

High Dynamic Range Novel View Synthesis with Single Exposure

Kaixuan Zhang¹ Hu Wang² Minxian Li^{1,3} Mingwu Ren^{1,3} Mao Ye² Xiatian Zhu⁴

Abstract

High Dynamic Range Novel View Synthesis (HDR-NVS) aims to establish a 3D scene HDR model from Low Dynamic Range (LDR) imagery. Typically, multiple-exposure LDR images are employed to capture a wider range of brightness levels in a scene, as a single LDR image cannot represent both the brightest and darkest regions simultaneously. While effective, this multiple-exposure HDR-NVS approach has significant limitations, including susceptibility to motion artifacts (e.g., ghosting and blurring), high capture and storage costs. To overcome these challenges, we introduce, for the first time, the *single-exposure* HDR-NVS problem, where only single exposure LDR images are available during training. We further introduce a novel approach, Mono-HDR-3D, featuring two dedicated modules formulated by the LDR image formation principles, one for converting LDR colors to HDR counterparts and the other for transforming HDR images to LDR format so that unsupervised learning is enabled in a closed loop. Designed as a meta-algorithm, our approach can be seamlessly integrated with existing NVS models. Extensive experiments show that Mono-HDR-3D significantly outperforms previous methods. Source code is released at <https://github.com/prinasi/Mono-HDR-3D>.

1. Introduction

Compared to common Low Dynamic Range (LDR) imaging, High Dynamic Range (HDR) imaging enables the capture and representation of a broader range of luminance / brightness levels, thereby providing more realistic and visually

¹Nanjing University of Science and Technology. ²University of Electronic Science and Technology of China. ³State Key Laboratory of Intelligent Manufacturing of Advanced Construction Machinery. ⁴University of Surrey. Correspondence to: Minxian Li <minxianli@njust.edu.cn>.



Figure 1. Examples of (a, b) underexposure and (c, d) overexposure. Δt : Exposure time.

appealing representations of real-world scenes (Cai et al., 2024). It can encompass both the darkest shadows and the brightest highlights within a single frame. This capability is crucial in a number of fields such as creative media production, photography, virtual reality, and augmented reality that require precise color reproduction, detailed shadow and highlight information, and enhanced visual realism (Wang & Yoon, 2021). This enhanced dynamic range not only facilitates more realistic and visually appealing representations of complex scenes (Liu et al., 2023), but also improves the performance of various computer vision tasks, such as object recognition, scene segmentation, and depth estimation, by providing richer and more detailed visual information (Yan et al., 2023).

Novel View Synthesis (NVS) refers to the process of generating new views of a scene from arbitrary viewpoints, given a set of input images captured from different perspectives (Duan et al., 2024). This involves understanding and modeling the underlying 3D structure of a scene, as well as accurately rendering the appearance of the scene and ob-

jects. Most NVS works focus on LDR image models which fall short in those domains requiring HDR rendering.

Indeed, a couple of recent works (Cai et al., 2024; Huang et al., 2022) have studied the HDR-NVS problem by capturing multiple exposures of LDR images per view about the same scene. However, multiple exposure-based approaches remain vulnerable to motion artifacts, ghosting effects, and demand precise alignment of images captured under varying exposure settings (Eilertsen et al., 2017b). Specifically, longer exposure frames tend to accumulate object or camera movement, leading to blurred details—an issue that becomes more pronounced when exposure durations differ significantly (Kalantari et al., 2017). During HDR synthesis, pixel-wise fusion (e.g., weighted averaging) can superimpose differing object positions onto the same region, producing semi-transparent or duplicated contours that are especially evident in dynamic scenes with significant object displacement (Reinhard, 2020). Furthermore, variations in exposure times often yield discrepancies in brightness distribution, local contrast, and overall appearance, complicating conventional registration algorithms. Finally, in rapidly changing environments or when using mobile devices, capturing multiple exposures in quick succession may prove impractical, limiting the applicability of these methods.

To address these issues, we propose a more deployable yet more challenging task, namely *HDR-NVS with single-exposure LDR images*, which eliminates the reliance on multiple exposures and thus avoids the aforementioned limitations. However, as illustrated in Fig. 1, single-exposure images frequently suffer from overexposure or underexposure, posing significant challenges for HDR-NVS. Furthermore, we present a novel HDR 3D scene modeling framework, **Mono-HDR-3D**, characterized by learning to approximate the underlying LDR image formation process of camera imaging. Specifically, we start with learning an LDR 3D scene model from single-exposure LDR training images, followed by lifting the color space to HDR with a dedicated color transformation module in a per-channel manner. This from-LDR-to-HDR design is opposite to previous models since single-exposure LDR images are insufficient for deriving an HDR model. We further introduce a closed-loop design by augmenting a process of converting HDR images to LDR images, allowing additional supervision even in the case of no access to HDR ground-truth training data.

Our **contributions** can be summarized as follows. (I) We introduce a new HDR-NVS problem where only single-exposure LDR images are needed so that the data acquisition process is made significantly easier and generic, as well as eliminating those intrinsic limitations with multiple exposures. (II) We propose a generic framework, Mono-HDR-3D, that learns to capture the underlying camera imaging process for bridging LDR and HDR space effectively under

the challenging single exposure scenario. Designed as a generic approach, our method can be integrated with different 3D scene models such as NeRF (Mildenhall et al., 2021) or 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023). (III) Extensive experiments validate that our method significantly outperforms previous alternatives.

2. Related Work

High Dynamic Range Imaging Conventional HDR imaging primarily relies on specialized high-end cameras to capture HDR images (Tiwari & Rani, 2015). However, the high cost of these cameras renders them inaccessible to general consumers. An alternative approach involves reconstructing HDR images from imagery captured by LDR cameras using algorithms (Wang & Yoon, 2021). Before NeRF (Mildenhall et al., 2021) was proposed, two primary approaches have been extensively explored. The first generates HDR content by merging multiple LDR images of the same scene taken at varying exposure levels (Kalantari et al., 2017; Yan et al., 2020). However, the necessity of capturing LDR images with different exposures demands specific software and hardware capabilities, which is not only costly but also brings in various issues, as discussed earlier. Thus, the second focuses on synthesizing HDR imagery from single-exposure LDR images (Eilertsen et al., 2017a). Without the challenges associated with multi-exposure capture, it is more feasible for generating HDR images in scenarios where multiple exposures are impractical or datasets are limited (Hanji et al., 2022).

Recent deep learning based methods try to capture the mapping relationship between LDR and HDR images, often achieving state-of-the-art performances (Dille et al., 2025; Kim et al., 2024). However, these methods mainly focus on individual 2D imagery, lacking 3D perception capabilities and are not suitable for the novel view HDR image rendering problem.

