iNVS: Repurposing Diffusion Inpainters for Novel View Synthesis

Yash Kant[1,2] @yash2kant
Aliaksandr Siarohin[1] @NoTwitterYet
Michael Vasilkovsky[1] @NoTwitterYet
Rıza Alp Güler[1] @NoTwitterYet
Jian Ren[1] @JianRen_
Sergey Tulyakov[1] @SergeyTulyakov
Igor Gilitschenski[2] @igilitschenski

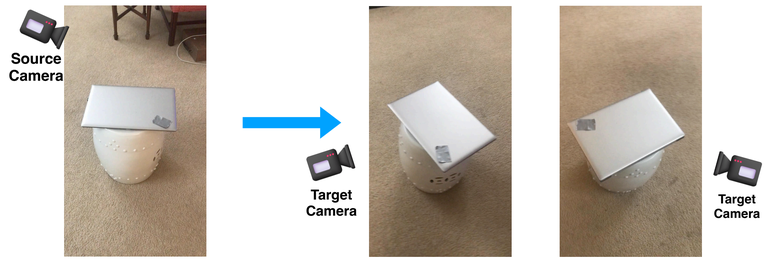
Snap Inc[1]
University of Toronto[2]

SIGGRAPH ASIA 2023
SYDNEY | 12 – 15 DEC

SCAN ME

Overview

NVS Task: Given single image of an object, we want to synthesise novel views.



Contributions and Takeaways:

- Maximise reuse of source view by unprojecting pixels in 3D.
- Splat 3D points from target viewpoint to create a partial view.
- Train 3D-aware Inpainter to fill-in newly discovered regions.

Related Work

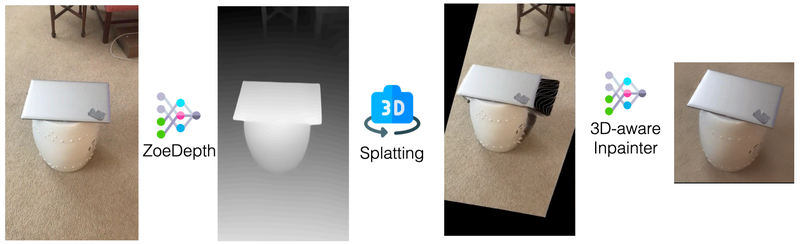
Zero-1-to-3:

- Insufficient reuse of input — details are garbled.
- Camera encoding allows only for coarse control.

Shap-E:

- Unstable training — hypernetwork formulation.
- Input details are not preserved.

Method and Results

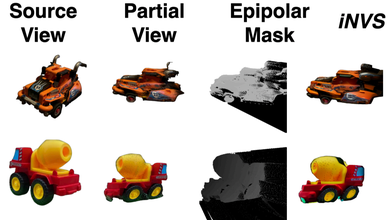


Partial View: Monocular depth based reprojection of source pixels.

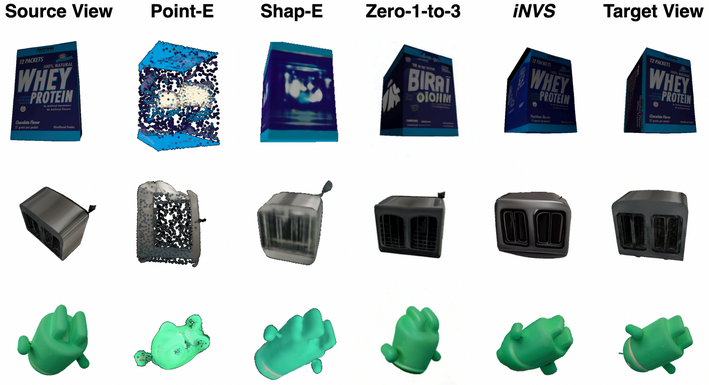
Inpainter: Trained on Objaverse to learn 3D completion priors.

Epipolar and Pose-aware Mask:

- Inpainter only generates regions occluded in source view.
- Smooth inpainting mask conveys the relative angle between source and target camera ray.



iNVS preserves features and object pose well compared to baselines.



Results

iNVS outperforms baselines on scanned and real-world datasets.

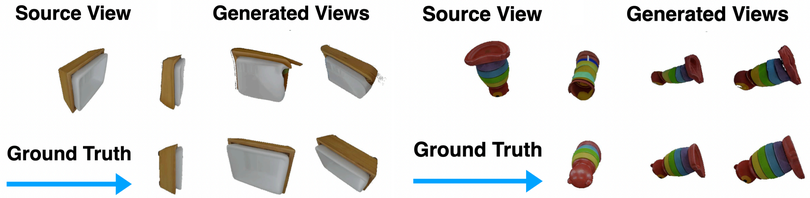
Method	PSNR ↑	SSIM ↑	LPIPS ↓
iNVS	**18.95**	0.30	**0.24**
Zero-1-to-3	14.74	**0.34**	0.25

Google Scanned Objects

Method	PSNR ↑	SSIM ↑	LPIPS ↓
iNVS	**17.58**	**0.33**	**0.36**
Zero-1-to-3	12.32	**0.33**	0.42

Common Objects in 3D

iNVS can generate consistent views from multiple viewpoints



Failure Modes

We find four different failure modes of ***iNVS*** — caused primarily due to imprecise monocular depth; downsampled inpainting mask; or flipped pixels in partial view.