

Supplementary Material for ReconDreamer++: Harmonizing Generative and Reconstructive Models for Driving Scene Representation

Guosheng Zhao^{1, 2, 3*} Xiaofeng Wang^{1, 2, 3*} Chaojun Ni^{3, 4*} Zheng Zhu^{3*†}
Wenkang Qin³ Guan Huang³ Xingang Wang^{1, 5‡}

¹Institute of Automation, Chinese Academy of Sciences

²School of Artificial Intelligence, University of Chinese Academy of Sciences

³GigaAI ⁴Peking University ⁵Luoyang Institute for Robot and Intelligent Equipment

Scene	Start Frame	End Frame
segment-10359308928573410754_720_000_740_000_with_camera_labels.tfrecord	120	159
segment-11450298750351730790_1431_750_1451_750_with_camera_labels.tfrecord	0	39
segment-12496433400137459534_120_000_140_000_with_camera_labels.tfrecord	110	149
segment-15021599536622641101_556_150_576_150_with_camera_labels.tfrecord	0	39
segment-16767575238225610271_5185_000_5205_000_with_camera_labels.tfrecord	0	39
segment-17860546506509760757_6040_000_6060_000_with_camera_labels.tfrecord	90	129
segment-3015436519694987712_1300_000_1320_000_with_camera_labels.tfrecord	40	79
segment-6637600600814023975_2235_000_2255_000_with_camera_labels.tfrecord	70	109

Table 1. Eight scenes from the Waymo dataset [6] featuring high interactive activity, numerous vehicles, and complex driving trajectories.

In the supplementary material, we provide further details about the datasets and an explanation of the evaluation metrics. Additionally, we present more extensive qualitative results conducted on these datasets.

1. Datasets and Evaluation Metrics

Waymo. For the selection of Waymo [6] scenes, we follow the approach used in ReconDreamer [5]. These eight scenes are characterized by their rich dynamic foregrounds and large-scale variations. The specific scene names are listed in Tab. 1.

nuScenes. Similarly, we select eight scenes from the nuScenes dataset [1], each featuring complex traffic environments with diverse and challenging conditions. The specific scenes are as follows: *0037*, *0040*, *0050*, *0062*, *0064*, *0086*, *0087*, *0100*.

PandaSet. The selection of scenes from the PandaSet dataset [8] follows the approach used in UniSim [10] and NeuRAD [7]. The specific scenes are *001*, *011*, *016*, *028*, *053*, *063*, *084*, *106*, *123*, and *158*.

EUVS. EUVS [3] is a dataset specifically designed for extrapolated urban views, where all results across different lanes are captured from real-world observations. However, since the data for different lanes are not collected simultaneously, the dataset primarily focuses on static backgrounds, requiring the foreground to be ignored during evaluation. Consequently, when calculating metrics such as PSNR, SSIM, and LPIPS, the foreground must first be masked out to ensure accurate assessment of the static scene reconstruction. The four scenes from Level 1 are: *vegas_location_1*, *vegas_location_2*, *vegas_location_15*, *vegas_location_22*.

Metrics. As mentioned in the main text, we utilize Novel Trajectory Agent Intersection over Union (NTA-IoU) and Novel Trajectory Lane Intersection over Union (NTL-IoU) to assess the quality of the rendered video, both metrics proposed in DriveDreamer4D [11]. These metrics are specifically designed to evaluate the spatiotemporal coherence of foreground agents and background lanes, respectively.

The NTA-IoU processes images rendered under new trajectories using the YOLO11 [4] detector to extract 2D bounding boxes of vehicles. Meanwhile, by applying geometric transformations to the 3D bounding boxes from the original trajectories, they can be accurately projected onto the new trajectory perspective, thus obtaining the ground

*These authors contributed equally to this work.

†Corresponding authors. zhengzhu@ieee.org, xingang.wang@ia.ac.cn.

