Supplementary Material:
3DRealCar: An In-the-wild RGB-D Car Dataset with 360-degree Views

Xiaobiao Du [1,2,3] Yida Wang [3] Haiyang Sun [3] Zhuojie Wu [2] Hongwei Sheng [2]
Shuyun Wang [2] Jiaying Ying [2] Ming Lu [2] Tianqing Zhu [4] Kun Zhan [3] Xin Yu [2*]

[1] University of Technology Sydney [2] The University of Queensland
[3] Li Auto Inc. [4] City University of Macau

1. Broader Impacts Statement

The introduction of our 3DRealCar dataset has profound effects on self-driving research. We expect this dataset can encourage extensive research to promote the advancement of the community.

Research Impacts. By providing dense 360-degree views of cars with point clouds as initialization, our 3DRealCar can be used to reconstruct high-quality 3D real cars for 3D printing and simulation in corner-case scenes. By providing detailed car parsing map annotations, our dataset can be leveraged to segment 2D car components or point clouds. Note that our 3DRealCar is the first dataset providing 3D car parsing annotations. In our 3D reconstruction benchmarking experiments, the reflective and dark lighting conditions of our dataset bring challenges to existing methods to reconstruct 3D cars under awful lighting conditions. We expect our dataset to encourage widespread collaboration and accelerate the exploration of 3D real car reconstruction, parsing, and simulation.

Societal Impacts. We collect our 3DRealCar dataset with the consent of the owners. In addition, we blur license plates and other private information. We try our best to hide and preserve the privacy of owners. Therefore, our dataset would not have any privacy violation problems. Due to our dataset focusing on a car class, we believe our dataset has the potential to be employed in future self-driving research and improve self-driving systems further.

2. Limitation and Discussion

Although our 3DRealCar is the largest dataset for the 3D real car dataset so far (2500 car instances with annotations), its scale is still limited compared to other datasets in the computer vision community. Therefore, we will further extend our dataset in the future. Moreover, our 3DRealCar dataset only provides the exterior views of cars without interior views. It is very crucial to reconstruct both exterior and interior views of cars for car marketing agencies. We will collect both exterior and interior views in the future to further extend our 3D real car dataset for intact 3D car models.

*Corresponding author.

3. Experimental Settings

Note that all models used in this work are publicly available. Each model we use is linked below:

1. **3D Reconstruction:** Instant-NGP [8], 3DGS [5], GaussianShader [4], and 2DGS [3].
2. **2D Car Parsing:** MMsegmentation. This repository includes all 2D segmentation models [7, 10, 15, 16] we used in this work.
3. **Novel View Synthesis:** Zero-123-XL [6].
4. **3D Generation:** DreamCraft3D [12].
5. **Corner-case Simulation:** YOLOv5 and YOLOv8 [14], YOLOv12 [13], CO-DETR [17], and libcom [9]. Specifically, we use YOLOv5 and YOLOv8 serial models, YOLOv12, and CO-DETR as detectors and libcom for the simulation of corner-case scenes.

We express great appreciation to the authors of the aforementioned repositories for their invaluable contributions. For the GPU specification, we use 8 A100 GPUs for 3D reconstruction, 3D generation, and novel view synthesis. We utilize 2 3090 GPUs for other tasks. We use the default hyperparameters for training.

4. Detailed Simulation Process and Additional Visualizations

In this section, we show how we simulate corner-case scenes. As shown in Fig. 1, we use images from Nuscenes [1] as backgrounds and leverage ViT-Adapter [2] to segment entire scenes for road masks. Then, we copy and paste the rendered images from the reconstructed high-quality 3D cars into the backgrounds with the guidance of road masks. In particu-

