

LEVERAGING NERF-RENDERED IMAGES FOR 3D GAUSSIAN SPLATTING

Mizuki Morikawa, Yuta Shimizu, Chunyu Li, Yusuke Monno, Masatoshi Okutomi

Institute of Science Tokyo, Tokyo, Japan

ABSTRACT

Neural radiance field (NeRF) and 3D Gaussian splatting (3DGS) are two mainstream approaches for novel view synthesis. They often show complementary performance, i.e., 3DGS demonstrating faster rendering speed and NeRF demonstrating higher rendering quality. Motivated by this, we propose leveraging NeRF-rendered images for 3DGS. Specifically, we target street scenes and utilize a pre-trained street-specific NeRF method to produce training images for a target 3DGS method. In our 3DGS training, NeRF-rendered images are used to remove transient objects in street-level input views and to generate bird’s-eye views as additional views, inheriting the higher-quality rendering of NeRF into 3DGS. We further incorporate a diffusion-based image enhancement to improve the image quality of the additional views. Experimental results on one synthetic and two real datasets demonstrate that our proposed method improves street-scene rendering while preserving the speed of 3DGS and the quality of NeRF.

1. INTRODUCTION

Novel view synthesis (NVS) is the task of rendering images from arbitrary viewpoints by modeling an explicit or implicit 3D representation of the scene from observed multi-view images. 3D Gaussian splatting (3DGS) [1] and neural radiance fields (NeRF) [2] are currently two mainstream approaches for NVS. It is widely understood that 3DGS has a strong advantage in rendering speed and can achieve real-time NVS, whereas NeRF usually cannot. On the other hand, whether 3DGS or NeRF achieves higher rendering quality is inconclusive due to their differences in 3D representations and rendering mechanisms.

In this study, we target the situation where 3DGS and NeRF provide complementary performance, i.e., 3DGS achieves real-time rendering, and NeRF demonstrates higher rendering quality. In this situation, we propose leveraging NeRF-rendered images to train 3DGS. The top of Fig. 1 shows the general and conceptual flow of our proposed method. We first train a NeRF method, which is called reference NeRF, in a standard manner using input images and corresponding camera poses. Then, we render additional-view images from generated camera poses using the reference NeRF model. They are subsequently used for training

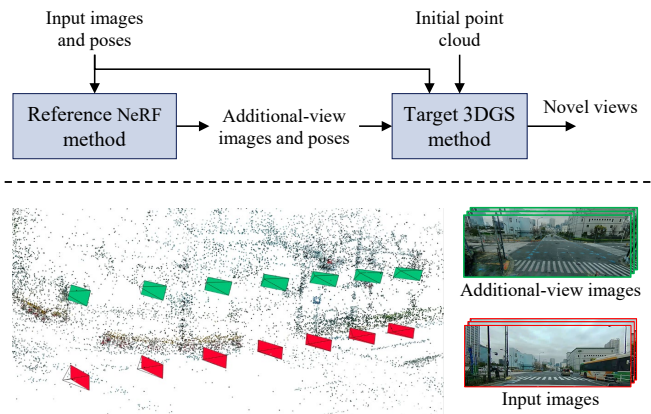


Fig. 1: The top part shows the general and conceptual flow of our proposed method, where we utilize NeRF-rendered images for additional views to train a 3DGS method. The bottom part shows the examples of generated additional views targeting the NVS of a street scene.

a 3DGS method, which is called target 3DGS, as well as the original input images and an initial point cloud estimated by structure from motion (SfM) [3]. In this pipeline, the reference NeRF plays a role in augmenting training views for the target 3DGS, exploiting its higher rendering quality.

Similar to this study, some recent studies addressed combining NeRF and 3DGS. Viewing direction Gaussian splatting (VDGS) introduces viewing direction encoding of NeRF into the opacity modeling of 3D Gaussians [4]. NeRF-GS proposes a framework to jointly train NeRF and 3DGS to realize feature-level coupling between them [5]. While these studies design a specific network model for each method, we propose a more general approach that exploits a pre-trained reference NeRF model to augment training images for 3DGS, allowing the inclusion of arbitrary 3DGS-based methods as the target 3DGS.

