EUROGRAPHICS 2025 / A. Bousseau and A. Dai
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 44 (2025), Number 2

Bracket Diffusion: HDR Image Generation by Consistent LDR Denoising

M. Bemana[1] T. Leimkühler[1] K. Myszkowski[1] H.P. Seidel[1] and T. Ritschel[2]

[1] Max-Planck-Institut für Informatik, Germany
[2] University College London, United Kingdom

Figure 1: *Existing denoising diffusion models (top row) generate images with low-dynamic range (LDR) on a certain exposure in the center. When re-exposed to other levels, bright parts like the lamps do not retain their contrast, and dark areas do not reveal details as in the shadow below the table. In our high-dynamic range (HDR) approach (bottom), diffusion is performed at multiple exposure brackets, such that the lamps retain their contrast and the details in the animals' bodies are produced without noise (see insets). An example application is an HDR display, where high pixel values map to high physical intensity.*

Abstract

We demonstrate generating HDR images using the concerted action of multiple black-box, pre-trained LDR image diffusion models. Common diffusion models are not HDR as, first, there is no sufficiently large HDR image dataset available to re-train them, and, second, even if it was, re-training such models is impossible for most compute budgets. Instead, we seek inspiration from the HDR image capture literature that traditionally fuses sets of LDR images, called "exposure brackets", to produce a single HDR image. We operate multiple denoising processes to generate multiple LDR brackets that together form a valid HDR result. To this end, we introduce a brackets consistency term into the diffusion process to couple the brackets such that they agree across the exposure range they share. We demonstrate HDR versions of state-of-the-art unconditional and conditional as well as restoration-type (LDR2HDR) generative modeling.

1. Introduction

Images generated by modern denoising diffusion models [RBL*22, SDWMG15] have shown an unprecedented combination of user control and image quality. Unfortunately, the resulting images are LDR while in computer graphics, several applications, such as physically-based simulation and rendering [Deb98, RWP*10], scene reconstruction with significant shadows and specular highlights [JSYJYBO22, HZF*22, MHMB*22], as well as advanced television displays [LZH*24, SIS11, SHS*04], and emerging virtual reality systems [ZJY*21, ZMW*20], rely on the capabilities of HDR imaging.

We propose to close this gap by introducing a simple and effective method to upgrade a black-box denoising diffusion model from LDR to HDR image generation.

This poses two main challenges: first, the *limited scale of the available HDR training data*, which is orders of magnitude lower than its LDR counterpart, and second, the fact that for most users, it is *impossible to re-train the denoiser due to the sheer compute requirements*. We overcome the first challenge by avoiding producing HDR directly. Instead, we produce a set of individual *brackets*, i.e., LDR images, which can be merged into an HDR image. This allows us to circumvent the first challenge by never operating the denoiser on HDR images, and hence, also overcome the second challenge, as we circumvent the need to re-train the denoiser in HDR. Our method does not need any fine-tuning or training and considers the denoiser a black box.

Figure 2: *Recalling HDR merging: LDR brackets are shown on the left; right, the weights for each bracket, for simplicity in binary. White means this pixel will contribute to the final HDR.*

Instead, the task is to produce brackets that are meaningful, i.e., meaningful on their own and meaningful in combination with other brackets (Fig. 2). To be plausible on its own, a bracket should have all details, without noise, in the range of values it represents. To work as a combination, a value in one bracket must match its value re-exposed to another bracket and ultimately when they are merged. We achieve these properties by deriving a diffusion process based on ideas from diffusion posterior sampling (DPS) [CKM*22] that operates between multiple brackets jointly.

2. Background: Multi-exposure HDR imaging

HDR images directly register scene radiance, typically up to a scale factor, so that image details in the darkest and brightest scene regions are readily available. As sensors with HDR capabilities are relatively rare and expensive, typically, a stack of differently exposed LDR photographs (refer to Fig. 2) is merged into an HDR image [DM97, MN99, RBS03, WSP*23b]. By transforming each pixel value through an inverted camera response and then dividing by the exposure time, a measurement of the scene radiance can be derived [RHD*10]. As such, per-pixel measurements are the most reliable in the middle range of the camera response [DM97]; an accordingly weighted average of the measurements can be computed for all exposures. Fig. 2-right shows a simplified version of such weights for exposure brackets EV-1, EV+0, and EV+1, where EV+x denotes multiplying with 2^x in the linear radiance space. Note that the radiance ranges below the black level and over 1 are covered just in a single exposure EV+1 and EV-1, respectively, while for EV+0, radiance information is clamped on both sides of the range. Dark image regions are also contaminated with sensor noise, whose characteristics may differ between exposures, which makes consistent denoising difficult [MKM*20, CFXL20, CBM*22]. Some camera manufacturers introduce hard clamping at a black-level radiance, assuming that there is no reliable image information below this threshold due to noise. Finally, the performance of the multi-exposure methods might be limited for large scene/camera motion that causes ghosting that is further aggravated by simultaneous image saturation [KR17, YGS*19, YWL*20, WXTT18]. The latter problem can be reduced through consistent image hallucination using adversarial training [NWL*21, LWW*22] or conditional diffusion [YHS*23] components.

In this work, we aim to use diffusion [HJA20, SDWMG15, CKM*22] to generate consistent multiple exposures. In this process, we need to account for missing information due to clamping and, when relevant, denoise.

3. Previous Work

In this section, we discuss previous work on deep single-image HDR reconstruction methods and the use of diffusion models in HDR imaging that are central to this work. A broader perspective on other aspects of deep learning for HDR imaging can be found in a recent survey [WY22].

Deep single-image HDR reconstruction (LDR2HDR) An alternative solution to multi-exposure techniques (Sec. 2) relies on restoring HDR information from a single LDR image. Traditional methods are extensively covered by Banterle et al. [BADC17], and here, we focus on recent machine-learning solutions. Single-image HDR reconstruction can be performed directly [EKD*17, MBRHD18, SRK20, LLC*20, ZA21, YLL*21, CWL22], or, alternatively, by first producing a stack of different exposures that are then merged into an HDR image [EKM17, LAK18a, LAK18b, LJAK20, JLAK21]. Instead of producing LDR stacks with fixed predefined EVs, Chen et al. [CYL*23] propose generating LDR stacks at continuous arbitrary values to achieve higher quality. Specialized solutions are required when an observation EV+0 is captured in dark conditions, where denoising is a key problem [CCXK18, WYY*23]. Text conditioning driven by a contrastive language-image pre-training (CLIP) model [RKH*21] can be used for the generation of a well-exposed LDR environment map that is then transformed into its HDR counterpart by a fully supervised network [CWL22]. Even though some methods employ adversarial training [ZA21, LAK18b], the key problem remains limited performance in reconstructing clamped regions. Those methods mostly require LDR and HDR image pairs for training, which is problematic due to limited datasets. Recently, GlowGAN [WSP*23a] addressed the latter two problems by fully unsupervised learning a generative model of HDR images exclusively from in-the-wild LDR images. As this approach is based on StyleGAN-XL [SSG22], it requires GAN training on narrow domains (e.g., lightning, fireworks) to capture the respective HDR image distribution.

Diffusion models in HDR imaging Denoising diffusion probabilistic models (DDPMs) [HJA20, SDWMG15] demonstrate huge capacity in modeling complex distributions and typically outperform other generative models in terms of image realism, diversity, and detail reproduction [DN21]. DDPMs also proved useful for solving linear [SSDK*21] and non-linear [CKM*22] inverse imaging problems that are common in image restoration and enhancement tasks guided by the degraded input image. Image inpainting [LDR*22], deblurring [KEES22], and super-resolution [SHC*23] are examples of such restoration tasks, where the degradation models are typically linear and known [FLP*23]. In HDR imaging tasks, the degradation model is more complex, and existing solutions based on DDPMs are more sparse. Wang et al. [WYY*23] propose low-light image enhancement using exposure diffusion that is directly initialized with the noisy low-light image instead of Gaussian noise, which greatly simplifies denoising and consequently reduces the network complexity and the required number of inference steps. The method

