VersatileGaussian: Real-time Neural Rendering for Versatile Tasks using Gaussian Splatting (Supplementary Material)

Renjie Li[1⋆], Zhiwen Fan[2,†,⋆], Bohua Wang[3], Peihao Wang[2], Zhangyang Wang[2], and Xi Wu[4]

[†] Project Lead
[1]THU, [2]UT Austin, [3]Baidu, [4]CUIT
https://VersatileGaussian.github.io

1 Introduction

In Sec. 2, we provide the implementation details of our model. In Sec. 3, we visualize the attention map from the Task Correlation Attention. We also provide evaluations on the extra SceneNet dataset. In Sec. 4, we provide more qualitative results of our method on ScanNet, Replica and SceneNet.

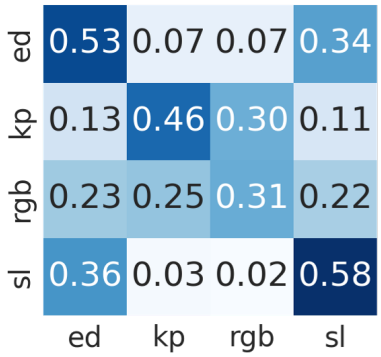

Fig. 1: Visualization of Attention Map. We visualize the attention map by averaging over all attention heads. ED, KP, RGB, and SL stand for edge detection, keypoint detection, RGB image synthesis, and semantic segmentation. We conduct the experiment on ScanNet and the attention maps are averaged over all scenes.

2 Implementation Details

Similar to 3D Gaussian Splatting [1], We optimize VersatileGaussian by 30k iterations and densify the Gaussians from the 500^{th} iteration to the 15000^{th} iteration every 100 iterations with gradient threshold set to 0.0002. We reset the opacity every 3000 iterations. The learning rate for Gaussian position decays from 1.6×10^{-4} to 1.6×10^{-6}. The learning rate for the features attached to each Gaussian and the Task Correlation Attention is set to be 0.0025. The learning rate for the opacity, scaling, and rotation is 0.05, 0.005, and 0.001 respectively.

3 Additional Experiments

Visualization of the Attention Map. We conduct further analysis of the Task Correlation Attention. We visualize the attention map in Fig. 1 by averaging over

⋆ Equal Contribution.

all heads. We experiment on Scannet, involving the first four tasks consisting of ED, KP, RGB, SL. We then train ED and KP with other tasks and compare the performance to that trained alone. We observe that, for the ED task, among KP, RGB and SL, SL brings the most performance improvement, where the L1 error is boosted from 0.027 to 0.021 (the lower the better), a 22% improvement. For the KP task, RGB brings the most performance improvement (from 0.016 to 0.009, a 43% improvement). Notablely, as can be found in Fig. 1, except the diagonal elements, ED is of largest attention score with SL (0.34), and KP with RGB (0.30), indicating that the higher attention value correlates with a stronger mutual correlation.

Comparison on SceneNet. We evaluate VersatileGaussian on the SceneNet [2] testing set provided by MuvieNeRF [4]. The SceneNet testing set consists of 4 scenes, each of which consists of 32 source views and 8 target views. We train VersatileGaussian on all source views and evaluate on target views. We only evaluate VersatileGaussian and the performances of the rest approaches are derived from [4]. The overall drop measures how much the other approaches underperform VersatileGaussian in multi-task overall accuracy, using the metric defined by Eq. 11 in the main draft. As shown in Tab. 1. VersatileGaussian performs best on most tasks. The best existing method underperforms VersatileGaussian by 18.99% regarding the overall performance. The qualitative results on SceneNet are provided in Fig. 4.

Table 1: Comparison SceneNet Dataset. The comparison is conducted on the SceneNet testing datasets used in MuvieNeRF [4]. The performances of the rest approaches are derived from MuvieNeRF [4]. RGB, KP, ED, SN, and SL stand for RGB image synthesis, key point detection, edge detection, surface normal estimation, and semantic segmentation.

