SCIGaussian-D: Dynamic Scene Reconstruction from a Single Snapshot Compressive Image

Supplementary Material

6. Details of Binary Masks

In this paper, we apply binary masks to encode the input images, then generate the SCI images. As we indicated in the main paper, we follow the ablation studies of Li et al. [20] to control the mask overlapping rate to be 0.25. Specifically, the mask overlapping rate is defined as:

$$\mathbf{OR}(x,y) = \frac{\sum_{i=1}^{N} \mathbf{M}_i(x,y)}{N}, \tag{15}$$

where $\mathbf{OR}$ denotes mask overlapping rate, $\mathbf{M}_i$ is the i-th mask and (x,y) indicates pixel coordinate, and N is total number of compressed images. When we generate synthetic datasets and capture real datasets, we follow the mask generation procedure proposed by Li et al. [20] to obtain the 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

7. Computational Efficiency of Different Methods

Existing SOTA methods generally require extensive training time. For instance, EfficientSCI, one of the the top-performing prior methods, requires more than 100 hours (nearly five days) to train its Transformer-based model, and its inference speed is limited to only 2.6 FPS. The SCINeRF training process takes approximately 5 hours, and the rendering speed is just at 0.25 FPS, falling short of real-time requirements. SCISplat incorporates 3DGS, which greatly accelerates training time to less than an hour, with over 200FPS rendering speed. As for our method, it achieves less than 2.5 hours for training with 56 FPS on rendering speed, which remains real-time rendering capabilities. The slower training and rendering speed of our method compared to SCISplat can be attributed to the introduction of the deformation MLP, which takes extra training and inference time. Nevertheless, our method can still perform real-time rendering. Table 3 shows the details of quantitative comparisons.

Model	Training Time (hrs)↓	Inference Speed (FPS)↑
GAP-TV [51]	N/A	0.13
PnP-FFDNet [52]	N/A	0.01
PnP-FastDVDNet [54]	N/A	0.01
EfficientSCI [44]	≈ 100	2.6
SCINeRF [20]	≈ 5	0.25
SCISplat [19]	< 1	205
Ours	< 2.5	56

Table 3. **Quantitative comparisons of training time (in hrs, if applicable) and inference/rendering speed (FPS) of different methods on synthetic dataset.** Our proposed method maintains the real-time rendering capabilities of 3DGS, while achieves significantly faster training speed and inference speed compared to prior SCI image reconstruction methods and SCINeRF.

8. Additional Experimental Results

Figure 5 shows the additional qualitative results on synthetic datasets. Our method is capable of reconstructing dynamic 3D scenes from a single compressed images, as indicated by Figure 5. Table 4 and Figure 6 shows the quantitative and qualitative results of our method against prior SOTA methods on SCINeRF synthetic datasets, where all compressed 3D scenes are static. Our method surpasses existing SOTA SCI image recovery methods and SCINeRF, and reaches comparable performance with SCISplat, which is the current SOTA SCI static 3D representation method. Figure 7 shows the qualitative results on scenes in SCINeRF real datasets. The experimental results on SCINeRF dataset further demonstrate that our method not only effectively reconstructs the dynamic parts of the scene but also maintains high performance and rendering quality in the static regions.

We further evaluate the 3D scene representation quality of our approach against a set of naive baselines. In these baselines, restored images from prior SOTA SCI reconstruction methods (including GAP-TV [51], PnP-FFDNet [52], PnP-FastDVDNet [54], and EfficientSCI [44]) are directly fed into the original deformable 3DGS framework [50] for scene reconstruction. However, due to the limited high-frequency details and the lack of multi-view consistency in their reconstructions, these images cannot sup-

