

# Bi-box Regression for Pedestrian Detection and Occlusion Estimation

Chunluan Zhou<sup>1,2</sup>[0000–0003–0284–6256] and Junsong Yuan<sup>2</sup>[0000–0002–7324–7034]

<sup>1</sup> Nanyang Technological University, Singapore

<sup>2</sup> The State University of New York at Buffalo, USA  
czhou002@e.ntu.edu.sg, jsyuan@buffalo.edu

**Abstract.** Occlusions present a great challenge for pedestrian detection in practical applications. In this paper, we propose a novel approach to simultaneous pedestrian detection and occlusion estimation by regressing two bounding boxes to localize the full body as well as the visible part of a pedestrian respectively. For this purpose, we learn a deep convolutional neural network (CNN) consisting of two branches, one for full body estimation and the other for visible part estimation. The two branches are treated differently during training such that they are learned to produce complementary outputs which can be further fused to improve detection performance. The full body estimation branch is trained to regress full body regions for positive pedestrian proposals, while the visible part estimation branch is trained to regress visible part regions for both positive and negative pedestrian proposals. The visible part region of a negative pedestrian proposal is forced to shrink to its center. In addition, we introduce a new criterion for selecting positive training examples, which contributes largely to heavily occluded pedestrian detection. We validate the effectiveness of the proposed bi-box regression approach on the Caltech and CityPersons datasets. Experimental results show that our approach achieves promising performance for detecting both non-occluded and occluded pedestrians, especially heavily occluded ones.

**Keywords:** Pedestrian detection · Occlusion handling · Deep CNN

## 1 Introduction

Pedestrian detection has a wide range of applications including autonomous driving, robotics and video surveillance. Many efforts have been made to improve its performance in recent years [3, 8, 17, 6, 33, 40, 5, 39, 41, 37, 34, 4]. Although reasonably good performance has been achieved on some benchmark datasets for detecting non-occluded or slightly occluded pedestrians, the performance for detecting heavily occluded pedestrians is still far from being satisfactory. Take the Caltech dataset [9] for example. One of the top-performing approaches, SDS-RCNN [4], achieves a miss rate of about 7.4% at 0.1 false positives per image (FPPI) for non-occluded or slightly occluded pedestrian detection, but its miss rate increases dramatically to about 58.5% at 0.1 FPPI for heavily occluded



