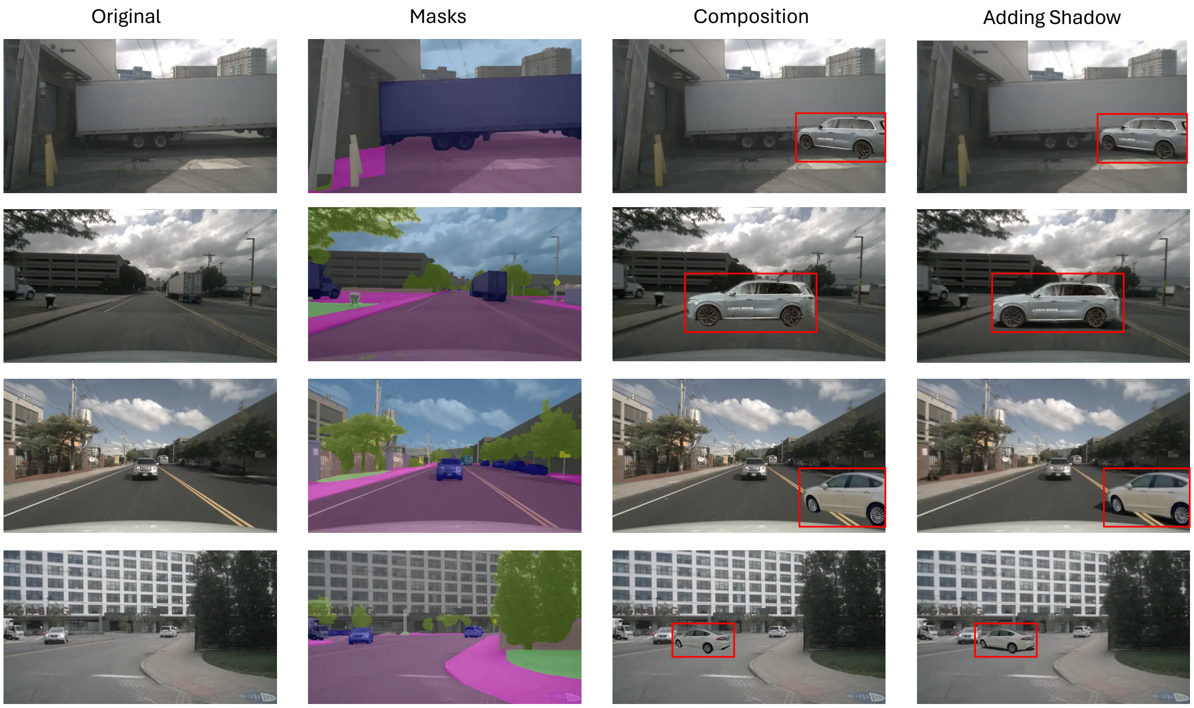

Figure 1. **Visualizations of ablating simulation procedures.** We use a red rectangle to highlight the simulated vehicles.