Method	Overall Drop $\Delta_m(\%)$	RGB PSNR↑	KP $\mathcal{L}_1 \downarrow$	ED $\mathcal{L}_1 \downarrow$	SN $\mathcal{L}_1 \downarrow$	SL mIoU↑
VersatileGaussian	0.00	**33.48**	0.0050	**0.0129**	**0.0126**	0.9506
MuvieNeRF [4]	-18.99	29.56	0.0050	0.0189	0.0173	**0.9556**
SS-NeRF [3]	-22.79	29.18	0.0052	0.0197	0.0182	0.9510
SemanticNeRF [5]	-23.57	28.85	0.0051	0.0198	0.0186	0.9417

4 More Qualitative Results

We provide more qualitative results of our method on Replica, ScanNet, and SceneNet in Fig. 2 and Fig. 3, Fig. 4, respectively. These results demonstrate that VersatileGaussian is capable of generating high-quality versatile labels at novel views.

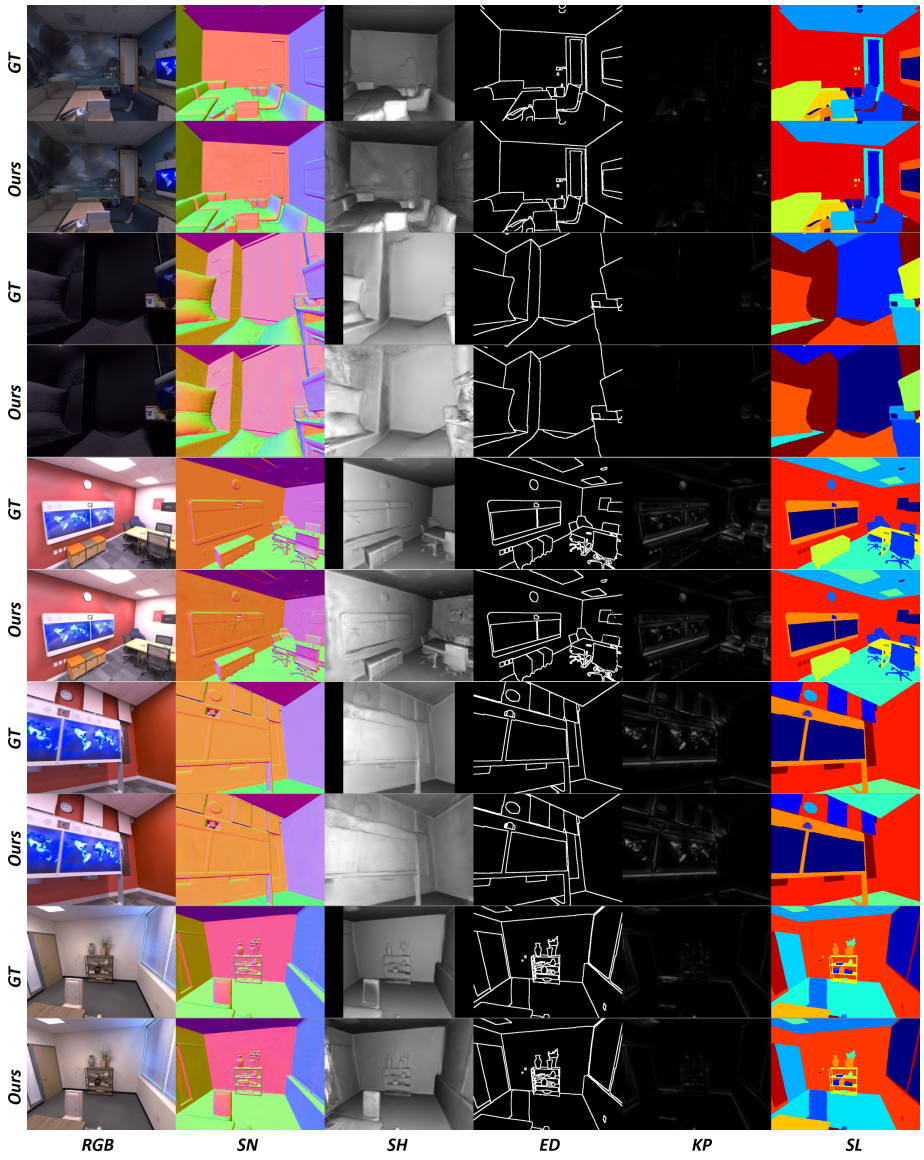

Fig. 2: Visualization on Replica. We provide more visulization on Replica. RGB, KP, SH, ED, SN, and SL stand for RGB image synthesis, key point detection, reshading, edge detection, surface normal estimation, and semantic segmentation.

