Completing Explicit 3D Reconstruction via View Extrapolation with Diffusion Priors

Seunghoon Jeong shoon0602@snu.ac.kr
Eunho Lee eunho1124@snu.ac.kr
Myung-Hwan Jeon mhjeon@illinois.edu
Ayoung Kim ayoungk@snu.ac.kr

S. Jeong | RPM SNU

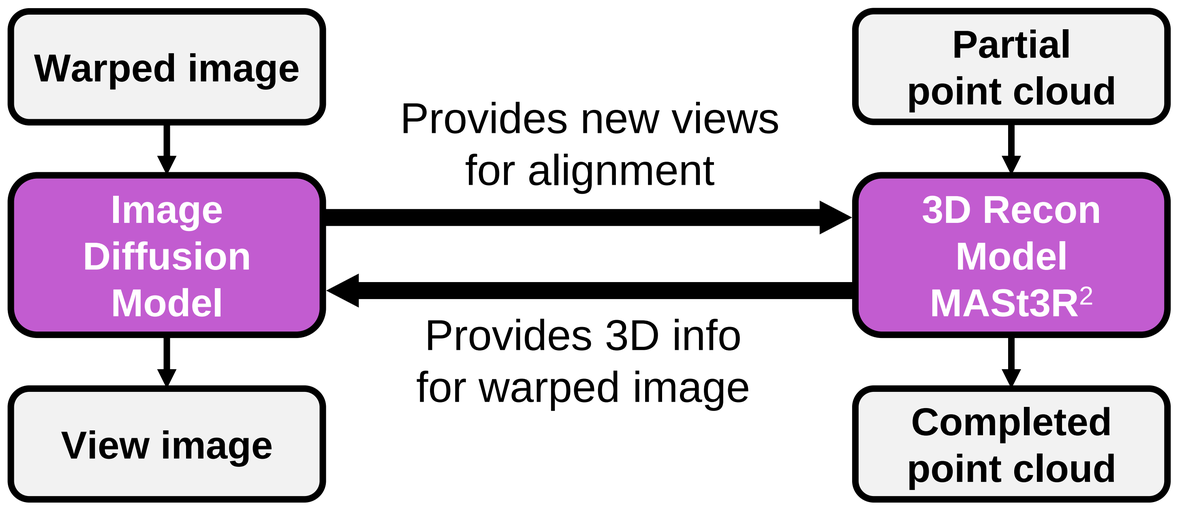
Sparse view 3D Scene Completion

Optimization + Generation Problem

Existing 3DGS[1]-based methods:

