International Journal of Automotive Technology, Vol. 24, No. 2, pp. 469–481 (2023)
DOI 10.1007/s12239–023–0039–0

pISSN 1229–9138/ eISSN 1976–3832

HIGH DEFINITION MAP AIDED OBJECT DETECTION FOR AUTONOMOUS DRIVING IN URBAN AREAS

Yuki Endo[1)*], Ehsan Javanmardi[2)], Yanlei Gu[3)] and Shunsuke Kamijo[2)]

[1)]Graduate School of Information Science and Technology, Department of Information & Communication Engineering, The University of Tokyo, Tokyo 153-8505, Japan
[2)]The Institute of Industrial Science (IIS), The University of Tokyo, Tokyo 153-8505, Japan
[3)]College of Information Science and Engineering, Ritsumeikan University, Shiga 525-0058, Japan

(Received 28 December 2021; Revised 22 June 2022; Accepted 26 September 2022)

ABSTRACT–Detecting object locations and semantic classes in an image, such as traffic signs, traffic lights, and guide signs, is the crucial problem for autonomous driving, known as object detection. However, stable object detection in complex real-world environments, such as urban environments, is still challenging because of clutter, time of day, blur etc., even with modern deep convolutional neural networks (DCNNs). On the other hand, a high definition (HD) map is a pre-built information resource for autonomous driving tasks, especially for controls. Besides controls, HD map utilization for detection tasks has been gaining attention in recent years, enabling us to stabilize detection even in complex real-world environments. However, it is challenging to use object information from an HD map as detection directly because the self-localization error affects the transformed object locations on the image coordinate system from the HD map's coordinate system. This paper explores incorporating HD map information into deep feature maps of a DCNN-based model. Two proposed modules implicitly make the feature extraction efficient and stable by utilizing HD map information. As a result of the experiments, the proposed module improved a modern model for challenging images of the urban area Shinjuku by 37 % in mAP, even in self-localization errors.

KEY WORDS : Autonomous driving, Object detection, Deep learning, High definition map, Self-localization

1. INTRODUCTION

Detecting object locations and semantic classes in a traffic image is one of the most crucial tasks in autonomous driving, called object detection. In recent years, deep convolutional neural networks (DCNNs) have achieved promising performance in object detection via efficient deep feature extractors and detector heads (Ren *et al*., 2015; Liu *et al*., 2016; Redmon *et al*., 2015). However, because images taken in a natural environment have clutter, blur, weather variations, and time of day variations, developing robust object detectors is still challenging (Liu *et al*., 2020). Especially in urban areas, object detection problems have become more complicated.

On the other hand, a high definition (HD) map has been attracted as a critical technology to enhance operations in autonomous driving (Seif and Hu, 2016; Zang *et al*., 2018; Takeuchi *et al*., 2018; Yang *et al*., 2018; Ravi Kiran *et al*., 2018; Bender *et al*., 2014). An HD map contains physical and semantic information about environments in the 3D space, such as lanes, traffic objects, and elevation information. In recent years, mapping companies, such as TomTom, HERE, and Baidu, have been developing HD maps for various environments (TomTom International BV., 2021; HERE, 2021; Baidu, 2021). Many works using HD maps have been done on path planning, but HD maps can also be utilized in perception (Yang *et al*., 2018; Possatti *et al*., 2019). By using the traffic objects information contained in the HD map, it is expected to be able to solve the object detection problem in real complex environments.

For HD map utilization in object detection, HD map information such as object shapes and semantic labels should be projected onto an image coordinate system using a self-localization result. However, the projection may cause errors because self-localization results often have uncertainty, especially in urban areas (Gu *et al*., 2016), even with map-based self-localization (Javanmardi *et al*., 2018; Endo *et al*., 2021). For this reason, recent studies only use HD map information for inspection at post-processing in perception tasks (Possatti *et al*., 2019; Hirabayashi *et al*., 2019). Even though HD maps contain rich information about objects, they do not use that information to improve the accuracy and robustness of object detection. Using prior object information for feature extraction in DCNN may improve robustness against noisy

**Corresponding author*. e-mail: engmech1@sina.com