‡Project Page: <https://recondreamer-plus.github.io>

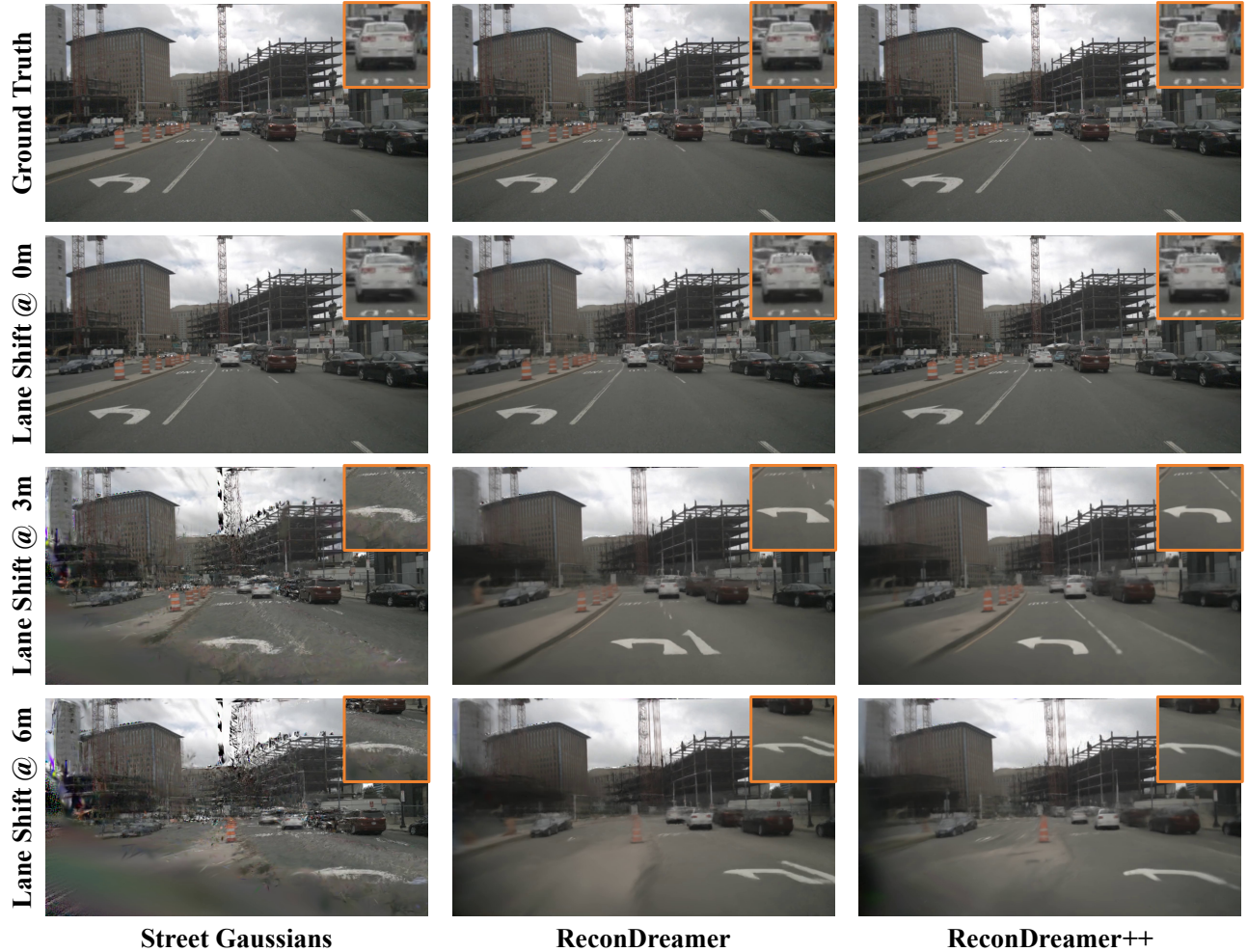


Figure 1. Qualitative comparisons of different trajectory renderings on nuScenes [1]. The orange boxes highlight that ReconDreamer++ significantly enhances the rendering quality with Street Gaussians [9] and ReconDreamer [5].

truth 2D bounding boxes in the new trajectory view. Each projected 2D bounding box will find the nearest 2D bounding box generated by the detector and compute their Intersection over Union (IoU).

Similarly, the NTL-IoU employs the TwinLiteNet [2] model to detect lane in the images rendered under the new trajectories, and the lane from the original trajectories will also be projected onto the new trajectory through corresponding geometric transformations. Finally, the mean Intersection over Union (IoU) between the projected and detected lane lines is calculated.

2. More Qualitative Results

The visualization and comparison results for all datasets can be found in the *.video_comparison* folder. This folder contains detailed comparisons for each dataset under both original and novel trajectories, along with ablation study videos

that highlight the contributions and impact of different components.

nuScenes. The visualization results for the nuScenes dataset are presented in Fig. 1, offering a comprehensive comparison of the performance of different methods. The experimental findings demonstrate that ReconDreamer++ achieves results on the original trajectory that are comparable to traditional reconstruction techniques such as Street Gaussians [9]. Notably, ReconDreamer++ even outperforms Street Gaussians in certain fine-grained details, showcasing its ability to capture intricate structures within the scene. In contrast, ReconDreamer [5] exhibits limitations, producing relatively blurry and less detailed rendering results on the original trajectory, which highlights the challenges faced by earlier approaches. When evaluating novel trajectories, the visualizations reveal the superior capabilities of ReconDreamer++. Specifically, the method demonstrates exceptional performance in maintaining high



Figure 2. Qualitative comparisons of different trajectory renderings on PandaSet [8]. The orange boxes highlight that ReconDreamer++ significantly enhances the rendering quality with Street Gaussians [9] and ReconDreamer [5].

geometric consistency of structured elements, particularly the ground surface, with respect to the ground truth. As highlighted by the orange bounding boxes in Fig. 1, ReconDreamer++ accurately renders left-turn arrows on the ground under novel trajectories, maintaining strong geometric fidelity. In contrast, both ReconDreamer and Street Gaussians struggle to render these ground structures with sufficient accuracy, often leading to distorted or incomplete representations. Such inaccuracies can significantly impact downstream tasks in autonomous driving, where precise un-

derstanding of road markings and other structured elements is essential for safe navigation.