Novel View Synthesis (NVS) is essential for applications such as virtual/augmented reality, gaming, and 3D reconstruction (Avidan & Shashua, 1997; Gao et al., 2023). Traditional methods, including Structure-from-Motion (SfM) (Schonberger & Frahm, 2016) and Multi-View Stereo (MVS) (Rosu & Behnke, 2022), rely on multi-view geometry to reconstruct 3D scenes but often struggle with occlusions, textureless regions, and high computational costs (Jiang, 2023). Recent advances in NVS leverage deep learning to learn continuous scene representations. Notably, NeRF (Mildenhall et al., 2021) encodes color and density in a neural network, enabling novel view synthesis by querying 3D coordinates and viewing directions. Extensions such as Mip-NeRF (Barron et al., 2021), FastNeRF (Garbin et al., 2021), and transformer-based models (Lin et al., 2023; Miyato et al., 2023) have improved efficiency, scalability, and

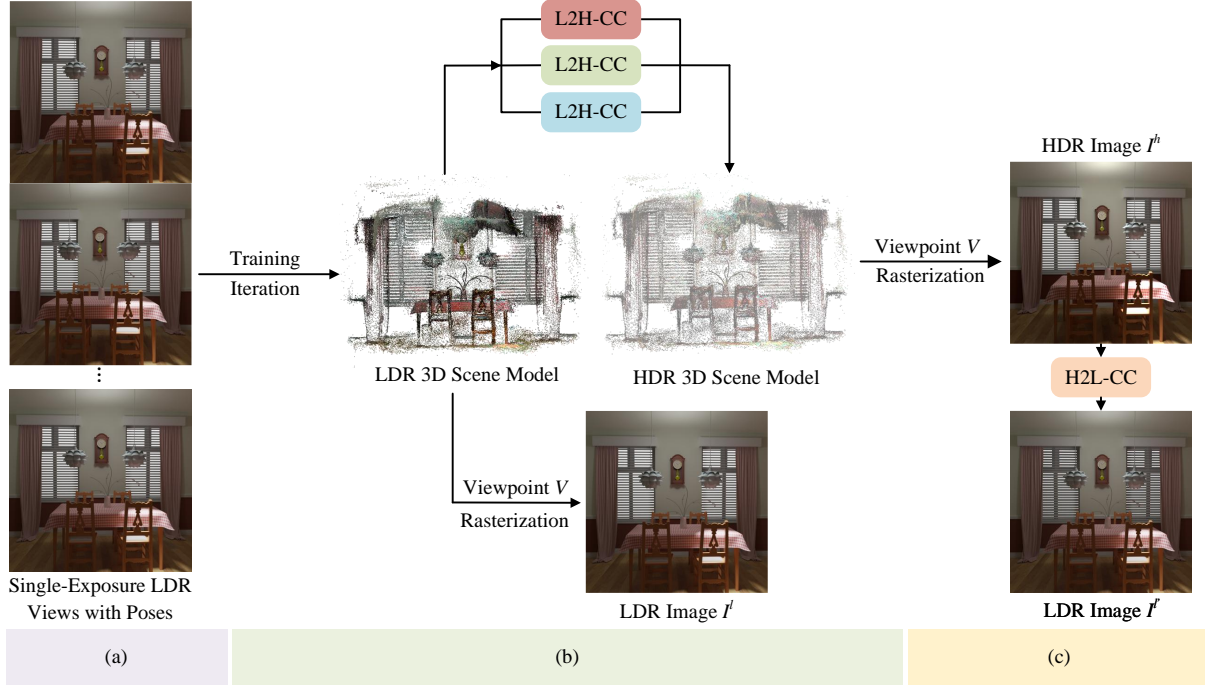


Figure 2. Overview of Mono-HDR-3D. (a) Given single exposure LDR training images with camera poses, we learn an LDR 3D scene model (e.g., NeRF or 3DGS). (b) Importantly, this LDR model is lifted up to an HDR counterpart via a camera imaging aware LDR-to-HDR Color Converter (L2H-CC). (c) Further, a closed-loop design is formed by converting HDR images back to LDR counterparts with a latent HDR-to-LDR Color Converter (H2L-CC). This enables optimizing the HDR model even with LDR training images, particularly useful in case of no access to HDR training data. During inference, only the HDR or LDR 3D scene model is needed, taking the novel camera view as the input and outputting the corresponding image rendering.

quality. However, most techniques focus on LDR outputs, limiting their applicability for HDR rendering. Alternatively, 3D Gaussian Splatting (3DGS) represents scenes using learnable 3D Gaussians optimized with multi-view supervision (Kerbl et al., 2023) which bypasses volumetric integration and heavy network optimizations, achieving faster training and inference, facilitating real-time rendering.

HDR Novel View Synthesis (HDR-NVS) aims to generate novel view HDR images from LDR observations, crucial for scenes with large brightness variations and rich details. Huang *et al.* proposed the first HDR-NVS model, HDR-NeRF (Huang et al., 2022), by extending the standard NeRF (Mildenhall et al., 2021) to learn an implicit mapping from physical radiance to HDR color. However, this method is costly in both model training and inference. Taking the advantage of 3DGS (Kerbl et al., 2023), Cai et al. (2024) addressed this issue by learning an MLP-based tone-mapper between LDR and HDR models. Despite their promising results, these methods rely on multiple-exposure LDR training imagery, limiting their applicability in cases with dynamic environments or limited image capturing conditions. To address this limitation, we introduce single-exposure HDR-NVS, which leverages only single-exposure LDR images.

3. Method

3.1. Problem formulation

For each of N distinct viewpoints $\mathbf{V} = \{V_1, V_2, \dots, V_N\}$, we capture a set of single-exposure LDR images denoted as $\mathbf{I}_{\mathbf{V}}^l = \{I_{V_1}^l, I_{V_2}^l, \dots, I_{V_N}^l\}$. The objective is to learn a 3D scene model \mathcal{F} that can synthesize an HDR image $\mathbf{I}_{V_{new}}^h$ for any given novel viewpoint V_{new} :

$$\mathcal{F} : (\mathbf{I}_{\mathbf{V}}^l, V_{new}) \rightarrow \mathbf{I}_{V_{new}}^h. \quad (1)$$

The synthesized HDR image $\mathbf{I}_{V_{new}}^h$ needs to exhibit an expanded dynamic range compared to LDR training imagery, while maintaining geometric coherence with the underlying 3D structure of the scene (Reinhard, 2020). Let G represent the 3D geometry inferred from $\mathbf{I}_{\mathbf{V}}^l$, then $\mathcal{F}(\mathbf{I}_{\mathbf{V}}^l, V_{new})$ must align with G at a viewpoint V_{new} . In addition, the HDR synthesis must preserve consistent lighting and color across different views (Debevec & Malik, 2023).

