

Parallel Feature Pyramid Network for Object Detection

Seung-Wook Kim^[0000–0002–6004–4086], Hyong-Keun Kook, Jee-Young Sun,
Mun-Cheon Kang, and Sung-Jea Ko

School of Electrical Engineering, Korea University, Seoul, South Korea
fswkim, hkkook, jysun, mckang@dal.ikorea.ac.kr, sjko@korea.ac.kr

Abstract. Recently developed object detectors employ a convolutional neural network (CNN) by gradually increasing the number of feature layers with a pyramidal shape instead of using a featurized image pyramid. However, the different abstraction levels of CNN feature layers often limit the detection performance, especially on small objects. To overcome this limitation, we propose a CNN-based object detection architecture, referred to as a parallel feature pyramid (FP) network (PFPNet), where the FP is constructed by widening the network width instead of increasing the network depth. First, we adopt spatial pyramid pooling and some additional feature transformations to generate a pool of feature maps with different sizes. In PFPNet, the additional feature transformation is performed in parallel, which yields the feature maps with similar levels of semantic abstraction across the scales. We then resize the elements of the feature pool to a uniform size and aggregate their contextual information to generate each level of the final FP. The experimental results confirmed that PFPNet increases the performance of the latest version of the single-shot multi-box detector (SSD) by mAP of 6.4% AP and especially, 7.8% AP_{small} on the MS-COCO dataset.

Keywords: Real-Time Object Detection, Feature Pyramid.

1 Introduction

Multi-scale object detection is a difficult and fundamental challenge in computer vision. Recently, object detection has achieved a considerable progress thanks to a decade of advances in convolutional neural networks (CNNs).

The early CNN-based object detectors utilize a deep CNN (DCNN) model as part of an object detection system. OverFeat [38] applies a CNN-based classifier to an image pyramid in a sliding window manner [5, 7]. The regions with CNN features (R-CNN) method [10] adopts a region-based approach (also known as a two-stage scheme), where the image regions of object candidates are provided for a CNN-based classifier. Recent region-based detectors such as Fast R-CNN [9] and Faster R-CNN [35] utilize a single-scale feature map, which is transformed by a DCNN model, as shown in Fig. 1(a) (top). In [35], using this single-scale feature, a complete object detection system is formed as an end-to-end CNN model and exhibits state-of-the-art performance.