With the above feature of our method in mind, in this paper, we design a method specifically tailored for NVS in street scenes. For this purpose, we employ a recent segmentation-guided NeRF method for street scenes [6] (we call it Street-NeRF in short) and generate bird’s-eye views as additional views, as shown at the bottom of Fig. 1, where red cameras are original input views from a car-mounted camera and green cameras are generated additional views. By adding the

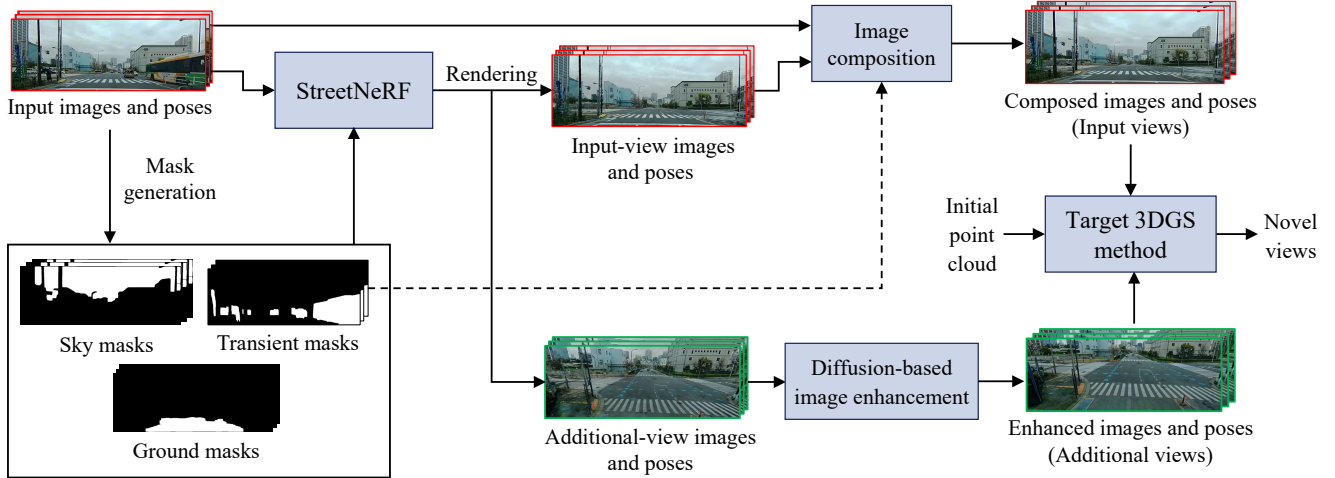


Fig. 2: The overview of our proposed method using StreetNeRF as the reference NeRF method, targeting NVS in street scenes.

bird’s-eye views rendered from StreetNeRF in high quality, our method improves the NVS quality of the target 3DGS while preserving its real-time performance. Furthermore, inspired by the recent study of [7], we apply a diffusion-based image enhancement to the additional views as an option, which further enhances the NVS quality. In experiments, we apply our proposed method to three representative 3DGS methods, i.e., Splatfacto [8] (3DGS implementation in Nerfstudio [9]), Scaffold-GS [10], and 2DGS [11], and evaluate them using one synthetic and two real datasets. The results demonstrate that our method consistently improves the NVS quality of the above target 3DGS methods, especially for novel views downward to the street, which are unobservable from street-level input views.

2. PROPOSED METHOD

2.1. Overview

Figure 2 illustrates the overview of our proposed method using StreetNeRF [6] as the reference NeRF method, targeting NVS of street scenes. Firstly, following StreetNeRF, three types of binary masks are generated using Grounded SAM [12] to identify the sky, the ground, and the transient objects in each input image. Then, StreetNeRF is trained using those masks, input images, and camera poses estimated by SfM [3]. The trained NeRF model is then used to render images at the input views and newly defined bird’s-eye additional views. For the input views, image composition is performed using the transient masks to generate transient-object-removed composed images. For the additional views, a diffusion-based image enhancement [7] is optionally applied to generate enhanced images. Finally, the composed images and the enhanced images are used as the inputs to train a target 3DGS method, enabling the generation of novel view images.

2.2. Review of StreetNeRF

StreetNeRF is a derivative of Zip-NeRF [13] specifically adapted to street scenes by introducing segmentation-guided region decomposition and region-specific processing to address challenging areas that typically lead to performance degradation in street scenes. Using the generated three types of binary masks mentioned before, specialized treatments are applied to each region, including infinite-depth color modeling for the sky, plane regularization for the ground, and disabling loss computation in transient-object regions. In addition, to mitigate artifacts caused by inconsistencies in illumination conditions across input images, color transformations for each image are trained for appearance embedding.

From the aspect of our proposed method, StreetNeRF provides the following benefits.

- **Input-view images:** Since StreetNeRF excludes the transient-object regions from loss computation during its training, the rendered images for input views (i.e., training views) are not affected by transient objects. Because these images also contain information behind moving objects, they enable the target 3DGS method to be trained with richer visual information.
- **Additional-view images:** Additional information not directly observed from the input views can be supplied to the target 3DGS method. By generating effective additional views for street scenes, a higher-quality rendering of StreetNeRF, according to the plane regularization and the sky color modeling, can be inherited into the target 3DGS method.

The detailed processes are explained below.