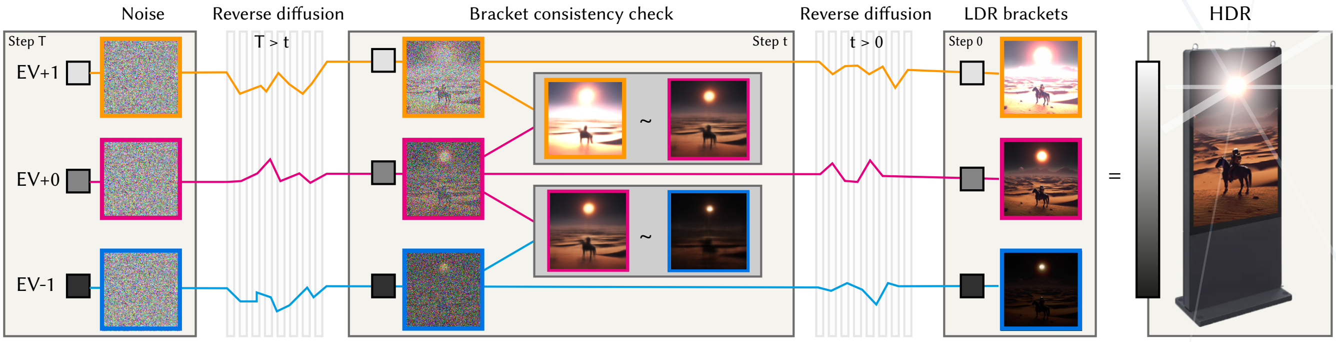

Figure 3: *Overview of our approach. Diffusion occurs from left to right and across multiple exposure levels (brackets), shown vertically. We show an example with three brackets. The process starts with three independent noises. At each diffusion step (one is shown), denoising is guided by an brackets consistency term (middle block). In this term, first, a denoised estimate of the current noisy images is computed (Eq. 3), then brackets are made consistent when re-exposed (∼ symbol) using Eq. 4 and Eq. 5. When diffusion has finished, the brackets form an HDR image under a common HDR fusion technique.*

can be trained using pairs of low-light and normally-exposed photographs, as well as synthetic data using different noise models. Fei et al. [FLP*23] employ a pre-trained DDPM and propose the Generative Diffusion Prior (GDP) for unsupervised modeling of the natural image posterior distribution. They demonstrate the utility of this framework for low-light image enhancement and HDR image reconstruction by merging low, medium, and high exposures. A similar task, but with explicit emphasis on large motion between the three exposures and severe clamping at the same time, is addressed in Yan et al. [YHS*23]. Lyu et al. [LTH*23] train a DDPM to capture the distribution of natural HDR environment maps, but are limited to rather narrow classes (e.g., urban streets) due to scarcity of available HDR training data. Dalal et al. [DVSR23] train a DDPM on LDR–HDR image pairs (roughly 2,000 images, from the HDR-Real [LLC*20] and HDR-Eye [NKHE15] datasets) and reconstruct HDR images from single LDR images.

Our work follows Chung et al. [CKM*22] and relies on off-the-shelf pre-trained diffusion models [DN21, NDR*21] that feature better domain generalizability due to intensive training on large datasets than explicit training on small datasets of LDR–HDR image pairs [DVSR23, LTH*23]. Our solution does not require any HDR images at the training stage. Instead, we implicitly leverage the exposure statistics of real-world photographs used for DDPM training, which allows the model to reason on the underlying radiance distributions. In single-image reconstruction, we require as the input just one LDR exposure and then generate a stack of different spatially consistent LDR exposure brackets. This way, we avoid possible problems with large motion inherent for time-sequential capturing [FLP*23, YHS*23].

Optionally, the hallucinated HDR content in saturated regions can be conditioned on text prompts [NDR*21]. Such text prompts can also be used as the only input to generate standalone HDR images. Histograms with the desired pixel color distribution, possibly derived from existing images, can guide global contrast relations in generated HDR content and can optionally be combined with text prompts. Tab. 1 summarizes all text conditioning and image/histogram guidance combinations we explore. With respect to non-diffusion methods such as GlowGAN [WSP*23a], we benefit from an overall better quality of generated images by diffusion models [DN21, NDR*21] and avoid a lossy inversion of an input LDR exposure into a latent code as required by GANs.

Our approach also differs from existing methods that enforce consistency between multiple joint diffusion instances to create seamless high-resolution panoramas by blending colors, features [BTYLD23, Jim23], maintaining style and content [LKKS23], or ensuring semantic coherence [QPCC24]. In contrast, our work focuses on bracket consistency requirements specifically for HDR reconstruction. In Fig. 12, we demonstrate how HDR-specific conditions can also be combined with panorama stitching consistency.

4. Our Approach

We will first briefly recall the mechanics of sample generation using DDPMs with a guiding term (Sec. 4.1), before presenting our idea (Sec. 4.2).

4.1. Guided Diffusion

Data generation with a pre-trained DDPM [HJA20, SDWMG15] amounts to gradual denoising of a sample $\mathbf{x} \in \mathbb{R}^u$ using

$$\mathbf{x}_{t-1} := \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - (1-\alpha_t)\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\Big) + \mathbf{z}_t. \quad (1)$$

This update rule involves a noise schedule $\alpha_t \in \mathbb{R}_+$, random vectors $\mathbf{z}_t \in \mathbb{R}^u$, and, at its core, a score function $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$. Optionally, the score can be conditioned on a signal $\mathbf{c} \in \mathbb{R}^v$, such as a text prompt embedding, to yield $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{c})$. In modern DDPMs, scores are typically approximated by a neural network $\mathbf{s}_\theta(\mathbf{x}_t, \mathbf{c}, t) \in (\mathbb{R}^u \times \mathbb{R}^v \times \mathbb{Z}) \to \mathbb{R}^u$. Please refer to Yang et al. [YZS*23] for an in-depth treatise.

In the framework of diffusion posterior sampling (DPS) [CKM*22], an additional guiding signal $\mathbf{y} \in \mathbb{R}^w$, such as a partial observation of $\mathbf{x}$, is incorporated into the denoising process to arrive at the posterior score

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{c},\mathbf{y}) \approx \mathbf{s}_\theta(\mathbf{x}_t,\mathbf{c},t) - \lambda\nabla_{\mathbf{x}_t} C(\hat{\mathbf{x}}_t,\mathbf{y}). \quad (2)$$

Here, $C \in (\mathbb{R}^u \times \mathbb{R}^w) \to \mathbb{R}$ is a problem-specific measurement term that drives the denoising process towards solutions that incorporate the guiding signal $\mathbf{y}$, and $\lambda \in \mathbb{R}_+$ is a balancing term. For increased stability, Chung et al. [CKM*22] propose to feed the current estimate of the clean sample

$$\hat{\mathbf{x}}_t = \frac{1}{\sqrt{\bar{\alpha}_t}}\Big(\mathbf{x}_t + (1-\bar{\alpha}_t)\mathbf{s}_\theta(\mathbf{x}_t, \mathbf{c}, t)\Big) \tag{3}$$

to C, where $\bar{\alpha}_t$ is derived from α_t.

4.2. Exposure diffusion

The above equations Eq. 1 and Eq. 2 are valid for producing a single LDR result image $\mathbf{x}$. Our idea is to produce HDR by diffusing multiple LDR results. Hence, we operate (Fig. 3) on a set of LDR images $\{\mathbf{x}^{-m}, \dots, \mathbf{x}^0, \dots, \mathbf{x}^n\}$, called "brackets". Positive and negative superscripts denote positive and negative EVs, respectively. All brackets are initialized to noise with mean zero and standard deviation one. They, further, need to be gamma-corrected sRGB LDR images, as we consider the score function a black box that cannot be retrained to work on linear HDR.

Score term The first term in Eq. 2 is the common score function that points from the current solution into the direction of a more plausible one. It may or may not be conditioned as per the second column of Tab. 1, leading to different application scenarios. It is a black box we do not need to know any details of, nor differentiate, as it already encodes a gradient. We only need to know its noise schedule α_t to also use $\hat{\mathbf{x}}$ from Eq. 3. The score function is hence simply computed on each bracket independently.

Posterior term The second term in Eq. 2 is very specific to our problem, the bracket consistency term. The consistency of two brackets measures how much $\hat{\mathbf{x}}^i$, a free variable, is compatible with another bracket $\hat{\mathbf{x}}^r$ that is assumed fixed. For each bracket $\hat{\mathbf{x}}^i$, the reference bracket $\hat{\mathbf{x}}^r$ is exposed to another bracket (that can both be higher or lower EV), and the resulting differences are checked using the function `braco`, defined as

$$\texttt{braco}(\hat{\mathbf{x}}^r \rightarrow \hat{\mathbf{x}}^i) := \mathit{CRF}_\gamma\left(\min\!\left(\frac{\alpha^i}{\alpha^r} \odot \mathit{CRF}_\gamma^{-1}(\hat{\mathbf{x}}^r), 1\right)\right) - \hat{\mathbf{x}}^i,$$

where $\mathit{CRF}_\gamma(x) = x^\gamma$ with $\gamma = \frac{1}{2.2}$ represents the camera response function, and its inverse is given by $\mathit{CRF}_\gamma^{-1}(x) = x^{1/\gamma}$. We first apply inverse CRF, as the solution exists in non-linear space for the black box score. Next, we scale by the ratio between the exposure times (α) and then clamp and apply CRF again to simulate the behavior of a real camera.

