International Journal of Computer Vision (2023) 131:3332–3352
https://doi.org/10.1007/s11263-023-01869-9

GLENet: Boosting 3D Object Detectors with Generative Label Uncertainty Estimation

Yifan Zhang[1] · Qijian Zhang[1] · Zhiyu Zhu[1] · Junhui Hou[1] · Yixuan Yuan[2]

Received: 20 July 2022 / Accepted: 26 July 2023 / Published online: 15 August 2023

Abstract

The inherent ambiguity in ground-truth annotations of 3D bounding boxes, caused by occlusions, signal missing, or manual annotation errors, can confuse deep 3D object detectors during training, thus deteriorating detection accuracy. However, existing methods overlook such issues to some extent and treat the labels ass deterministic. In this paper, we formulate the label uncertainty problem as the diversity of potentially plausible bounding boxes of objects. Then, we propose GLENet, a generative framework adapted from conditional variational autoencoders, to model the one-to-many relationship between a typical 3D object and its potential ground-truth bounding boxes with latent variables. The label uncertainty generated by GLENet is a plug-and-play module and can be conveniently integrated into existing deep 3D detectors to build probabilistic detectors and supervise the learning of the localization uncertainty. Besides, we propose an uncertainty-aware quality estimator architecture in probabilistic detectors to guide the training of the IoU-branch with predicted localization uncertainty. We incorporate the proposed methods into various popular base 3D detectors and demonstrate significant and consistent performance gains on both KITTI and Waymo benchmark datasets. Especially, the proposed GLENet-VR outperforms all published LiDAR-based approaches by a large margin and achieves the top rank among single-modal methods on the challenging KITTI test set. The source code and pre-trained models are publicly available at https://github.com/Eaphan/GLENet.

Keywords 3D object detection · Label uncertainty · Conditional variational autoencoders · Probabilistic object detection · 3D point cloud

1 Introduction

As one of the most practical application scenarios of computer vision, 3D object detection has been attracting much academic and industrial attention in the current deep learning era with the rise of autonomous driving and the emergence of large-scale annotated datasets (e.g., KITTI (Geiger et al., 2012), and (Sun et al., 2020)).

In the current community, despite the proliferation of various deep learning-based 3D detection pipelines, it is observed that mainstream 3D object detectors are typically designed as deterministic models, without considering the critical issue of the ambiguity of annotated ground-truth labels. However, different aspects of ambiguity/inaccuracy inevitably exist in the ground-truth annotations of object-level bounding boxes, which may significantly influence the overall learning process of such deterministic detectors. For example, in the data

Communicated by Jiri Matas.

This work was supported in part by the Hong Kong Research Grants Council under Grant 11202320 and Grant 11219422, and in part by the Hong Kong Innovation and Technology Fund under Grant MHP/117/21.

✉ Junhui Hou
jh.hou@cityu.edu.hk

Yifan Zhang
yzhang3362-c@my.cityu.edu.hk

Qijian Zhang
qijizhang3-c@my.cityu.edu.hk

Zhiyu Zhu
zhiyuzhu2-c@my.cityu.edu.hk

Yixuan Yuan
yxyuan@ee.cuhk.edu.hk

1 Department of Computer Science, City University of Hong Kong, Hong Kong, China

2 Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China