2.3. Image composition for input views

At the input views, replacing the real images entirely with the images rendered from the trained StreetNeRF model can

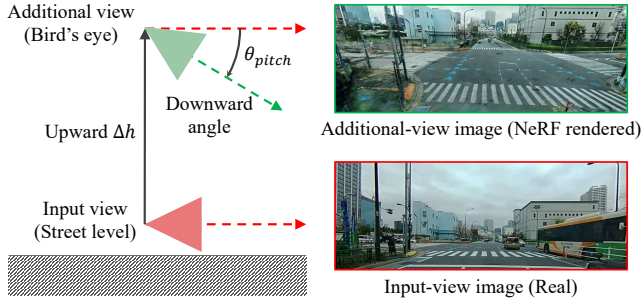


Fig. 3: Schematic illustration of additional view generation.

suppress the influence of transient objects. However, this may reduce the reliability of the image content in static regions. To balance these factors, we compose the two types of images to construct better input-view images. For this purpose, we reuse the binary transient mask to replace only the regions corresponding to the transient objects in the real images with those rendered by the NeRF model. This allows us to utilize high-confidence images for training the target 3DGS method while eliminating the influence of the transient objects.

2.4. Generation of additional views

In street scenes, typical viewpoints that cannot be observed using an on-board camera, but are important for NVS applications such as street map viewing, are the viewpoints looking down at the street from the top. Thus, we generate bird's-eye views as additional views, as shown in Fig. 3. The parameters we set are the upward distance Δh from the input views and the downward angle θ_{pitch} for the bird's-eye views.

As an optional process to further enhance the image quality, an image enhancement diffusion model provided in [7] is applied to the NeRF-rendered additional-view images. This process sharpens the regions that are blurry in the images.

3. EXPERIMENTAL RESULTS

3.1. Results on synthetic dataset

Dataset and settings: We used the synthetic Block Small dataset in MatrixCity [14], as shown in Fig. 4. The left images are examples of the input images. The top-right figure shows the point cloud and the camera poses. The Block Small dataset only provides street-level ($3m$ height and 0° pitch) training and evaluation views shown as the red cameras, which are insufficient to evaluate the rendering quality from novel views far from the training views. Thus, we constructed additional evaluation views, shown as the yellow cameras in the bottom-right figure, for which the cameras were circularly moved at two different heights and pitches, i.e., ($10m$, 25°) and ($13m$, 30°). We call these sets of evaluation views bird's-eye evaluation views 1 and 2 (BEV_1 and BEV_2), respectively. In our method, we set $\Delta h = 10m$ and $\theta_{pitch} = 30^\circ$ for

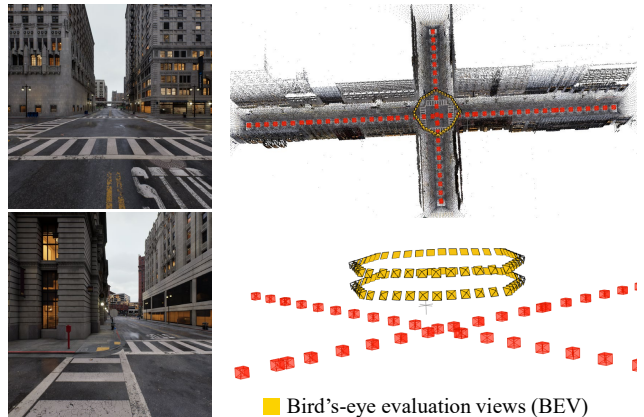


Fig. 4: Synthetic dataset of MatrixCity Block Small.

Table 1: Quantitative comparisons on Block Small dataset. The bold represents the best, and the underline represents the second best for each target 3DGS method.

	PSNR \uparrow			SSIM \uparrow		
	Street	BEV_1	BEV_2	Street	BEV_1	BEV_2
StreetNeRF [6]	23.56	22.02	21.43	0.759	0.722	0.701
Splatfacto [8]	22.47	18.71	16.33	<u>0.782</u>	0.658	0.608
Ours w/o diffusion (Splat.)	23.49	<u>22.24</u>	<u>21.75</u>	0.787	0.748	0.729
Ours w/ diffusion (Splat.)	<u>22.93</u>	22.56	21.87	0.781	<u>0.746</u>	<u>0.725</u>
Scaffold-GS [10]	24.39	21.70	20.99	0.839	0.739	0.696
Ours w/o diffusion (Scaff.)	24.14	<u>22.18</u>	<u>21.51</u>	<u>0.826</u>	<u>0.762</u>	<u>0.735</u>
Ours w/ diffusion (Scaff.)	<u>24.19</u>	22.92	22.38	0.817	0.764	0.742
2DGS [11]	22.80	19.98	17.75	0.788	0.699	0.665
Ours w/o diffusion (2DGS)	23.14	<u>22.36</u>	<u>21.77</u>	<u>0.771</u>	<u>0.753</u>	<u>0.735</u>
Ours w/ diffusion (2DGS)	<u>22.91</u>	22.64	22.09	0.767	0.755	0.739

generating additional views, which are not identical to BEV_1 and BEV_2, because the additional views are generated according to the training views, as illustrated in Fig. 3. We applied our proposed method (w/o and w/ diffusion) to three target 3DGS methods, i.e., Splatfacto [8], Scaffold-GS [10], and 2DGS [11].

