Mono-HDR-3D: High Dynamic Range Novel View Synthesis with Single Exposure

https://github.com/prinasi/Mono-HDR-3D

Kaixuan Zhang[1], Hu Wang[2], Minxian Li[1 3], Mingwu Ren[1 3], Mao Ye[2], Xiatian Zhu[4]

[1]Nanjing University of Science and Technology
[2]University of Electronic Science and Technology of China
[3]State Key Laboratory of Intelligent Manufacturing of Advanced Construction Machinery
[4]University of Surrey

Background

- Low Dynamic Range (LDR) images are constrained by limited luminance ranges (usually from 0 to 255, 8-bit), often leading to loss of detail in shadows and highlights due to device or sensor limitations.
- High Dynamic Range (HDR) images, however, capture and display a significantly broader dynamic range (beyond 8-bits), preserving intricate details in both extreme dark and bright regions.

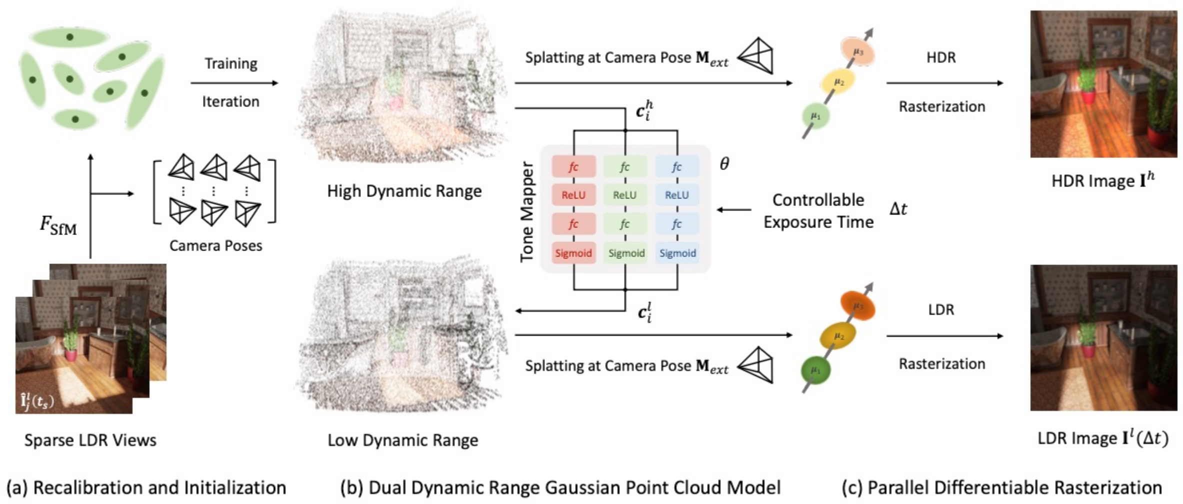
This expanded capability enables HDR to closely mimic human visual perception, delivering superior contrast, naturalness, and immersive quality compared to traditional imaging approaches.