Since negative EVs primarily involve hallucinating saturated content and positive EVs focus on denoising, our posterior term behaves slightly differently for positive, negative, and zero EV brackets. The posterior for decreasing exposure (negative EVs) is

$$\begin{aligned} C_\downarrow(\hat{\mathbf{x}}^i, \hat{\mathbf{x}}^r) = &\,||\texttt{sat}(\hat{\mathbf{x}}^r)\cdot\max(\texttt{braco}(\hat{\mathbf{x}}^r \rightarrow \hat{\mathbf{x}}^i), 0)||_2 + \\ &\lambda_\mathrm{s}\cdot||(1-\texttt{sat}(\hat{\mathbf{x}}^r))\cdot(\texttt{braco}(\hat{\mathbf{x}}^r \rightarrow \hat{\mathbf{x}}^i))||_2, \end{aligned} \tag{4}$$

while the one to increase exposure (positive EVs) is

$$\begin{aligned} C_\uparrow(\hat{\mathbf{x}}^i, \hat{\mathbf{x}}^r) = &\,||\texttt{dark}(\hat{\mathbf{x}}^r)\cdot(\texttt{braco}(\hat{\mathbf{x}}^r \rightarrow \hat{\mathbf{x}}^i))||_2 + \\ &\lambda_\mathrm{d}\cdot||(1-\texttt{dark}(\hat{\mathbf{x}}^r))\cdot(\texttt{braco}(\hat{\mathbf{x}}^r \rightarrow \hat{\mathbf{x}}^i))||_2, \end{aligned} \tag{5}$$

where λ_s and λ_d are the balancing weights. The `sat` and `dark` are the mask functions for saturated and near-zero pixels, respectively, and zero otherwise. However, in practice, we use linear functions $\texttt{sat}(\mathbf{x}) = \mathbf{x}$ and $\texttt{dark}(\mathbf{x}) = 1 - \mathbf{x}$ instead of conventional binary masking [KR17] to make our cost functions smooth and tractable. The possible combinations of consistency and up or down direction are discussed with an example in Fig. 4.

The max operation in Eq. 4 is responsible for generating plausible content in saturated areas. To clarify its role, consider $\hat{\mathbf{x}}^i$ as the optimized EV-1 bracket for $\hat{\mathbf{x}}^r$. In regions where $\hat{\mathbf{x}}^r$ is saturated (e.g., the blue dots in the top row of Fig. 4), there is a feasible range of values that $\hat{\mathbf{x}}^i$ can take, such that when exposed to $\hat{\mathbf{x}}^r$, they are clamped to 1. For the EV-1 case, this range is from 0.5 to 1. This constraint is enforced by the term $\texttt{sat}(\hat{\mathbf{x}}^r)\cdot\max(\hat{\mathbf{x}}^r/2 - \hat{\mathbf{x}}^i, 0)$ (assuming an identity CRF in this didactical example). The max term encourages the optimized bracket $\hat{\mathbf{x}}^i$ to be any value above $\hat{\mathbf{x}}^r/2$. Consequently, $\hat{\mathbf{x}}^r/2 - \hat{\mathbf{x}}^i$ becomes negative, resulting in a zero cost.

The weighting factor λ_s in Eq. 4 is set to 1; however, in Eq. 5, we weigh the two terms differently, with $\lambda_\mathrm{d} = 2$, to account for the noise removal effect. The darker regions (e.g., the red dots in the bottom row of Fig. 4) are often noisy or less reliable, so we apply a smaller coefficient to impose less data term prior in these areas compared to brighter regions (i.e., $1.0 - \texttt{dark}(\hat{\mathbf{x}}^r)$).

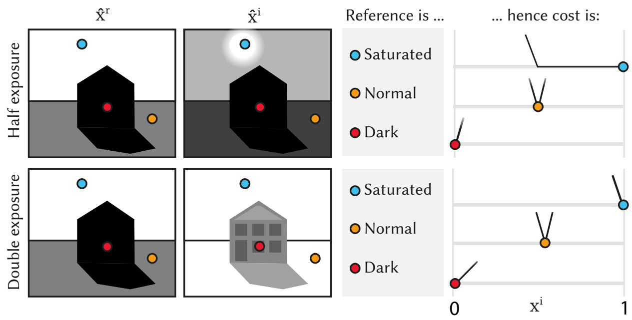

Figure 4: *Posterior based on bracket consistency cost for optimizing lower exposure (top row) and higher exposure (bottom row). The horizontal axis in the cost plot represents the pixel values in the current solution $\hat{\mathbf{x}}^i$, and dots are placed where their value in the reference $\hat{\mathbf{x}}^r$ is. The vertical axis shows the cost values, with horizontal lines representing zero cost. Depending on the exposure direction, this results in different costs for choices in $\hat{\mathbf{x}}^i$. When going down in exposure (top row), for the saturated region, we allow $\hat{\mathbf{x}}^i$ to take any value within a feasible range, such that when exposed to $\hat{\mathbf{x}}^r$, they will be clamped to 1. For higher exposure (bottom row), the consistency term is relaxed (indicated by a lower steepness of the penalty cost) for dark areas compared to other regions.*

Finally, we can also define an optional posterior term on the original image by applying a function f:

$$C_0(\hat{\mathbf{x}}^i, \mathbf{y}) = \lambda_c \cdot ||f(\hat{\mathbf{x}}^i) - \mathbf{y}||_2. \tag{6}$$

First, if f is, for example, the identity, and $\mathbf{y}$ an LDR image (the third column in Tab. 1), this becomes a reconstruction task. In that case, the solution for $\hat{\mathbf{x}}^i$ is immediately set to $\mathbf{y}$. As a second alternative,

Table 1: *Our method supports various applications through different combinations of score conditioning (text or null) and guidance (image, histogram, or none). For reconstruction tasks, the EV+0 is fixed to the input LDR image. The final column specifies the diffusion backbone used. Please note our approach is model-agnostic, meaning it can be adapted to different diffusion models based on the application. For instance, we utilize GLIDE's conditional model [NDR*21] for text-conditioned experiments and Stable Diffusion [RBL*22] for generating high-resolution samples.*

Application	Cond. **c**	Guide **y**	EV+0 fix?	Example	Backbone model
Generation	Text	—	×	Fig. 5, 13	[NDR*21, RBL*22]
Generation	—	Histo.	×	Fig. 6	[NDR*21]
Generation	Text	Histo.	×	Fig. 7	[NDR*21]
Recons.	—	Image	✓	Fig. 8, 9, 10, 12	[DN21]
Recons.	Text	Image	✓	Fig. 11	[NDR*21]

we explore using conversion to an LDR histogram as f. In this case, the parameter λ_c is set to 10.

Combining all together, we arrive at our final cost C:

$$C(\hat{\mathbf{x}}^i, \mathbf{y}) = \begin{cases} C_\downarrow(\hat{\mathbf{x}}^i, \hat{\mathbf{x}}^{i+1}) & \text{, if } i < 0 \text{, see Eq. 4,} \\ C_\uparrow(\hat{\mathbf{x}}^i, \hat{\mathbf{x}}^{i-1}) & \text{, if } i > 0 \text{, see Eq. 5 and} \\ C_0(\hat{\mathbf{x}}^i, \mathbf{y}) & \text{, if } i = 0 \text{, see Eq. 6.} \end{cases} \quad (7)$$

Eq. 7 is the expression for a single exposure bracket $\hat{\mathbf{x}}^i$. As per Eq. 2, this expression gets differentiated with respect to its first argument. The subtlety is that this is now done for multiple brackets, but they depend on each other. In our implementation, during one optimization step, however, for each bracket, the other bracket $\hat{\mathbf{x}}^r$ is considered a constant, so the second argument of $C_\downarrow$, $C_\uparrow$, and C_0 is "detached" in PyTorch parlance. Note that this is different from greedily optimizing each bracket sequentially.

5. Results

We begin by describing our experimental setup in Sec. 5.1. We then showcase the application of our method to HDR generation (Sec. 5.2) and reconstruction (Sec. 5.3), providing quantitative as well as qualitative results for both tasks.

5.1. Experimental setup

For our reconstruction experiments, specifically the LDR2HDR task, we utilize the pre-trained image-domain unconditional diffusion model of Dhariwal et al. [DN21]. Our input images are down-sampled to 256×256 before they are fed to this model, and we perform T=1,000 denoising steps to produce our results. In tasks involving text-conditioning or histogram guidance, we use the OpenAI GLIDE [NDR*21] diffusion model, which is text-conditional and generates images at a resolution of 64×64 using a classifier-free guidance strategy. Subsequently, an upsampling diffusion model is applied to increase the resolution to 256×256. In this case, we apply our DPS approach only to the text-conditional model and perform T=500 steps to produce the results. Once the exposure brackets are generated, they are individually upsampled using GLIDE's pre-trained upsampling module.

The hyper-parameter λ in Eq. 2 balances between the diffusion prior and our posterior term. It is worth noting that saturated regions are also included in our posterior term (Eq. 4), and since λ determines the weight of this term, its value directly affects the hallucinated content. We set $\lambda = 1.5$ when employing the conditional diffusion model [NDR*21]. However, in our experiments with the unconditional diffusion model [DN21], we observe that a constant λ sometimes leads to unrealistic hallucinations for saturated regions, as shown in Fig. 8. To achieve more consistent hallucinations, we adopt a time-dependent weight $\lambda = \lambda_0 \cdot (1 - t/T)^2$ with $\lambda_0 = 6$. Intuitively, each bracket is initialized randomly at the beginning, making it difficult for the data consistency term to provide the correct gradient. Therefore, we reduce its influence at the beginning ($t = T$) and gradually increase it as the denoising progresses.