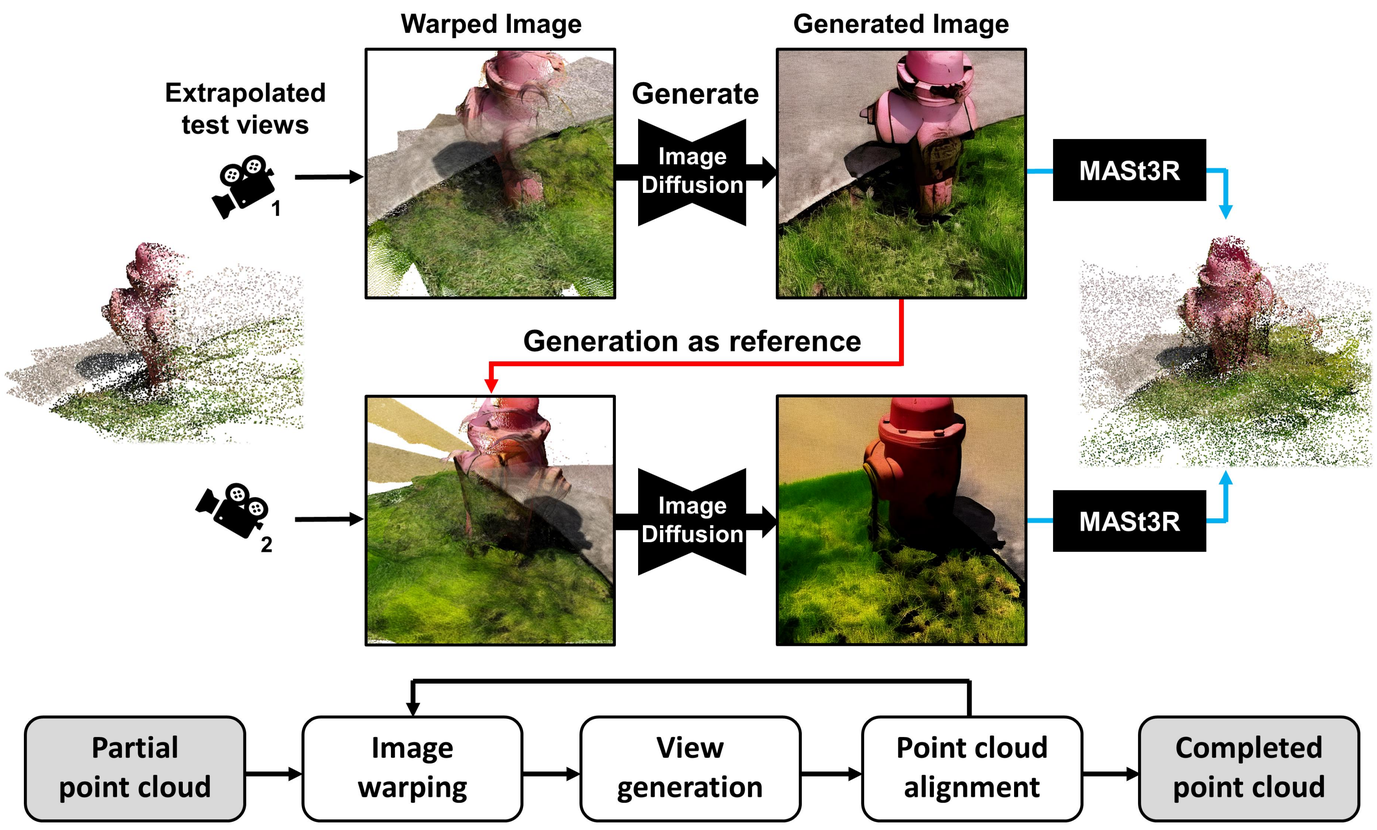
- Easily **overfit** to input views
- Difficult to produce **extrapolated** views
- Refining can't fully leverage **the diffusion prior**

Iterative Generation & Alignment



Ensure 3D Consistency & Realistic Extrapolation

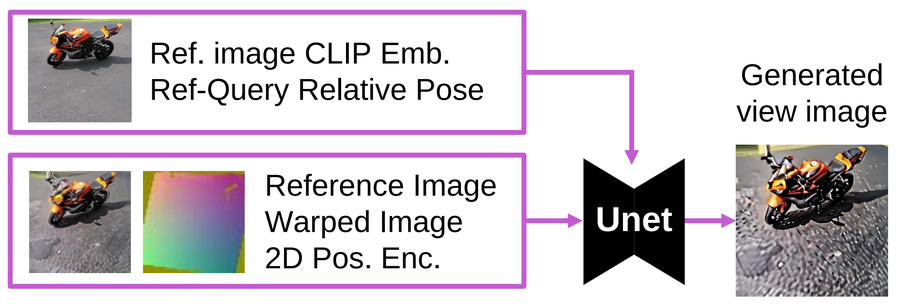
Proposed Method



1. Depth-aware Warped Image Generation

$$I_{\text{warp}} = H_{warp} I_{ref} = K\, P_{\text{query}}\, P_{\text{ref}}^{-1}\, D_{\text{ref}}\, K^{-1}\, I_{\text{ref}}$$