Formally, we need to ensure the following constraint holds for each 3D scene point:

$$C(\mathbf{I}_{V_{new}}^h(\pi_{V_{new}}(X))) \approx C(\mathbf{I}_{V_i}^l(\pi_{V_i}(X))), \forall X \in G, \forall V_i \in \mathbf{V} \quad (2)$$

where C denotes the color information of images, X represents the 3D point's coordinates, and $\pi_V(X)$ is a projection

function mapping the 3D scene point X onto the 2D image coordinates corresponding to viewpoint V .

These constraints require the model \mathcal{F} to effectively utilize the limited information from single-exposure inputs to compensate for the absence of multi-exposure sequences.

3.2. Mono-HDR-3D

Architecture. To address the proposed problem as in Sec. 3.1, we propose a novel single-exposure HDR-NVS framework, *Mono-HDR-3D*. Specifically, given single-exposure LDR images (with corresponding camera poses) as input, we first learn an LDR 3D scene model (e.g., NeRF (Mildenhall et al., 2021) or 3DGS (Kerbl et al., 2023)). This is because single-exposure LDR images provide insufficient information to fully recover an HDR scene. Then, we elevate this LDR model to an HDR counterpart via our camera-imaging-aware LDR-to-HDR Color Converter (L2H-CC). Additionally, we introduce a latent HDR-to-LDR Color Converter (H2L-CC) as a closed-loop component, enabling the optimization of HDR features even when only LDR training images are available, which ensures the framework to be robust in the absence of ground-truth HDR data. The overall architecture of Mono-HDR-3D is depicted in Fig. 2.

Camera imaging mechanism. We embark with the seminal LDR image formation formula (Hasinoff et al., 2010):

$$I_l = \begin{cases} \Delta t/g \cdot I_h + I_0 + \epsilon, & \text{Unsaturatation;} \\ I_{\max}, & \text{Saturation} \end{cases} \quad (3)$$

where I_l denotes the LDR color, Δt is the exposure time, g is the sensor gain, I_h represents the corresponding HDR pixel value, and I_0 is the constant offset current with ϵ denoting the sensor noise. Unsaturatation refers to those pixels that can be accurately represented by the LDR image after the camera’s imaging pipeline processing, while saturation occurs when the sensor reaches its limit, causing the pixel value to be capped at a maximum saturation value I_{\max} .

Let the saturated pixel value in Eq. (3) of LDR images as

$$I_{\max} = I_{\text{ideal}} - I_{\text{overflow}} \quad (4)$$

where I_{ideal} and I_{overflow} represent the pixel values captured by an infinitely capable camera and the overflow values between the ideal and real cameras, respectively. For unsaturation pixels, obviously $I_{\text{overflow}} = 0$.

By integrating Eq. (4) with Eq. (3), the formation process of LDR images can be unified as:

$$I_l = \underbrace{\Delta t/g \cdot I_h}_{D(\cdot)} + \underbrace{I_0 + \epsilon - I_{\text{overflow}}}_{B(\cdot)}, \quad (5)$$

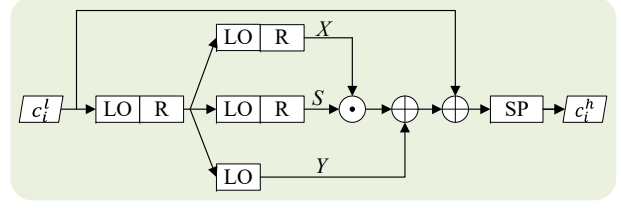


Figure 3. Structure of our camera imaging aware LDR-to-HDR Color Converter (L2H-CC). c_l^l/c_l^h : LDR/HDR color; LO: Linear Operation, R: ReLU, SP: Softplus. \odot and \oplus : Element-wise multiplication and addition.

where the term $D(\cdot)$ is responsible for linearly scaling the brightness values of HDR images to fit within the representation range of LDR images, while the term $B(\cdot)$ is to learn the offset and correction of LDR image brightness values.

By reversing Eq. (5), the HDR value can be obtained as:

$$I_h = \underbrace{g/\Delta t \cdot (I_l - I_0 + I_{\text{overflow}})}_{X(\cdot)} - \underbrace{g/\Delta t \cdot \epsilon}_{Y(\cdot)}, \quad (6)$$

where the term $X(\cdot)$ serves as a scaling factor that linearly amplifies the brightness values of the LDR image to match the range of HDR images, the term $S(\cdot)$ adjusts and corrects the amplified LDR brightness values, and the term $Y(\cdot)$ performs noise correction on the adjusted HDR brightness values.

L2H-CC. Simulating the above camera imaging formula Eq. (6), we design an LDR-to-HDR color converter, L2H-CC, that learns to approximate the inherent camera response characteristics and facilitates accurate HDR color estimation:

$$c_l^h = f_{12h}(c_l^l), \quad (7)$$

where c_l^h and c_l^l represent the HDR color and the LDR color, respectively. This is challenging as only the LDR color I_l is known whilst all the rest are not, resulting in vast modeling freedom.

To address this challenge, we impose network architectural prior in the spirit of camera imaging. That being said, L2H-CC consists of three dedicated modules organized in a way that approximates the camera’s color conversion behavior (Eq. (6)), as shown in Fig. 3. Given an LDR color c_l^l , a linear layer with a ReLU activation is first used to convert LDR colors into a latent feature space. To simulate the three terms $S(\cdot)$, $X(\cdot)$ and $Y(\cdot)$, we adopt a simple MLP with ReLU for efficient non-linear computation. The ReLU activation ensures nonnegative outputs, aligning with the underlying physical constraints of these parameters (Hasinoff et al., 2010). Note that no activation function is applied to the $Y(\cdot)$ module, as the noise component ϵ is inherently

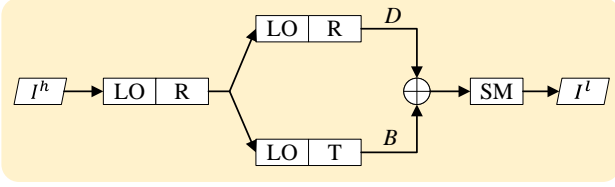


Figure 4. Structure of our camera imaging aware HDR-to-LDR Color Converter (H2L-CC). I^h/I^l : HDR/LDR image; LO: Linear Operation, R: ReLU, T: Tanh, SM: Sigmoid. \oplus : Element-wise addition.

random. We also adopt a residual structure (He et al., 2016), which stabilizes the learning process by capturing subtle discrepancies between the LDR input and the HDR output, thereby preserving fine-grained color details more effectively.