Quantitative comparisons: Table 1 shows the quantitative results, where we evaluate peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [15]. For the original street-level evaluation views (Street), our method does not consistently improve the original methods. This is because these views are close to the training views, resulting in sufficient quality even by the original methods. This is evident from that the numerical performance on BEV_1 and BEV_2 significantly drops for the original methods. In contrast, our methods w/o and w/ diffusion consistently improve the original methods on BEV_1 and BEV_2 and alleviate the performance drops compared with Street. Regarding the rendering speed, our method with Splatfacto maintains its original speed and achieves a frame-per-second of 50.86 using an NVIDIA RTX 4090 GPU, which is much faster than that of StreetNeRF (0.115). These results indicate that our

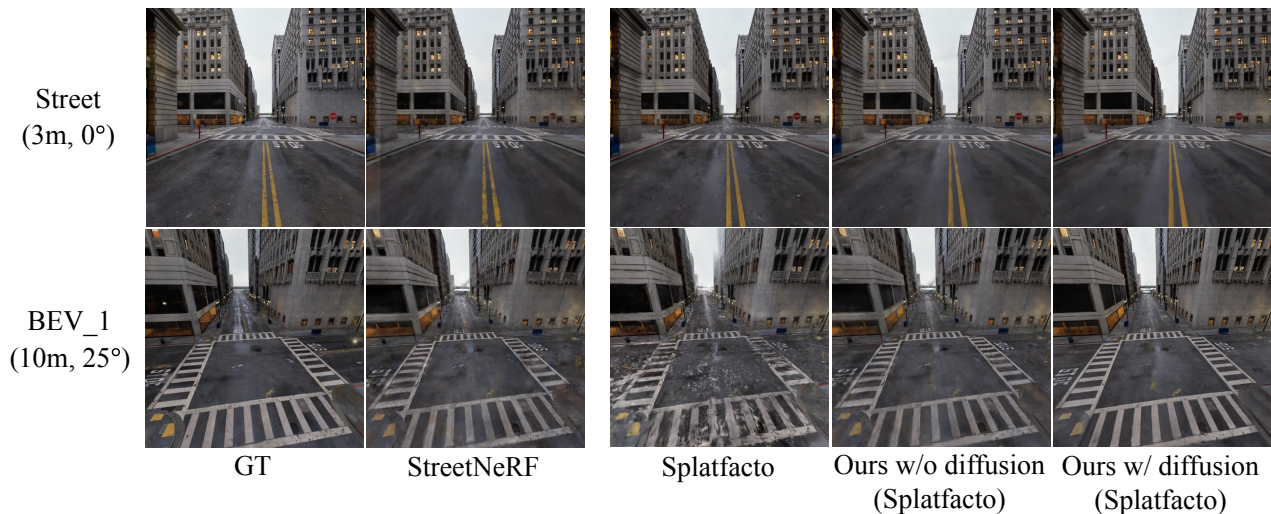


Fig. 5: Visual comparisons on the Block Small dataset using Splatfacto.

Table 2: Ablation study on Splatfacto.

Add. views from 3DGS	Add. views from NeRF	Diffusion	PSNR \uparrow		
			Street	BEV_1	BEV_2
			22.47	18.71	16.33
✓			<u>22.97</u>	18.87	16.30
✓			22.73	20.27	18.22
	✓		23.49	<u>22.24</u>	<u>21.75</u>
	✓	✓	22.93	22.56	21.87

method realizes both high quality and high speed by combining NeRF and 3DGS.

Visual comparisons: Figure 5 shows the visual comparisons using Splatfacto. The visual results support a consistent interpretation with the numerical results. For Street on the top, all methods produce similar results sufficiently close to the ground truth (GT). In contrast, for BEV_1, Splatfacto generates significant artifacts and black-colored holes on the road. It is confirmed that the reference StreetNeRF succeeds in reducing those artifacts and holes, owing to the plane regularization of the road. Our method w/o diffusion produces a similar result with StreetNeRF, validating that our method can successfully inherit the rendering quality of StreetNeRF to the target Splatfacto. Finally, our method w/ diffusion demonstrates the best visual result, owing to the strong capability of the diffusion model to enhance image quality. The visual comparisons using Scaffold-GS and 2DGS are included in the supplementary material, where a similar tendency with Splatfacto can be observed.