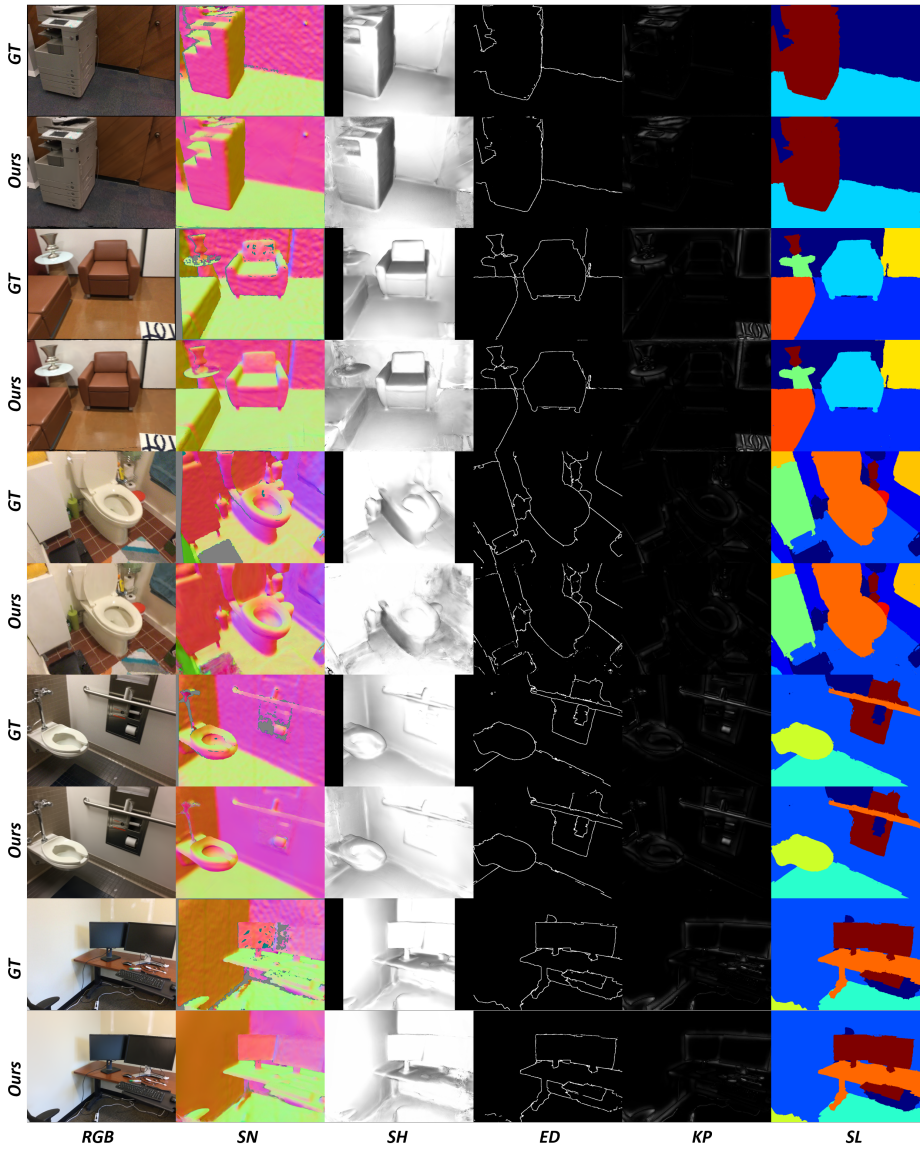

Fig. 3: Visualization on ScanNet. We provide more visulization on ScanNet. RGB, KP, SH, ED, SN, and SL stand for RGB image synthesis, key point detection, reshading, edge detection, surface normal estimation, and semantic segmentation.

