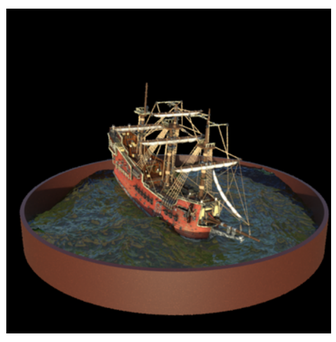
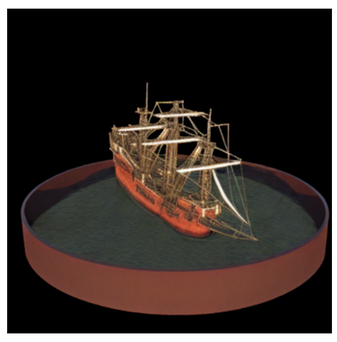
Investigating Geometric Conditioning for Neural Rendering Artifact Correction

Alvaro Lopez, Mohammadreza Tavasoli Naeini, Oleksii Nakhod

Department of Computer Science, University of Toronto

Motivation

- **The Problem:** NeRFs and ***3DGS*** struggle with sparse views, producing "floaters" and blurry textures.
- **Existing Solution:** Methods like ***Difix3D+*** use generative diffusion to "heal" renders.
- **The Gap:** Unconstrained diffusion **hallucinates details** (e.g., changing object geometry) or creates inconsistent textures.

Target	Hallucination
	
	
	

- **Our Contribution:** We constrain the refinement using ***depth*** and ***Canny edge T2I-Adapters*** to ensure visual improvements respect the underlying 3D geometry.

Related Work

1. **Real-Time Rendering:** *NeRF [1]* and ***3DGS*** **[2]**.
2. **Generative Refinement:** *GSFix3D [3]* and ***Difix3D+*** **[4]**.
3. **Controlled Generation:** *ControlNet [5]* and ***T2I-Adapter*** **[6]**.

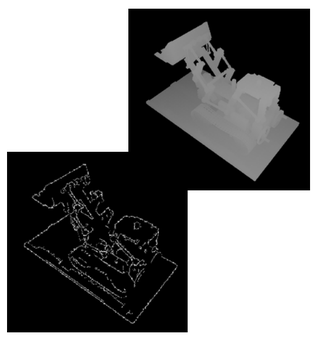
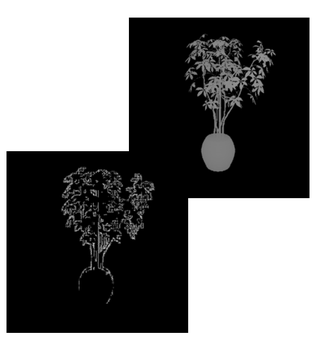
References

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NERF: Representing scenes as neural radiance fields for view synthesis," arXiv.org, 2020.
[2] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," arXiv.org, 2023.
[3] J. Wei, S. Leutenegger, and S. Schaefer, "GSFIX3D: Diffusion-Guided Repair of Novel Views in Gaussian Splatting," arXiv.org, 2025.
[4] J. Z. Wu et al., "Difix3D+: Improving 3D Reconstructions with Single-Step Diffusion Models," arXiv.org, 2025.
[5] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to Text-to-Image diffusion models," arXiv.org, 2023.
[6] C. Mou et al., "T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models", arXiv.org, 2023.

New Technique

- We integrate ***T2I-Adapters*** into the Difix3D+ refinement loop:

1. **Render:** Rasterize RGB (I_{ren}) and Depth (D_{ren}) from the current 3DGS. Extract Canny edges (E_{ren}) from I_{ren} .
2. **Guided Refinement:** Process through ***SD-Turbo*** with adapters injecting features into the U-Net: $\epsilon_\theta\left(x_t, t, prompt, D_{ren}, E_{ren}\right)$
3. **Update:** The refined image acts as a pseudo-ground-truth to update the 3D Gaussians via L2 and LPIPS loss.

Input	Conditioning	Output
		
		

Experimental Results

- **Dataset:** We evaluate on the **NeRF Synthetic Dataset** (8 scenes). To leverage ground truth depth for adapter training, which is absent in standard training splits, we perform a 90/10 split on the official test set.
- **Qualitative Results:** Our approach yields a higher fidelity and perceptual quality compared to the baseline during stage 1 of the pipeline.

- **Quantitative Results:** Canny edge-conditioned Difix outperforms the baseline across all metrics in stage 1. The best aggregate result is highlighted in **bold**.

Model	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Base	31.8619	0.9632	0.0279	15.5294
Depth	31.8545	0.9624	0.0282	15.9502
Canny	**32.1287**	**0.9642**	**0.0272**	**14.7017**
Depth + Canny	32.1070	0.9638	0.0273	15.5019