Ablation study: To investigate the effectiveness of combining NeRF-rendered images and diffusion-based image enhancement, we conducted an ablation study shown in Table 2 using Splatfacto. The first row shows the result of the original Splatfacto. The second row represents the method that utilizes the same additional views as ours, but they are rendered

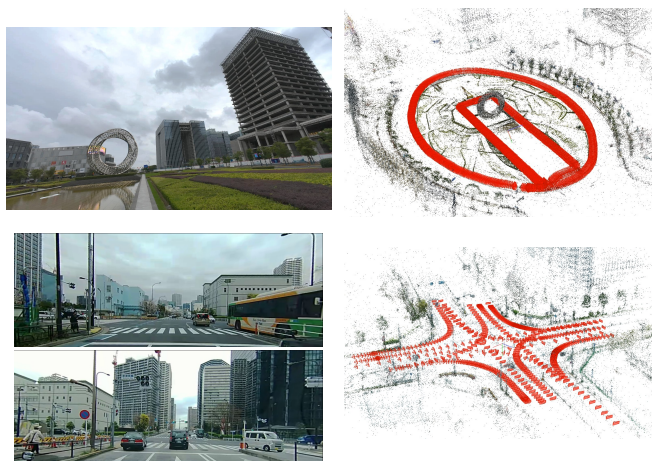


Fig. 6: Real datasets of Park (top) and Shinagawa (bottom).

from the original Splatfacto. The third row is the method that further adopts the diffusion-based image enhancement to the additional views rendered from the Splatfacto. Compared with the first row, simply adding views rendered by the original Splatfacto (second row) yields no significant difference. While applying the diffusion-based image enhancement to these 3DGS-rendered views (third row) brings an improvement, our proposed methods leveraging NeRF-rendered additional views (the fourth and fifth rows) achieve a substantially larger performance improvement for BEV_1 and BEV_2. These results validate the importance of utilizing the NeRF method as the reference to generate the additional views.

3.2. Results on real datasets

Datasets and settings: We used two real datasets, as shown in Fig. 6. The first dataset is the Park dataset provided in [16].

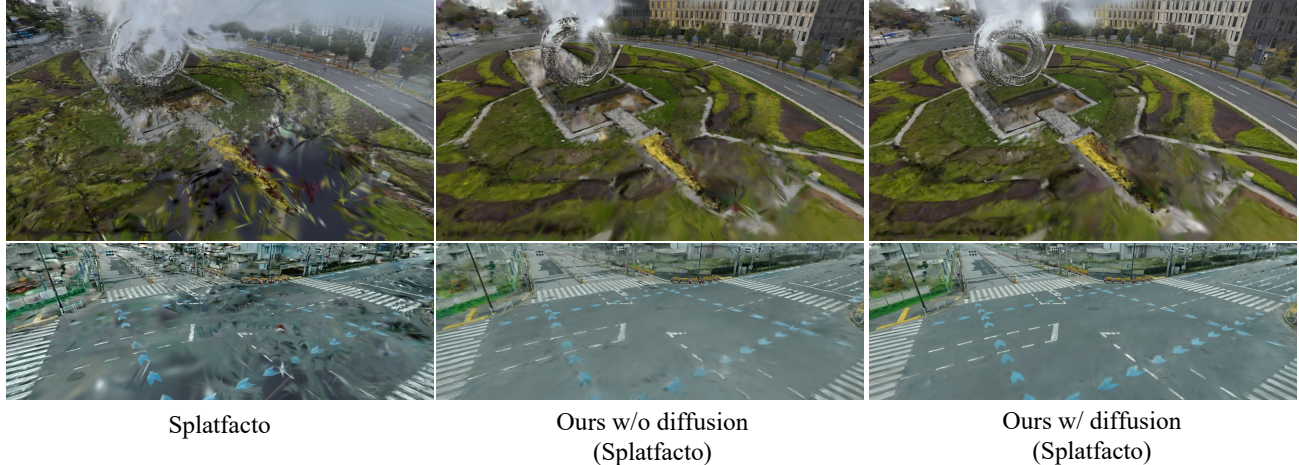


Fig. 7: Visual comparisons on the Park (top) and the Shinagawa (bottom) datasets using Splatfacto.

While the Park dataset contains street-level and aerial images, we only used the street-level images for the training, according to our assumed situation in this study. Using only street-level images, we performed SfM by using COLMAP [3] to estimate the camera poses and the point cloud. The parameters to generate additional views were empirically set as $\Delta h = 5\text{m}$ and $\theta_{pitch} = 25^\circ$.

Another dataset shown in the bottom part of Fig. 6 is our constructed dataset, named the Shinagawa dataset, capturing an intersection of Shinagawa City using a dashcam. The dataset includes a total of 12 video clips, i.e., four straight directions (South to North, North to South, West to East, and East to West) and eight turning directions (e.g., South to West, South to East, etc.). The camera poses and the point cloud were estimated by SfM using COLMAP. The parameters to generate additional views were empirically set as $\Delta h = 5\text{m}$ and $\theta_{pitch} = 15^\circ$.