H2L-CC. We further introduce a closed-loop design, H2L-CC, that converts the rendered HDR images back to LDR for enabling HDR model optimization even when only LDR training data is available. Similarly, we formulate this component according to the camera imaging principle expressed in Eq. (5), formally denoted as:

$$I^l = f_{h2l}(I^h), \quad (8)$$

where I^l and I^h denote the rendered LDR and HDR images, respectively.

Concretely, H2L-CC is composed of two modules that approximate the terms of Eq. (5), as shown in Fig. 4. We first transform the HDR image colors into a latent feature space with a linear layer followed by ReLU activation. To simulate each term $D(\cdot)$ and $B(\cdot)$, a specific linear layer with activation is employed, with ReLU for $D(\cdot)$ and Tanh for $B(\cdot)$. This choice of activation functions is made by their physical meanings (Hasinoff et al., 2010) as discussed earlier, ensuring that the network can effectively simulate the HDR-to-LDR conversion process.

3.3. Model optimization and instantiation

The overall objective loss function of Mono-HDR-3D can be generally expressed as

$$\mathcal{L} = \mathcal{L}_{ldr} + \alpha \mathcal{L}_{hdr} + \beta \mathcal{L}_{h2l}, \quad (9)$$

where \mathcal{L}_{ldr} denotes the standard loss function of the underlying 3D representation model used (e.g., NeRF (Mildenhall et al., 2021) or 3DGS (Kerbl et al., 2023)), \mathcal{L}_{hdr} for matching the HDR ground-truth images if available, and \mathcal{L}_{h2l} is used to train the proposed H2L-CC in the same function as \mathcal{L}_{ldr} . The two hyper-parameters, α and β , control the relative importance among the terms.

Mono-HDR-GS is obtained by integrating Mono-HDR-3D with 3DGS (Kerbl et al., 2023). For the LDR branch, we adopt:

$$\mathcal{L}_{ldr} = \mathcal{L}_1(I^l, \hat{I}^l) + \lambda \cdot \mathcal{L}_{D-SSIM}(I^l, \hat{I}^l), \quad (10)$$

where the \mathcal{L}_1 loss and D-SSIM loss (Wang et al., 2004) are balanced by λ . \hat{I}^l denotes the ground-truth LDR images. To optimize HDR generation, following HDR-GS we use a \mathcal{L}_2 loss in the μ -law LDR (Kalantari et al., 2017) domain as

$$\mathcal{L}_{hdr} = \left\| \frac{\log(1 + \mu \cdot \text{norm}(I^h))}{\log(1 + \mu)} - \frac{\log(1 + \mu \cdot \text{norm}(\hat{I}^h))}{\log(1 + \mu)} \right\|_2^2, \quad (11)$$

where μ denotes the amount of compression, \hat{I}^h represents the ground-truth HDR images, and $\text{norm}(\cdot)$ specifies the min-max normalization.

Mono-HDR-NeRF is formed by integrating Mono-HDR-3D with NeRF (Mildenhall et al., 2021). In this case, we adopt the Mean Square Error (MSE) based loss function same as HDR-NeRF:

$$\mathcal{L}_{ldr} = \mathcal{L}_{hdr} = \mathcal{L}_{h2l} = \text{MSE}(I^l, \hat{I}^l). \quad (12)$$

4. Experiments

Datasets. Following HDR-GS and HDR-NeRF, we use the multi-view image dataset with 8 synthetic scenes created by the software Blender (Blender Foundation, 2025) and 4 real scenes captured by a camera, where each scene contains 35 images captured under 5 different exposure times $\{t_1, t_2, t_3, t_4, t_5\}$. We use the same training and test data, where images at 18 views with the exposure time randomly selected from $\{t_1, t_3, t_5\}$ are used for training, while the other 17 views at the same exposure time and HDR images are used for testing. Under our proposed single exposure setting, *only a specific exposure time* is selected for model training and evaluation for one experiment. All methods are compared fairly using the same training and test sets.

Evaluation metrics. We employ the PSNR and SSIM as quantitative metrics. We utilize LPIPS as a perceptual metric, where lower values signify better perceptual quality. Similar to HDR-GS (Cai et al., 2024), we also quantitatively evaluate the rendered HDR images in the tone-mapped domain and qualitatively show HDR results tone-mapped by Photomatrix Pro (HDRsoft Team, 2025).

Implementation details. Both models are trained with the Adam optimizer with the same parameters as HDR-NeRF and HDR-GS. For Eq. (9), we set β to 0.01/0.05, while $\alpha = 0.6$ for Mono-HDR-NeRF/Mono-HDR-GS. We set the learning rate of L2H-CC/H2L-CC to $5 \times 10^{-4}/1 \times 10^{-3}$, and the decays to $5 \times 10^{-5}/5 \times 10^{-4}$ by cross-validation.

Table 1. Quantitative results on the synthetic datasets. For the LDR results, we report averaged results across exposure times t_1 , t_3 , and t_5 and all scenes.

Method	Inference Speed (fps)	LDR result			HDR result		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
HDR-NeRF	0.26	30.62	0.658	0.285	13.76	0.511	0.443
Mono-HDR-NeRF (Ours)	0.26	38.78	0.936	0.048	32.86	0.940	0.068
HDR-GS	147.45	39.48	0.977	0.018	35.30	0.965	0.030
Mono-HDR-GS (Ours)	136.97	41.68	0.983	0.009	38.57	0.975	0.012

