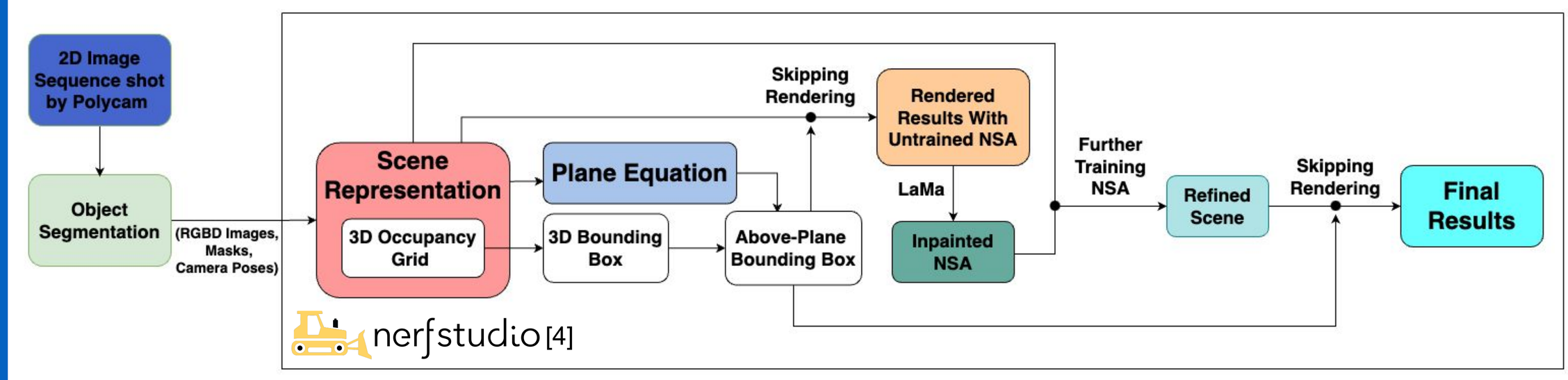
ETH zürich

Let Objects Vanish!
NeRaser: NeRF-based 3D Object Eraser

Group Members:
Liuxin Qing, Shi Chen,
Shao Zhou, Xichong Ling

Supervisor:
Sandro Lombardi



- Generate segmentation masks from image stream using Space-Time Memory Networks [1].
- Skip object in rendering using 3D bounding box generated from occupancy grid.
- Identify & Inpaint 'Never-Seen Area' (NSA) with LaMa [5], assuming object placed on a locally flat surface.
- Further train NeRF model using NSA-inpainted image with object completely erased while preserving consistency.

1 Introduction

- **Motivation**

  Neural Radiance Fields (NeRFs) have gained popularity as an innovative method for novel view synthesis. An important and challenging task of NeRF scene editing involves removing unwanted objects from a 3D environment in a way that ensures that the replaced area appears visually consistent with the surrounding context. Existing works [2][3] replace the unwanted object in 2D images with the result of inpainting, passing these synthesized images to NeRF training. This approach does not explicitly guarantee visual and temporal consistency of the NSA. We aim to address this limitation, so that the edited region in the 3D environment is not only better restored in texture but is also consistent both spatially and temporally.

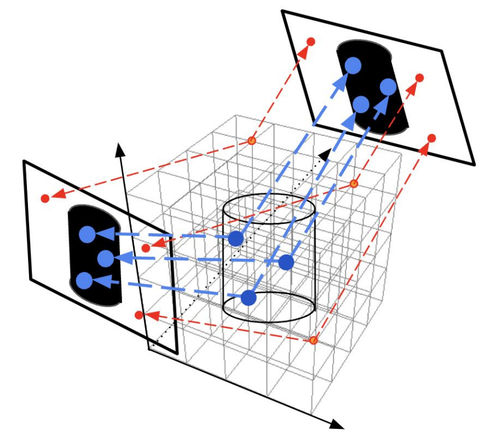
- **Contribution**

  We implement a novel NeRF-editing method which we call *NeRaser,* that can visually remove an object in 3D while retaining spatial and temporal consistency of the occluded area.

2 Implementation Details

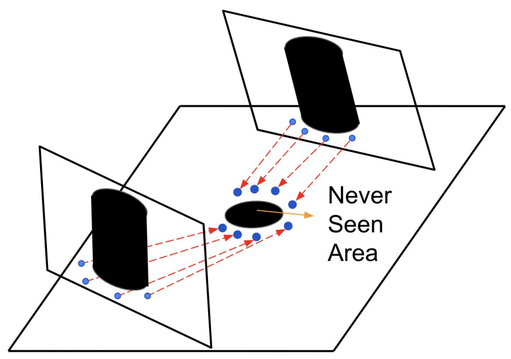
- **Occupancy Grid**

❑ Generate occupancy grid by projecting grid vertices onto 2D mask images.
❑ Calculate proportion of point falling in mask area to get occupancy value.
❑ Filter out grid vertices with occupancy value lower than a pre-set threshold to get occupancy grid.