Low Dynamic Range image

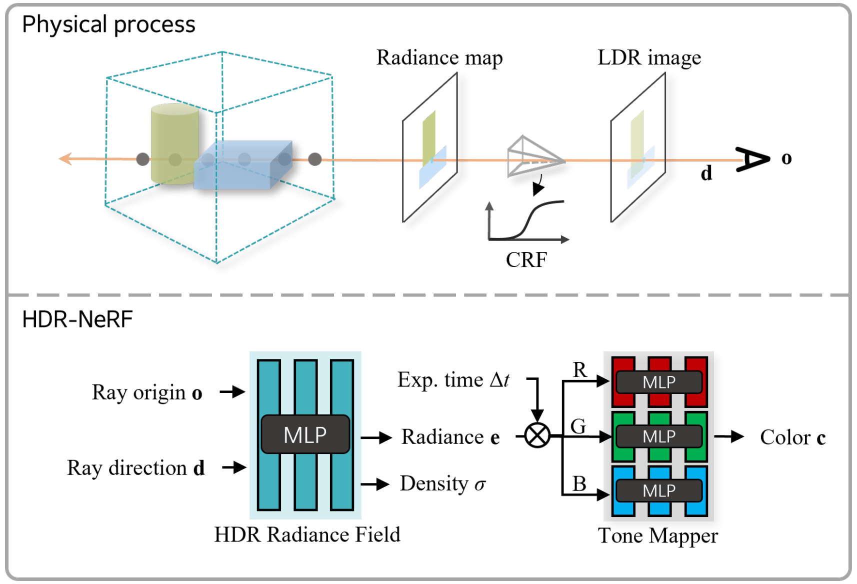
High Dynamic Range image

Motivation

High Dynamic Range Novel View Synthesis (HDR-NVS) aims to establish a 3D scene HDR model from Low Dynamic Range (LDR) imagery. However, current HDR-NVS approaches typically rely on multi-exposure LDR images as input to reconstruct 3D HDR scenes, which have significant limitations, including susceptibility to motion, high capture and storage costs.



Overview of HDR-GS



Overview of HDR-NeRF

Cai Y, Xiao Z, Liang Y, et al. HDR-GS: Efficient high dynamic range novel view synthesis at 1000x speed via gaussian splatting[J]. Advances in Neural Information Processing Systems, 2024, 37: 68453-68471.
