A object detection algorithm based on pyramid Convolutional Neural Networks (CNN) and feature map fusion model

Nu Wen [a], *, Biao He [a], Zhilu Yuan [a], Yong Fan [a],

[a] Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen
University, Shenzhen, China; whu_hebiao@hotmail.com, zhiluyuan@szu.com.cn, gisfanyong@163.com
* Corresponding author: wennu1989@126.com

Keywords: Convolutional Neural Networks; pyramid model; feature fusion; object detection;

Abstract:

The aim of this paper is to solve two problems: object detection of small objects and multi-view scenes. First, in practical applications, the collected traffic video is affected by the resolution, viewing angle, focal length and model of the front-end acquisition device. The object size, shape and attitude of the video to be detected are different, resulting in the overall detection performance of the algorithm recognition. In particular, for traffic intersections, the size of the vehicle is related to the distance between the vehicle and the camera, and the object resolution of the vehicle near the intersection is relatively high. As the relative distance increases, the resolution of the object gradually decreases, resulting in feature extraction of the detection object to be detected. And identification becomes more and more difficult, and the probability of the object being detected is greatly reduced. Secondly, there are usually many ways to collect traffic data, such as fixed-position camera, high-altitude camera, and cruising UAV (Unmanned Aerial Vehicle). These video sources collected at different viewing angles and locations pose challenges to the stability, robustness, and generalization capabilities of the detection algorithms. Therefore, design a new algorithm and optimizing model parameters and training samples of different source data is extremely important for multi-view object detection.

An object detection algorithm based on pyramid Convolutional Neural Networks(CNN) and feature map fusion method was proposed, and the deep learning technology and the object detection algorithm are used to detect and identify the video objects of multiple viewing angles and different resolution scenes in traffic field. By mixing the lower and deeper feature map model, the algorithm can detect a smaller object in multiple viewing angles and different resolution scenes. Meanwhile, an image block and multi-threading technology was used to avoid scale limit of input image. The experiments show that it can be more efficient and accurate in practical applications of traffic detection filed.

The method can be used for the existing network model (VGG16, ResNet101, etc.) to build the skeleton of the object detection algorithm. A new object detection algorithm is developed for these goals, which contain small object recognition and multi-view recognition of traffic video, and it can enable it to extract the lower features of the object and effectively realize multi-object recognition of different scenes. Using the pyramid CNN model, it is possible to effectively combine low-level features and high-level features to achieve feature extraction and fusion of the object, and to solve the problem of small object recognition accuracy to a certain extent. Meanwhile, in view of the shortcomings of the existing object detection algorithm to re-compress the image size, the image block and multi-threading technology are used to restore the original resolution of the image. By using this technology, the accuracy of image object to be detected can be improved.