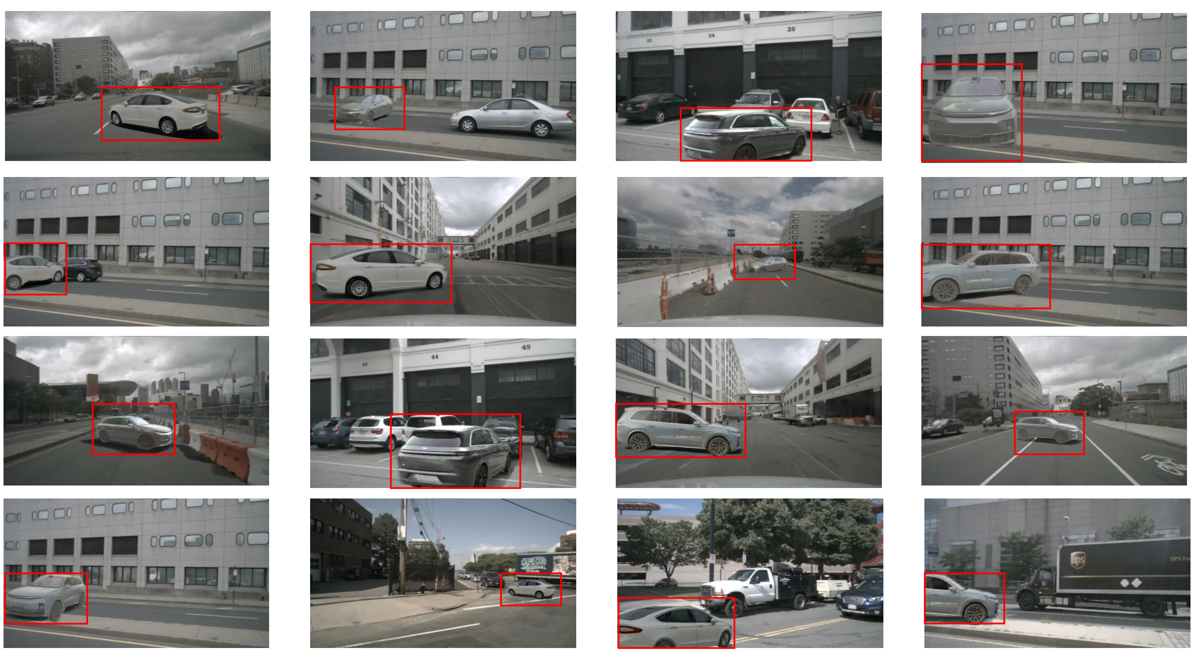

Figure 2. **More visualizations of simulated corner-case scenes.** We use a red rectangle to highlight simulated vehicles. These corner-case scenes show some vehicles have potential risks to traffic safety.

lar, we blur the edge between simulated foregrounds and backgrounds and then we use a color transfer algorithm [11] to make the whole simulated scene look harmonious. Finally, we use the shadow generation method in libcom [9] to add shadow for the simulated cars such that the entire scene looks photorealistic. However, this simulation method would generate some unreasonable scenes. Therefore, we manually intervene to select photorealistic corner-case scenes. Additional simulation results are shown in Figure 2.

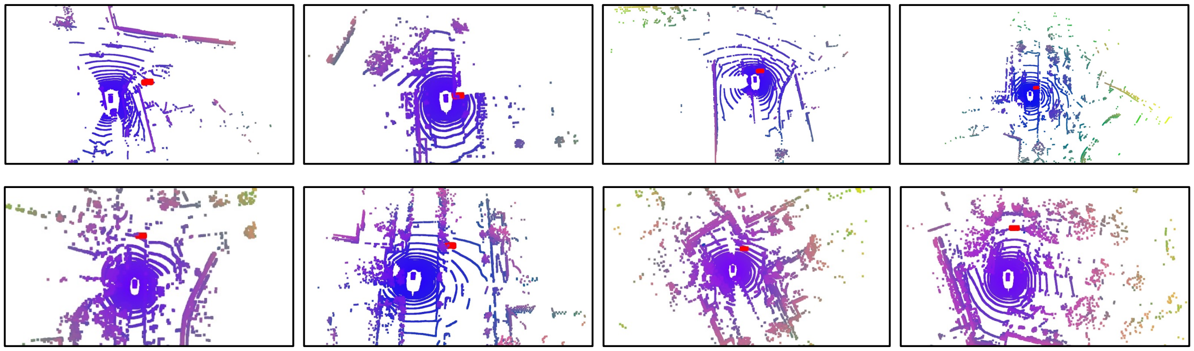

Figure 3. **Visualizations of point cloud inserting.** We use red color to annotate inserting vehicular point clouds with high density for better differentiation.