For all results, we compute five exposure brackets: EV-4, EV-2, EV+0, EV+2, and EV+4, unless otherwise specified. These exposure brackets are merged using the standard technique [DM97] to create our HDR image. For Fig. 9, 11, and 10, we show the result by applying the tonemapping of Mantiuk et al. [MMS06] while in all other results, we directly show the optimized brackets. We release our code and provide the results in an HDR format on our webpage: https://bracketdiffusion.mpi-inf.mpg.de/

Figure 5: *Text-based HDR generation. Text prompts are on the left, alongside low (EV-4), medium (EV+0), and high exposures (EV+4).*

5.2. Generation

Image generation is a premiere ability of diffusion models, which we extend to HDR. Image generation without any conditioning or guidance frequently results in scenes that, in reality, do not exhibit high dynamic ranges. Therefore, capitalizing on the generality of our framework, we consider generation conditioned on text prompts, guided by RGB color histograms, and a combination thereof (first three rows in Tab. 1).

Text-based Here, we consider the task of text-conditioned generation, where the score function takes a conditioning signal **c** in the form of a text embedding. We omit C_0, i.e., the generation is free to synthesize any consistent brackets following the text prompt. Results of this application are shown in Fig. 5. The low exposures present detailed depictions of visible light sources, such as the structure of candle flames, including glares typically found around strong light sources. In the daylight scenes, most of the details are properly exposed for the medium exposure (EV+0), while in the night scenes, a high exposure (EV+4) is required to see sufficient detail.

Figure 6: *Histogram-based HDR generation. The first column shows the input image and its histogram. The other columns show our generated brackets. Note that the method never sees the input image (left), only its histogram.*

Histogram-based Here, we explore guided generation using a target histogram. In our experiments, we first compute an LDR histogram with 10 bins per color channel of an input image as our guiding signal **y** (Fig. 6, first column). Then, we utilize C_0 to direct the generation process towards producing an EV+0 bracket that matches this histogram (Fig. 6, third column), using a differentiable histogram function with soft bin assignments as f. Our framework produces consistent brackets of HDR content (Fig. 6, second to fourth column).

Text & histogram-based In Fig. 7, we combine the control modalities of the previous two paragraphs. In the first three rows, we apply constraints where 50%, 25%, and 1% of saturated pixels are enforced on the histograms of the EV+0 bracket, all while utilizing the same text prompt. We observe that our approach enables the generation of different HDR contents that faithfully reflect the queries. In the last row, a guiding histogram is extracted from an input image.

5.3. Reconstruction

We now turn to one of the supreme disciplines of HDR imaging: LDR2HDR restoration. There are two major challenges involved in this task. Firstly, we need to fill the saturated (white) regions in the LDR image **y** with appropriate content. Secondly, dark regions in **y** often contain strong noise that needs to be removed. Our approach

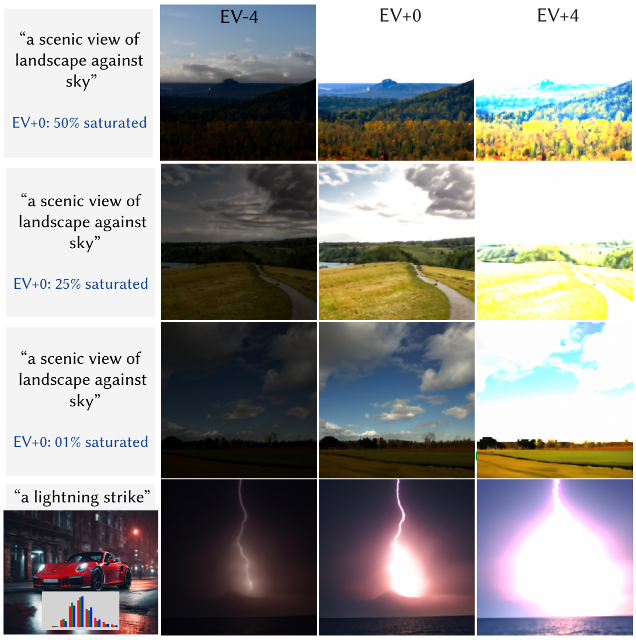

Figure 7: *Text- and histogram-based HDR generation. The first column is the query, and the other three columns are our results. Additional results are provided in our supplementary.*

Figure 8: *The effect of different λ setting in Eq. 2 on the LDR2HDR task. The reconstructed (tone-mapped) HDR results are shown on the right for a given input LDR image (left). A constant λ value often leads to reconstructions with artifacts, whereas our proposed time-dependent setting, $\lambda = \lambda_0 \cdot (1 - t/T)^2$ (See Sec. 5.1), produces significantly better results.*

naturally supports this task by setting f in Eq. 6 to be the identity function. We demonstrate both unconditional and text-conditioned reconstruction (last two rows in Tab. 1).

Methods and dataset We compare our approach for the LDR2HDR task with **CERV** [CYL*23] and **GlowGAN** [WSP*23a], which are recent state-of-the-art methods. Additionally, we evaluate against two other top-performing methods, **MaskHDR** [SRK20] and **HDRCNN** [EKD*17], as identified in recent studies [BMBRD24, WSP*23a]. Note that the only other generative approach, **GlowGAN**, requires

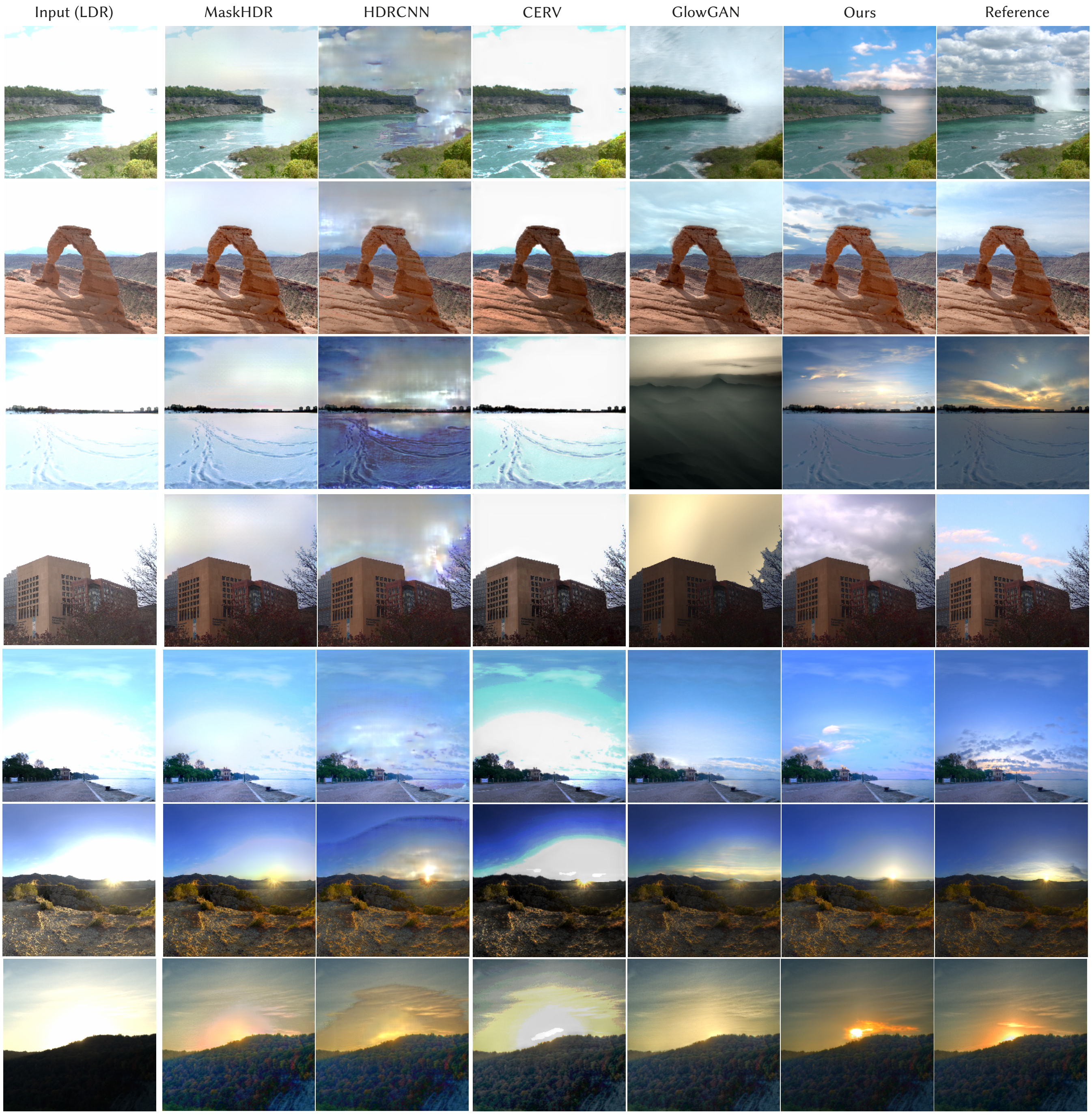

Figure 9: *LDR2HDR reconstruction for our method and competitors given an input LDR images (first column). All HDR images (right columns) are tone-mapped using the same tone-mapper, whose parameters are tuned for each row to achieve the best visual appearance of the corresponding reference HDR image.*

training a domain-specific model. Thus, for a fair comparison, we limit our evaluation to landscape images, as a pre-trained **GlowGAN** model is available for this category. Specifically, we curate a dataset comprising 75 HDR images sourced from various online platforms, which will be made available on publication.