Huang X, Zhang Q, Feng Y, et al. HDR-NeRF: High dynamic range neural radiance fields[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18398-18408.

南京理工大学 NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY　电子科技大学 University of Electronic Science and Technology of China　UNIVERSITY OF SURREY　ICML International Conference On Machine Learning

Problem Formulation

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, single-exposure images frequently suffer from overexposure or underexposure, posing significant challenges for HDR-NVS.

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

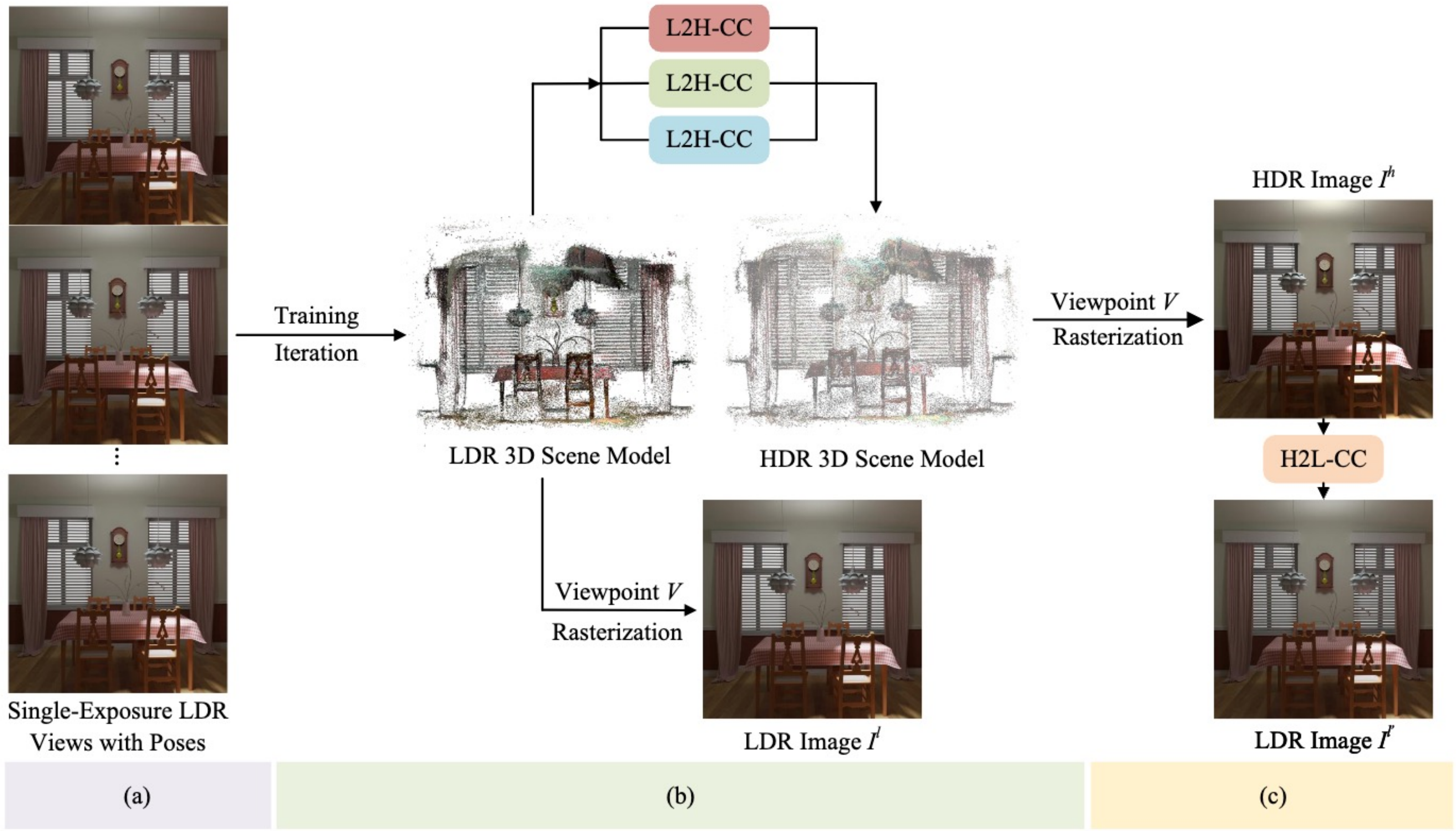
Definition:

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

$$\mathcal{F} : (\boldsymbol{I}^l_{\boldsymbol{V}}, \boldsymbol{V}_{new}) \rightarrow \boldsymbol{I}^h_{V_{new}}.$$

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

Overview of proposed Mono-HDR-3D:



- LDR-to-HDR Color Converter (L2H-CC)
- HDR-to-LDR Color Converter (H2L-CC)

Module Design

Camera imaging mechanism:

$$I_l = \begin{cases} \Delta t/g \cdot I_h + I_0 + \epsilon, & \text{Unsaturation;} \\ I_{\max}, & \text{Saturation} \end{cases} \tag{1}$$

Let the saturated pixel value of LDR images in Eq. (1) as

$$I_{\max} = I_{\text{ideal}} - I_{\text{overflow}} \tag{2}$$

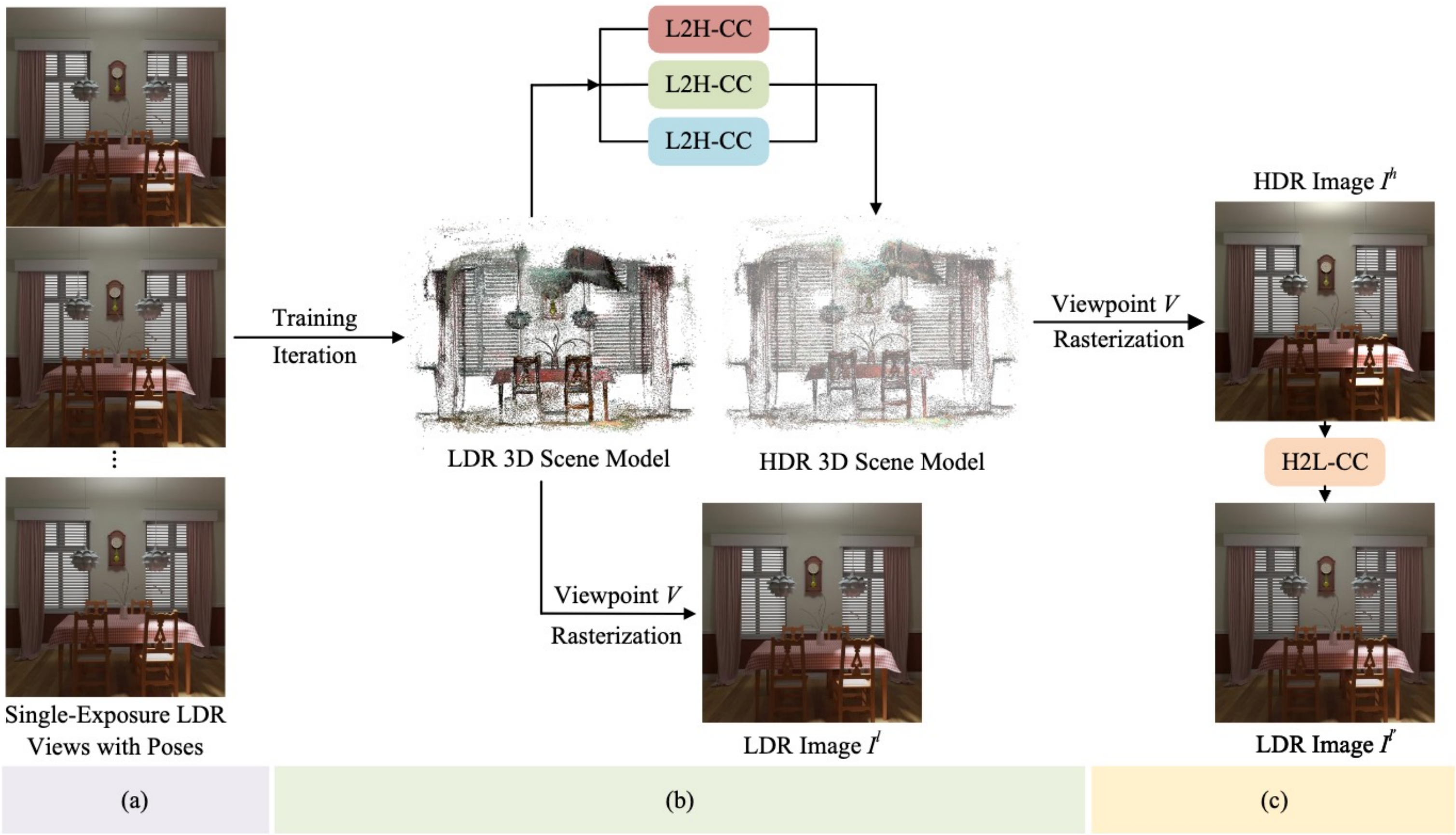
By integrating Eq. (2) with Eq. (1), the formation process of LDR images can be unified as:

$$I_l = \underbrace{\Delta t/g \cdot I_h}_{D(\cdot)} \underbrace{+ I_0 + \epsilon - I_{\text{overflow}}}_{B(\cdot)}, \tag{3}$$

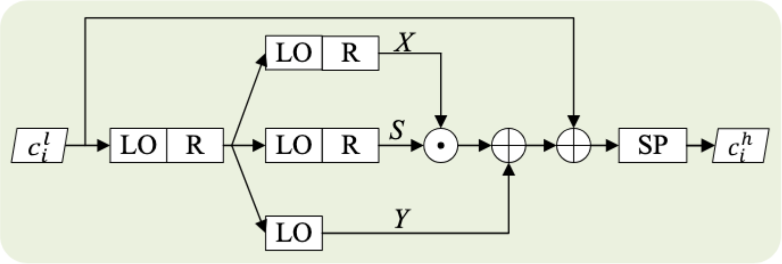
By reversing Eq. (3), the HDR value can be obtained as:

$$I_h = \underbrace{g/\Delta t}_{X(\cdot)} \cdot \underbrace{(I_l - I_0 + I_{\text{overflow}})}_{S(\cdot)} \underbrace{- g/\Delta t \cdot \epsilon}_{Y(\cdot)}, \tag{4}$$

