PAPR Up-close: Close-up Neural Point Rendering without Holes

Supplementary Material

6. Video Results

Please find the video results in the supplementary zip file, which includes 360-degree renderings of various scenes produced using our method. We also provide renderings of the same scenes generated with PAPR [46] for comparison.

7. Additional Results

7.1. Close-up Real Dataset

Fig. 9 shows renderings from distant to close-up views of real-world scenes in the close-up real dataset, comparing PAPR Up-close with PAPR[46]. This comparison shows that the proposed method significantly enhances the model's ability to fill holes in close-up renderings while preserving high-quality renderings of distant views in real-world scenarios.

7.2. Tanks and Temples Dataset

Fig. 10 shows qualitative comparisons of our method, PAPR Up-close, with the baselines on a subset of the Tanks & Temples [15], following the same data pre-processing steps as in [41]. For all experiments, the models are trained using the original distant training views in the dataset. During testing, close-up results are generated by moving the camera closer along a testing view direction, while maintaining a fixed focal length. As ground truth images for close-up views are unavailable, we show zoomed-in views of the most similar ground truth images from the same view directions. These images are presented solely for comparison purposes within the figure. The results demonstrate that PAPR Up-close outperforms in producing images that are both hole-free and rich in details. In comparison, the baseline methods frequently encounter difficulties in achieving hole-free renderings, for example, SNP and PAPR on the Family scene and Point-NeRF on the Caterpillar scene, and tend to introduce either artifacts or blurriness, for example, 3DGS and Mip-Splatting on all the scenes.

8. Additional Sensitivity Analysis

8.1. Hyperparameters in DPS

In the proposed Dynamic Point Selection (DPS) method, we classify the points by the q^{th} 10-quantiles of the K attention weights, where $q \in \{1, 2, \ldots, 10\}$, as mentioned in Sec. 3.3. In Fig. 11, we show how different values of q affect the quantitative results on the Chair, Hotdog, and Lego scenes in the close-up synthetic dataset. As shown, DPS achieves the best rendering quality when $q = 2$.

8.2. Hyperparameters in RPS

In Table 2 and Fig. 12 we evaluate how different values of k, which controls the magnitude of perturbation of the rays, affect the quantitative results on Chair, Hotdog, and Lego scenes from the close-up synthetic dataset. The results show that both excessively small and excessively large values of k may result in a decline in rendering quality. Additionally, larger k values can lead to a decline in the rendering quality of high-frequency details, such as the patterns on the chair. Based on these observations, we select either $k = 1$ or $k = 5$ for our experiments, preferring smaller values of k for scenes with more high-frequency details (e.g., Chair and Lego).

9. Data Collection

9.1. Close-up Real Dataset

We capture two scenes, Bowser and Cup, from views surrounding the object using a Oneplus 8T smartphone at two distances. The camera is approximately 1 meter from the object for distant views, and 0.2 meter from the object for close-up views. We then use COLMAP [30, 31] to reconstruct the point cloud and camera poses. For the Bowser scene, we capture 165 images from distant views as the training set, and 51 images from close-up views as the testing set. For the cup scene, we capture 141 images from distant views as the training set, and 26 images from close-up views as the testing set. We then digitally remove the background from each captured image. Samples from the dataset are shown in Fig. 13.

9.2. Close-up Synthetic Dataset

As described in Sec. 4.1, we create a close-up synthetic dataset by rendering scenes in the NeRF-synthetic dataset [21] from close-up views using Blender. For each training view in the original NeRF-synthetic dataset, we first calculate the minimum depth d of a crop with size $\frac{H}{4} \times \frac{W}{4}$ centering at the original ground truth depth map, where the size of the depth map is $H \times W$. Intuitively, d gives us an estimation of how close we can move the camera towards the object, such that everything is still in front of the camera. For each training view, we translate the camera center along the optical axis by $d - 1$ to get the close-up camera and rendering in our new dataset, where the 1 is the default distance between the camera center and image plane in the NeRF-synthetic dataset, results in around $3\times$ closer camera views. Samples from the close-up synthetic dataset are shown in Fig. 14.

Figure 9. Qualitative comparison of novel view synthesis of our method and PAPR [46] on Bowser and Cup scenes. The images are rendered from the cameras located at different distances L from the object along the same view direction. As shown in the figure, the proposed PAPR Up-close significantly improves the model's hole-filling ability for the close-up renderings, while maintaining high-quality renderings of distant views compared to PAPR.

