*International Journal of Automotive Technology*, Vol. 24, No. 3, pp. 773–786 (2023)
DOI 10.1007/s12239–023–0064–z

pISSN 1229–9138/ eISSN 1976–3832

# CNN-BASED OBJECT DETECTION AND DISTANCE PREDICTION FOR AUTONOMOUS DRIVING USING STEREO IMAGES

Jin Gyu Song and Joon Woong Lee*

Department of Industrial Engineering, Chonnam National University, Gwangju 61186, Korea

*(Received 26 August 2022; Revised 28 November 2022; Accepted 21 December 2022)*

**ABSTRACT**–Convolutional neural networks (CNNs) have been successful for tasks such as object detection; however, they involve time-consuming processes. Therefore, there are difficulties in applying these CNNs to autonomous driving. Moreover, most autonomous driving technologies require both object detection and distance prediction. However, CNNs that predict distance involve more time-consuming processes than object detection models. In addition, the applications for autonomous driving require object detection and distance prediction accuracy. This paper proposes an end-to-end trainable CNN that can meet these requirements. The proposed CNN accurately implements object detection and distance prediction in real time using stereo images. We demonstrate the superiority of the proposed CNN using stereo images from the KITTI 3D object detection dataset.

**KEY WORDS** : Autonomous driving, Object detection, Distance prediction, Real-time processing

## 1. INTRODUCTION

To effectively adopt image processing to technologies such as autonomous driving and robotics, it is necessary to detect objects in a captured image and to estimate their distance from the camera in real-time (Zaarane *et al.*, 2020). Adaptive cruise control (ACC) is an autonomous driving technology that requires both object detection and distance measurement (Han *et al.*, 2016; Yurtsever *et al.*, 2020). In this paper, we propose an end-to-end trainable CNN that can accurately perform object detection and distance prediction from stereo images in real time for application to an autonomous driving technology such as ACC. We aim to detect a vehicle in the form of a two-dimensional (2D) bounding box (BB) and predict the shortest distance from the camera to the detected vehicle.

Recently, with the advent of various deep learning algorithms, there has been significant progress in research on object detection in images. Generally, 2D object detection is divided into two processes. The first process finds all possible regions where objects can be located in a given image. The regions identified in this process are called region proposals, and they are expressed with 2D BBs. The second process classifies whether the proposed regions from the previous process belong to one of the classes of interest or to the background.

Various CNN-based deep learning models from the Region-based Convolutional Neural Network (R-CNN) (Girshick *et al.*, 2014) and Fast R-CNN (Girshick, 2015) to You Only Look Once version 4 (YOLOv4) (Bochkovskiy *et al.*, 2020) have been proposed for 2D object detection. The earliest object detection algorithms, such as R-CNN and Fast R-CNN, search for areas where objects are likely to exist in a given image using a technique such as selective search. Then, the CNN determines whether the searched area contains an object. In the algorithms that were proposed after Fast R-CNN, the CNN performs a series of processes such as extracting features from an input image, searching for the location of an object, and classifying the object. Models that perform such functions include Faster R-CNN (Ren *et al.*, 2016), Single Shot MultiBox Detector (SSD) (Liu *et al.*, 2016), and the YOLO series (Redmon *et al.*, 2016; Redmon and Farhadi, 2018; Bochkovskiy *et al.*, 2020). Among these models, the R-CNN series and SSD have difficulty with real-time object detection, whereas YOLOv4 has the ability to detect objects in real time.

Recently, CNNs have been used to predict disparities from a pair of stereo images. The first method to attempt disparity prediction using a CNN was MC-CNN (Zbontar and LeCun, 2016). Since MC-CNN was proposed, various CNN-based disparity prediction algorithms, such as GC-Net (Kendall *et al.*, 2017), PSM Net (Chang and Chen, 2018) and GWC Net (Guo *et al.*, 2019), have been proposed.

In general, when the absolute distance is predicted using stereo images, the corresponding pixels between the left and right images are first searched, and the disparity is

*Corresponding author. e-mail: joonlee@chonnam.ac.kr