Supplementary Material for SCINeRF: Neural Radiance Fields from a Snapshot Compressive Image

Yunhao Li[1,2] Xiaodong Wang[1,2] Ping Wang[1,2] Xin Yuan[2] Peidong Liu[2†]
[1]Zhejiang University [2]Westlake University
{liyunhao, wangxiaodong, wangping, xyuan, liupeidong}@westlake.edu.cn

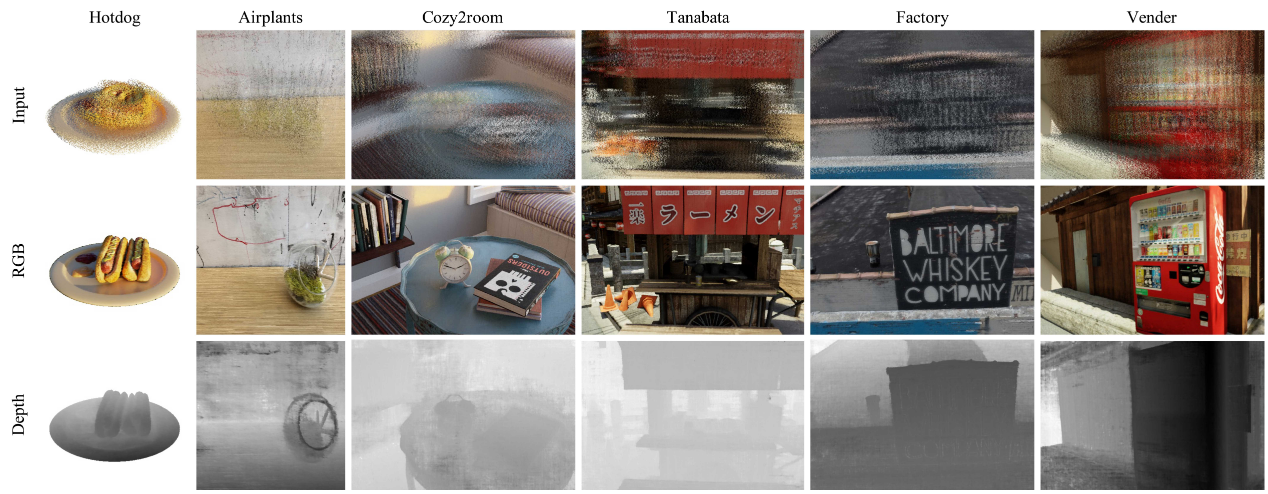

Figure 1. **Qualitative results of our SCINeRF method with synthetic dataset.** Our SCINeRF takes a snapshot compressive image (top row) as input, and recovers the underlying 3D scene. Leveraging strong novel view image synthesis capabilities of NeRF, our SCINeRF can render high-quality novel view images (middle row) and depth maps (bottom row).

In this supplementary material, we present the method we used to generate binary masks for image modulation, and additional qualitative evaluations between our SCINeRF and state-of-the-art methods [2–5] on synthetic and real datasets.

1. Generation Process of Binary Masks

In the SCI image formation process, the SCI encoder employs a series of binary masks to modulate input images. We generate the masks randomly. Algorithm 1 shows the mask generation process, where $\mathbf{OR}$ denotes the mask overlapping rate, H, W are image height and width. N indicates the number of masks, which is equivalent to the number of compressed images, and $\mathbf{M}$ are generated masks.

Algorithm 1 Mask Generation

Require: $H, W, N, \mathbf{OR}$.
Initial $i = 1, j = 1, n = \mathbf{OR} \times N$, $\mathbf{M} = \mathbf{0} \in \mathbb{R}^{H \times W \times N}$.
while $i \leq H$ **do**
  while $j \leq W$ **do**
    Randomly select n indices k_n from N frames.
    $\mathbf{M}_{k_n}(i, j) = 1$.
    $j = j + 1$
  end while
  $i = i + 1$
end while

2. Additional Qualitative Evaluations

As we introduced before, our SCINeRF takes a single snapshot compressive image as input and estimates the 3D scene. From the estimated 3D scene, we can render high-quality novel-view images and depth maps. Fig. 1 shows the rendered RGB images and depth maps from the estimated 3D scene in synthetic dataset.

In this supplementary material, we compared the image reconstruction quality between our SCINeRF and state-of-the-art SCI image restoration methods qualitatively. We compared synthesized novel-view images from our SCINeRF against that of vanilla NeRF [1] with reconstructed images from prior state-of-the-art methods, i.e., NeRF+SOTA. Fig. 2 shows the qualitative comparisons between our SCINeRF and NeRF+SOTA approaches. We also present additional qualitative evaluations on real dataset. Fig. 3 presents the results respectively. It further demonstrates the superior performance of our method against prior state-of-the-art approaches.