Metrics We employ four different metrics to assess restoration performance. Firstly, we employ the full-reference metric HDR-VDP-3 [MHH23], which evaluates reconstruction fidelity without considering that saturated regions in an LDR image may allow for multiple, different HDR solutions. Secondly, to gauge overall plausibility, we utilize the no-reference HDR image metric PU21-

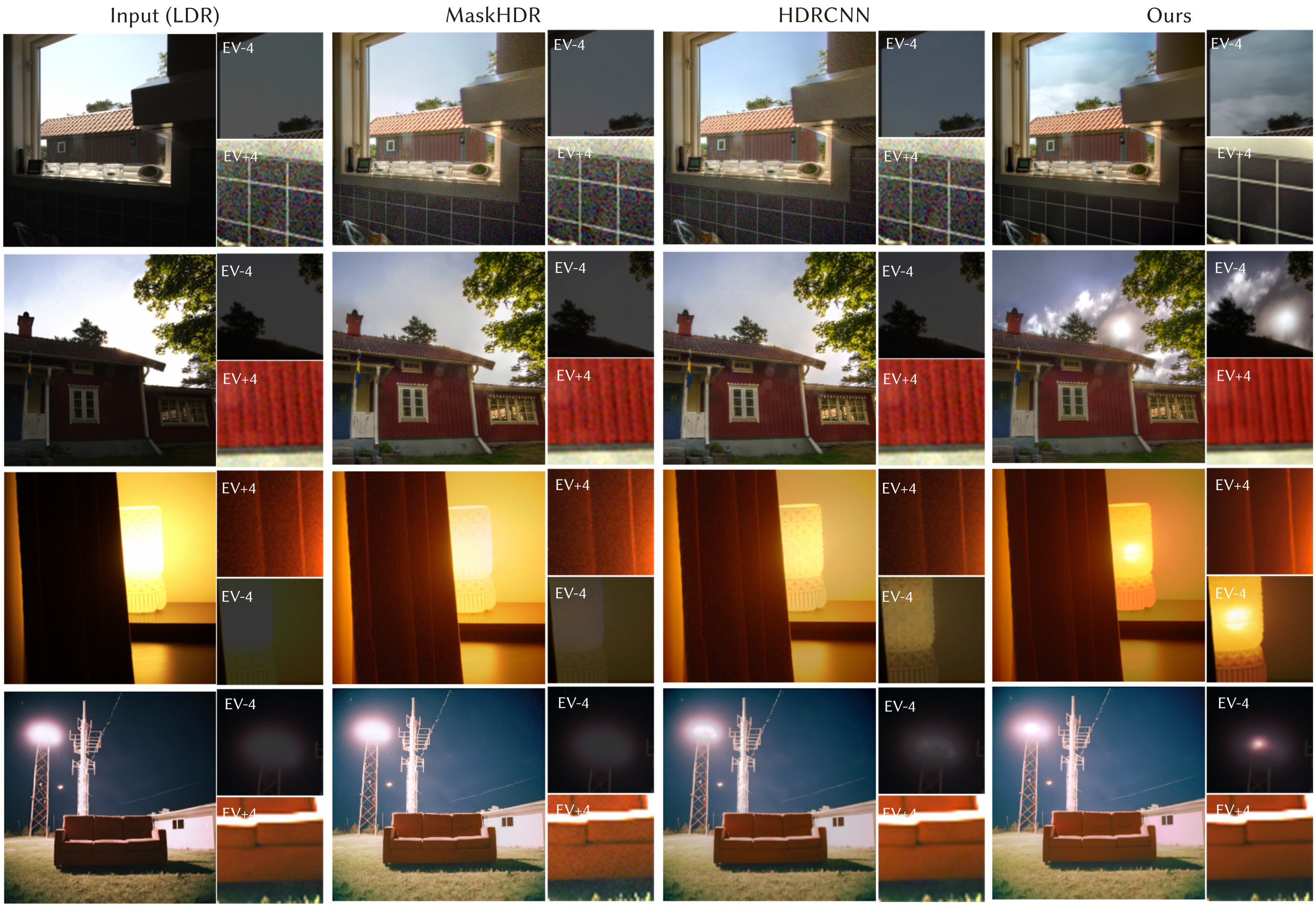

Figure 10: *LDR2HDR reconstruction for MaskHDR, HDRCNN, and Ours methods guided by the input LDR images (left column). Insets show dark, and hence noisy, as well as bright, partially saturated input regions. Other methods can remove some noise, but ours not only gets the semantics right in saturated areas (e.g., for the lamp or sun), but also removes noise in dark areas. The images in the first three rows are examples from the SI-HDR dataset [HME*22], while the input image in the last row is an AI-generated image with Stable-Diffusion.*

PIQE [HME*22]. This metric, however, is agnostic of the expected distribution of hallucinated contents in our narrow domain.

To address these considerations, we also employ two additional metrics: DreamSim [FTS*23] and FID [HRU*17]. DreamSim evaluates high-level visual similarities and differences between image pairs, providing insights into perceptual alignment. Meanwhile, the FID score, widely used in generative settings, measures discrepancies between distributions of generated and reference images, serving as a reliable measure of generative quality. However, since FID relies on a vision model [KSH12] pre-trained on LDR images, it cannot be directly applied to HDR content. Rather, we seek to produce a representative distribution of LDR images derived from the HDR content, accounting for uncalibrated and unnormalized pixel values across methods. We opt to apply the auto-exposure method by Shim et al. [SLK14] to each HDR image. This technique helps determine the EV0 bracket, from which we derive EV±2 and EV±4 brackets. Subsequently, we select 100 random 64×64-pixel crops from each image. We maintain consistency in selecting crop locations across methods [CGS*22]. This precaution is necessary because having small bright light sources, such as the sun, in some patches in one method but not in another could disproportionately bias the measurement. Our protocol leads to stable estimates based on 7.5k patches per bracket and 37.5k patches in total.

Table 2: *Reconstruction task performance. The first and second best-performing methods are highlighted in bold and underlined, respectively.* `Ours`† *refers to a version of our method with a more complex camera response function (see Sec. 6).*

	FID↓								
Method	EV-4	EV-2	EV+0	EV+2	EV+4	All.	DreamSim↓	No-Ref.↓	Full-Ref.↑
`MaskHDR`	14.36	09.44	**04.13**	**01.14**	**02.81**	03.63	<u>0.053</u>	51.7 ± 7.5	05.87 ± 1.6
`HDRCNN`	14.54	16.89	13.06	03.73	03.27	06.54	0.082	<u>47.2 ± 7.1</u>	**06.67 ± 1.2**
`CERV`	21.83	16.63	10.04	08.29	16.22	08.00	0.129	75.1 ± 9.6	05.14 ± 1.5
`GlowGAN`	<u>08.59</u>	<u>06.94</u>	05.32	03.61	08.09	<u>03.08</u>	0.078	**45.5 ± 8.6**	<u>06.57 ± 1.5</u>
`Ours`†	10.13	09.45	06.43	03.23	06.63	03.41	0.081	50.8 ± 8.1	06.46 ± 1.3
`Ours`	**06.25**	**06.48**	<u>04.65</u>	<u>01.28</u>	<u>02.89</u>	**02.05**	**0.048**	51.7 ± 7.6	06.51 ± 1.2

Results Our quantitative evaluation results are presented in Tab. 2. We observe that our approach outperforms the baselines in terms of overall FID (denoted as "All") and excels in the challenging cases of negative EV where content needs to be hallucinated. Addition-

Table 3: *Performance comparison in terms of runtime and GPU memory usage using a single NVIDIA Quadro RTX 8000 GPU for a 256×256 resolution input.*

Method	Runtime	Memory(GB)
`HDRCNN`	0.03 s	2.5
`MaskHDR`	0.53 s	0.5
`CERV`	0.32 s	0.2
`GlowGAN`	15 min	8.0
`Ours` w/ [DN21]	22 min	23.0
`Ours` w/ [NDR*21]	2 min	9.3

ally, our method achieves the best performance across all baselines when evaluated using the DreamSim metric. Results for the other two metrics remain inconclusive due to statistical insignificance. Note that the full-reference metrics (included here only to follow the previous practice) favor blurriness in hallucinated content and poorly evaluate its naturalness. FID, a standard metric for generative methods, clearly shows that our solution consistently outperforms all other approaches.

