IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT

Automatic Generation of Synthetic LiDAR Point Clouds for 3-D Data Analysis

Fei Wang, Yan Zhuang, Member, IEEE, Hong Gu, and Huosheng Hu, Senior Member, IEEE

Abstract—The recent success of deep learning in 3-D data analysis relies upon the availability of large annotated data sets. However, creating 3-D data sets with point-level labels are extremely challenging and require a huge amount of human efforts. This paper presents a novel open-sourced method to extract light detection and ranging point clouds with ground truth annotations from a simulator automatically. The virtual sensor can be configured to simulate various real devices, from 2-D laser scanners to 3-D real-time sensors. Experiments are conducted to show that using additional synthetic data for training can: 1) achieve a visible performance boost in accuracy; 2) reduce the amount of manually labeled real-world data; and 3) help to improve the generalization performance across data sets.

Index Terms—Deep learning, semantic segmentation, synthetic light detection and ranging (LiDAR) point clouds.

I. INTRODUCTION

DEEP learning technology has become the research focus of various light detection and ranging (LiDAR)-based perception tasks, such as recognition [1], semantic segmentation [2], and scene understanding [3]. These methods often require a large amount of labeled training data which means a huge amount of human efforts. Tracing accurate object boundaries in 3-D is very difficult. This is particularly true for data sets acquired with real-time LiDAR sensors. As a result, there are very limited LiDAR point cloud data sets available. To solve this problem, one possible way is to use simulators to automatically generate annotated LiDAR point clouds.

The idea of using synthetic data starts from computer vision communities. Different methods have been presented to extract synthetic images from simulators [4], [5]. In terms of LiDAR simulation, Yue et al. [6] presented a framework to generate synthetic point clouds from a video game for road objects segmentation. They observed a 9% improvement in accuracy when augmenting training data with the synthetic data. In [7], more realistic data were generated by using point clouds collected from the real world as a static background. But it required human interactions to generate dynamic traffic scenes. In [9], synthetic data were used to train a deep model for segmentation of road objects in urban environments.

To the best of our knowledge, limited works have been done on automatic generation of simulated LiDAR point clouds with point-level labels for 3-D data analysis. The main contributions of this paper are as follows.

1) An open-sourced method to automatically generate 3-D annotated LiDAR point clouds with highly configurable parameters to simulate various real devices. The code is available at https://github.com/ZhuangYanDLUT/carla.

2) Showing by experiments that using synthetic data can: 1) achieve a visible performance boost of a deep model; 2) significantly reduce the amount of manually labeled data for training; and 3) help to avoid the data set bias problem when real data contains limited types of scenes.

II. GENERATION OF SYNTHETIC 3-D DATA

Our synthetic data generation method is based on CARLA [9], a simulator with various digital assets (such as urban layouts, buildings, and vehicles). The virtual LiDAR sensor uses ray casting to simulate a laser ray. The ray casting API, provided by the engine, takes the sensor location and the 3-D coordinates of the ending point of a ray as inputs and returns descriptions about the first point it hits. If a ray does not hit anything, it is filtered out in the outputs. Three-dimensional coordinates of all hit points are then transformed from the world frame to the sensor frame and sent to the client.

We use the returned descriptions of ray casting to extract point-level labels. The descriptions include a pointer to an actor/component that the hit point belongs to. The actor is the base class for an object that can be placed in the simulation. An actor may have several components to control its behaviors or rendering processes. Both actors and components contain an array of tags. These tags are generated by the simulator during rendering. They are divided into 13 categories, e.g., buildings, fences, and others. Thus, we extract these tags as semantic labels for the hit points.

The virtual sensors allow for flexible configurations to simulate various types of real devices. Typical configurable parameters are shown in Fig. 1. Users can set these parameters to generate synthetic data of different sensors and settings, such as two vertically side-faced 2-D scanners (used in Oakland data set), a Velodyne HDL-64 laser scanner (used in KITTI data set), or a high-resolution static laser scanner (used in Semantic3D.net data set). Fig. 2 shows the examples of the synthetic data generated from these settings.
Fig. 1. Configurable parameters of the virtual sensor. $\mu$ and $\nu$ denote vertical and horizontal FOV; $\alpha$ and $\beta$ are vertical and horizontal resolutions.

Besides, the scenes are also configurable. Users can specify the car model, number of cars, number of pedestrians, and weather and time of day in the simulation. The virtual world map can also be modified or customly built if it is needed.

Our method is fully automatic and no human interaction is needed during data collection. Once the parameters are configured, a vehicle equipped with sensors will drive around autonomously. If it crashes into other objects, the simulation will restart within a few seconds. Measurements are saved automatically when the vehicle has moved a certain distance from the last saved point.

III. EXPERIMENTAL RESULTS

In our experiments, we use the same sensor configurations as those used in KITTI data set to collect a synthetic data set. The simulations run three times with a different number of dynamic objects to make the data more diverse. Totally 1703 frames are stored. It takes about 15 h to generate the entire synthetic data set.

The usage of synthetic data is mainly tested on semantic segmentation. A deep model, similar to [2], is jointly trained on the synthetic and real-world data set. The performances are evaluated in terms of overall accuracy (OA), mean of per class accuracy (mACC), and mean of per class intersection over union (mIoU).

A. Limited Real Data

In this experiment, we use different amounts of real data to train a deep model. The results are shown in Fig. 3. For models trained only on real data, significant performance degradation is observed when reducing the size of the data set. If synthetic...
Besides, with 1/8 of real data, our model can achieve similar performance to the other approaches trained on the entire data set if synthetic data are used during training.

### IV. Conclusion

In this paper, we presented a new method to automatically extract 3-D LiDAR point clouds with point-level ground truth labels from an autonomous driving simulator for 3-D data analysis. Comprehensive evaluation results demonstrate that: 1) training with synthetic data can improve the performance of a deep model and this improvement becomes more prominent as the real-world data set becomes smaller; 2) using synthetic data can significantly reduce the amount of manually labeled real data that required to train a deep model and thus release human efforts to create of 3-D data sets; and 3) using synthetic data also helps to improve generalization performance when real data contains limited types of scenes.