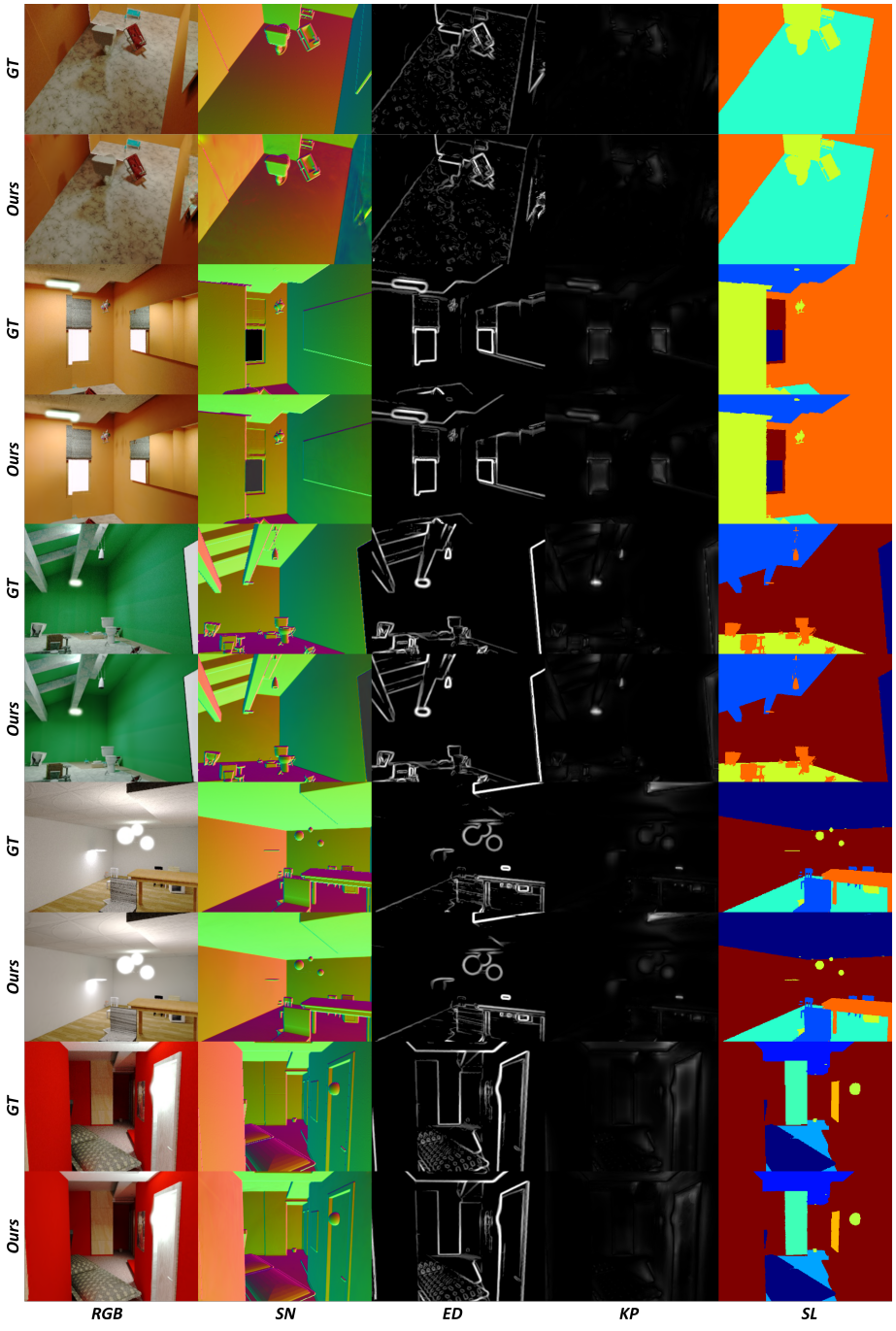

Fig. 4: Visualization on SceneNet. We provide more visulization on SceneNet. RGB, KP, ED, SN, and SL stand for RGB image synthesis, key point detection, edge detection, surface normal estimation, and semantic segmentation.

References

1. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (ToG) **42**(4), 1–14 (2023)
2. McCormac, J., Handa, A., Leutenegger, S., Davison, A.J.: Scenenet rgb-d: 5m photorealistic images of synthetic indoor trajectories with ground truth. arXiv e-prints pp. arXiv–1612 (2016)
3. Zhang, M., Zheng, S., Bao, Z., Hebert, M., Wang, Y.X.: Beyond rgb: Scene-property synthesis with neural radiance fields. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 795–805 (2023)
4. Zheng, S., Bao, Z., Hebert, M., Wang, Y.X.: Multi-task view synthesis with neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21538–21549 (2023)
5. Zhi, S., Laidlow, T., Leutenegger, S., Davison, A.J.: In-place scene labelling and understanding with implicit scene representation. In: ICCV (2021)