Learnable Infinite Taylor Gaussian for Dynamic View Rendering

Bingbing Hu[1*] Yanyan Li[2*] Rui Xie[1] Bo Xu[2] Haoye Dong[2] Junfeng Yao [1,3†] Gim Hee Lee[2]
[1] School of Film, School of Informatics, Xiamen University
[2] National University of Singapore
[3] Key Laboratory of Digital Protection and Intelligent Processing of Intangible Cultural Heritage of Fujian and Taiwan Ministry of Culture and Tourism

A. Novel View Rendering

In the evaluation section, we have designed a series of comprehensive experiments to assess the performance of our method. Here, we present a set of **visual results** to further validate the effectiveness of our approach more thoroughly.

A.1. Qualitative Analysis of Details

In the quantitative analysis of novel view rendering algorithms, we focused on several key evaluation metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) and Perceptual Image Patch Similarity (LPIPS) [3]. These metrics help us quantify the details and overall quality of the reconstructed images. By comparing these metrics, we can more accurately assess the performance differences between different algorithms. To gain a more comprehensive understanding, we also incorporated qualitative analysis, examining the detailed performance of various algorithms in the reconstructed images, leading to a deeper evaluation. Through visual presentation, we can further assess the strengths and weaknesses of the algorithms, ensuring a multidimensional and comprehensive understanding of their performance.

As shown in Figure 1, we can see that our algorithm demonstrates superior performance in detail reconstruction compared to others. However, D3DGS and 4DGS face challenges such as artifacts and distortions during the reconstruction process. We provide a detailed explanation of each row in Figure 1:

First Row: Overall, the texture reconstruction quality of D3DGS and 4DGS is below the standard. Additionally, as seen in the red box (curtain reconstruction) and the blue box (wall reconstruction), the detail reconstruction performance of D3DGS and 4DGS is also poor.

Second Row: The overall reconstruction performance of D3DGS is poor. Both 4DGS and D3DGS exhibit issues in detail reconstruction, such as blurred shadows in the yellow box, significant reflections and artifacts on the leather stool in the red box, and additional artifacts appearing in the blue box for D3DGS.

Third Row: D3DGS performs poorly in both overall reconstruction quality and detail representation. 4DGS also has some issues in detail reconstruction, such as unexplained black spots above the white bottle in the green box and unexplained light appearing on the left side of the blue box.

Fourth Row: Both D3DGS and 4DGS exhibit color distortions. Additionally, shadows appear in certain areas (e.g., red, yellow, and blue boxes), and the image in the blue box lacks contrast. There are also extraneous elements in the green box of D3DGS.

A.2. Qualitative Analysis of Ablation Experiments

In our ablation experiments on the *Sear Steak* class in the N3DV dataset, we conducted an in-depth qualitative analysis to evaluate the impact of ablating different modules on the performance of reconstructed images and the representation of fine details. By systematically comparing images reconstructed after the ablation of various modules, we were able to uncover their respective strengths and limitations in handling complex scenes.

First, significant differences were observed in rendering quality across the ablations of different modules. Ablating specific modules reduced the ability to capture geometric details of objects, as shown in Figure 2, such as surface textures and edge contours. For example, as highlighted by the blue bounding box, both *w/o Peano remainder* and *w/o Time-opacity* failed to accurately capture geometric details, leading to missing geometric information. Similarly, as shown in the green bounding box, *w/o Peano remainder*, *w/o Time-opacity*, and *w/o Time-scale* exhibited poor performance in reconstructing surface textures, producing artifacts such as shadowing and linear streaks. Additionally, *w/o Peano remainder* and *w/o Time-opacity* demonstrated a weaker capability in capturing edge contours, resulting in blurred or muddled details during reconstruction.