PandaSet. The visualization results for PandaSet [8] are shown in Fig. 2, with comparisons that align with the experimental findings on Waymo [6] and nuScenes [1]. ReconDreamer++ demonstrates a comprehensive improvement over ReconDreamer [5] in rendering quality for both original and novel trajectories. Notably, it not only ensures rendering performance on par with traditional reconstruction methods, such as Street Gaussians [9], for original tra-



Figure 3. Qualitative comparisons of different trajectory renderings on EUVS [3]. The orange boxes highlight that ReconDreamer++ significantly enhances the rendering quality with Street Gaussians [9] and ReconDreamer [5].

jectories, but also achieves state-of-the-art (SOTA) results for novel trajectories. In particular, ReconDreamer++ excels in rendering structured elements such as lane markings on the ground, achieving exceptional accuracy and consistency. As highlighted by the orange bounding boxes in Fig. 2, the method accurately reconstructs these critical road features even under challenging novel viewpoints. These results further validate the effectiveness and robustness of ReconDreamer++ across diverse datasets and scenarios.

EUVS. The experimental results for EUVS [3] are shown in Fig. 3, demonstrating the superior performance of ReconDreamer++ on both the training set and the test set. Since this dataset consists of data collected from the same scene at different times, the dynamic object regions—including cars, pedestrians, and other moving elements—in the rendered results are masked to facilitate a clearer and more accurate comparison. Specifically, as highlighted by the orange bounding boxes in Fig. 3, ReconDreamer++ maintains an exceptionally high level of consistency with the ground truth in terms of detail rendering.

References

- [1] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multi-modal dataset for autonomous driving. In *CVPR*, 2020. 1, 2, 3
- [2] Quang-Huy Che, Dinh-Phuc Nguyen, Minh-Quan Pham, and Duc-Khai Lam. Twinlitenet: An efficient and lightweight model for driveable area and lane segmentation in self-driving cars. In *MAPR*, 2023. 2
- [3] Xiangyu Han, Zhen Jia, Boyi Li, Yan Wang, Boris Ivanovic, Yurong You, Lingjie Liu, Yue Wang, Marco Pavone, Chen Feng, and Yiming Li. Extrapolated urban view synthesis benchmark. *arXiv preprint arXiv:2412.05256*, 2024. 1, 4
- [4] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024. 1
- [5] Chaojun Ni, Guosheng Zhao, Xiaofeng Wang, Zheng Zhu, Wenkang Qin, Guan Huang, Chen Liu, Yuyin Chen, Yida Wang, Xueyang Zhang, et al. Recondreamer: Crafting world models for driving scene reconstruction via online restoration. *arXiv preprint arXiv:2411.19548*, 2024. 1, 2, 3, 4
- [6] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, 2020. 1, 3
- [7] Adam Tonderski, Carl Lindström, Georg Hess, William Ljungbergh, Lennart Svensson, and Christoffer Petersson. Neurad: Neural rendering for autonomous driving. In *CVPR*, 2024. 1
- [8] Pengchuan Xiao, Zhenlei Shao, Steven Hao, Zishuo Zhang, Xiaolin Chai, Judy Jiao, Zesong Li, Jian Wu, Kai Sun, Kun Jiang, et al. Pandaset: Advanced sensor suite dataset for autonomous driving. In *ITSC*, 2021. 1, 3
- [9] Yunzhi Yan, Haotong Lin, Chenxu Zhou, Weijie Wang, Haiyang Sun, Kun Zhan, Xianpeng Lang, Xiaowei Zhou, and Sida Peng. Street gaussians for modeling dynamic urban scenes. *arXiv preprint arXiv:2401.01339*, 2024. 2, 3, 4
- [10] Ze Yang, Yun Chen, Jingkan Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. Unisim: A neural closed-loop sensor simulator. In *CVPR*, 2023. 1
- [11] Guosheng Zhao, Chaojun Ni, Xiaofeng Wang, Zheng Zhu, Guan Huang, Xinze Chen, Boyuan Wang, Youyi Zhang, Wenjun Mei, and Xingang Wang. Drivedreamer4d: World models are effective data machines for 4d driving scene representation. *arXiv preprint arXiv:2410.13571*, 2024. 1