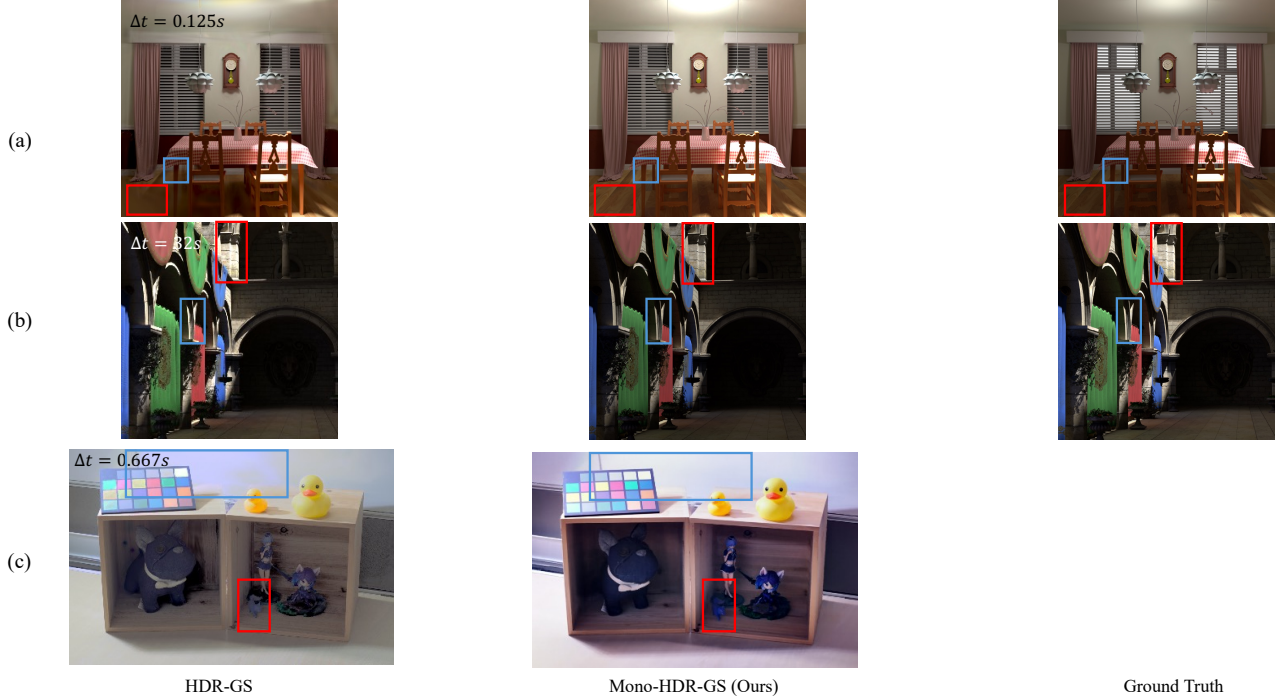


Figure 5. Comparison of HDR-NVS on both (a/b) synthetic and (c) real datasets. Δt : Exposure time.

4.1. Quantitative evaluation

Competitors. We compare Mono-HDR-3D with two latest state-of-the-art approaches: (1) HDR-NeRF (Huang et al., 2022), the first to synthesize HDR images of novel views using the implicit NeRF model, and (2) HDR-GS (Cai et al., 2024), which leverages the efficient representation of 3DGS to build an HDR representation model. To the best of our knowledge, these are the only existing methods specifically designed to synthesize HDR novel views from LDR training imagery. Whilst designed for multi-exposure LDR setting, they can be also applied to our proposed single-exposure setting. We used their official repositories to conduct the experiments for ensuring their own optimal performance.

Tab. 1 presents the quantitative results on the synthetic datasets for both LDR and HDR-NVS. Notably, HDR-NVS

results are the most important, as they encapsulate the core objective of this study. The reported results are averaged across three different exposure times to ensure the completeness and reliability of the performance metrics.

In addition to visual quality assessment, model inference speed (fps) is also included. We highlight the following key points:

(I) In terms of HDR-NVS results, our models significantly outperform all alternatives, particularly HDR-NeRF, in generation quality. This advantage arises because HDR-NeRF struggles to converge without multiple exposure LDR training data, often producing images that are entirely black or white. This highlights the greater challenges associated with the proposed single exposure setting while also validating the efficacy and superiority of our model design in

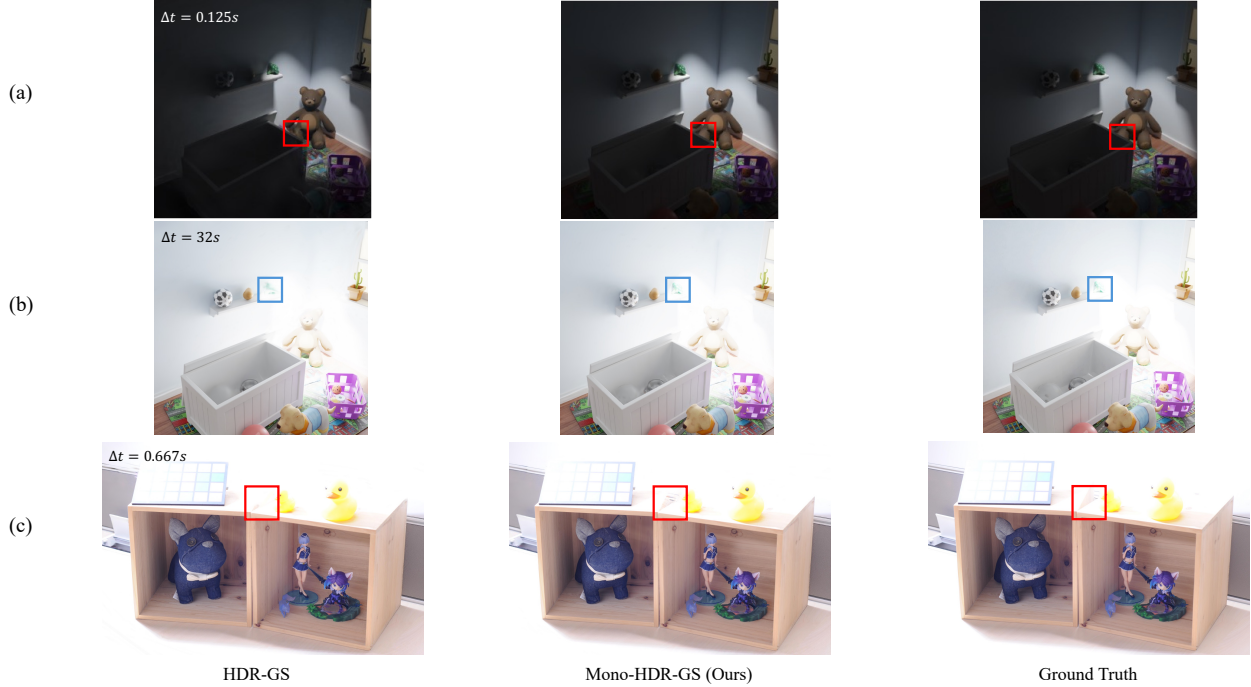


Figure 6. Comparison of LDR-NVS on both (a/b) synthetic and (c) real datasets. Δt : Exposure time.

addressing such challenges.

(II) Regarding the LDR-NVS results, we observe a similar performance advantage with our models. This indicates that directly learning an HDR model from single-exposure LDR data, as competitors do, would be inferior due to the absence of multiple exposure observations. This also partly explains our stronger HDR-NVS results, which somehow are dependent on the quality of the LDR output.

(III) In terms of efficiency, our models perform comparably to alternatives using either NeRF or 3DGS as the representation model. This suggests that our models do not sacrifice efficiency for the sake of quality.