	Airplants			Hotdog			Cozy2room			Tanabata			Factory			Vender		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
GAP-TV	22.85	0.406	0.499	22.35	0.766	0.318	21.77	0.432	0.603	20.42	0.426	0.625	24.05	0.566	0.0.515	20.00	0.368	0.688
PnP-FFDNet	27.79	0.912	0.182	29.00	0.977	0.051	28.98	0.892	0.984	29.17	0.903	0.119	31.75	0.897	0.114	28.70	0.923	0.131
PnP-FastDVDNet	28.18	0.909	0.175	29.93	0.972	0.052	30.19	0.913	0.079	29.73	0.933	0.098	32.53	0.916	0.105	29.68	0.940	0.045
EfficientSCI	30.13	0.942	0.112	30.75	0.956	0.046	31.47	0.932	0.047	32.30	0.958	0.060	32.87	0.925	0.070	33.17	0.940	0.104
SCINeRF	30.69	0.933	0.072	31.35	0.987	0.031	33.23	0.949	0.044	33.61	0.963	0.037	36.60	0.963	0.022	36.40	0.984	0.029
SCISplat	**31.45**	**0.951**	**0.036**	**32.67**	**0.991**	**0.021**	**35.26**	**0.972**	**0.011**	**37.86**	**0.985**	**0.005**	**38.92**	0.975	**0.010**	**39.49**	**0.992**	**0.004**
Ours	30.91	0.942	0.040	32.54	0.966	0.026	34.86	0.958	0.025	36.92	0.975	0.008	38.07	**0.981**	0.018	38.95	0.972	0.005

Table 4. **Quantitative SCI image reconstruction comparisons on SCINeRF dataset** The results are computed from the rendered images from estimated scenes via our proposed method, and recovered images from state-of-the-art SCI image restoration methods on six scenes in static SCINeRF dataset: *Airplants, Hotdog, Cozy2room, Tanabata, Factory* and *Vender*. We apply conventional evaluation metrics including PSNR, SSIM and LPIPS. The experimental results demonstrate that our proposed SCIGaussian-D can render images with comparable quality with SCISplat, the current state-of-the-art method on static SCI datasets. The results on static datasets further prove that our method maintains high quality on reconstructing static regions from a single SCI measurement.

	Clock			Lego			Jump			Punch			Tank		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
GAP-TV+Deformable3DGS	18.78	0.460	0.501	17.71	0.461	0.536	18.95	0.358	0.541	20.71	0.461	0.416	23.67	0.669	0.372
PnP-FFDNet+Deformable3DGS	26.92	0.893	0.100	24.91	0.907	0.135	31.77	0.962	0.194	28.30	0.881	0.166	29.32	0.905	0.129
PnP-FastDVDNet+Deformable3DGS	27.34	0.887	0.099	24.57	0.893	0.112	31.12	0.952	0.141	28.61	0.900	0.147	31.61	0.955	0.120
EfficientSCI+Deformable3DGS	32.34	0.939	0.062	30.34	0.972	0.066	34.58	0.982	0.087	29.22	0.916	0.130	32.36	0.962	0.093
ours	**34.76**	**0.970**	**0.014**	**33.73**	**0.987**	**0.015**	**37.24**	**0.991**	**0.013**	**33.74**	**0.962**	**0.024**	**34.66**	**0.980**	**0.022**

Table 5. **Quantitative SCI image reconstruction comparisons on the synthetic datasets with naive baselines** The results are computed from the rendered images from estimated scenes via our proposed method, and rendered images from original deformable 3DGS with reconstructed images from existing SOTA methods as input. The experimental results demonstrate that our proposed method outperforms existing naive two-stage approaches.

port reliable SfM for estimating camera poses and point clouds. To make these baselines operational, we initialize deformable 3DGS using camera poses and point clouds computed from ground-truth images with COLMAP [38]. This setting gives the baselines an inherent advantage, as they exploit accurate poses that are unavailable in a real test scenario, thereby making the comparison unfavorable to our method. Despite this bias, the quantitative results in Table 5 and qualitative results in Figure 8 show that our method still delivers superior performance in 3D scene representation compared with 2-stage naive baselines, underscoring its robustness and effectiveness even under disadvantageous conditions.