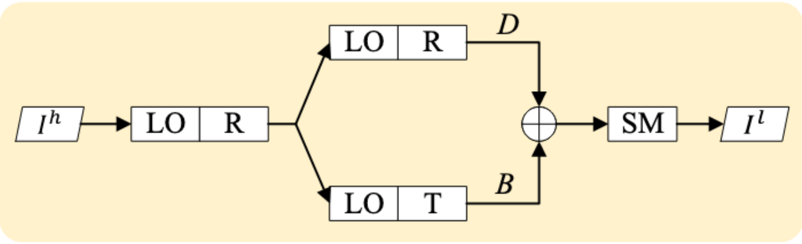
Module Design



- L2H-CC:



- H2L-CC:



Experiments

- Synthetic Datasets:

Method	Inference Speed (fps)	LDR result			HDR result		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
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**

- Real Datasets:

Method	LDR result		
	PSNR↑	SSIM↑	LPIPS↓
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**

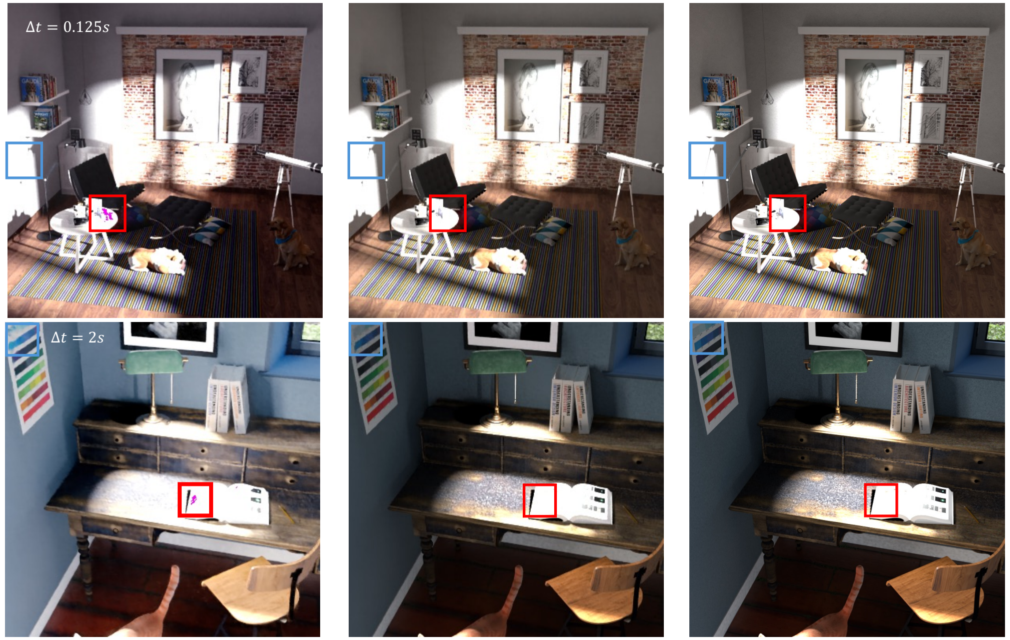
Visualization of Mono-HDR-GS and HDR-GS:

- HDR imaging comparison on the (a/b) synthetic and (c) real datasets:

- LDR imaging comparison on the (a/b) synthetic and (c) real datasets:

Visualization of Mono-HDR-NeRF and HDR-NeRF:

➢ HDR imaging comparison on the synthetic datasets:



HDR-NeRF | Mono-HDR-NeRF (Ours) | Ground Truth

➢ LDR imaging comparison on both (a/b) synthetic and (c) real datasets:

HDR-NeRF | Mono-HDR-NeRF (Ours) | Ground Truth