Result analysis on real data. While less important, Tab. 2 presents the quantitative results of LDR-NVS on the real datasets, as there are no ground-truth HDR images available. The results are averaged across three distinct exposure times and encompass all scenes. It is evident that no method clearly stands out in synthesis quality.

4.2. Qualitative evaluation

Numerical metrics such as PSNR, SSIM, and LPIPS may not fully reflect the perceived quality of images. Therefore, a qualitative evaluation through visual comparison is essential. For HDR-NVS results on both synthetic and real datasets, as shown in Fig. 5, HDR-GS struggles to accurately reconstruct the darkest and brightest details, whereas our Mono-HDR-GS excels in rendering more intricate struc-

Table 2. Quantitative results on the real datasets. We report the results averaged across all scenes and exposure times t_1 , t_3 , and t_5 . Note, *HDR results cannot be reported due to no ground-truth*.

Method	LDR result		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
HDR-NeRF	32.50	0.948	0.069
Mono-HDR-NeRF (Ours)	32.52	0.948	0.069
HDR-GS	35.34	0.966	0.019
Mono-HDR-GS (Ours)	35.81	0.967	0.017

tures. Regarding LDR-NVS results, as illustrated in Fig. 6, HDR-GS tends to produce blurry and visually unappealing results (*e.g.*, the synthetic data case) or fails in rendering extremely bright and contrastive regions (*e.g.*, the real data case). In contrast, our Mono-HDR-GS can successfully recover smoother color details and present the brightness properly for such challenging cases.

We make similar observations when comparing NeRF-based models, as reflected in Fig. 7 (Also see Fig. 8 in Appendix A), where HDR-NeRF produces color artifacts and blurry outputs, while Mono-HDR-NeRF achieves superior color consistency and detail preservation.

4.3. Ablation studies

We conduct an ablation study with the most efficient model, Mono-HDR-GS, on the synthetic datasets.

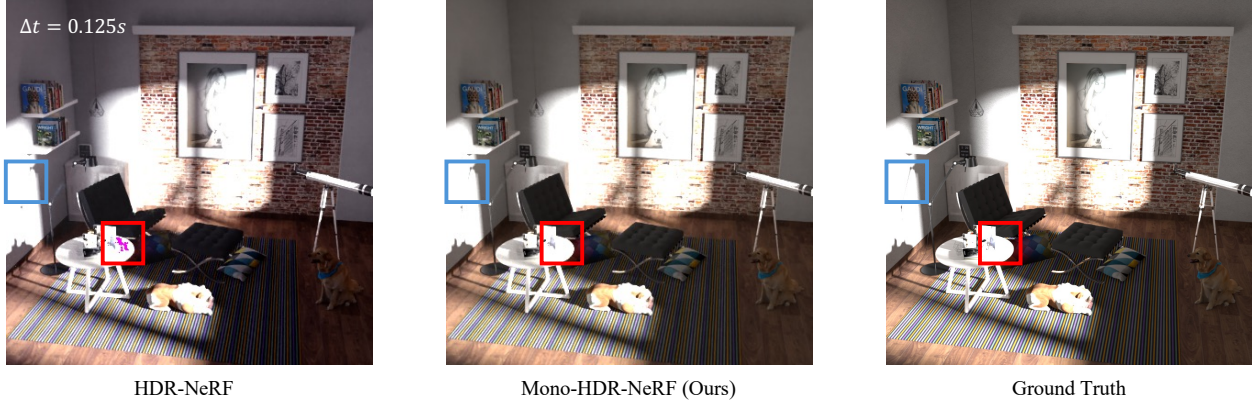


Figure 7. HDR reconstruction comparison on synthetic datasets. Δt : Exposure time.

Table 3. Ablation analysis on the synthetic datasets. The results are averaged across exposures and scenes. MLP: Alternative design with similar amount of parameters.

Row	L2H-CC		H2L-CC		HDR result		
	MLP	Ours	MLP	Ours	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
1	✓			✓	19.02	0.778	0.327
2		✓	✓		38.43	0.974	0.015
3		✓		✓	38.57	0.975	0.012

Table 4. Ablation analysis of closed-loop design on the synthetic datasets. The results are averaged across exposures and scenes.

Row	L2H-CC	H2L-CC	HDR result		
			PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
1	✓	×	38.19	0.974	0.015
2	✓	✓	38.57	0.975	0.012

Module design. To evaluate the design of the proposed L2H-CC and H2L-CC modules, we compare them with an alternative MLP with a similar number of parameters. We make several observations from Tab. 3:

(I) **Row 1 vs. 3:** With a plain MLP to replace L2H-CC, the model performance will degrade significantly, validating the importance of our simulating the camera imaging process (see Sec. 3.2).

(II) **Row 2 vs. 3:** When replacing H2L-CC with MLP, we also observe a performance drop, although quite slight, suggesting that our closed-loop design is useful even when HDR ground truth is available.

Effect of closed-loop design. Following the above design, we further look into the effect of our closed-loop design with H2L-CC with HDR training data. The results in Tab. 4 indicate that this design brings a positive impact of 0.38 dB increase in PSNR for HDR-NVS, demonstrating that the closed-loop framework contributes significantly to enhancing the quality of the reconstructed HDR images.

Table 5. Ablation analysis of different loss components on the synthetic datasets. Results are averaged across exposures and scenes.

Row	Loss	HDR result		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
1	L_{ldr}	-	-	-
2	L_{hdr}	33.93	0.925	0.050
3	L_{h2l}	11.87	0.504	0.371
4	$L_{ldr} + L_{hdr}$	38.19	0.974	0.015
5	$L_{ldr} + L_{h2l}$	13.50	0.507	0.359
6	$L_{hdr} + L_{h2l}$	33.58	0.934	0.058
7	$L_{ldr} + L_{hdr} + L_{h2l}$	38.57	0.975	0.012

Loss contributions. Based on our proposed Mono-HDR-GS, we conduct systematic ablation studies to evaluate the contribution of each loss component in Eq. (9). The results in Tab. 5 reveal three key observations:

(I) **Row 2 vs. Row 7:** HDR loss L_{hdr} serves as the foundational component for HDR-NVS performance. When used alone, Mono-HDR-GS achieves moderate metrics (33.93 dB PSNR, 0.925 SSIM), while its absence leads to severe degradation (e.g., Row 5 shows 13.50 dB PSNR without L_{hdr}). This validates its critical role in reconstructing high dynamic range scenes.