* These authors contributed equally to this work
† Corresponding author

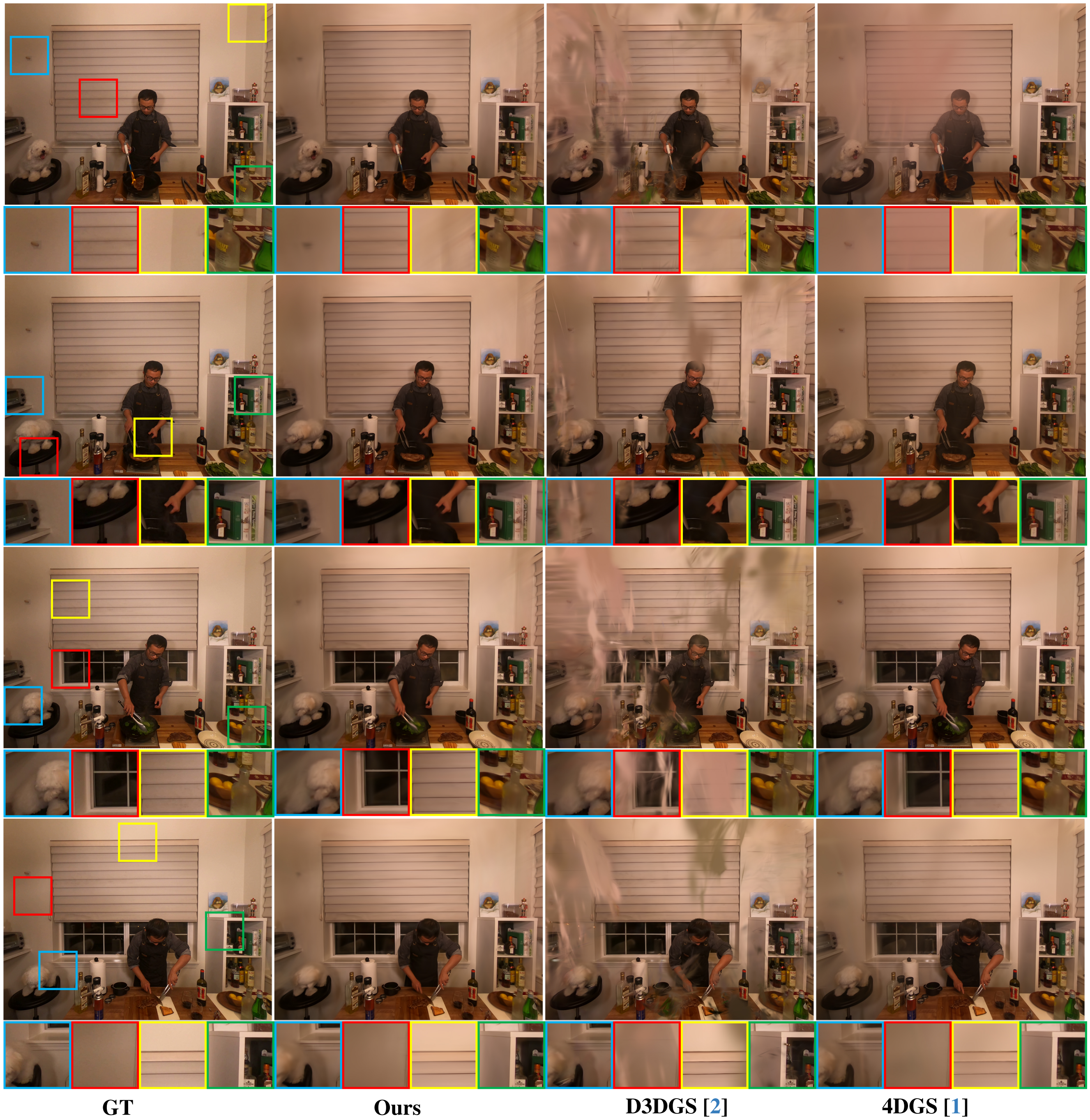

Figure 1. Qualitative analysis of novel view rendering on the N3DV dataset, comparing the detail information of reconstructed images from different algorithms.

In other cases, module ablations introduced noticeable noise or over-smoothing in specific details. For instance, as illustrated in the red bounding box, *w/o Time-motion* and *w/o Time-rotation* introduced significant noise when reconstructing fine details compared to the original images. These differences were particularly pronounced when processing *Sear Steak* samples with rich geometric features, highlighting the critical role of these modules in maintaining reconstruction fidelity. Furthermore, we evaluated the impact of different module ablations on handling complex scenes. As shown in the yellow bounding box, the reconstruction quality of *w/o Time-motion*, *w/o Time-rotation*, and *w/o Time-scale* was relatively blurry, with increased noise and excessive smoothing, ultimately degrading the overall visual quality. This underscores the importance of these modules in accurately capturing fine details in complex scenes.

In summary, this qualitative analysis not only revealed the impact of ablating specific modules on reconstruction quality but also provided deeper insights into their effectiveness in capturing fine details. These findings hold significant implications for optimizing novel view rendering algorithms and improving image quality. By identifying the strengths and limitations of each module, we can better target algorithmic improvements to achieve more accurate and high-quality novel view rendering outcomes.

B. Performance Analysis of Large-Scale Data Scene Reconstruction

To better evaluate the performance of Large-Scale Data Scene Reconstruction, we conducted a qualitative analysis on the Technicolor Light Field Dataset. For a more in-depth assessment, we explored the model's visual performance in handling complex scenes, focusing on its ability to capture fine details and reconstruct object surface textures. By visually comparing the model's outputs with real-world scenes, we gained deeper insights into its strengths and limitations in practical applications.

As highlighted in the boxed regions, it is evident that our method can render higher-quality images. As shown in Figure 3, 4, we can see that in the Birthday scene, our reconstruction captures details better compared to other models. Several issues are observed in the reconstructions of 4DGS and FSGS: in the area marked by the blue box, both methods exhibit reconstruction blurriness; in the area marked by the red box, neither 4DGS nor FSGS successfully reconstructs the yellow object near the person's nose bridge, and the images generated by both methods have relatively lower resolution. Additionally, FSGS introduces motion blur artifacts. In the area marked by the yellow box, both methods make errors in reconstruction, mistakenly generating a red object beneath the green leaf. Lastly, in the area marked by the green box, both 4DGS and FSGS fail to accurately reconstruct the text along the edges.

