luo, Zhengning Wang, Ping Tan, Bing Zeng, and Shuaicheng Liu. Learning efficient meshflow and optical flow from event cameras. *IEEE TPAMI*, 48(2):1355–1372, 2026. 2

[31] Kede Ma, Zhengfang Duanmu, Hanwei Zhu, Yuming Fang, and Zhou Wang. Deep guided learning for fast multi-exposure image fusion. *IEEE TIP*, 29:2808–2819, 2019. 2

[32] Rafał Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM TOG*, 30(4):1–14, 2011. 6

[33] Rafal K Mantiuk, Dounia Hammou, and Param Hanji. Hdr-vdp-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625*, 2023. 6

[34] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *Comput. Graph. Forum*, pages 37–49. Wiley Online Library, 2018. 2, 3, 6

[35] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *PG*, pages 382–390, 2007. 1, 2

[36] Hiromi Nemoto, Pavel Korshunov, Philippe Hanhart, and Touradj Ebrahimi. Visual attention in ldr and hdr images. In *VPQM*, page 6, 2015. 5, 6

[37] Mang Ning, Mingxiao Li, Jianlin Su, Albert Ali Salah, and Itir Onal Ertugrul. Elucidating the exposure bias in diffusion models. In *ICLR*, 2024. 2

[38] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions. *IEEE TIP*, 30: 3885–3896, 2021. 2

[39] Tae-Hyun Oh, Joon-Young Lee, Yu-Wing Tai, and In So Kweon. Robust high dynamic range imaging by rank minimization. *IEEE TPAMI*, 37(6):1219–1232, 2014. 2

[40] K Ram Prabhakar, Gowtham Senthil, Susmit Agrawal, R Venkatesh Babu, and Rama Krishna Sai S Gorthi. Labeled from unlabeled: Exploiting unlabeled data for few-shot deep hdr deghosting. In *CVPR*, pages 4875–4885, 2021. 2

[41] Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In *ICCV*, pages 10721–10733, 2023. 3

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015. 5

[43] Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. Single image hdr reconstruction using a cnn with masked features and perceptual loss. *ACM TOG*, 2020. 2

[44] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *ACM TOG*, 31(6):203, 2012. 2

[45] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 3, 6

[46] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 3

[47] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023. 3

[48] Steven Tel, Zongwei Wu, Yulun Zhang, Barthélémy Heyrman, Cédric Demonceaux, Radu Timofte, and Dominique Ginhac. Alignment-free hdr deghosting with semantics consistent transformer. In *ICCV*, 2023. 2

[49] Chao Wang, Francesco Banterle, Bin Ren, Radu Timofte, Xin Lu, Yufeng Peng, Chengjie Ge, Zhijing Sun, Ziang Zhou, Zihao Li, et al. Aim 2025 challenge on inverse tone mapping report: Methods and results. In *ICCV*, pages 5571–5584, 2025. 5, 6, 8

[50] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The thrity-seventh asilomar conference on signals, systems & computers, 2003*, pages 1398–1402. IEEE, 2003. 6

[51] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 6

[52] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *ECCV*, pages 117–132, 2018. 1, 2

[53] Shaoan Xie, Zhifei Zhang, Zhe Lin, Tobias Hinz, and Kun Zhang. Smartbrush: Text and shape guided object inpainting with diffusion model. In *CVPR*, pages 22428–22437, 2023. 3

[54] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *CVPR*, pages 1751–1760, 2019. 1, 2

[55] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 6

[56] Zhixing Zhang, Ligong Han, Arnab Ghosh, Dimitris N Metaxas, and Jian Ren. Sine: Single image editing with text-to-image diffusion models. In *CVPR*, pages 6027–6037, 2023. 3

[57] Chu Zhou, Hang Zhao, Jin Han, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. Unmodnet: Learning to unwrap a modulo image for high dynamic range imaging. In *NeurIPS*, pages 1559–1570, 2020. 2

[58] Tianhao Zhou, Haipeng Li, Ziyi Wang, Ao Luo, Chen-Lin Zhang, Jiajun Li, Bing Zeng, and Shuaicheng Liu. Recdiffusion: Rectangling for image stitching with diffusion models. In *CVPR*, pages 2692–2701, 2024. 3