2. Warped Image-conditioned Diffusion



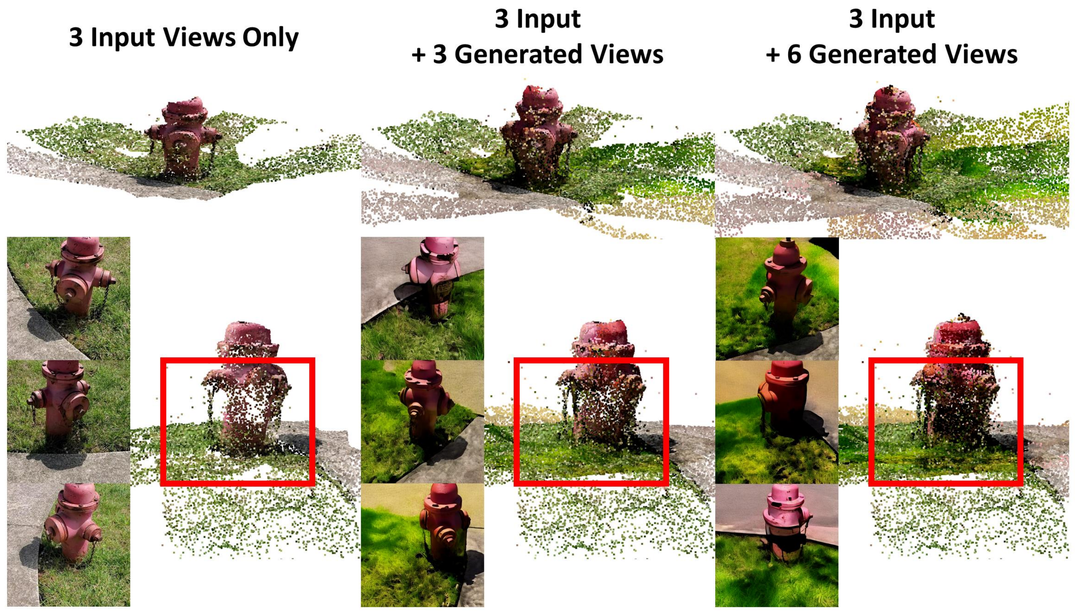
- **Fine-tune ZeroNVS[3] model** to generate views
- **Add warped Image** / Positional Encoding inputs
- **Per-scene LoRA** via cross-input view generation

3. Incremental 3D Reconstruction

- **Optimize** only the new images' **depth maps**

Results

		LPIPS ↓		DISTS ↓		Masked PSNR ↑		Masked SSIM ↑		Perfrom Time ↓	
		3-view	6-view	3-view	6-view	3-view	6-view	3-view	6-view	3-view	6-view
3DGS based	3DGS	0.707	0.640	0.385	0.351	12.759	14.342	0.583	0.621	291.7	335.7
	FSGS	0.666	0.596	0.384	0.338	14.784	15.218	0.684	0.673	108.8	118.8
	InstantSplat	0.588	**0.481**	0.327	0.276	**16.792**	**18.149**	**0.750**	**0.753**	71.1	103.6
	DNGaussian	0.723	0.644	0.417	0.354	13.759	14.463	0.683	0.677	159.5	141.4
Diffusion based	ZeroNVS	0.694	0.652	0.339	0.311	12.641	12.812	0.606	0.569	22.6	37.0
	MegaScenes	0.603	0.526	0.279	**0.245**	14.352	14.783	0.646	0.631	38.0	59.3
	Ours	**0.577**	0.525	**0.266**	0.246	14.397	14.694	0.645	0.616	38.5	59.4



- Proposed method achieves strong results on **LPIPS** (**realism**) and **DISTS** (**geometric consistency**).
- Proposed model can **generate extrapolated views** that 3D reconstruction model **can confidently rely on**.
- Previously **unseen parts** of the scene **are gradually filled in** by the generated views.

* This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) No. RS-2023-0024175

Reference

[1] Kerbl, Bernhard, et al. "3d gaussian splatting for real-time radiance field rendering." *ACM Trans. Graph.* 42.4 (2023): 139-1.
[2] Leroy, Vincent, Yohann Cabon, and Jérôme Revaud. "Grounding image matching in 3d with mast3r." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024.
[3] Sargent, Kyle, et al. "Zeronvs: Zero-shot 360-degree view synthesis from a single image." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024..