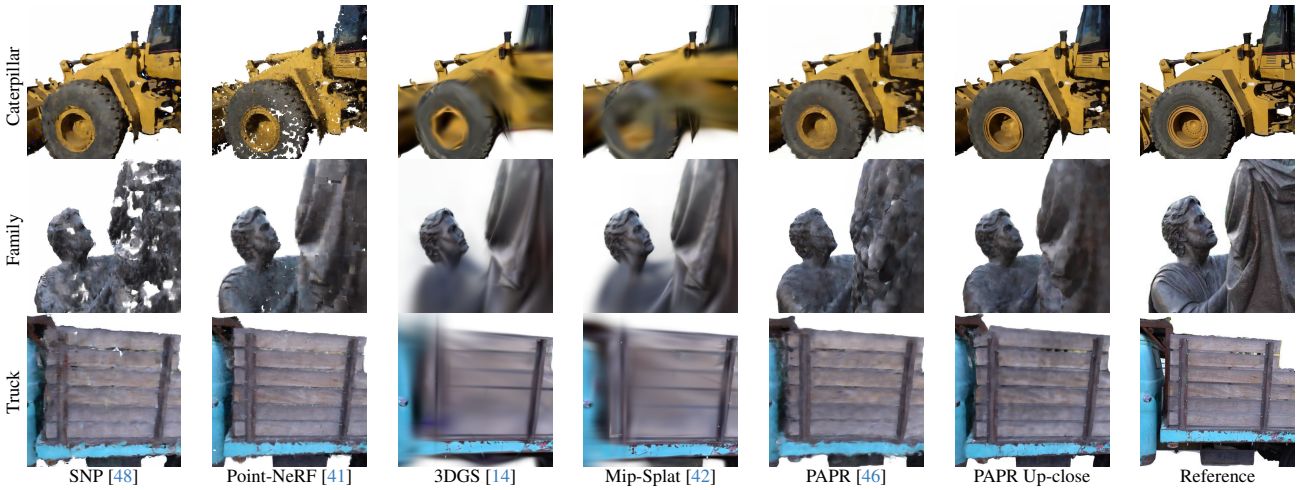

Figure 10. Qualitative comparison of novel view synthesis between our method, PAPR Up-close, and the more competitive baselines on Tanks & Temples [15] subset. All baselines use a total number of $30,000$ points. The figure shows close-up results generated by adjusting the camera closer along a testing view direction with the focal length fixed. We use the most similar zoomed-in ground truth images from the same view directions as references for the true close-up views. The results show that our method, PAPR Up-close, achieves the best rendering quality without creating holes evaluating under unseen close-up views in challenging real-world scenes.