3. Additional Evaluations on Camera Pose Estimation

Synthetic Dataset	Airplants	Hotdog	Cozy2room	Factory	Tanabata	Vender
Absolute Trajectory Error ↓	0.0321	0.0218	0.1256	0.0205	0.0967	0.0948

Table 1. **Pose estimation performance of SCINeRF on synthetic dataset.** These results are in the ATE metric.

We further evaluate the accuracy of estimated camera trajectories on synthetic datasets. We exploit the commonly used Absolute Trajectory Error (ATE) metric for the evaluation. The experimental results are presented in Table 1. The results demonstrate that our method can also deliver accurate pose estimations.

References

[1] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 3

[2] Lishun Wang, Miao Cao, and Xin Yuan. Efficientsci: Densely connected network with space-time factorization for large-scale video snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18477–18486, 2023. 1, 3, 4

[3] Xin Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *2016 IEEE International conference on image processing (ICIP)*, pages 2539–2543. IEEE, 2016. 1, 3, 4

[4] Xin Yuan, Yang Liu, Jinli Suo, and Qionghai Dai. Plug-and-play algorithms for large-scale snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1447–1457, 2020. 1, 3, 4

[5] Xin Yuan, Yang Liu, Jinli Suo, Fredo Durand, and Qionghai Dai. Plug-and-play algorithms for video snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7093–7111, 2021. 1

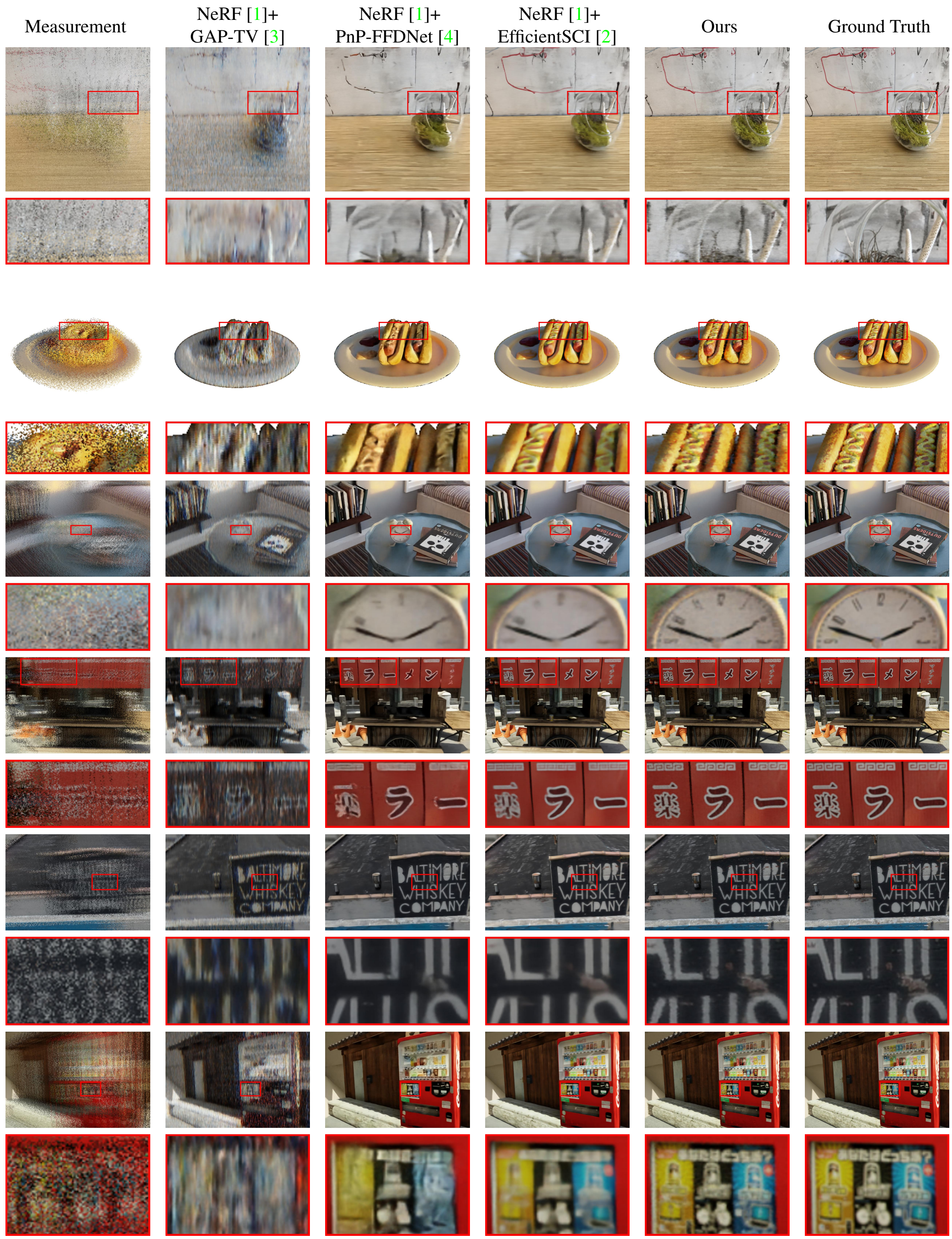

Figure 2. **Qualitative evaluations of our method against NeRF+SOTA methods.** We compared the quality of synthesized novel-view images from our SCINeRF against that of vanilla NeRF with reconstructed images from state-of-the-art methods. Top to bottom shows different scenes including *Airplants, Hotdog, Cozy2room*, *Tanabata*, *Factory* and *Vender*. The qualitative comparisons demonstrate that our SCINeRF outperforms existing approaches.