Visual comparisons: Because there are no suitable evaluation views for the real datasets, we only performed visual comparisons for those datasets. Specifically, the aerial views in the original Park dataset are too high in height and placed outside the range of reconstruction areas from street-level views, making them unusable for evaluation views. Also, there are no views other than street-level views for the Shinagawa dataset because the images were captured using a dashcam mounted on a real car.

Figure 7 presents visual comparisons on the Park (top) and the Shinagawa (bottom) datasets using Splatfacto. For both datasets, our proposed methods w/o and w/ diffusion successfully improve the original Splatfacto, significantly reducing the holes on the ground and the road, which is attributed to the incorporation of the plane regularization in StreetNeRF. Even though the Shinagawa dataset contains a large number of transient objects, artifacts caused by such objects are not apparent in the results of our methods, indicating that high-quality street-scene representations are successfully

achieved, which is attributed to the exclusion of transient objects in the input views.

3.3. Video results

Video results can be confirmed from our project page¹, including additional results using Scaffold-GS for real datasets. The overall trends of the results are consistent with the results in this paper, clearly highlighting the better rendering quality of our proposed methods in both cases w/o and w/ diffusion.

4. CONCLUSION

In this paper, we have proposed a method for integrating a reference NeRF method to improve a target 3DGS method. As the reference NeRF method, we adopted StreetNeRF, which incorporates street-specific processing to achieve high-quality NVS. We used the trained StreetNeRF model to generate both input-view images and additional-view images. Image composition and diffusion-based image enhancement are applied to these images, respectively, and the resulting images are then used to train the target 3DGS method. Both quantitative and qualitative evaluations have demonstrated that our proposed method achieves NVS quality comparable to or surpassing that of StreetNeRF, while maintaining the rendering speed of the original 3DGS method. One of the limitations of our method is that we currently generate the poses of additional views in an empirical manner. The generation of optimal poses for additional views is one of our future works.

Acknowledgment This research was conducted as part of a collaborative research project with MICWARE CO., LTD.

¹<http://www.ok.sc.e.titech.ac.jp/res/NVS/NeRFtoGS.html>

5. REFERENCES

- [1] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis, “3D Gaussian splatting for real-time radiance field rendering,” *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 139–1, 2023.
- [2] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng, “NeRF: Representing scenes as neural radiance fields for view synthesis,” in *Proc. of European Conference on Computer Vision (ECCV)*, 2020, pp. 405–421.
- [3] Johannes Lutz Schönberger and Jan-Michael Frahm, “Structure-from-motion revisited,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4104–4113.
- [4] Dawid Malarz, Weronika Smolak-Dyżewska, Jacek Tabor, Sławomir Tadeja, and Przemysław Spurek, “Gaussian splatting with NeRF-based color and opacity,” *Computer Vision and Image Understanding*, vol. 251, pp. 104273–1–11, 2025.
- [5] Shuangkang Fang, I Shen, Takeo Igarashi, Yufeng Wang, ZeSheng Wang, Yi Yang, Wenrui Ding, Shuchang Zhou, et al., “NeRF is a valuable assistant for 3D Gaussian splatting,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025, pp. 26230–26240.
- [6] Yizhou Li, Masatoshi Okutomi, Yuuichi Tanaka, Seiichi Kataoka, and Teruaki Kosiba, “Segmentation-guided neural radiance fields for novel street view synthesis,” in *Proc. of International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 2025, pp. 591–597.
- [7] Jay Zhangjie Wu, Yuxuan Zhang, Haithem Turki, Xuanchi Ren, Jun Gao, Mike Zheng Shou, Sanja Fidler, Zan Gojcic, and Huan Ling, “DIFIX3D+: Improving 3D reconstructions with single-step diffusion models,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 26024–26035.
- [8] Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa, “gsplat: An open-source library for gaussian splatting,” *Journal of Machine Learning Research*, vol. 26, no. 34, pp. 1–17, 2025.
- [9] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David Mcallister, Justin Kerr, and Angjoo Kanazawa, “Nerfstudio: A modular framework for neural radiance field development,” in *Proc. of ACM SIGGRAPH*, 2023, pp. 1–12.
- [10] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai, “Scaffold-GS: Structured 3D Gaussians for view-adaptive rendering,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 20654–20664.
- [11] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao, “2D Gaussian splatting for geometrically accurate radiance fields,” in *Proc. of ACM SIGGRAPH*, 2024, pp. 1–11.
- [12] Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, Zhaoyang Zeng, Hao Zhang, Feng Li, Jie Yang, Hongyang Li, Qing Jiang, and Lei Zhang, “Grounded SAM: Assembling open-world models for diverse visual tasks,” in *arXiv preprint 2401.14159*, 2024.
- [13] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman, “Zip-NeRF: Anti-aliased grid-based neural radiance fields,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 19697–19705.
- [14] Yixuan Li, Lihan Jiang, Linning Xu, Yuanbo Xiangli, Zhenzhi Wang, Dahua Lin, and Bo Dai, “MatrixCity: A large-scale city dataset for city-scale neural rendering and beyond,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 3205–3215.
- [15] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [16] Lihan Jiang, Kerui Ren, Mulin Yu, Linning Xu, Junting Dong, Tao Lu, Feng Zhao, Dahua Lin, and Bo Dai, “Horizon-GS: Unified 3D Gaussian splatting for large-scale aerial-to-ground scenes,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 26789–26799.
- [17] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 586–595.