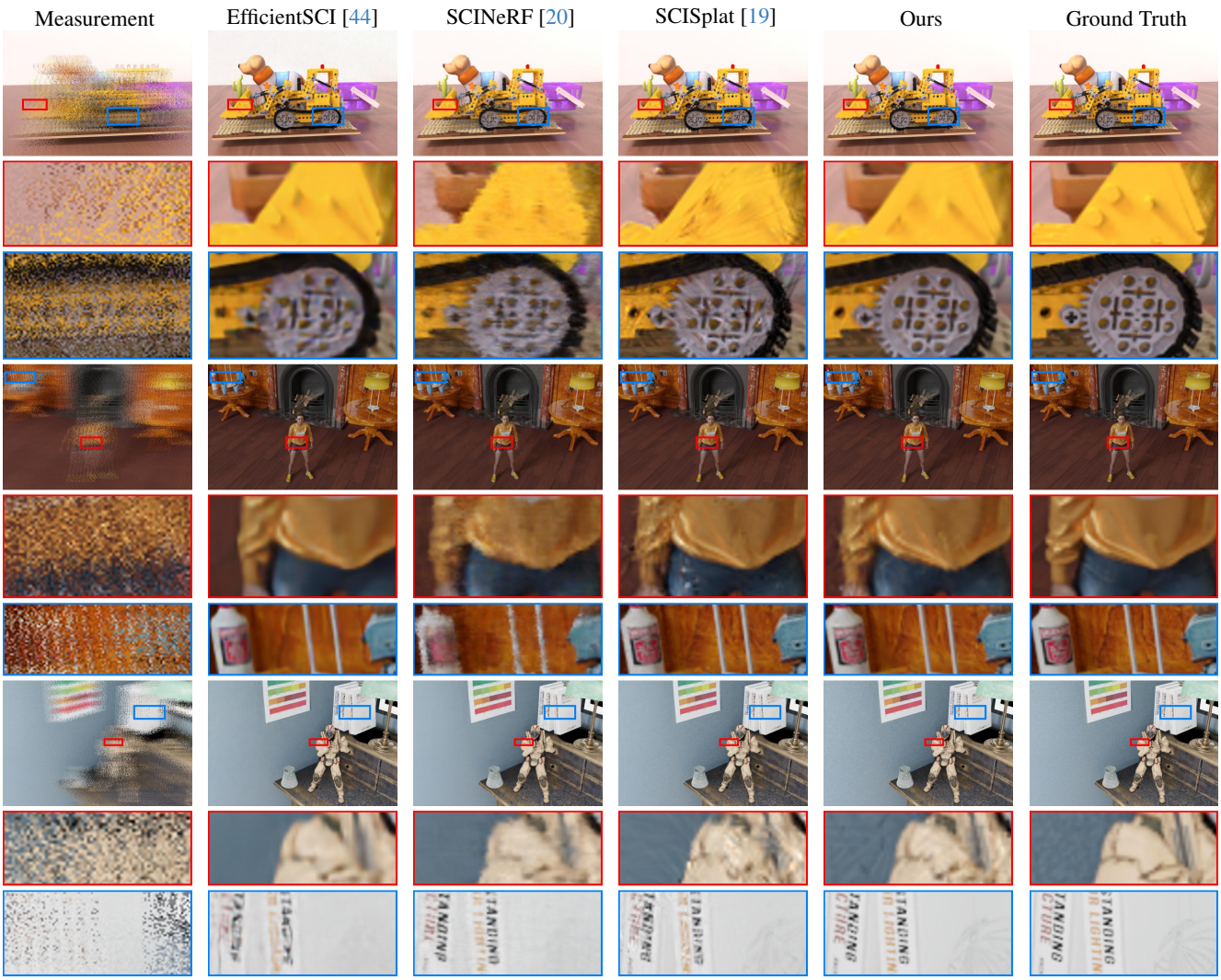

Figure 5. **Additional Qualitative evaluations of our method against SOTA SCI image restoration methods on the synthetic dataset.** Top to bottom shows the results for different scenes, including *Lego*, *Jump* and *Punch*. The experimental results demonstrate that our method achieves superior performance on dynamic image restoration from a single compressed image (the far-left column).

