

A. Impact Statements

This paper presents work whose goal is to advance the field of Novel View Synthesis. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

B. More Implementation Details of Gaussian Progressive Propagation

Neighboring view selection. In Gaussian progressive propagation, we need to select neighboring views to evaluate color or geometric consistency and determine the propagated depth and normal. For video inputs in the Waymo dataset, we directly choose two consecutive frames before and after the current frame as neighboring views. For an unordered collection of images in MipNeRF360, we determine neighboring frames based on the number of SfM (Structure from Motion) points shared between different frames. During propagation, to improve the accuracy of depth and normals, we generate additional hypotheses for each local plane by randomly perturbing the values of distance and normal.

Derivation of homography transformation. The derivation of homography transformation \mathbf{H} used in Sec. 4.3 is as follows. For a pixel with its coordinate \mathbf{p} and its depth z , we first project it back to the 3D space as a 3D point \mathbf{x} :

$$\mathbf{x} = z\mathbf{K}^{-1}\tilde{\mathbf{p}}, \quad (11)$$

where \mathbf{K} is the camera intrinsic, $\tilde{\mathbf{p}}$ is the homogeneous coordinate of \mathbf{p} . Based on the relative transformation $[\mathbf{W}_{\text{rel}}, \mathbf{t}_{\text{rel}}]$ between the current view and the neighboring view, we transform point \mathbf{x} from the current viewpoint to the neighboring viewpoint, and then project it back to the 2D space as \mathbf{p}' :

$$\mathbf{p}' \simeq \mathbf{K}(\mathbf{W}_{\text{rel}}\mathbf{x} + \mathbf{t}_{\text{rel}}). \quad (12)$$

Since point \mathbf{x} lies on a local 3D plane parameterized as (d, \mathbf{n}) , which satisfies:

$$\mathbf{n}^\top \mathbf{x} + d = 0. \quad (13)$$

Substituting Eq. 13 into Eq. 12 as:

$$\begin{aligned} \tilde{\mathbf{p}}' &\simeq \mathbf{K} \left(\mathbf{W}_{\text{rel}}\mathbf{x} - \frac{\mathbf{t}_{\text{rel}}\mathbf{n}^\top \mathbf{x}}{d} \right) \\ &\simeq \mathbf{K} \left(\mathbf{W}_{\text{rel}} - \frac{\mathbf{t}_{\text{rel}}\mathbf{n}^\top}{d} \right) \mathbf{x} \\ &\simeq \mathbf{K} \left(\mathbf{W}_{\text{rel}} - \frac{\mathbf{t}_{\text{rel}}\mathbf{n}^\top}{d} \right) \mathbf{K}^{-1}\tilde{\mathbf{p}} \\ &\simeq \mathbf{H}\tilde{\mathbf{p}} \end{aligned} \quad (14)$$

C. More Rendering Results

In this section, we present more results on Waymo (Fig. 8) and MipNeRF360 (Fig. 9) datasets. Compared to 3DGS, our method achieves more accurate rendering results and improved geometric structures, particularly the large-scale scenes on Waymo. Additionally, the results on MipNeRF360 reveal that 3DGS already optimizes accurate geometry, especially depth, for small-scale scenes with rich textures. Therefore, the improvements achieved by our method on the MipNeRF360 dataset are limited.

D. Results of Each Scene in Waymo and MipNeRF360

Table 5-10 presents the evaluation metrics for each scene in Waymo and MipNeRF360. For the MipNeRF360 dataset, we include the metrics from the original 3DGS paper (3DGS*), as well as the metrics obtained by rerunning 3DGS with new poses generated using COLMAP.

In the Waymo dataset, we achieve state-of-the-art (SOTA) rendering results for each scene, which significantly outperform the baseline 3DGS. In the MipNeRF360 dataset, we achieve improvements over 3DGS in indoor scenes (the first four rows), which contain many structured planes. In natural scenes with rich textures but lacking structured planes, our results