Supplementary Material: **Leveraging NeRF-Rendered Images for 3D Gaussian Splatting**

6. ADDITIONAL RESULTS

Table S1 shows the full numerical comparison for the synthetic Block Small dataset, including the learned perceptual image patch similarity (LPIPS) [17]. The results of LPIPS for BEV_1 and BEV_2 indicate that our proposed method with diffusion enhancement shows the best performance for all 3DGS methods. Figures S1 to S3 show additional visual comparisons for the Block Small dataset, including the results using Scaffold-GS and 2DGS, as well as the results for BEV_2, which cannot be included in the main manuscript due to limited space. The overall trends of the results are consistent with those mentioned in the main manuscript.

Table S1: Quantitative comparisons on Block Small dataset. The bold represents the best, and the underline represents the second best for each target 3DGS method.

	PSNR \uparrow			SSIM \uparrow			LPIPS \downarrow		
	Street	BEV_1	BEV_2	Street	BEV_1	BEV_2	Street	BEV_1	BEV_2
StreetNeRF [6]	23.56	22.02	21.43	0.759	0.722	0.701	0.371	0.298	0.318
Splatfacto [8]	22.47	18.71	16.33	<u>0.782</u>	0.658	0.608	0.316	0.332	0.364
Ours w/o diffusion (Splat.)	23.49	<u>22.24</u>	<u>21.75</u>	0.787	0.748	0.729	<u>0.365</u>	<u>0.299</u>	<u>0.312</u>
Ours w/ diffusion (Splat.)	<u>22.93</u>	22.56	21.87	0.781	<u>0.746</u>	<u>0.725</u>	0.369	0.263	0.269
Scaffold-GS [10]	24.39	21.70	20.99	0.839	0.739	0.696	0.194	<u>0.215</u>	<u>0.249</u>
Ours w/o diffusion (Scaff.)	24.14	<u>22.18</u>	<u>21.51</u>	<u>0.826</u>	<u>0.762</u>	<u>0.735</u>	<u>0.261</u>	0.264	0.290
Ours w/ diffusion (Scaff.)	<u>24.19</u>	22.92	22.38	0.817	0.764	0.742	0.267	0.208	0.216
2DGS [11]	22.80	19.98	17.75	0.788	0.699	0.665	0.382	0.350	0.371
Ours w/o diffusion (2DGS)	23.14	<u>22.36</u>	<u>21.77</u>	<u>0.771</u>	<u>0.753</u>	<u>0.735</u>	<u>0.426</u>	<u>0.323</u>	<u>0.334</u>
Ours w/ diffusion (2DGS)	<u>22.91</u>	22.64	22.09	0.767	0.755	0.739	0.427	0.301	0.305