(II) **Row 4 vs. Row 6:** LDR loss L_{ldr} provides essential regularization for scene modeling. When combined with L_{hdr} , it improves PSNR by +4.26 dB (Row 2 vs. 4) and maintains structural fidelity. Notably, L_{ldr} alone fails to train (Row 1), but acts as a complementary constraint when paired with HDR-aware objectives.

(III) **Row 4 vs. 7:** Closed-loop loss L_{h2l} enhances photometric consistency across exposure domains. Its integration elevates PSNR by +0.38 dB while reducing perceptual distance. This demonstrates that bidirectional optimization between HDR and LDR spaces refines both radiance estimation and tone mapping.

Table 6. Ablation studies of different ratio of LDR / HDR images.

Method	LDR / HDR	HDR result		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
HDR-GS	1/1	35.30	0.965	0.030
Mono-HDR-GS		38.57	0.975	0.012
HDR-GS	2/1	35.26	0.963	0.033
Mono-HDR-GS		37.97	0.975	0.013
HDR-GS	3/1	35.16	0.958	0.035
Mono-HDR-GS		37.53	0.974	0.014
HDR-GS	5/1	34.89	0.961	0.027
Mono-HDR-GS		35.51	0.963	0.023
HDR-GS	0/1	33.46	0.936	0.075
Mono-HDR-GS		33.93	0.925	0.050
HDR-GS	1/0	10.51	0.503	0.350
Mono-HDR-GS		13.50	0.507	0.359

Performance of different LDR / HDR ratios. To evaluate the robustness of our Mono-HDR-GS framework under varying data availability, we conduct experiments with LDR / HDR ratios ranging from 0/1 (pure HDR) to 5/1 (dominant LDR), as shown in Tab. 6. The results demonstrate:

(I) **Data Efficiency:** Our model maintains strong performance even with sparse HDR data. For instance, at 5/1 ratio, Mono-HDR-GS retains 92.6% of its peak PSNR (35.51 dB vs. 38.57 dB at 1/1), suggesting effective knowledge transfer from LDR supervision.

(II) **LDR Supervision:** When HDR data becomes extremely scarce (5/1 vs. 0/1), the PSNR degradation of Mono-HDR-GS (from 35.51 dB to 33.93 dB) is significantly smaller than HDR-GS (from 34.89 dB to 33.46 dB). This confirms that LDR supervision provides a more robust geometric prior for HDR reconstruction.

(III) **HDR Criticality:** Pure HDR supervision (0/1) outperforms pure LDR supervision (1/0) by +23.4 dB PSNR, validating the irreplaceable role of HDR data in capturing radiance information. This finding emphasizes HDR’s fundamental importance for high-quality HDR novel view synthesis.

(IV) **Overall Superiority:** Across all ratios, Mono-HDR-GS consistently outperforms HDR-GS with statistically significant margins. At equal computational cost, our Mono-HDR-GS achieves up to +3.27 dB PSNR improvement (1/1 ratio) and maintains superior structural fidelity (SSIM 0.975 vs. 0.965). Notably, even with 100% LDR data (1/0 ratio), our method generates marginally better results than HDR-GS trained solely on HDR images, demonstrating the effectiveness of closed-loop design.

5. Conclusion

This paper pioneers the Single-Exposure HDR-NVS problem by introducing Mono-HDR-3D, a novel meta-algorithm designed to operate effectively with only single-exposure LDR images during training. Unlike conventional HDR-NVS approaches that rely on multiple-exposure imagery, Mono-HDR-3D addresses critical limitations such as motion artifacts, high capture and storage costs, and the need for precise exposure tuning. This not only enhances applicability but also simplifies deployment in dynamic and rapidly changing scenes. Extensive experimental evaluations demonstrate that Mono-HDR-3D significantly outperforms existing methods in generative quality under such more challenging conditions. Importantly, the seamless integration capability of Mono-HDR-3D with existing 3D representation models highlights its versatility and potential for widespread adoption, making advanced HDR techniques accessible to a broader audience and even future advancement in representation modeling.

This work opens new avenues for efficient and robust HDR scene modeling, especially in contexts where access to expensive, professional cameras for training data collection is limited or not possible. By democratizing the process of HDR imaging, we empower more individuals and organizations even with limited resources to engage with high-quality imaging technologies.

Acknowledgments

This work was supported in part by the Key Research and Development Plan of Jiangsu Province (Industry Foresight and Key Core Technology Project) under Grant No. BE2023008-2.

Impact Statement

In this work, we address the high demands for high-quality visual generation and immersion as required in many fields such as VR / AR, entertainment, creative media, broadcasting, TV, and gaming. This has great potential for democratizing both the academic research often featured with limited resources, as well as related industries with diverse backgrounds and contexts. Unlike existing methods requiring costly multi-exposure LDR imagery capture, our approach enables HDR scene reconstruction from single-exposure images with strong accuracy and realism.

While this research problem and our technology are still in the early stage, potential misuse risks (*e.g.*, malicious content generation) might warrant ethical considerations. We advocate for responsible deployment and emphasize that its benefits in advancing safer, high-fidelity digital environments outweigh foreseeable risks per se.