In Fig. 9, we show corresponding qualitative results with a focus on saturated regions; complete sets of images are provided in the supplemental. Our approach consistently generates arguably the highest-quality hallucinations in saturated regions. This is facilitated by the first term in Eq. 4, which gives the process the freedom to generate any content as long as it is bright enough. Notably, in the third row of Fig. 9, we present a particularly challenging case where one color channel is nearly entirely saturated across the image. In this instance, we observe how the baselines struggle to produce plausible content, even **`GlowGAN`**, which typically excels in generating realistic results due to its domain-specific generative capabilities. In the last two rows, we see that **`HDRCNN`** and **`MaskCNN`** struggle with image regions close to the sun, producing unnatural discontinuities and halo effects, respectively. **`CERV`** fails in almost all examples, which is not surprising given that the authors explicitly noted their method's inability to generate reasonable content in largely saturated regions. As anticipated, given the inherent ambiguities of the LDR2HDR restoration task, all methods, including ours, generate results that diverge from the reference.

Another challenging aspect of LDR2HDR reconstruction involves eliminating noise from regions that were initially very dark. A naïve scaling of the original image content leads to substantial noise, making these results practically unusable. In Fig. 10, we illustrate how our approach serves as an effective denoiser, yielding visually pleasing outcomes.

We also evaluate the runtime and GPU memory usage of our method against other baselines on a single NVIDIA Quadro RTX 8000 GPU for a 256×256 resolution input, with results presented in Tab. 3. The reported runtime for our method is based on generating five brackets. As expected, diffusion-based models are significantly slower than feed-forward methods. However, using modern GPUs like the NVIDIA Tesla A100 reduces the runtime for generating five brackets with Dhariwal et al. [DN21] model to approximately six minutes. Our approach also scales linearly with the number of brackets in terms of GPU memory usage. For example, using the GLIDE model [NDR*21], generating 3, 5, 7, and 9 brackets requires approximately 6.2, 9.3, 12.7, and 15.3 GB of GPU memory, respectively.

Text-based reconstruction Our framework offers a unique opportunity: the ability to dictate which content to hallucinate in saturated regions through text conditioning. This is demonstrated in Fig. 11, where, in addition to the guiding LDR signal $\mathbf{y}$, the user provides a text prompt conditioning signal $\mathbf{c}$. We see that this combination of control modalities enables precise HDR content generation. We emphasize that this task differs from typical inpainting in the LDR domain. Here, saturated pixel values are not replaced by darker ones but rather extended in dynamic range while forced to align with the LDR observation (Eq. 4).

Figure 11: *Text-based reconstruction. The LDR image on the left has ambiguous regions, e.g., the sky. The right three columns show what the sky could look like in a tone-mapped result on a reconstructed HDR. Each variant is conditioned on different text prompts shown on the top.*

6. Ablations

In this section, we analyze various aspects of our method, including the number of optimized brackets, the effect of the CRF model, the underlying pre-trained diffusion model, and different optimization strategies.

Number of brackets Our method is flexible with respect to the number of exposure brackets. We conduct two experiments to assess the impact of different numbers of brackets on output quality for the LDR2HDR task. In the first, we fix the dynamic range and vary the number of brackets, corresponding to different levels of overlap between exposures. In the second, we increase dynamic ranges while keeping the exposure ratio fixed. For both, we report FID and DreamSim scores. Additionally, to evaluate the effectiveness of our bracket consistency term, we compute the consistency between neighboring brackets by re-exposing all synthesized brackets to their neighboring ones using a process similar to our `braco` function and measuring the differences using the PSNR metric.

In the first experiment, we fix the exposure range from EV-4 to

Table 4: *Ablation study on the number of brackets used for LDR2HDR task. Here, we fix the exposure range and increase the overlap between the exposures. The final column reports the consistency between brackets using the PSNR metric.*

#EVs	FID↓ (All.)	DreamSim↓	Consist.↑ (dB)
3	03.09	0.063	39.1
5	02.05	0.048	39.4
7	03.36	0.055	37.4

Table 5: *Ablation study on the number of brackets used for LDR2HDR task. Here, we extend the dynamic ranges. Bracket consistency is measured in dB.*

	FID↓							
#EVs	EV-6	EV-4	EV-2	EV+0	EV+2	All.	DreamSim↓	Consist.↑
3	05.71	05.31	04.12	03.40	02.57	02.06	0.025	42.8
5	05.12	04.71	04.01	03.45	03.18	01.86	0.026	38.0
7	04.48	04.40	03.71	02.88	03.32	01.62	0.025	33.4

EV+4 and use 3, 5, and 7 brackets. The results are summarized in Tab. 4. Here, the FID score is measured using the same evaluation set as in Tab. 2. With only three exposures (EV-4, EV+0, EV+4), the optimization becomes more challenging due to inadequate sampling of the dynamic range. The best performance is achieved with five brackets, yielding the lowest FID (2.05) and DreamSim (0.048) scores, along with a bracket consistency of 39.4 dB. This level of consistency is comparable to the differences observed in high-quality JPEG compression, which is commonly used for HDR bracket fusion. However, increasing the number of brackets to seven does not improve HDR recovery. Our bracket consistency remains high overall; however, as the brackets are optimized recursively, with more brackets, consistency begins to decrease.

In the second experiment, we optimize for different dynamic ranges—EV-2 to EV+2, EV-4 to EV+4, and EV-6 to EV+6—with 3, 5, and 7 brackets and an EV-2 stop separation, respectively. In this experiment, we choose a subset of our evaluation set featuring an extremely high dynamic range (e.g., the presence of the sun). We report both per-exposure and overall FID scores in Tab. 5. We limit the results to exposures up to EV+2, as the outputs at EV+4 and EV+6 are nearly saturated. Overall, the findings indicate that increasing the number of brackets consistently enhances the recovery of higher dynamic ranges (e.g., EV-6). However, five brackets strike the best balance between computational efficiency and output quality, making it the practical choice for our method.

The effect of CRF The CRF maps raw sensor readings, which correspond to actual light intensity, to pixel values in the displayed image. In our experiments, we employ a commonly used CRF modeled as a simple gamma function, $CRF_\gamma(x) = x^\gamma$. Substituting this gamma function into the `braco` consistency expression (Sec. 4.2) yields:

$$\texttt{braco}(\hat{\mathbf{x}}^r \to \hat{\mathbf{x}}^i) := \left(\min(\frac{\alpha^i}{\alpha^r} \odot (\hat{\mathbf{x}}^r)^{1/\gamma}, 1)\right)^\gamma - \hat{\mathbf{x}}^i.$$

This expression can be further simplified to:

$$\texttt{braco}(\hat{\mathbf{x}}^r \to \hat{\mathbf{x}}^i) := \min((\frac{\alpha^i}{\alpha^r})^\gamma \odot \hat{\mathbf{x}}^r, 1) - \hat{\mathbf{x}}^i.$$

Here, we observe that the gamma function primarily scales the exposure ratio, leading to linearly scaled HDR values in the final output of our method. Since HDR reconstruction inherently suffers from a global scale ambiguity, this scaling does not pose a limitation. To further evaluate the impact of the CRF, we test a more complex model introduced by Eilertsen et al. [EKD*17], defined as:

$$CRF_{\beta,\gamma}(x) = \frac{(1+\beta)x^\gamma}{\beta + x^\gamma}, \quad (8)$$

where $\beta \sim \mathcal{N}(0.6, 0.1)$ and $\gamma \sim \mathcal{N}(0.9, 0.1)$ represent the distributions of the CRF parameters derived from the analysis of a large dataset of real-world images [WSP*23a]. We use the mean values of these parameters and re-run our method with this CRF model. The corresponding results, labeled as **`Ours`**† in Tab. 2, show no significant performance gains, suggesting that the simpler gamma model remains effective for our application.

Based on these findings, we argue that the choice of CRF does not significantly affect the performance of our method.

Figure 12: *Panoramic HDR generation at a 256×640 resolution given an AI-generated LDR image (middle row): To generate a panoramic image, we follow the diffusion composition technique from [Jim23] and simultaneously denoise three tiles of 256×256 resolution, each with a 64-pixel overlap, to ensure smooth transitions between them. The image-domain unconditional diffusion model [DN21] serves as our base model for this process.*

Extension to latent diffusion models The results presented so far are generated using the best-performing image-domain diffusion models. Although image-domain models have limited resolution, in Fig. 12, we demonstrate that producing high resolutions with these models is still possible given enough computing time. However, to further enhance both the quality and resolution of image generation, we employ our DPS approach directly on latent diffusion models (LDMs) [RBL*22], following the methodology outlined by Rout et al. [RRD*24]. In this context, we perform posterior sampling in the latent space, and accordingly, our prior and posterior scores in Eq. 2 are modified to:

$$\nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t|\mathbf{c},\mathbf{y}) \approx \mathbf{s}^*_\theta(\mathbf{z}_t,\mathbf{c},t) - \lambda \nabla_{\mathbf{z}_t} C(\mathbf{D}(\hat{\mathbf{z}}_t),\mathbf{y}). \quad (9)$$

The rest of the equations, Eq. 4 and Eq. 5, remain unchanged. Here, $\mathbf{z}$ represents the latent code, $\mathbf{s}^*_\theta$ is the score function of a pre-trained LDM, and $\mathbf{D}$ is the latent decoder that translates the latent code $\mathbf{z}$ back into pixel space as $\mathbf{x} = \mathbf{D}(\mathbf{z})$. Note Rout et al. [RRD*24] also introduces a "gluing term" to penalize inconsistencies at mask boundaries; however, we did not find it necessary for our purposes. In this experiment, we again apply the time-dependent λ with $\lambda_0 = 2$ and perform T = 500 iterations to generate results. Fig. 13 illustrates some examples for text-based generation at a resolution of 512×512 using the pre-trained Stable Diffusion v-1.5 [RBL*22].

