LiDAR-Video Dataset: Learning Driving Policies more Effectively

Anonymous CVPR submission

Paper ID ****

Abstract

Learning autonomous-driving policies is one of the most challenging but promising tasks for computer vision and machine learning. It's generally accepted that future research and applications should combine cameras, videos and GPS Registered laser scanners in order to obtain comprehensive semantic understanding of real traffic. However, current approaches only learn from large-scale videos, for the lack of demanding benchmarks that provide precise laser-scanner data. In this paper, we firstly propose a LiDAR-Video dataset, which provides large-scale high-quality point clouds picked up by a Velodyne laser, videos and standard drivers' behaviors. Our accompanying baseline experiments, as well as the state-of-the-art approaches, demonstrate that extra laser information help networks to determine driving policies indeed. Our benchmark is available online: https://github.com/XXX/LiDAR-Video

1. Introduction

- learning policies is a challenging problem
- Lack of dataset designed for decision making. BBD only contains video images.
- Firstly purpose LiDAR-Video dataset

2. Related Work

- 2 mainstream ways to learn self-driving policies. 1) Object Segmentation(only using cameras). 2) Utilizing laser scanner.
- Existent dataset. (Cityscape, KITTI, BBD)

[5] [7] [6] [1] [11] [2] [11] [4] [13] [12] [3] [9] [8] [10]

3. Dataset

3.1. Platform and Data Collection

Describe our data acquisition platform, our devices such as digital recorder, Velodyne laser and others. . . .

Fig. 1 [car and devices figure]

3.2. Data Preprocessing

Describe how to process the origin data, including frame merging, time matching, format converting and downsampling.

Fig. 2 [Pipeline of data preprocessing]

3.3. Hierarchy and Features

Show the structure of out dataset and give detailed explanation.

- 7000+ pairs(video+point clouds)
- Drivers(wheel-angle and driving speed)

Fig. 3 [Display point clouds and recorder scenes]
Fig. 4 [Dataset statistic features]

4. Tasks and Metrics

Define the tasks 1) predict the real-time car speed. 2) predict the real-time wheel angle. 3) combine together

The effectiveness and practicability of tasks. Giving theory demonstration.

5. Experimental Evaluation

CNN, CNN+RNN, pointnet, feature-map

Fig. 5,6 Examples of Prediction
Table. 1,2 Comparision of Results

6. Conclusion

7. Future Work

The utilization of LiDAR point clouds. Unsupervised segmentation of point clouds.

References

[1] P. Babahajiani, L. Fan, J.-K. Kämäräinen, and M. Gabbouj. Urban 3d segmentation and modelling from street view images and lidar point clouds. *Machine Vision and Applications*, May 2017. 1

[2] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba. End to end learning for self-driving cars. *CoRR*, abs/1604.07316, 2016. 1

[3] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. *CoRR*, abs/1604.01685, 2016. 1

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009. 1

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, Jun 2010. 1

[6] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013. 1

[7] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1

[8] S. Gupta, R. B. Girshick, P. Arbelaez, and J. Malik. Learning rich features from RGB-D images for object detection and segmentation. *CoRR*, abs/1407.5736, 2014. 1

[9] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. 1

[10] W. Lotter, G. Kreiman, and D. Cox. Deep predictive coding networks for video prediction and unsupervised learning. *CoRR*, abs/1605.08104, 2016. 1

[11] D. A. Pomerleau. Advances in neural information processing systems 1. chapter ALVINN: An Autonomous Land Vehicle in a Neural Network, pages 305–313. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1989. 1

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. Li. Imagenet large scale visual recognition challenge. *CoRR*, abs/1409.0575, 2014. 1

[13] H. Xu, Y. Gao, F. Yu, and T. Darrell. End-to-end learning of driving models from large-scale video datasets. *CoRR*, abs/1612.01079, 2016. 1