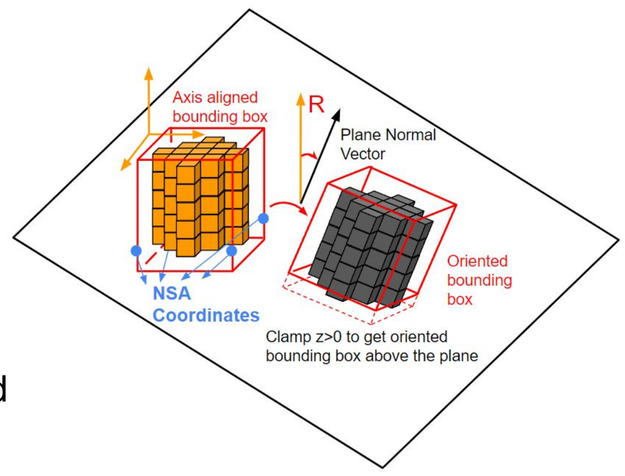
- **Plane Equation**

❑ Sample 2D pixels below 2D masks
❑ Render depth values of 2D pixels to get 3D point samples on the plane
❑ Apply linear regression on the 3D point samples to get plane equation.



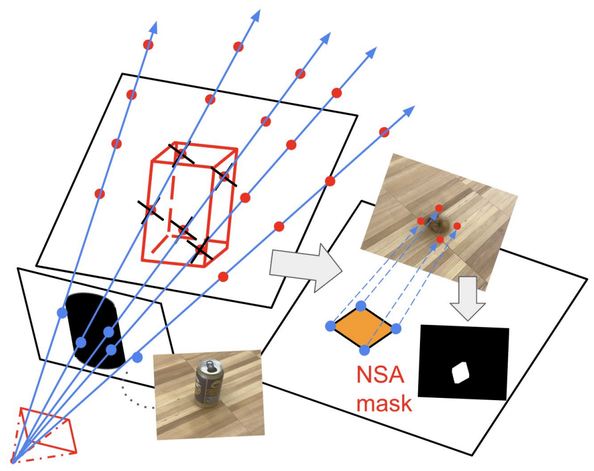
- **3D Bounding Box**

❑ Rotate the z-axis of the occupancy grid to the normal vector of the plane to get proper oriented bounding box.
❑ Get NSA coordinates by calculating intersections between the axis aligned bounding box and the plane.



- **NSA mask**

❑ Skipping the oriented bounding box while rendering to get object erased.
❑ Project NSA corners onto 2D image to get NSA mask used for refining NeRF model.



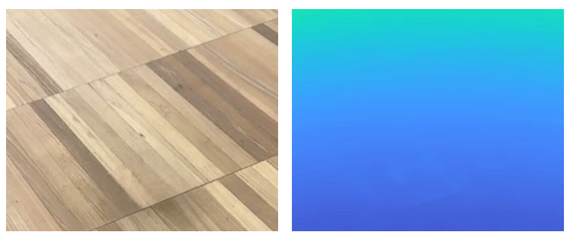
3 Experimental Results

❑ We scan objects using iOS app Polycam to generate RGBD sequences and corresponding camera poses

❑ Object is perfectly removed in novel view rendering

❑ Occluded areas are rendered with temporal and spatial consistency

Rendered RGB(Left) and Depth(Right) Images without object removed



Rendered RGB(Left) and Depth(Right) Images **with object removed**

References

[1] Cheng, Ho Kei, Yu-Wing Tai, and Chi-Keung Tang. "Rethinking space-time networks with improved memory coverage for efficient video object segmentation." Advances in Neural Information Processing Systems 34 (2021): 11781-11794.
[2] Mirzaei, Ashkan, et al. "SPIn-NeRF: Multiview segmentation and perceptual inpainting with neural radiance fields." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[3] Weder, Silvan, et al. "Removing objects from neural radiance fields." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[4] Tancik, Matthew, et al. "Nerfstudio: A Modular Framework for Neural Radiance Field Development"
[5] Suvorov, Roman, et al. "Resolution-robust Large Mask Inpainting with Fourier Convolutions"