Figure 13: *Text-based HDR generation using the recent latent diffusion model [RBL*22] as the backbone. More examples are provided in our supplementary material.*

Alternative solution to DPS We further investigate the alternative choice of score distillation sampling (SDS) [PJBM22] for HDR generation. The SDS method naturally allows for direct reconstruction of an HDR signal. In this approach, the optimized image can be represented by either a 2D-pixel grid or a neural network (NN); however, we found the NN provides better results than a simple pixel grid. During each optimization step, the HDR image is randomly exposed with EV+x, where x is drawn from a normal distribution with a mean of zero and a standard deviation of four. We compute the SDS loss on the exposed images and update the parameters of the HDR image accordingly. The SDS loss guides the current estimate of the exposed images towards the manifold of natural images learned by the pre-trained diffusion model [RBL*22]. In Fig. 14, we present our best-effort results. While this simpler approach can generate HDR content, achieving natural results remains challenging.

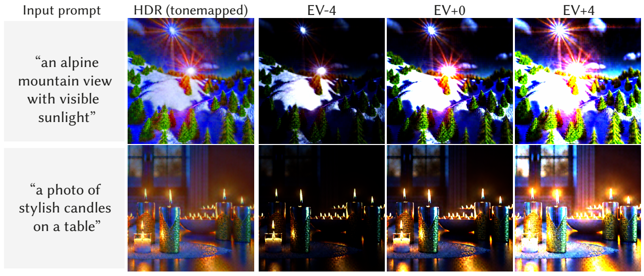

Figure 14: *HDR generation using SDS-based optimization [PJBM22]: the resulting images are HDR, but unfortunately not natural.*

7. Limitations

Inheriting the properties of diffusion models, our proposed approach is inherently slow, especially compared to feed-forward methods like HDRCNN and MaskHDR (Tab. 3). This limitation is further exacerbated in our framework, as we simultaneously denoise multiple brackets, making it slower than the original DPS. The DPS framework typically requires a large number of diffusion steps to converge, significantly contributing to the slower sampling speed. Incorporating advanced sampling strategies, such as those proposed by Song et al. [SVMK23] and Zhu et al. [ZZL*23], can help address this bottleneck. Another constraint is the GPU memory requirement, which limits the number of exposure brackets that can be processed.

8. Conclusion

We have suggested a novel method for generating HDR images using a black-box diffusion-based image generation model without the need for expensive retraining or fine-tuning. The key idea is to generate multiple LDR brackets in a synchronized and consistent manner. Our approach is simple to implement, intuitive, and capable of producing results with unprecedented quality in the highlight regions while effectively reducing noise in shadows. These capabilities have been validated through diverse applications of our method and comparisons with baseline techniques, demonstrating its effectiveness and versatility.

Extending our approach to HDR video can be an interesting direction for future work, particularly in scenarios where EV+0 exposure varies across frames due to auto-exposure adjustments. This introduces challenges such as ensuring temporal consistency across frames. Additionally, other frame-specific factors, including motion blur, defocus blur, depth-of-field blur, and varying noise characteristics, will likely necessitate modifications to the proposed consistency terms. A particularly challenging task would be reconstructing an all-in-focus HDR frame from an input LDR image impacted by these distortions. Building on the consistency terms proposed in this work, similar strategies could also be employed to generate focal or depth-of-field stacks.

References

[BADC17] Banterle F., Artusi A., Debattista K., Chalmers A.: *Advanced high dynamic range imaging*. AK Peters/CRC Press, 2017.

[BMBRD24] Banterle F., Marnerides D., Bashford-Rogers T., Debattista K.: Self-supervised high dynamic range imaging: What can be learned from a single 8-bit video? *ACM Transactions on Graphics 43*, 2 (2024), 1–16.

[BTYLD23] Bar-Tal O., Yariv L., Lipman Y., Dekel T.: Multidiffusion: fusing diffusion paths for controlled image generation. In *ICML* (2023).

[CBM*22] Cogalan U., Bemana M., Myszkowski K., Seidel H.-P., Ritschel T.: Learning hdr video reconstruction for dual-exposure sensors with temporally-alternating exposures. *Computers & Graphics 105* (2022), 57–72.

[CCXK18] Chen C., Chen Q., Xu J., Koltun V.: Learning to see in the dark. In *CVPR* (2018).

[CFXL20] Chang M., Feng H., Xu Z., Li Q.: Low-light image restoration with short- and long-exposure raw pairs, 2020.

[CGS*22] Chai L., Gharbi M., Shechtman E., Isola P., Zhang R.: Any-resolution training for high-resolution image synthesis. In *ECCV* (2022), pp. 170–188.

[CKM*22] Chung H., Kim J., Mccann M. T., Klasky M. L., Ye J. C.: Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687* (2022).

[CWL22] Chen Z., Wang G., Liu Z.: Text2light: Zero-shot text-driven HDR panorama generation. *ACM Trans. Graph. 41*, 6 (2022).

[CYL*23] Chen S.-K., Yen H.-L., Liu Y.-L., Chen M.-H., Hu H.-N., Peng W.-H., Lin Y.-Y.: Learning continuous exposure value representations for single-image hdr reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 12990–13000.

[Deb98] Debevec P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH* (1998), pp. 189–198.

[DM97] Debevec P. E., Malik J.: Recovering high dynamic range radiance maps from photographs. In *Proc. ACM SIGGRAPH* (1997), p. 369–378.

[DN21] Dhariwal P., Nichol A.: Diffusion models beat gans on image synthesis. In *NeurIPS* (2021).

[DVSR23] Dalal D., Vashishtha G., Singh P., Raman S.: Single image ldr to hdr conversion using conditional diffusion. In *ICIP* (2023), pp. 3533–3537.

[EKD*17] Eilertsen G., Kronander J., Denes G., Mantiuk R. K., Unger J.: HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph. (Proc SIGGRAPH) 36*, 6 (2017), 1–15.

[EKM17] Endo Y., Kanamori Y., Mitani J.: Deep reverse tone mapping. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 36*, 6 (2017).

[FLP*23] Fei B., Lyu Z., Pan L., Zhang J., Yang W., Luo T., Zhang B., Dai B.: Generative diffusion prior for unified image restoration and enhancement. In *CVPR* (2023), pp. 9935–9946.

[FTS*23] Fu S., Tamir N., Sundaram S., Chai L., Zhang R., Dekel T., Isola P.: Dreamsim: Learning new dimensions of human visual similarity using synthetic data. *arXiv preprint arXiv:2306.09344* (2023).

[HJA20] Ho J., Jain A., Abbeel P.: Denoising diffusion probabilistic models. *NeurIPS 33* (2020), 6840–6851.

[HME*22] Hanji P., Mantiuk R., Eilertsen G., Hajisharif S., Unger J.: Comparison of single image hdr reconstruction methods—the caveats of quality assessment. In *Proc. SIGGRAPH* (2022), pp. 1–8.

[HRU*17] Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NIPS 30* (2017).

[HZF*22] Huang X., Zhang Q., Feng Y., Li H., Wang X., Wang Q.: HDR-NeRF: High dynamic range neural radiance fields. In *CVPR* (2022), pp. 18398–18408.

[Jim23] Jiménez Á. B.: Mixture of diffusers for scene composition and high resolution image generation. *arXiv preprint arXiv:2302.02412* (2023).

[JLAK21] Jo S. Y., Lee S., Ahn N., Kang S.-J.: Deep arbitrary HDRI: Inverse tone mapping with controllable exposure changes. *IEEE Trans. Multimedia* (2021).

[JSYJYBO22] Jun-Seong K., Yu-Ji K., Ye-Bin M., Oh T.-H.: HDR-Plenoxels: Self-calibrating high dynamic range radiance fields. In *ECCV* (2022).

[KEES22] Kawar B., Elad M., Ermon S., Song J.: Denoising diffusion restoration models. In *NIPS* (2022).

[KR17] Kalantari N. K., Ramamoorthi R.: Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph. (Proc. SIGGRAPH) 36*, 4 (2017).

[KSH12] Krizhevsky A., Sutskever I., Hinton G. E.: Imagenet classification with deep convolutional neural networks. *NIPS 25* (2012).

[LAK18a] Lee S., An G. H., Kang S.-J.: Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access 6* (2018), 49913–49924.

[LAK18b] Lee S., An G. H., Kang S.-J.: Deep recursive HDRI: Inverse tone mapping using generative adversarial networks. In *ECCV* (2018), pp. 596–611.

[LDR*22] Lugmayr A., Danelljan M., Romero A., Yu F., Timofte R., Gool L. V.: Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR* (2022).