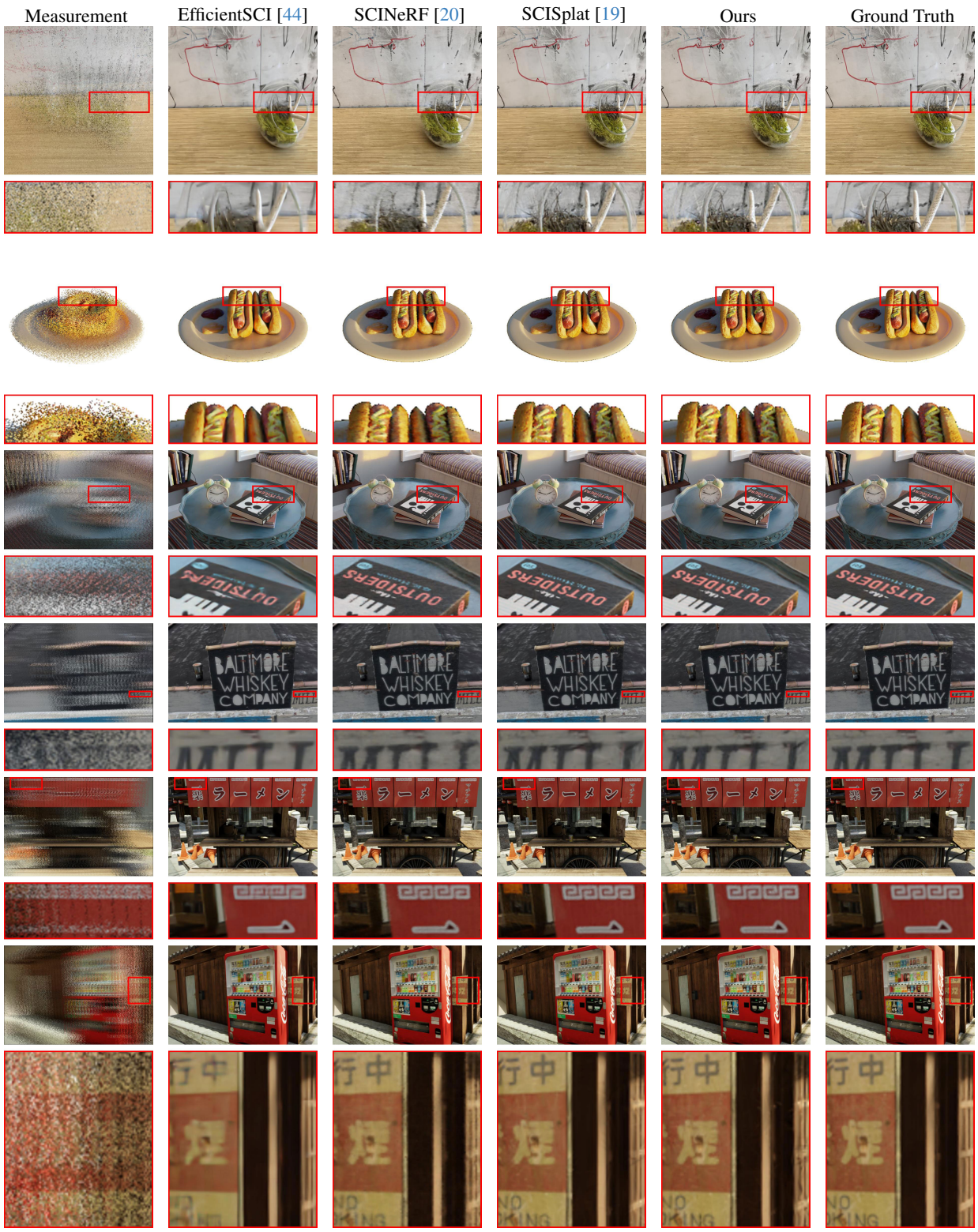

Figure 6. **Qualitative evaluations of our method against SOTA SCI image restoration methods on the static SCINeRF synthetic dataset.** Top to bottom shows the results for different scenes, including *Airplants*, *Hotdog*, *Cozy2room*, *Factory*, *Tanabata* and *Vender*. The experimental results demonstrate that our method achieves comparable performance on static scene representation with SCISplat, the current state-of-the-art static SCI 3D scene representation method.