In the Painter scene, it is evident that our model outperforms other models in reconstruction quality, while both 4DGS and FSGS exhibit the following issues: in the area marked by the blue box, noticeable hand deformation occurs; in the area marked by the red box, significant errors are observed in reconstructing the distance between the clothing and surrounding objects; in the area marked by the yellow box, the clothing texture shows clear differences compared to GT; and in the area marked by the green box, the highlights of the painting are not accurately reconstructed.

References

[1] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20310–20320, 2024. 2

[2] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20331–20341, 2024. 2

[3] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 1

Figure 2. *Sear Steak* Novel View Rendering on the N3DV Dataset: Qualitative Analysis of Ablation Experiments - Comparison of Reconstruction Quality and Detail Representation with Module Ablations.

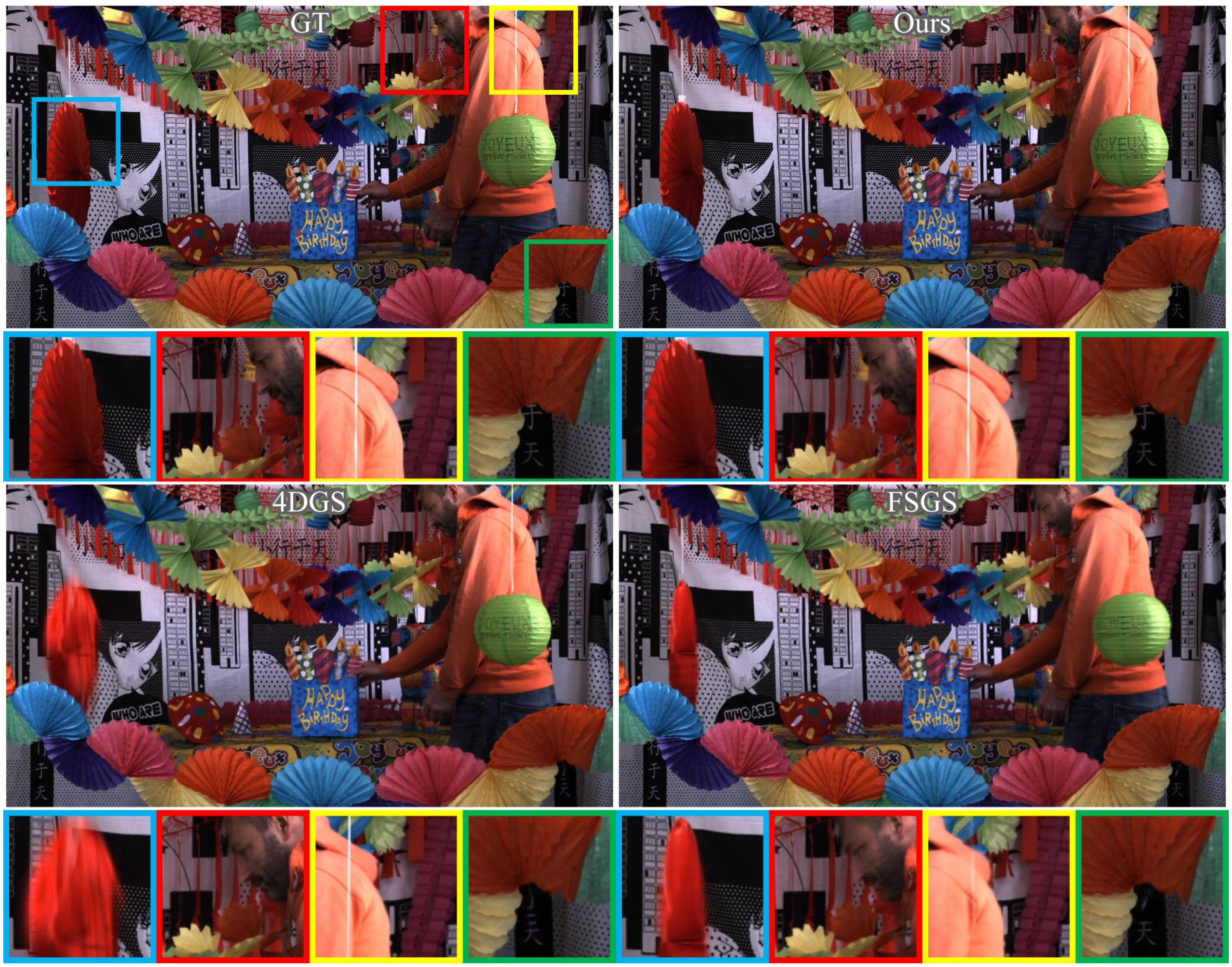

Figure 3. Qualitative analysis of novel view rendering on the Birthday dataset from the Technicolor, comparing the detailed reconstructions of different algorithms.

Figure 4. Qualitative analysis of novel view rendering on the Painter dataset from the Technicolor, comparing the detailed reconstructions of different algorithms.