References

- Avidan, S. and Shashua, A. Novel view synthesis in tensor space. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1034–1040, 1997.
- Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., and Srinivasan, P. P. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5855–5864, 2021.
- Blender Foundation. Blender, 2025. URL <https://www.blender.org>.
- Cai, Y., Xiao, Z., Liang, Y., Qin, M., Zhang, Y., Yang, X., Liu, Y., and Yuille, A. L. Hdr-gs: Efficient high dynamic range novel view synthesis at 1000x speed via gaussian splatting. *Advances in Neural Information Processing Systems*, 37:68453–68471, 2024.
- Debevec, P. E. and Malik, J. *Recovering High Dynamic Range Radiance Maps from Photographs*. Association for Computing Machinery, 1 edition, 2023.
- Dille, S., Careaga, C., and Aksoy, Y. Intrinsic single-image hdr reconstruction. In *European Conference on Computer Vision*, pp. 161–177. Springer, 2025.
- Duan, Y., Wei, F., Dai, Q., He, Y., Chen, W., and Chen, B. 4d gaussian splatting: Towards efficient novel view synthesis for dynamic scenes. *arXiv preprint arXiv:2402.03307*, 2024.
- Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R. K., and Unger, J. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics*, 36(6):1–15, 2017a.
- Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R. K., and Unger, J. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics*, 36(6), November 2017b. ISSN 0730-0301.
- Gao, K., Gao, Y., He, H., Lu, D., Xu, L., and Li, J. Nerf: Neural radiance field in 3d vision, a comprehensive review, 2023.
- Garbin, S. J., Kowalski, M., Johnson, M., Shotton, J., and Valentin, J. Fastnerf: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14346–14355, 2021.
- Hanji, P., Mantiuk, R., Eilertsen, G., Hajisharif, S., and Unger, J. Comparison of single image hdr reconstruction methods—the caveats of quality assessment. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–8, 2022.
- Hasinoff, S. W., Durand, F., and Freeman, W. T. Noise-optimal capture for high dynamic range photography. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 553–560. IEEE, 2010.
- HDRsoft Team. Photomatrix pro 6, 2025. URL <https://www.hdrsoft.com>.
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- Huang, X., Zhang, Q., Feng, Y., Li, H., Wang, X., and Wang, Q. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18398–18408, 2022.
- Jiang, L. *View transformation and novel view synthesis based on deep learning*. PhD thesis, Loughborough University, 2023.
- Kalantari, N. K., Ramamoorthi, R., et al. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):144–1, 2017.
- Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139–1, 2023.
- Kim, J., Zhu, Z., Bau, T., and Liu, C. Dcdr-unet: Deformable convolution based detail restoration via u-shape network for single image hdr reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5909–5918, 2024.
- Lin, K.-E., Lin, Y.-C., Lai, W.-S., Lin, T.-Y., Shih, Y.-C., and Ramamoorthi, R. Vision transformer for nerf-based view synthesis from a single input image. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 806–815, 2023.
- Liu, S., Zhang, X., Sun, L., Liang, Z., Zeng, H., and Zhang, L. Joint hdr denoising and fusion: A real-world mobile hdr image dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13966–13975, 2023.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- Miyato, T., Jaeger, B., Welling, M., and Geiger, A. Gta: A geometry-aware attention mechanism for multi-view transformers. *arXiv preprint arXiv:2310.10375*, 2023.

- Reinhard, E. High dynamic range imaging. In *Computer Vision: A Reference Guide*, pp. 1–6. Springer, 2020.
- Rosu, R. A. and Behnke, S. Neuralmvs: Bridging multi-view stereo and novel view synthesis. In *International Joint Conference on Neural Networks*, pp. 1–7. IEEE, 2022.
- Schonberger, J. L. and Frahm, J.-M. Structure-from-motion revisited. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4104–4113, 2016.
- Tiwari, G. and Rani, P. A review on high-dynamic-range imaging with its technique. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(9):93–100, 2015.
- Wang, L. and Yoon, K.-J. Deep learning for hdr imaging: State-of-the-art and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 8874–8895, 2021.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Process*, 13(4), 2004.
- Yan, Q., Zhang, L., Liu, Y., Zhu, Y., Sun, J., Shi, Q., and Zhang, Y. Deep hdr imaging via a non-local network. *IEEE Transactions on Image Processing*, 29:4308–4322, 2020.
- Yan, Q., Chen, W., Zhang, S., Zhu, Y., Sun, J., and Zhang, Y. A unified hdr imaging method with pixel and patch level. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22211–22220, June 2023.

A. Additional Visualization Comparisons of HDR-NeRF and Mono-HDR-NeRF

This section presents additional visual comparisons between HDR-NeRF and our Mono-HDR-NeRF, including LDR novel view rendering results on both synthetic and real datasets. These comparisons are illustrated in Fig. 8.

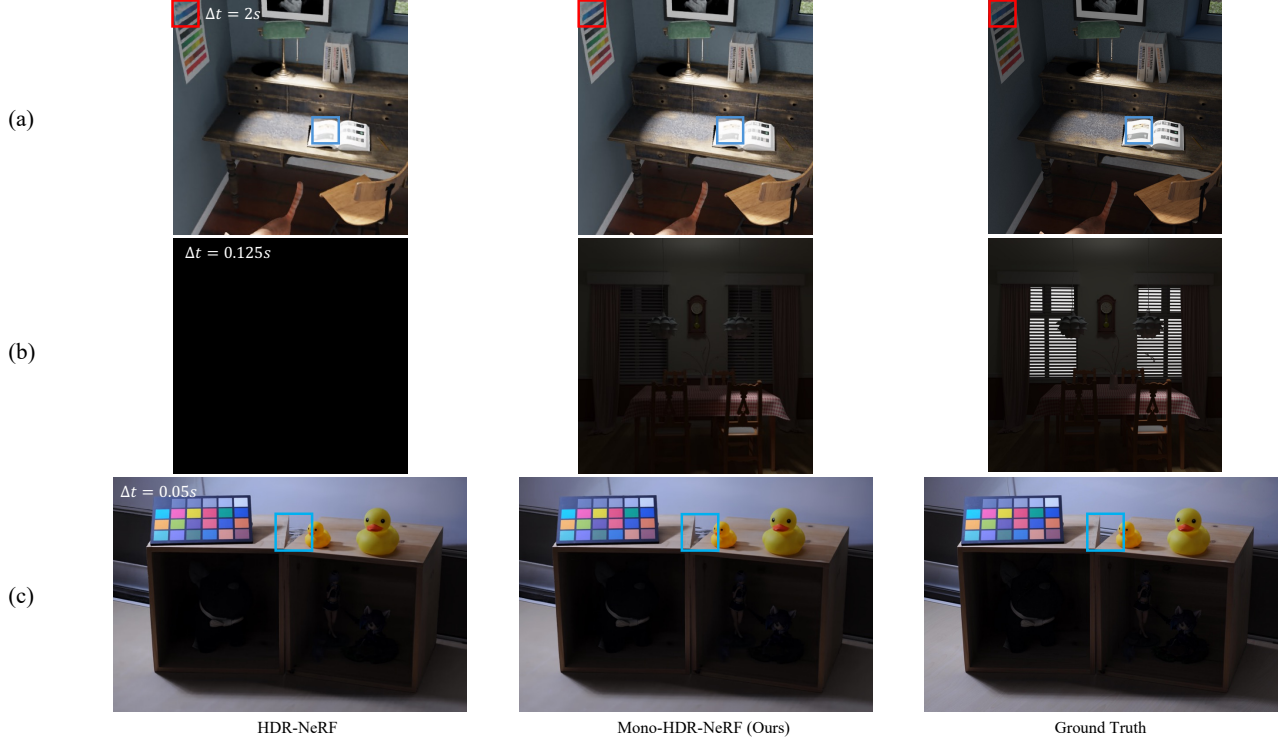


Figure 8. Comparison of LDR-NVS on both (a/b) synthetic and (c) real datasets. Δt : Exposure time.

It can be observed that, HDR-NeRF suffers from color artifacts and blurriness when rendering LDR images, and may fail to converge, producing black outputs without multi-exposure data. In contrast, Mono-HDR-NeRF achieves superior color consistency and detail preservation in LDR rendering.