DANGER: A Framework of Danger-Aware Novel Dataset Generator Extension for Robustness Test of Machine Learning

Shengjie Xu
University of California, Santa Cruz
sxu88@ucsc.edu

Leilani H. Gilpin
University of California, Santa Cruz
lgilpin@ucsc.edu

Abstract

Benchmark datasets for autonomous driving, such as KITTI, Argoverse, or Waymo are realistic, but they are designed to be too idealistic. These datasets do not contain errors, difficult driving maneuvers, or other corner cases. We propose a framework for perturbing autonomous vehicle datasets, the *DANGER* framework, which generates edge-case images on top of current autonomous driving datasets. The input to DANGER is a photorealistic datasets from real driving scenarios. We present the DANGER algorithm for vehicle position manipulation and the interface towards the renderer module, and present primitive generation cases applied to the virtual KITTI dataset. Our experiments prove that DANGER can be used as a framework for enlarging the current dataset to cover generative corner cases.

1 Motivation

Autonomous vehicles promise to decrease vehicle fatalities and increase safety in the modern automobile. However, the majority of datasets [3, 15, 13, 1, 7] and derived algorithms [4, 16, 8, 11, 10, 6, 2, 14, 18] are used for benchmarking. This causes a weakness: machine learning (ML) models cannot handle real-world unexpected road events [12]. To make ML models more robust, we describe an iterative procedure for creating out-of-domain examples for autonomous driving based on existing AV standard datasets. In summary, our key contributions are: (a) DANGER[1], a framework of Danger-Aware Novel dataset Generator Extension for Robustness test, with user input of vehicle driving trajectories and postures to complete a sequence of frames of data generation. (b) DANGER also supports the movement and deletion of vehicles in individual frames and can simulate illogical special camera failure modes.

2 Method

2.1 Framework Overview

The DANGER framework generates new samples using an object-aware rendering algorithm and a set of primitives. We include five predefined primitives in our DANGER implementation with additional user-defined primitive functionality. In the results section, we use 3D scene de-rendering networks (3D-SDN) [17], an optimal algorithm that generates photo-level realistic synthetic images, as the renderer module to demonstrate the feasibility of our framework. In practice, and the renderer/de-renderer module such as Panoptic Neural Fields can also be selected in practical implementation [9].

[1] Our code and results for the framework and experiments has been open-sourced under the MIT License and is available at: `https://github.com/jayhsu0627/DANGER`.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

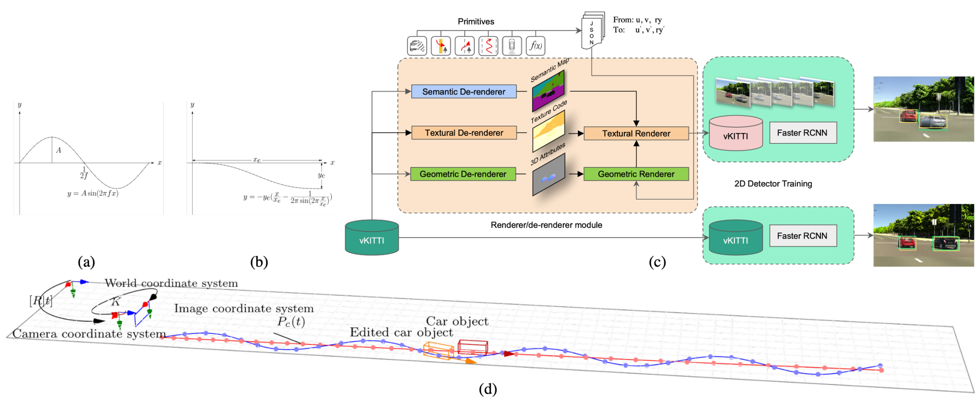

Figure 1: **DANGER architecture**. (a) slalom lane change function (b) cut-in function (c) DANGER contains a renderer/de-renderer module, primitive function module, and generated descriptive file that can help users to develop a wide variety of corner-case images based on any self-driving car datasets that support semantic and 3D annotations (d) the scene `0006` editing is shown in world coordinate.

The 3D-SDN employs an encoder-decoder architecture and has three branches: scene semantics, object geometry and 3D pose, the appearance of objects and the background. As shown in Fig. 1(c), the intentions of each branch are to learn the semantic segmentation of a scene, infer the object shape and 3D pose, and encode the appearance of each object and background segment. Disentangling 3D geometry and pose from the given scene enables 3D-aware scene manipulation with the given target location (u, v), pose r_y, and operations `(delete, modify)` in image coordinate.

Primitives We define a set of vehicle primitives to augment the input dataset. We represent five danger-aware primitives: `Disappear`, `Cut-in Opposite`, `Cut-in`, `Lane Change`, and `Reverse`. The point is that these primitives can be extended by composition and addition to cover larger error cases. For example, perhaps it has been identified that the computer vision algorithm is weak at predicting lane cuttings: when a car changes lanes quickly in front other another car. DANGER can be used to generate many of these cut-in behaviors to be used for training, re-training, and testing.