Fig. S1: Visual comparisons on the Block Small dataset using Splatfacto

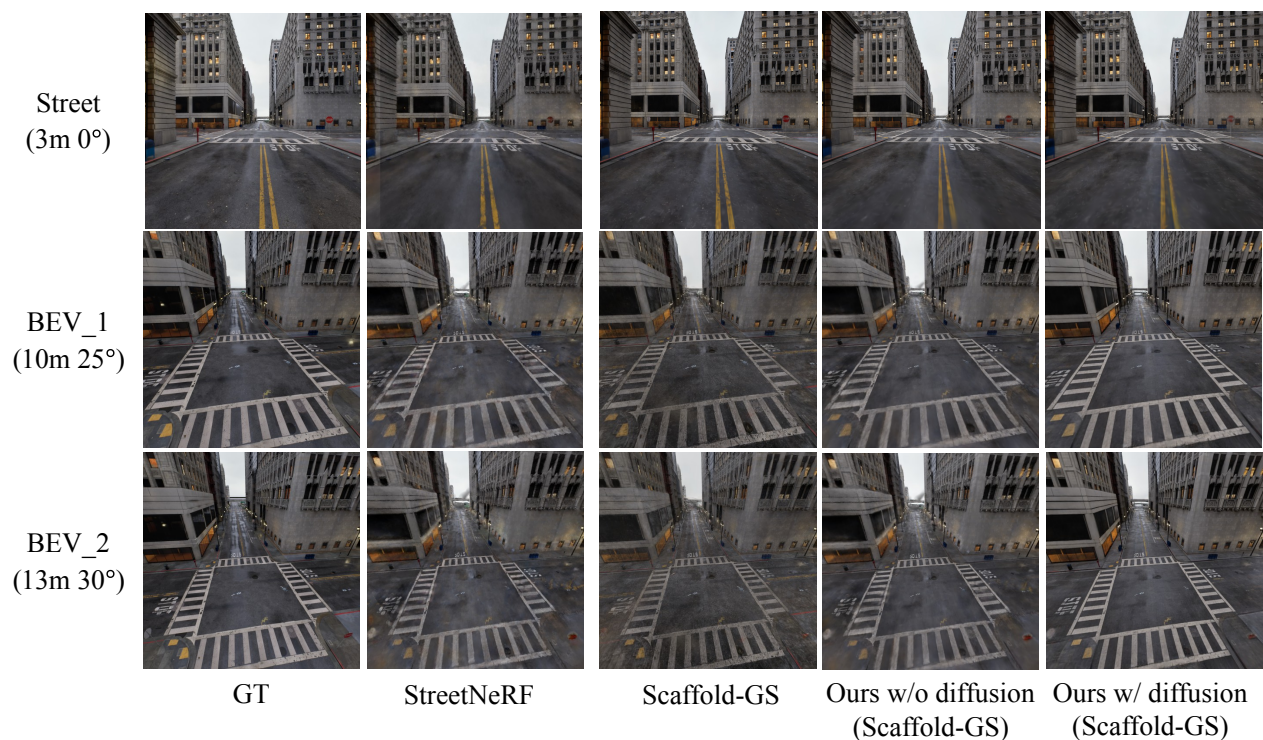


Fig. S2: Visual comparisons on the Block Small dataset using Scaffold-GS

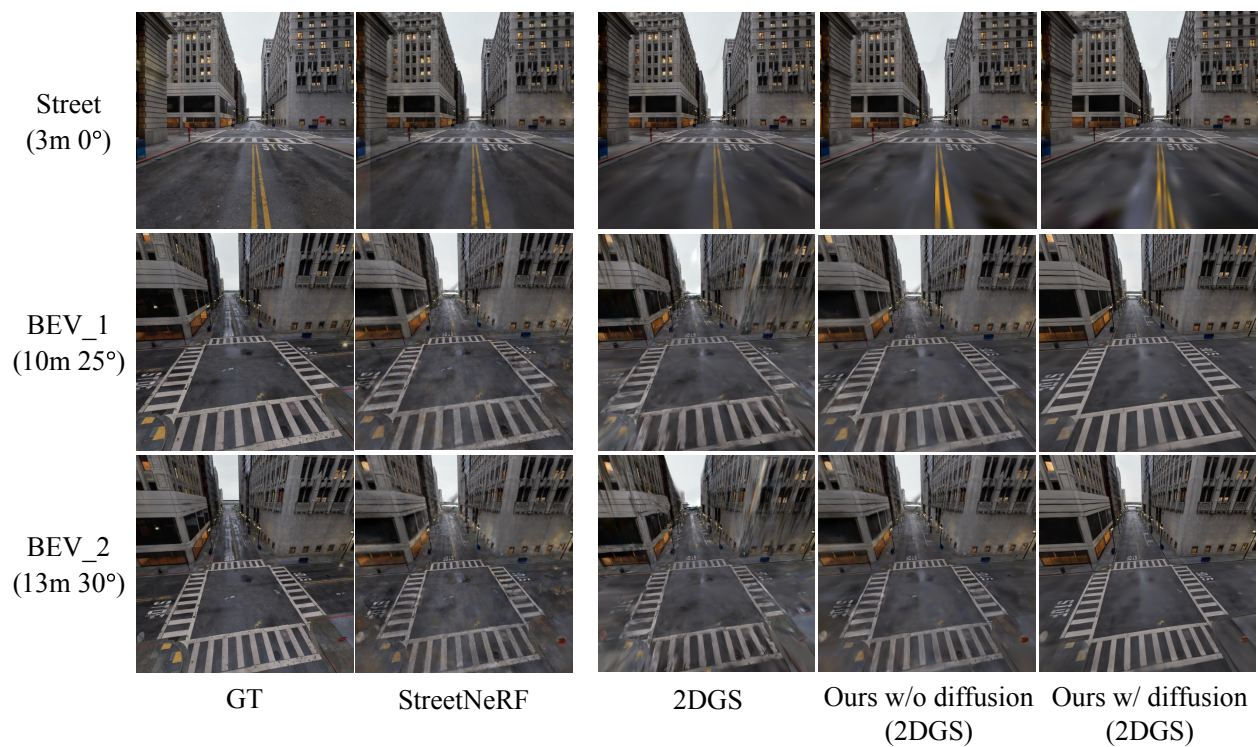


Fig. S3: Visual comparisons on the Block Small dataset using 2DGS