Vision and Learning Group

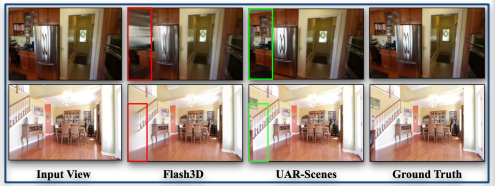
UAR-Scenes: Uncertainty-Aware Diffusion-Guided Refinement of 3D Scenes

Sarosij Bose, **Arindam** Dutta, **Sayak** Nag, **Junge** Zhang, **Jiachen** Li, **Konstantinos** Karydis, **Amit** K. Roy Chowdhury

University of California, Riverside (UCR)

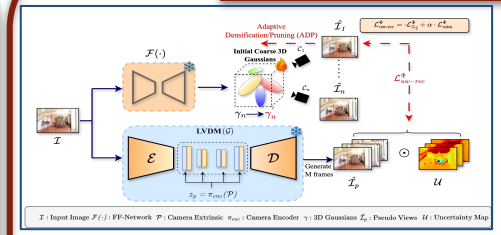
ICCV OCT 19-23, 2025 HONOLULU HAWAII

Problem Overview



How to refine scenes so that it retains high fidelity in regions it has seen and and generates consistent plausible completions for the unseen regions?

Method Overview

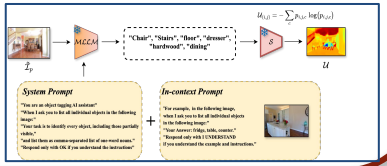


Building Blocks:-

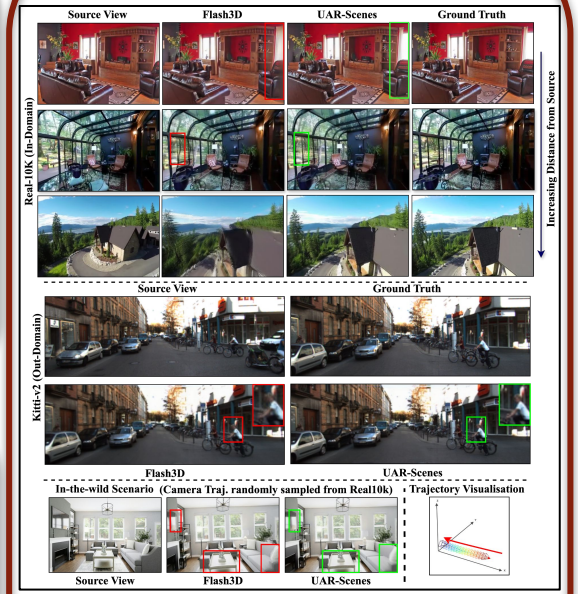
1.Obtain coarse scene gaussians obtained from feed-forward model.
2. Generate pseudo-views using camera controllable video diffusion model.
3.Improve regions using uncertainty mask guidance.

$$\mathcal{L}_{\text{uw-rec}} = \left\| \left(1 - \mathcal{U}_{(i,j)}\right) \odot \left(\tilde{\mathcal{I}} - \tilde{\mathcal{I}}_{\text{p}}\right) \right\|_2 + \alpha\, \text{SSIM}(\tilde{\mathcal{I}}, \tilde{\mathcal{I}}_{\text{p}}),$$

Simple Distillation of semantics with a MLLM driven prior is enough to estimate uncertainty!



Novel View Synthesis



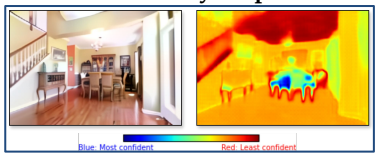
Qualitative results on benchmark datasets and in-the-wild scenarios

Contributions

- **Camera Controllable Video Diffusion-Guided Refinement** of 3D Scenes for occluded and out-of-view regions.
- **Uncertainty Estimation** by distilling entropy with segmentation and MLLM guidance.
- **Texture alignment on-the-fly** with for matching the color of regressed and generated images.

Results

Uncertainty Map $\mathcal{U}$



$$o_t = \mathcal{S}\left(\tilde{\mathcal{I}}_{\text{p}}, O\right) \in \mathbb{R}^{n \times (N+1)},$$

$$\mathcal{U}_{(i,j)} = -\sum_{c} p_{i,j,c} \log\left(p_{i,j,c}\right),$$

RealEstate-10K *(In-Domain)*

Table 1. **Novel View Synthesis.** Our model shows superior performance on RealEstate10k [26] for small, medium, and large baseline ranges. We highlight the best performance in **bold** and the second best performance in underline.

Model	5 frames			10 frames			$\mathcal{U}[-30,30]$		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Syn-Sin [52]	–	–	–	–	–	–	22.30	0.740	–
SV-MPI [53]	27.10	0.870	–	24.40	0.812	–	23.52	0.785	–
BTS [16]	–	–	–	–	–	–	24.00	0.755	0.194
Splatter Image [29]	24.15	0.894	0.110	25.60	0.760	0.240	23.10	0.730	0.290
MINE [54]	28.45	0.897	0.111	25.89	0.850	0.150	24.75	0.820	0.179
Flash3D [1]	28.46	0.899	0.100	25.94	0.857	0.133	24.93	0.833	0.160
UAR-Scenes	**28.67**	**0.902**	**0.095**	**26.54**	**0.861**	**0.112**	**27.81**	**0.887**	**0.107**

KITTI-v2 *(Out-Domain)*

Method	KITTI		
	PSNR↑	SSIM↑	LPIPS↓
LDI [56]	16.50	0.572	–
SV-MPI [53]	19.50	0.733	–
BTS [16]	20.10	0.761	0.144
MINE [54]	21.90	0.828	0.112
Flash3D [1]	21.96	0.826	0.132
UAR-Scenes	**22.31**	**0.844**	**0.128**

Acknowledgements: This work was supported by **USDA NRI grant** 021-67022-33453, **UC MRPI** grant A21-0101-S003, and **NSF** grant CMMI-2326309

Arxiv