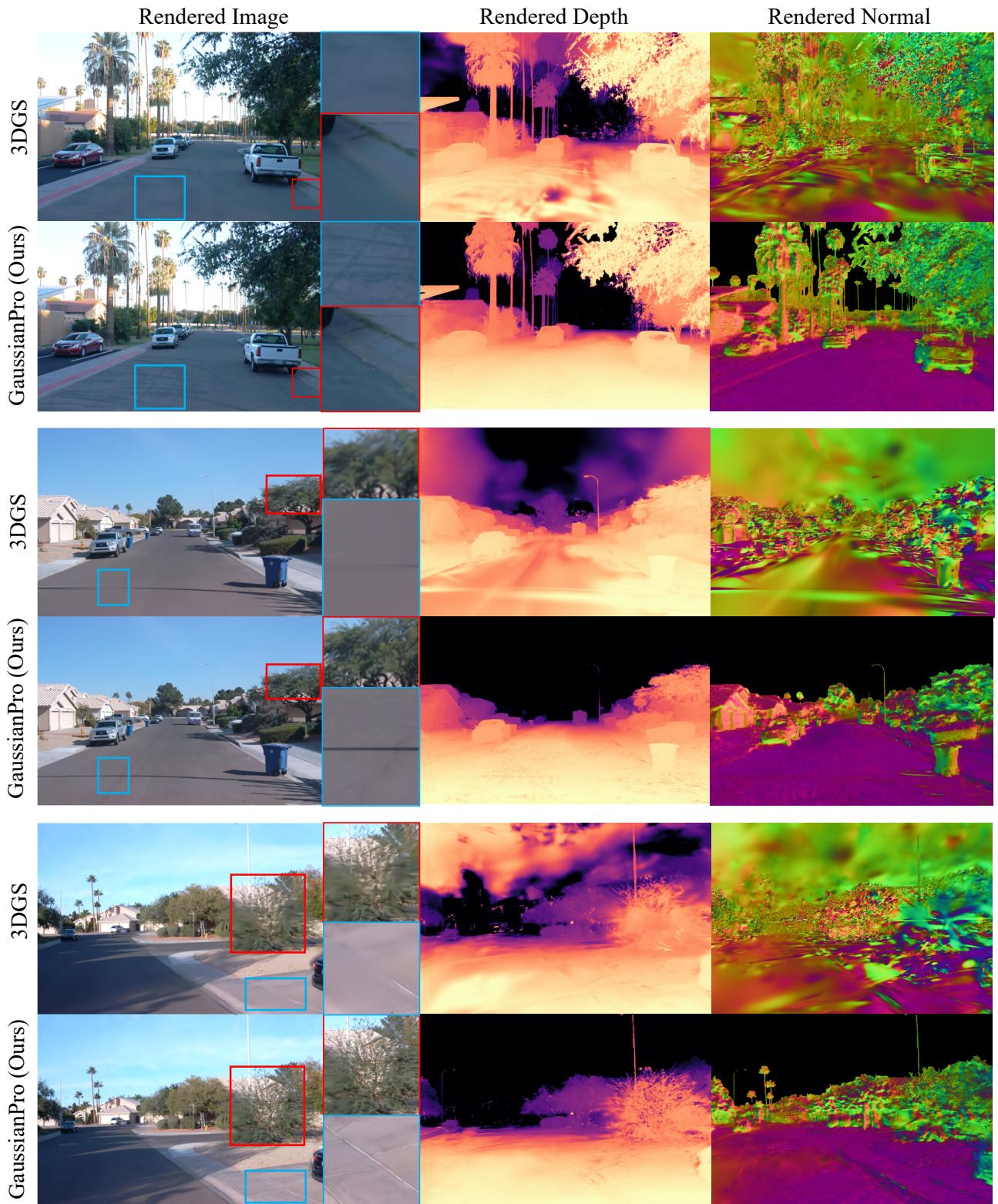


Figure 8: More rendering results on the Waymo dataset.