2.2 Scene Editing Computation

Given any defined function $f(x)$ as target, our approach projects a car object into a new location in the 2D plane under the world coordinate system. Given a image $I \in \mathbb{R}^{W \times H \times 3}$ with known camera intrinsic matrix $\mathbf{K} \in \mathbb{R}^{3 \times 4}$ and camera extrinsic matrix $[\mathbf{R}|\mathbf{t}]$, we can manipulate the position of an object in world, camera, or image coordinate system. $P_c \in \mathbb{R}^{4 \times 1}$ is the 3D point position in camera coordinates and $P_w \in \mathbb{R}^{4 \times 1}$ is the 3D point position in world coordinates. A camera extrinsic matrix $\mathbf{M} \in \mathbb{R}^{4 \times 4}$ is used to denote a projective mapping from world coordinates to pixel coordinates. The elements of object's center position vector P_c can be acquired from the MOT ground truth data of virtual KITTI [5], and the corresponding P_w will be easily obtained by apply the inverse of frame-dependent matrix $\mathbf{M}$. In the x-z plane, arbitrary vehicle poses can be generated according to the primitive function, where $\mathbf{r}'_y$ is a unit tangent vector to the curve at (x'_w, z'_w) representing the target orientation of the car object.

3 Results and conclusion

We conduct scene editing experiments on publicly available dataset virtual KITTI, as a proxy to KITTI, based on the descriptive files generated by the primitive functions. We show results on virtual KITTI here: `https://github.com/jayhsu0627/DANGER`. In this paper, we proposed a dataset expansion framework for generating hazardous driving scenarios. We hope DANGER can broadly accelerate AI research and value in improving ML model performance while also creating well-distributed, trusted datasets for ensuring safety-critical systems.

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.

[2] Yingfeng Cai, Lei Dai, Hai Wang, and Zhixiong Li. Multi-target pan-class intrinsic relevance driven model for improving semantic segmentation in autonomous driving. *IEEE Transactions on Image Processing*, 30:9069–9084, 2021.

[3] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8748–8757, 2019.

[4] Xuelian Cheng, Yiran Zhong, Mehrtash Harandi, Yuchao Dai, Xiaojun Chang, Hongdong Li, Tom Drummond, and Zongyuan Ge. Hierarchical neural architecture search for deep stereo matching. *Advances in Neural Information Processing Systems*, 33:22158–22169, 2020.

[5] Adrien Gaidon, Qiao Wang, Yohann Cabon, and Eleonora Vig. Virtualworlds as proxy for multi-object tracking analysis. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4340–4349, 2016.

[6] Aditya Ganeshan, Alexis Vallet, Yasunori Kudo, Shin-ichi Maeda, Tommi Kerola, Rares Ambrus, Dennis Park, and Adrien Gaidon. Warp-refine propagation: Semi-supervised auto-labeling via cycle-consistency. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15499–15509, 2021.

[7] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012.

[8] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[9] Abhijit Kundu, Kyle Genova, Xiaoqi Yin, Alireza Fathi, Caroline Pantofaru, Leonidas Guibas, Andrea Tagliasacchi, Frank Dellaert, and Thomas Funkhouser. Panoptic Neural Fields: A Semantic Object-Aware Neural Scene Representation. In *CVPR*, 2022.

[10] Cody Reading, Ali Harakeh, Julia Chae, and Steven L Waslander. Categorical depth distribution network for monocular 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8555–8564, 2021.

[11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[12] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. On a formal model of safe and scalable self-driving cars. *arXiv preprint arXiv:1708.06374*, 2017.

[13] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020.

[14] Xinshuo Weng, Jianren Wang, David Held, and Kris Kitani. Ab3dmot: A baseline for 3d multi-object tracking and new evaluation metrics. *arXiv preprint arXiv:2008.08063*, 2020.

[15] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next generation datasets for self-driving perception and forecasting. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021)*, 2021.

[16] Gangwei Xu, Junda Cheng, Peng Guo, and Xin Yang. Acvnet: Attention concatenation volume for accurate and efficient stereo matching. *arXiv preprint arXiv:2203.02146*, 2022.

[17] Shunyu Yao, Tzu Ming Harry Hsu, Jun-Yan Zhu, Jiajun Wu, Antonio Torralba, William T. Freeman, and Joshua B. Tenenbaum. 3d-aware scene manipulation via inverse graphics. In *Advances in Neural Information Processing Systems*, 2018.

[18] Xingyi Zhou, Vladlen Koltun, and Philipp Krähenbühl. Tracking objects as points. In *European Conference on Computer Vision*, pages 474–490. Springer, 2020.