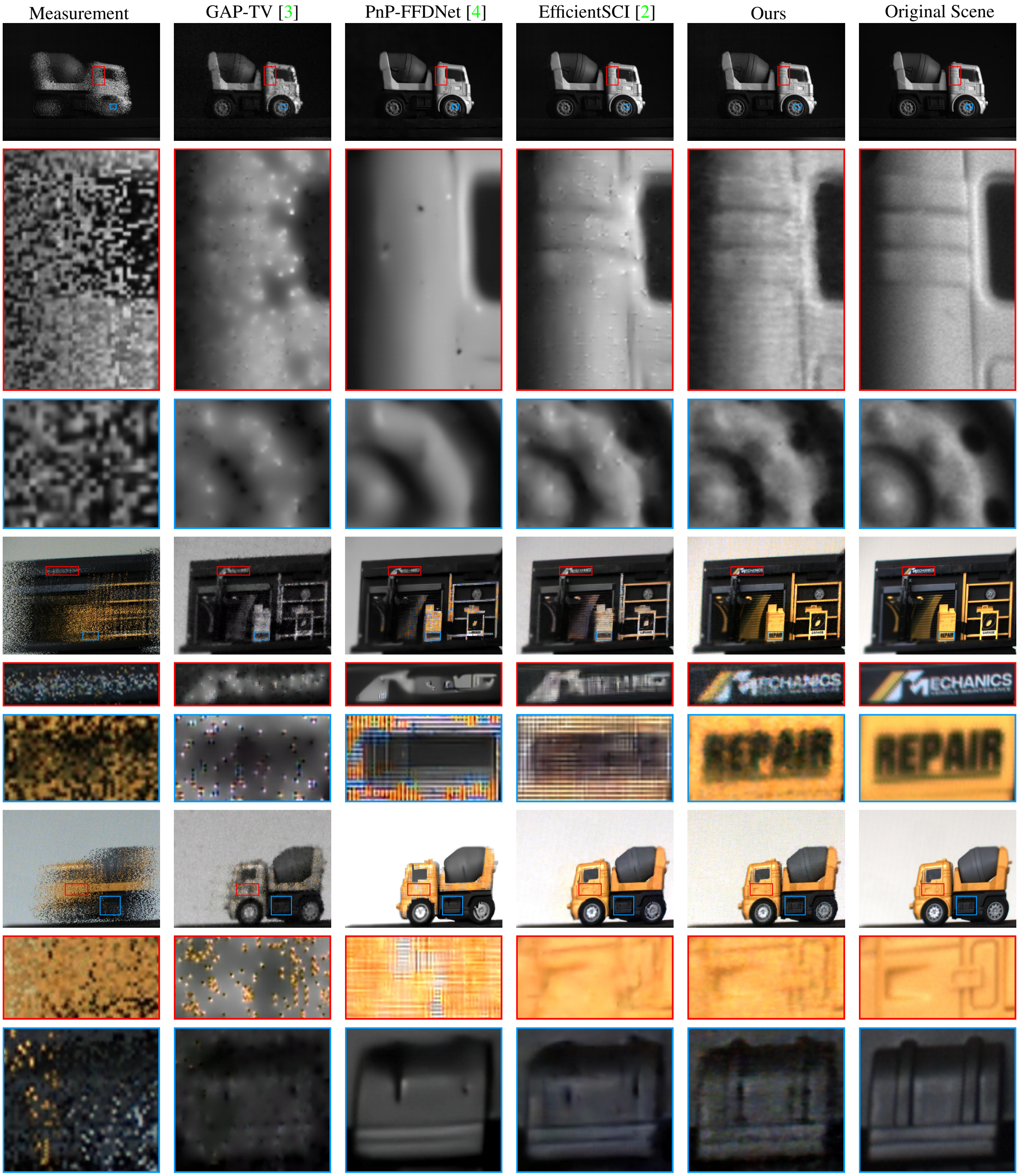

Figure 3. **Qualitative evaluations of our method against SOTA SCI image restoration methods with real dataset.** Top to bottom shows different scenes. Since compressed ground truth images are unavailable, we captured separate scene images after capturing SCI measurement. The qualitative comparisons demonstrate that our SCINeRF surpasses prior methods on image reconstruction with real dataset.