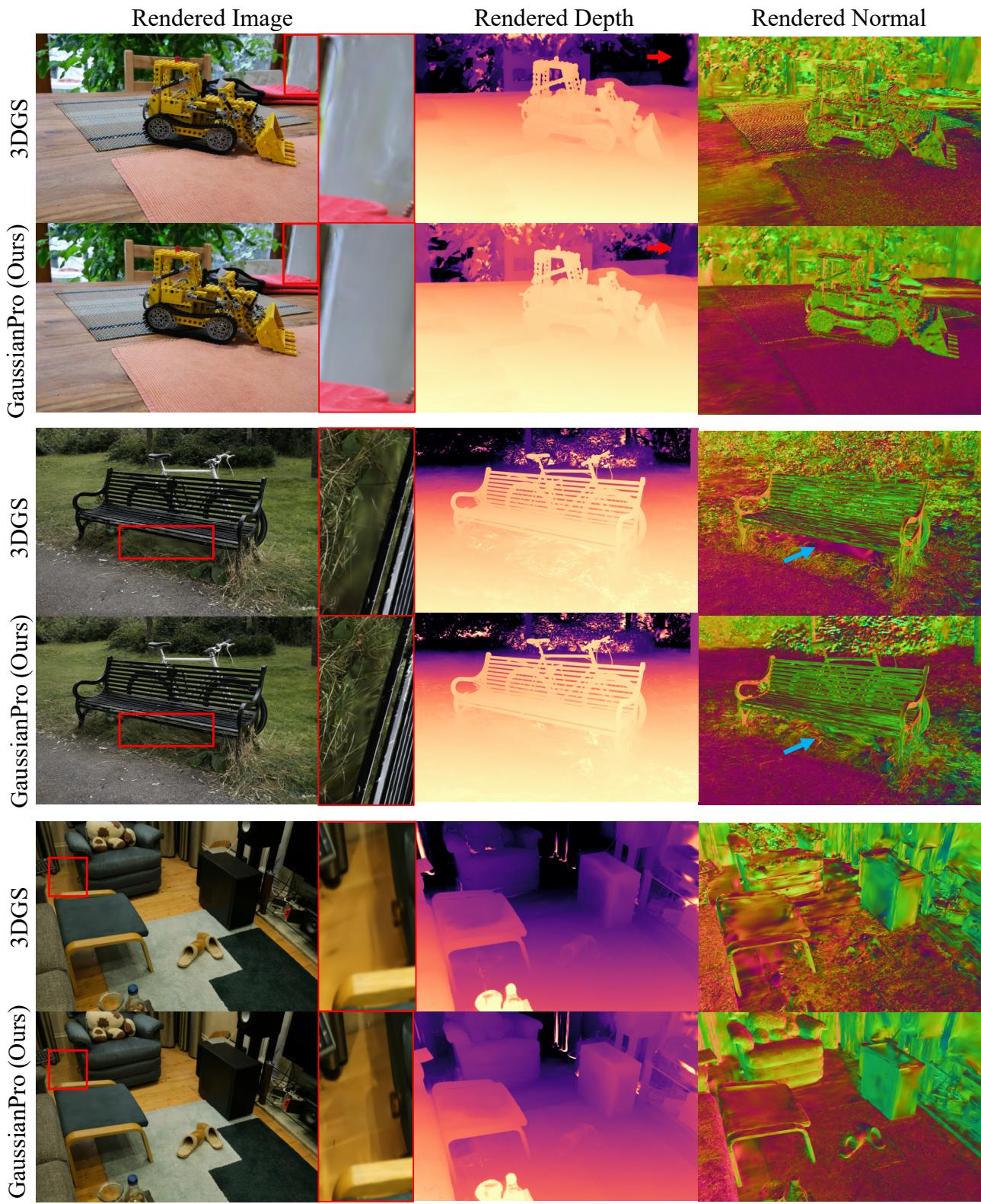


Figure 9: More rendering results on the MipNeRF360 dataset.

Table 5: SSIM results on the Waymo dataset.

Sequence	SSIM				
	NGP	Mip360	Zip	3DGS	Ours
Seg100613	0.823	0.927	<u>0.954</u>	0.947	0.958
Seg100275	0.917	0.952	<u>0.957</u>	0.950	0.960
Seg113792	0.905	0.934	<u>0.947</u>	0.933	0.948
Seg132384	0.935	0.948	0.964	0.957	<u>0.963</u>
Seg144248	0.852	0.762	0.896	<u>0.938</u>	0.939
Seg148697	0.856	0.885	<u>0.924</u>	0.913	0.934
Seg150623	0.928	0.937	<u>0.963</u>	0.960	0.968
Seg164701	0.855	0.909	<u>0.923</u>	0.917	0.927
Seg405841	0.899	0.924	0.927	<u>0.931</u>	0.940
Average	0.886	0.909	<u>0.939</u>	0.938	0.949

Table 6: PSNR results on the Waymo dataset.

Sequence	PNSR				
	NGP	Mip360	Zip	3DGS	Ours
Seg100613	33.67	31.07	<u>36.55</u>	35.71	36.83
Seg100275	32.38	34.27	36.17	35.05	<u>36.08</u>
Seg113792	32.63	32.67	35.76	33.26	<u>34.87</u>
Seg132384	33.03	32.90	37.10	35.52	<u>36.54</u>
Seg144248	28.92	21.17	<u>32.12</u>	34.54	35.21
Seg148697	27.95	27.20	<u>30.88</u>	29.75	30.96
Seg150623	33.19	31.16	<u>37.59</u>	36.94	38.23
Seg164701	27.35	29.40	<u>29.90</u>	29.38	30.48
Seg405841	29.70	30.93	<u>31.93</u>	31.63	32.91
Average	30.98	30.09	<u>34.22</u>	33.53	34.68

are comparable to 3DGS. In these scenes, SfM techniques usually provide a high-quality point cloud for initialization and the simple clone and split densification strategies don't show a bottleneck. Besides, these scenes contain intricate fine structures that cover a few pixels, such as the grass and leaves, making it challenging to accurately estimate surface normals. Although ZipNeRF achieves the best rendering quality on the MipNeRF360 dataset, its rendering speed is significantly slower compared to ours (0.09 FPS for ZipNeRF compared to 108 FPS for our method).

