

# Supplementary Materials: 3D Scene De-occlusion in Neural Radiance Fields: A Framework for Obstacle Removal and Realistic Inpainting

Anonymous Author(s)

## A SUMMARY

We provide details of our implementation in Appendix B. More results of the ablation studies are provided in Appendix C. Appendix D contains further qualitative experiments showing the inpainting performance of our method in different datasets.

## B IMPLEMENTATION DETAILS

We implement this framework in the PyTorch and Python 3.8 environment. Our model is trained and tested on a machine equipped with a GeForce RTX 3090 graphics card. For each scene, we iteratively train the NeRF model 100,000 times using the provided multi-view images and corresponding camera poses to obtain depth information for all training views in the scene. After that, we select one view in the training dataset to annotate the occluded area. The region encoding network in the object segmentation component follows the color network architecture in NeRF. It consists of 5 fully connected layers activated by ReLU activation functions, with 128 channels per layer. The segmentation network is able to produce masks for any view after 50,000 training iterations. We weight the terms in the loss function with  $\lambda_1 = 1$ ,  $\lambda_2 = \lambda_3 = 0.02$ .

## C ADDITIONAL RESULTS OF ABLATION STUDIES

In the object segmentation network, a joint optimization strategy of the three losses, including color loss  $\mathcal{L}_{rgb}$ , depth loss  $\mathcal{L}_{depth}$ , and area encoding loss  $\mathcal{L}_{area-encoding}$ , is used to ensure the accurate multi-view segmentation results. We verify its effectiveness by comparing the segmentation results with that referred by the combination of color loss and area encoding loss, as well as the combination of depth loss and area encoding loss, respectively. We show the results in Figure 1 and 2, where the combination of  $\mathcal{L}_{rgb}$  and  $\mathcal{L}_{area-encoding}$  is not adequate to locate the position and trace the contour of the obstacle mask. Although the combination of  $\mathcal{L}_{depth}$  and  $\mathcal{L}_{area-encoding}$  provides much better results, the segmentation is still not precise. In contrast, the joint optimization of all three

losses, which have been adopted in our method, can contribute to much accurate obstacle masks referring to the ground-truth.

## D ADDITIONAL EXPERIMENTS

Here, we provide additional qualitative examples to demonstrate the effectiveness of our 3D inpainting method. Qualitative comparisons between our method and four recent solutions on the LLFF [2] dataset and the our proposed dataset are shown in Figure 3 and 4 respectively. In Figure 3, a display board at the lower left corner is regarded as the obstacle in the input image. Referring to the supplementary view, a stand bar has been hidden behind the display board. After the de-occlusion and inpainting, only the proposed method successfully reveals the hidden bar in the input image. Similarly in Figure 4, the proposed method is able to recover more details realistically behind the obstacle.

## REFERENCES

- [1] Zhen Li, Cheng-Ze Lu, Jianhua Qin, Chun-Le Guo, and Ming-Ming Cheng. 2022. Towards An End-to-End Framework for Flow-Guided Video Inpainting. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 17541–17550. <https://doi.org/10.1109/CVPR52688.2022.01704>
- [2] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. (2021).
- [3] Ashkan Mirzaei, Tristan Aumentado-Armstrong, Konstantinos G. Derpanis, Jonathan Kelly, Marcus A. Brubaker, Igor Gilitschenski, and Alex Levinstein. 2023. SPIn-NeRF: Multiview Segmentation and Perceptual Inpainting with Neural Radiance Fields. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20669–20679. <https://doi.org/10.1109/CVPR52729.2023.01980>
- [4] I-Chao Shen, Hao-Kang Liu, and Bing-Yu Chen. 2024. NeRF-In: Free-Form Inpainting for Pretrained NeRF With RGB-D Priors. *IEEE Computer Graphics and Applications* 44, 2 (2024), 100–109. <https://doi.org/10.1109/MCG.2023.3336224>
- [5] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. 2022. Resolution-robust Large Mask Inpainting with Fourier Convolutions. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 3172–3182. <https://doi.org/10.1109/WACV51458.2022.00323>

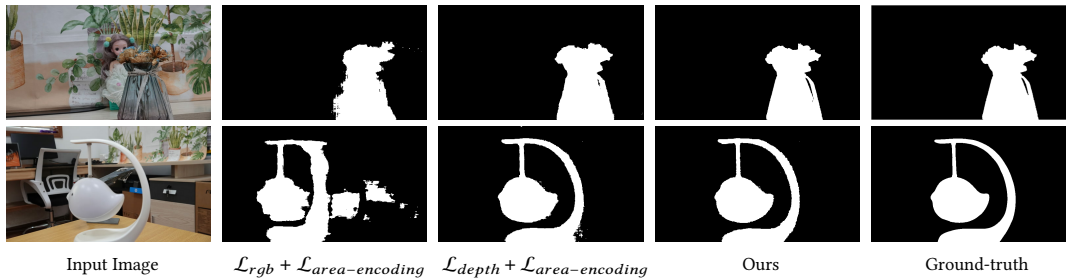


Figure 1: Ablation results in the proposed dataset.