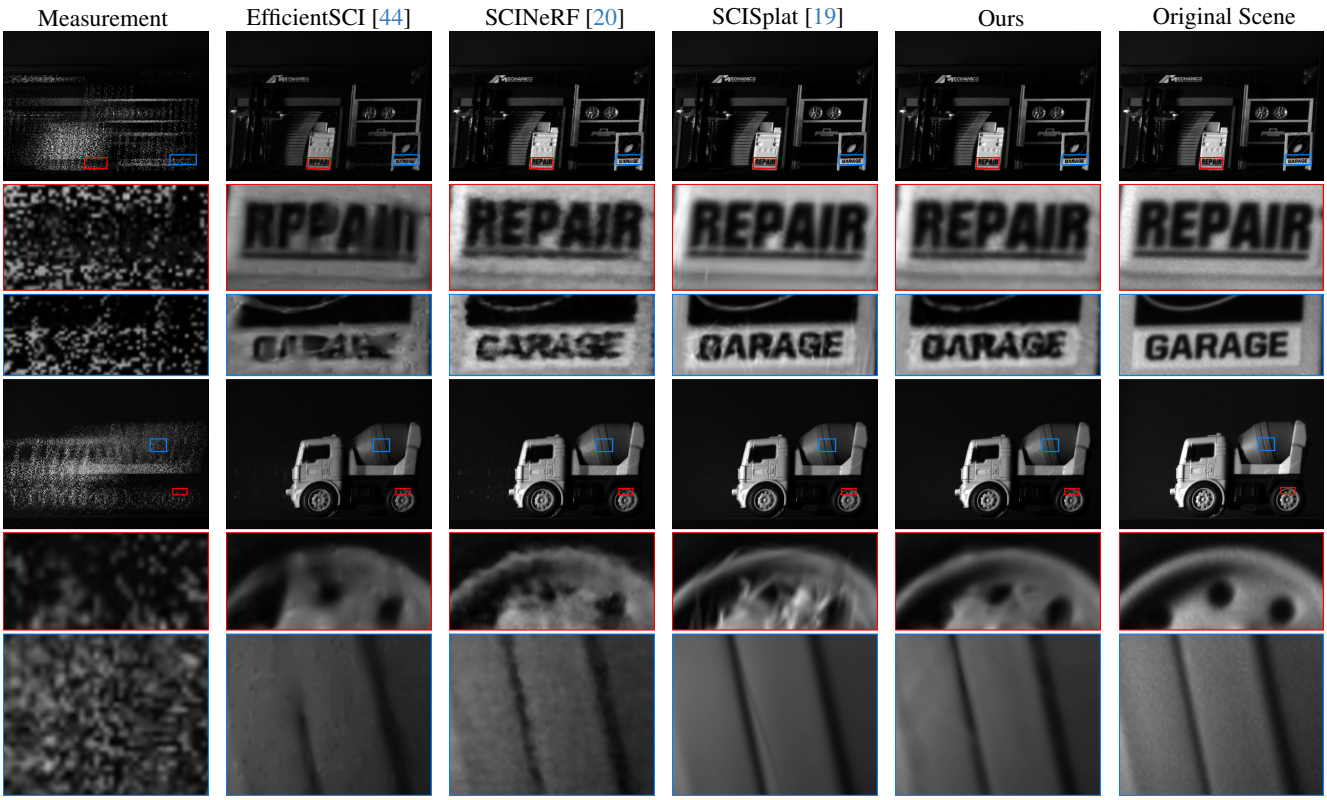

Figure 7. **Qualitative evaluations of our method against SOTA SCI image restoration methods on the SCINeRF real dataset.** Top to bottom shows different scenes. The results demonstrate that our proposed method surpasses existing image restoration methods on real datasets by effectively retrieving dynamic regions of the scene and maintaining high reconstruction quality on static regions of the scene.