Table 7: LPIPS results on the Waymo dataset.

Sequence	LPIPS				
	NGP	Mip360	Zip	3DGS	Ours
Seg100613	0.237	0.201	<u>0.178</u>	0.209	0.158
Seg100275	0.262	0.257	0.200	0.233	<u>0.204</u>
Seg113792	0.295	0.289	0.216	0.239	<u>0.233</u>
Seg132384	0.234	0.200	0.186	0.218	<u>0.190</u>
Seg144248	0.364	0.396	0.303	<u>0.239</u>	0.233
Seg148697	0.323	0.230	<u>0.185</u>	0.224	0.152
Seg150623	0.246	0.214	<u>0.191</u>	0.214	0.183
Seg164701	0.292	0.275	<u>0.194</u>	0.212	0.179
Seg405841	0.277	0.292	<u>0.195</u>	0.218	0.186
Average	0.281	0.262	<u>0.205</u>	0.223	0.191

Table 8: SSIM results on the MipNeRF360 dataset.

Sequence	SSIM					
	NGP	Mip360	Zip	3DGS*	3DGS	Ours
Room	0.871	0.913	<u>0.925</u>	0.914	0.919	0.927
Counter	0.817	0.894	0.902	0.905	<u>0.915</u>	0.916
Kitchen	0.858	0.920	0.928	0.922	<u>0.933</u>	0.935
Bonsai	0.906	0.941	<u>0.949</u>	0.938	0.945	0.952
Bicycle	0.512	0.685	0.769	<u>0.771</u>	0.810	0.810
Flowers	0.486	0.583	0.642	<u>0.605</u>	0.603	0.598
Garden	0.701	0.813	0.860	<u>0.868</u>	0.890	0.890
Stump	0.594	0.744	0.800	<u>0.775</u>	0.769	0.763
Treehill	0.542	0.632	0.681	0.638	0.636	0.631
Average	0.699	0.792	0.828	0.815	0.824	<u>0.825</u>

Table 9: PSNR results on the MipNeRF360 dataset.

Sequence	PSNR					
	NGP	Mip360	Zip	3DGS*	3DGS	Ours
Room	29.69	31.63	32.65	30.63	31.71	<u>31.98</u>
Counter	26.69	29.55	<u>29.38</u>	28.70	29.06	29.17
Kitchen	29.48	<u>32.23</u>	32.50	30.32	31.57	31.60
Bonsai	30.69	<u>33.46</u>	34.46	31.98	32.69	33.05
Bicycle	22.17	24.37	25.80	25.25	26.64	<u>26.60</u>
Flowers	20.65	21.73	22.40	21.52	21.64	21.54
Garden	25.07	26.98	28.20	27.41	28.77	<u>28.70</u>
Stump	23.47	26.40	27.55	26.55	<u>26.58</u>	26.42
Treehill	22.37	<u>22.87</u>	23.89	22.49	22.30	22.21
Average	25.59	27.69	28.54	27.21	27.88	<u>27.92</u>

Table 10: LPIPS results on the MipNeRF360 dataset.

Sequence	LPIPS					
	NGP	Mip360	Zip	3DGS*	3DGS	Ours
Room	0.261	0.211	<u>0.196</u>	0.220	0.192	0.192
Counter	0.306	0.204	<u>0.185</u>	0.204	<u>0.185</u>	0.175
Kitchen	0.195	0.127	<u>0.116</u>	0.129	0.113	0.113
Bonsai	0.205	0.176	0.173	0.205	<u>0.169</u>	0.163
Bicycle	0.446	0.301	0.208	0.205	0.194	0.195
Flowers	0.441	0.344	0.273	<u>0.336</u>	0.346	0.342
Garden	0.257	0.170	0.118	0.103	<u>0.107</u>	<u>0.107</u>
Stump	0.421	0.261	0.193	<u>0.210</u>	0.236	0.244
Treehill	0.450	0.339	0.242	<u>0.317</u>	0.337	0.344
Average	0.331	0.237	0.189	0.214	0.209	<u>0.208</u>