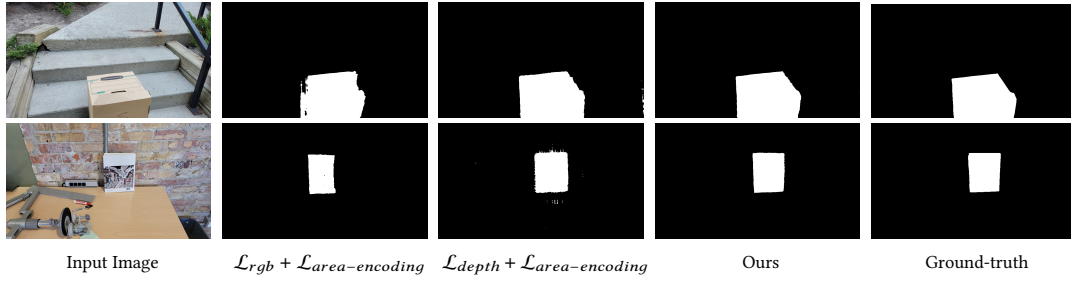


Figure 2: Ablation results in the SPIn-NeRF [3] dataset.

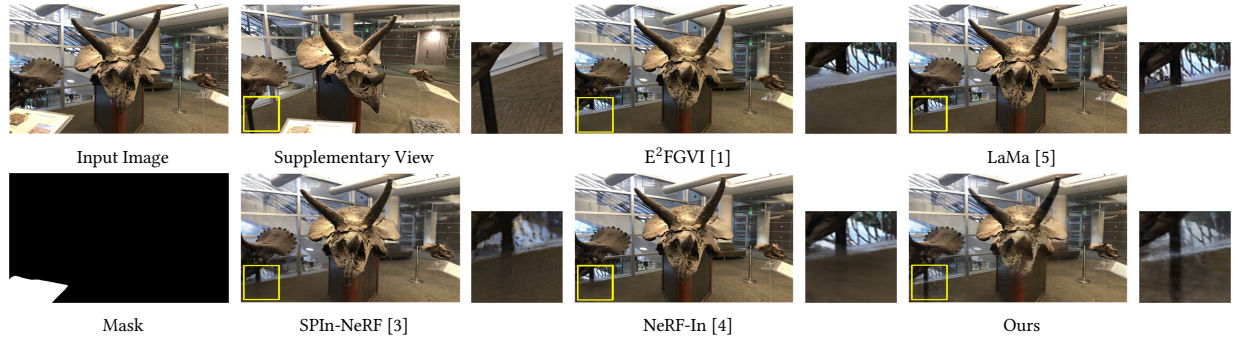


Figure 3: Additional qualitative comparison of de-occlusion results in the scene of the LLFF [2] dataset.

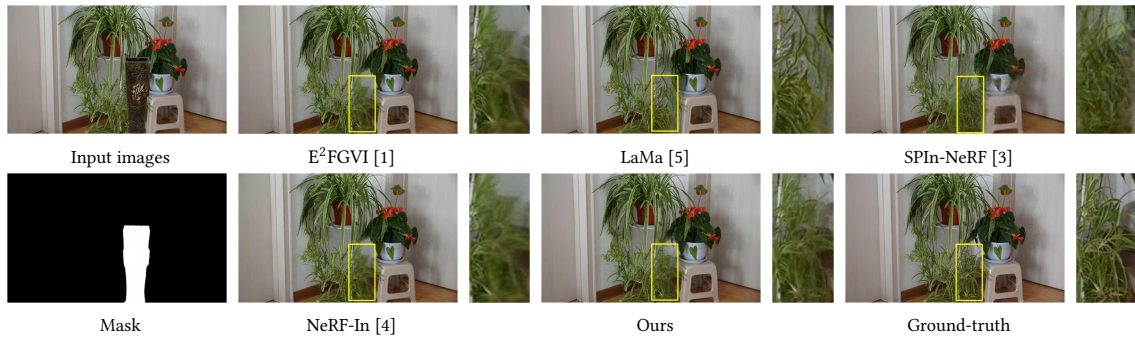
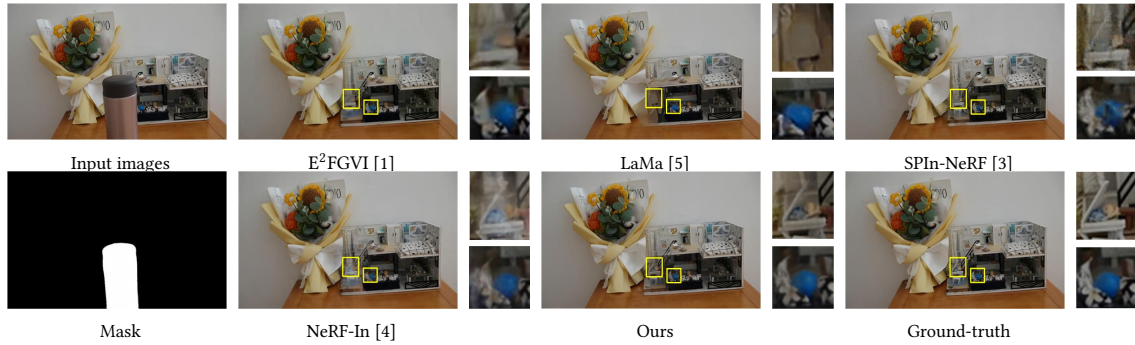


Figure 4: Additional qualitative comparison of de-occlusion results in the scene of the proposed dataset.