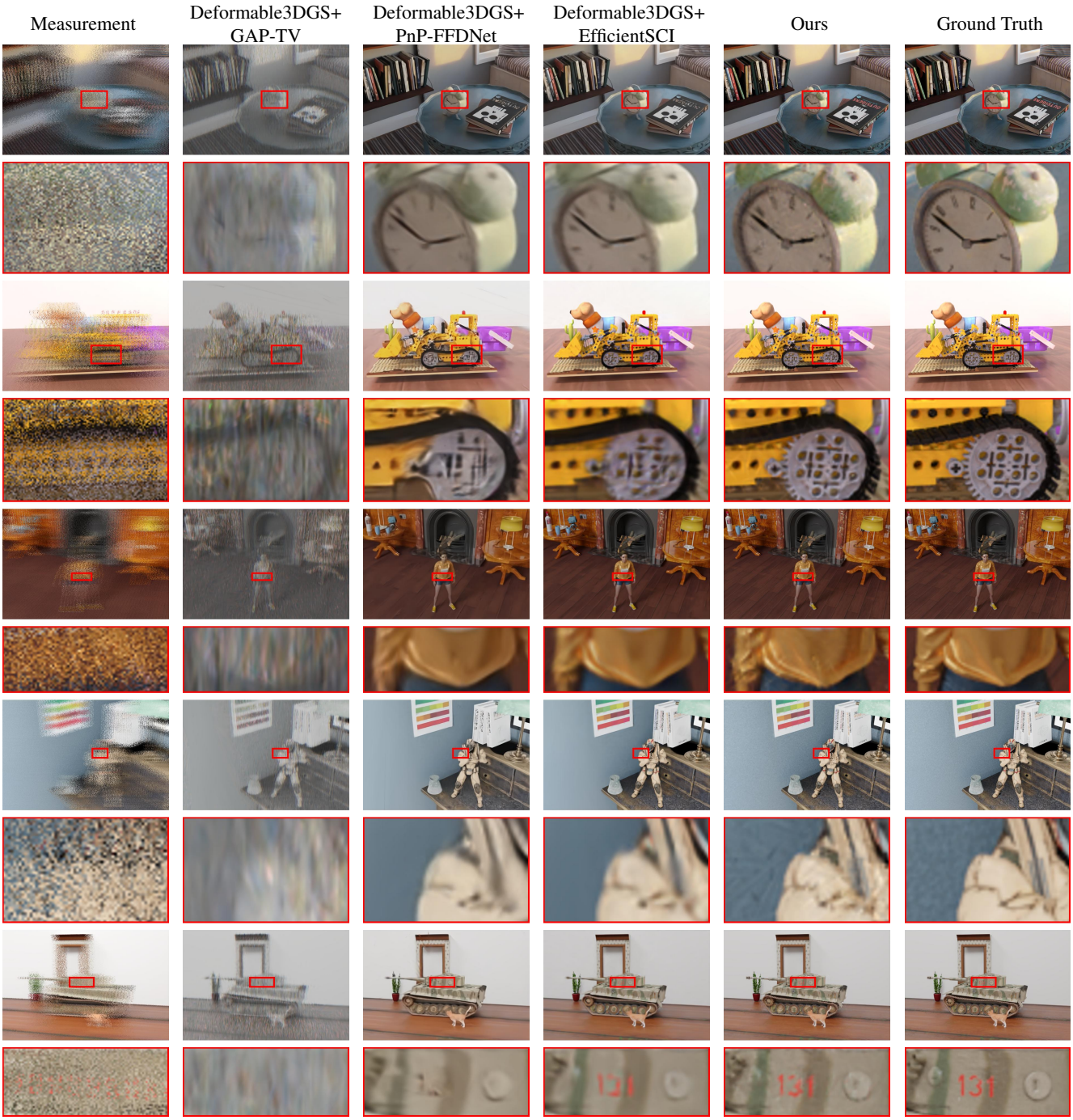

Figure 8. **Qualitative evaluations of our methods against naive two-stage baselines.** We compared the quality of synthesized images from our methods against that of original deformable 3DGS with images from SOTA methods as input. Top to bottom shows different scenes. The qualitative comparisons demonstrate that our methods outperforms the prior naive two-stage approaches.