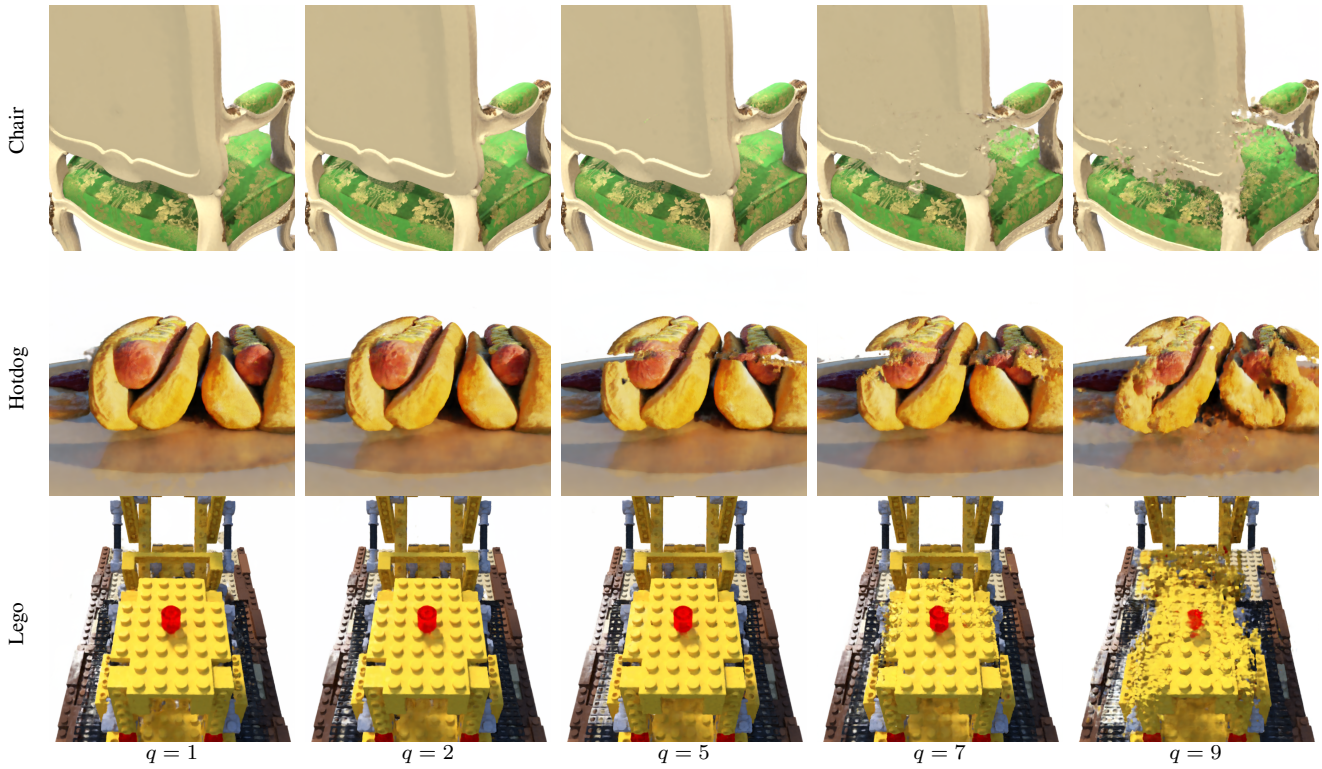

Figure 11. Quantitative results for varying values of q in the Dynamic Point Selection (DPS) method on Chair, Hotdog, and Lego scenes from the close-up synthetic dataset.

	Chair			Hotdog			Lego		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
w/o RPS	23.45	0.871	0.265	25.72	0.889	0.272	24.48	0.842	0.341
$k = 0.5$	23.46	0.873	0.268	26.42	0.898	0.270	24.62	0.843	0.352
$k = 1$	23.44	0.875	0.268	26.50	0.896	0.276	24.55	0.842	0.360
$k = 2$	23.33	0.876	0.276	26.57	0.898	0.284	24.49	0.842	0.372
$k = 5$	23.20	0.870	0.284	26.65	0.900	0.288	24.33	0.839	0.381
$k = 10$	23.18	0.871	0.286	26.52	0.899	0.290	24.40	0.842	0.381

Table 2. Quantitative comparison of the impact of varying values of k, which determine the extent of ray perturbation, on Chair, Hotdog, and Lego scenes from the close-up synthetic dataset.

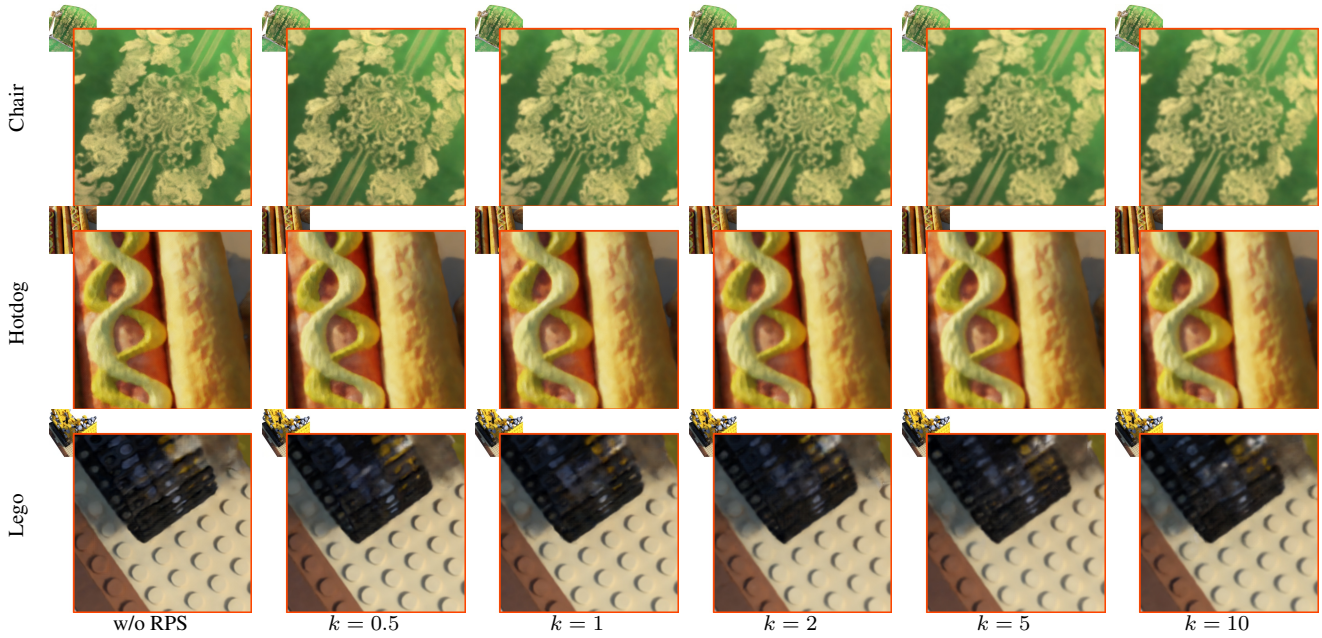

Figure 12. Qualitative comparison of the impact of varying values of k, which determine the extent of ray perturbation, on Chair, Hotdog, and Lego scenes from the close-up synthetic dataset.

Figure 13. Sample images from our close-up real dataset.

Figure 14. Sample images from our close-up synthetic dataset, generated from the same scenes in the NeRF synthetic dataset [21].