[LJAK20] Lee S., Jo S. Y., An G. H., Kang S.-J.: Learning to generate multi-exposure stacks with cycle consistency for high dynamic range imaging. *IEEE Trans. Multimedia 23* (2020), 2561–2574.

[LKKS23] Lee Y., Kim K., Kim H., Sung M.: Syncdiffusion: Coherent montage via synchronized joint diffusions. In *NIPS* (2023).

[LLC*20] Liu Y.-L., Lai W.-S., Chen Y.-S., Kao Y.-L., Yang M.-H., Chuang Y.-Y., Huang J.-B.: Single-image HDR reconstruction by learning to reverse the camera pipeline. In *CVPR* (2020), pp. 1651–1660.

[LTH*23] Lyu L., Tewari A., Habermann M., Saito S., Zollhöfer M., Leimkühler T., Theobalt C.: Diffusion posterior illumination for ambiguity-aware inverse rendering. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 42*, 6 (2023).

[LWW*22] Li R., Wang C., Wang J., Liu G., Zhang H.-Y., Zeng B., Liu S.: Uphdr-gan: Generative adversarial network for high dynamic range imaging with unpaired data. *IEEE Trans. Circuits and Systems for Video Technology 32*, 11 (2022), 7532–7546.

[LZH*24] Lei J., Zhu H., Huang X., Lin J., Zheng Y., Lu Y., Chen Z., Guo W.: Mini-led backlight: Advances and future perspectives. *Crystals 14*, 11 (2024), 922.

[MBRHD18] Marnerides D., Bashford-Rogers T., Hatchett J., Debattista K.: Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *Comp. Graph. Forum* (2018), vol. 37, pp. 37–49.

[MHH23] Mantiuk R. K., Hammou D., Hanji P.: HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv:2304.13625* (2023).

[MHMB*22] Mildenhall B., Hedman P., Martin-Brualla R., Srinivasan P. P., Barron J. T.: NeRF in the dark: High dynamic range view synthesis from noisy raw images. In *CVPR* (2022), pp. 16190–16199.

[MKM*20] Mustaniemi J., Kannala J., Matas J., Särkkä S., Heikkilä J.: LSD_2-joint denoising and deblurring of short and long exposure images with convolutional neural networks. In *BMVC* (2020).

[MMS06] Mantiuk R., Myszkowski K., Seidel H.-P.: A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept. 3*, 3 (2006), 286–308.

[MN99] Mitsunaga T., Nayar S. K.: Radiometric self calibration. *CVPR 1* (1999), 374–380.

[NDR*21] Nichol A., Dhariwal P., Ramesh A., Shyam P., Mishkin P., McGrew B., Sutskever I., Chen M.: Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741* (2021).

[NKHE15] Nemoto H., Korshunov P., Hanhart P., Ebrahimi T.: Visual attention in ldr and hdr images. In *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)* (2015).

[NWL*21] Niu Y., Wu J., Liu W., Guo W., Lau R. W.: HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions. *IEEE Trans. Image Processing 30* (2021), 3885–3896.

[PJBM22] Poole B., Jain A., Barron J. T., Mildenhall B.: Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988* (2022).

[QPCC24] Quattrini F., Pippi V., Cascianelli S., Cucchiara R.: Merging and splitting diffusion paths for semantically coherent panoramas, 2024.

[RBL*22] Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B.: High-resolution image synthesis with latent diffusion models. In *CVPR* (2022).

[RBS03] Robertson M. A., Borman S., Stevenson R. L.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *J Electronic Imaging 12*, 2 (2003), 219–228.

[RHD*10] Reinhard E., Heidrich W., Debevec P., Pattanaik S., Ward G., Myszkowski K.: *High dynamic range imaging: Acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[RKH*21] Radford A., Kim J. W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., et al.: Learning transferable visual models from natural language supervision. In *ICML* (2021), pp. 8748–8763.

[RRD*24] Rout L., Raoof N., Daras G., Caramanis C., Dimakis A., Shakkottai S.: Solving linear inverse problems provably via posterior sampling with latent diffusion models. *Advances in Neural Information Processing Systems 36* (2024).

[RWP*10] Reinhard E., Ward G., Pattanaik S., Debevec P., Heidrich W., Myszkowski K.: *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*, 2 ed. Elsevier (Morgan Kaufmann), Burlington, MA, 2010.

[SDWMG15] Sohl-Dickstein J., Weiss E., Maheswaranathan N., Ganguli S.: Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML* (2015), PMLR, pp. 2256–2265.

[SHC*23] Saharia C., Ho J., Chan W., Salimans T., Fleet D. J., Norouzi M.: Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell. 45*, 4 (2023), 4713–4726.

[SHS*04] Seetzen H., Heidrich W., Stuerzlinger W., Ward G., Whitehead L., Trentacoste M., Ghosh A., Vorozcovs A.: High dynamic range display systems. *ACM Trans. Graph. 23*, 3 (2004), 760–768.

[SIS11] Shimazu S., Iwai D., Sato K.: 3d high dynamic range display system. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality* (2011), IEEE, pp. 235–236.

[SLK14] Shim I., Lee J.-Y., Kweon I. S.: Auto-adjusting camera exposure for outdoor robotics using gradient information. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2014), pp. 1011–1017.

[SRK20] Santos M. S., Ren T. I., Kalantari N. K.: Single image HDR reconstruction using a cnn with masked features and perceptual loss. *ACM Trans. Graph. 39*, 4 (2020).

[SSDK*21] Song Y., Sohl-Dickstein J., Kingma D. P., Kumar A., Ermon S., Poole B.: Score-based generative modeling through stochastic differential equations. In *ICLR* (2021).

[SSG22] Sauer A., Schwarz K., Geiger A.: Stylegan-xl: Scaling StyleGAN to large diverse datasets. In *Proc. SIGGRAPH* (2022), pp. 1–10.

[SVMK23] Song J., Vahdat A., Mardani M., Kautz J.: Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations* (2023).

[WSP*23a] Wang C., Serrano A., Pan X., Chen B., Myszkowski K., Seidel H.-P., Theobalt C., Leimkühler T.: Glowgan: Unsupervised learning of hdr images from ldr images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 10509–10519.

[WSP*23b] Wang C., Serrano A., Pan X., Wolski K., Chen B., Seidel H.-P., Theobalt C., Myszkowski K., Leimkühler T.: A neural implicit representation for the image stack: Depth, all in focus, and high dynamic range. *ACM Trans. Grap. 42*, 6 (2023).

[WXTT18] Wu S., Xu J., Tai Y.-W., Tang C.-K.: Deep high dynamic range imaging with large foreground motions. In *ECCV* (2018).

[WY22] Wang L., Yoon K.-J.: Deep learning for hdr imaging: State-of-the-art and future trends. *IEEE Trans. Pattern Anal. Mach. Intell. 44* (2022), 8874–8895.

[WYY*23] Wang Y., Yu Y., Yang W., Guo L., Chau L.-P., Kot A. C., Wen B.: Exposurediffusion: Learning to expose for low-light image enhancement. In *ICCV* (2023).

[YGS*19] Yan Q., Gong D., Shi Q., van den Hengel A., Shen C., Reid I., Zhang Y.: Attention-guided network for ghost-free high dynamic range imaging. In *CVPR* (2019).

[YHS*23] Yan Q., Hu T., Sun Y., Tang H., Zhu Y., Dong W., Van Gool L., Zhang Y.: Towards high-quality hdr deghosting with conditional diffusion models. *IEEE Transactions on Circuits and Systems for Video Technology* (2023).

[YLL*21] Yu H., Liu W., Long C., Dong B., Zou Q., Xiao C.: Luminance attentive networks for HDR image and panorama reconstruction. In *Comp. Graph. Forum* (2021), vol. 40, pp. 181–192.

[YWL*20] Yan Q., Wang B., Li P., Li X., Zhang A., Shi Q., You Z., Zhu Y., Sun J., Zhang Y.: Ghost removal via channel attention in exposure fusion. *Computer Vision and Image Understanding 201* (2020), 103079.

[YZS*23] Yang L., Zhang Z., Song Y., Hong S., Xu R., Zhao Y., Zhang W., Cui B., Yang M.-H.: Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys 56*, 4 (2023), 1–39.

[ZA21] Zhang Y., Aydin T.: Deep HDR estimation with generative detail reconstruction. In *Comp. Graph. Forum* (2021), vol. 40, pp. 179–190.

[ZJY*21] Zhong F., Jindal A., Yöntem Ö., Hanji P., Watt S., Mantiuk R.: Reproducing reality with a high-dynamic-range multi-focal stereo display. *ACM Transactions on Graphics 40*, 6 (2021), 241.

[ZMW*20] Zhao Y., Matsuda N., Wang X., Zannoli M., Lanman D.: High dynamic range near-eye displays. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR)* (2020), vol. 11310, SPIE, pp. 268–279.

[ZZL*23] Zhu Y., Zhang K., Liang J., Cao J., Wen B., Timofte R., Van Gool L.: Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 1219–1229.