Lidar Point Cloud Guided Monocular 3D Object Detection

Liang Peng[1,2], Fei Liu[3], Zhengxu Yu[1], Senbo Yan[1,2], Dan Deng[2], Zheng Yang[2], Haifeng Liu[1], and Deng Cai[1,2] ✉

[1] State Key Lab of CAD&CG, Zhejiang University, China
{pengliang, senboyan, haifengliu}@zju.edu.cn yuzxfred@gmail.com
dengcai@cad.zju.edu.cn
[2] Fabu Inc., Hangzhou, China
{dengdan, yangzheng}@fabu.ai
[3] State Key Lab of Industrial Control and Technology, Zhejiang University, China
liufei21@zju.edu.cn

Abstract. Monocular 3D object detection is a challenging task in the self-driving and computer vision community. As a common practice, most previous works use manually annotated 3D box labels, where the annotating process is expensive. In this paper, we find that the precisely and carefully annotated labels may be unnecessary in monocular 3D detection, which is an interesting and counterintuitive finding. Using rough labels that are randomly disturbed, the detector can achieve very close accuracy compared to the one using the ground-truth labels. We delve into this underlying mechanism and then empirically find that: concerning the label accuracy, the 3D location part in the label is preferred compared to other parts of labels. Motivated by the conclusions above and considering the precise LiDAR 3D measurement, we propose a simple and effective framework, dubbed LiDAR point cloud guided monocular 3D object detection (LPCG). This framework is capable of either reducing the annotation costs or considerably boosting the detection accuracy without introducing extra annotation costs. Specifically, It generates pseudo labels from unlabeled LiDAR point clouds. Thanks to accurate LiDAR 3D measurements in 3D space, such pseudo labels can replace manually annotated labels in the training of monocular 3D detectors, since their 3D location information is precise. LPCG can be applied into any monocular 3D detector to fully use massive unlabeled data in a self-driving system. As a result, in KITTI benchmark, we take the first place on both monocular 3D and BEV (bird's-eye-view) detection with a significant margin. In Waymo benchmark, our method using 10% labeled data achieves comparable accuracy to the baseline detector using 100% labeled data. The codes are released at https://github.com/SPengLiang/LPCG.

Keywords: monocular 3D detection, LiDAR point cloud, self-driving.

1 Introduction

3D object detection plays a critical role in many applications, such as self-driving. It gives cars the ability to perceive the world in 3D, avoiding collisions with other