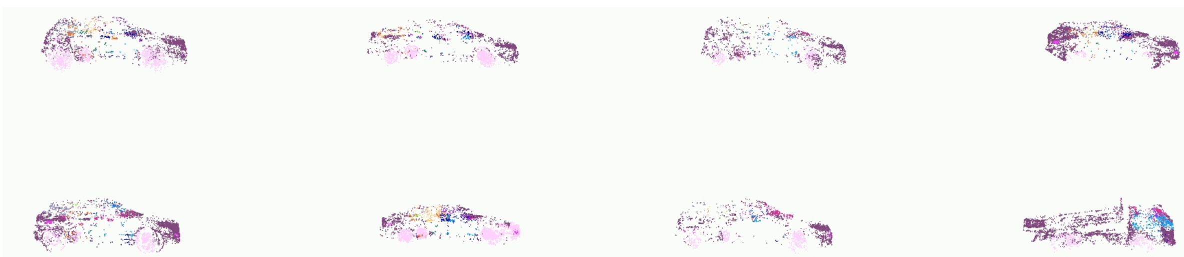

Figure 4. **Visualizations of 3D point cloud parsing.** With 2D car parsing map annotations, we lift the 2D car parsing maps into 3D point clouds and segment car components.

5. Simulated Lidar Scenes

As depicted in Fig. 3, we can insert our car point clouds into lidar scenes to simulate corner-case scenes, like a car passing or parking horizontally in front of the ego car. To better differentiate the inserted cars, we set them with dense point clouds and red color. In a practical scene, the vehicular point clouds should be sparse and only have one size that could be scanned by the lidar. Therefore, when we apply the inserted vehicular point clouds into a scene, we should make the vehicular point clouds sparse and only contain one side. By training on a variety of simulated scenarios, including rare or dangerous situations that are difficult to collect in real life, the self-driving system can learn to handle unexpected events more effectively.

6. 3D Car Parsing

As shown in Fig. 4, our dataset is the first to provide 3D car parsing annotations for parsing car components in 3D space. Thanks to that we provide 2D car parsing maps for every instance in our 3DRealCar dataset, we can lift 2D parsing maps to 3D and segment each component for point clouds and meshes. The primary purpose of these 3D car parsing maps is to enable precise and comprehensive analysis of vehicle structures, which is crucial for applications such as autonomous driving, vehicle design, vehicle editing, and virtual reality simulations. By using these detailed 3D parsing maps, developers and researchers can improve object recognition algorithms and enhance collision detection systems. Furthermore, this dataset facilitates the training of machine learning models to better understand the spatial relationships and physical attributes of car components, leading to more advanced and reliable automotive technologies.

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 multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 1

[2] Zhe Chen, Yuchen Duan, Wenhai Wang, Junjun He, Tong Lu, Jifeng Dai, and Yu Qiao. Vision transformer adapter for dense predictions. *arXiv preprint arXiv:2205.08534*, 2022. 1

[3] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. *arXiv preprint arXiv:2403.17888*, 2024. 1

[4] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. *arXiv preprint arXiv:2311.17977*, 2023. 1

[5] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*, 42(4): 1–14, 2023. 1

[6] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9298–9309, 2023. 1

[7] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*, 2024. 1

[8] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv:2201.05989*, 2022. 1

[9] Li Niu, Wenyan Cong, Liu Liu, Yan Hong, Bo Zhang, Jing Liang, and Liqing Zhang. Making images real again: A comprehensive survey on deep image composition. *arXiv preprint arXiv:2106.14490*, 2021. 1, 2

[10] Huihui Pan, Yuanduo Hong, Weichao Sun, and Yisong Jia. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes. *IEEE Transactions on Intelligent Transportation Systems*, 24(3):3448–3460, 2022. 1

[11] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001. 2

[12] Jingxiang Sun, Bo Zhang, Ruizhi Shao, Lizhen Wang, Wen Liu, Zhenda Xie, and Yebin Liu. Dreamcraft3d: Hierarchical 3d generation with bootstrapped diffusion prior. *arXiv preprint arXiv:2310.16818*, 2023. 1

[13] Yunjie Tian, Qixiang Ye, and David Doermann. Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*, 2025. 1

[14] Ultralytics. YOLOv8: A cutting-edge and state-of-the-art (sota) model that builds upon the success of previous yolo versions. `https://github.com/ultralytics/ultralytics?tab=readme-ov-file`, 2023. 1

[15] Wenhai Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14408–14419, 2023. 1

[16] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021. 1

[17] Zhuofan Zong, Guanglu Song, and Yu Liu. Detrs with collaborative hybrid assignments training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6748–6758, 2023. 1