

Universal-Prototype Enhancing for Few-Shot Object Detection: Supplementary Material

Aming Wu¹ Yahong Han^{2,3,4} Linchao Zhu⁵ Yi Yang⁵

¹School of Electronic Engineering, Xidian University, Xi'an, China

²College of Intelligence and Computing, Tianjin University, Tianjin, China

³Tianjin Key Lab of Machine Learning, Tianjin University, Tianjin, China

⁴Peng Cheng Laboratory, Shenzhen, China ⁵ReLER Lab, AAIL, University of Technology Sydney

amwu@xidian.edu.cn, yahong@tju.edu.cn, {Linchao.Zhu, yi.yang}@uts.edu.au

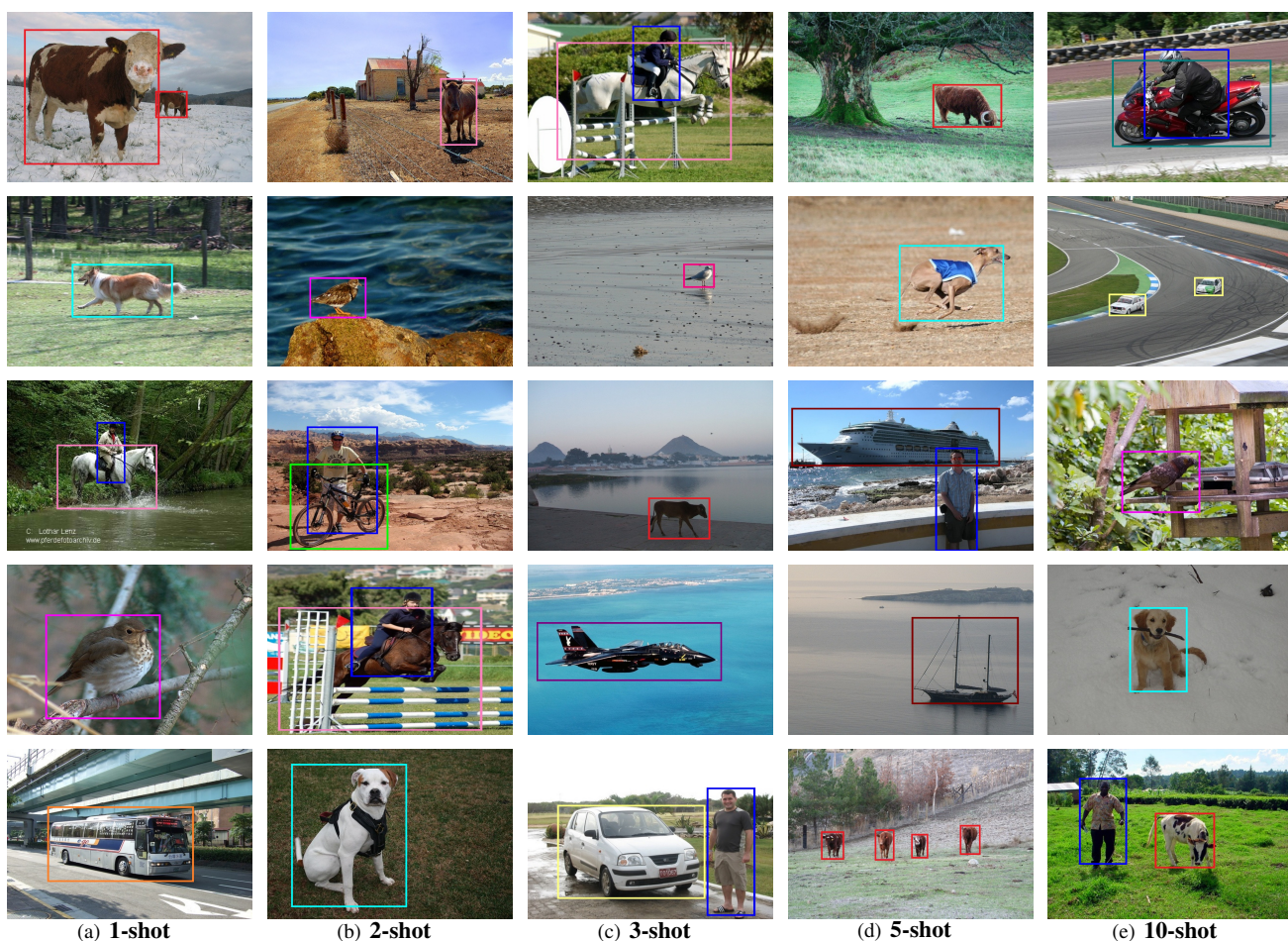


Figure 1: Detection results with different shots based on VOC Novel Set 1. Our method localizes and recognizes objects existing in images accurately, which demonstrates the effectiveness of the proposed method.

1. Analysis of Detection Results

In Fig. 1, we show many detection results of our method. We can see that based on different number of samples with

novel categories, our method detects objects accurately. Particularly, for images that contain multiple objects, our method accurately localizes and recognizes these objects, which further demonstrates that our method is effective for

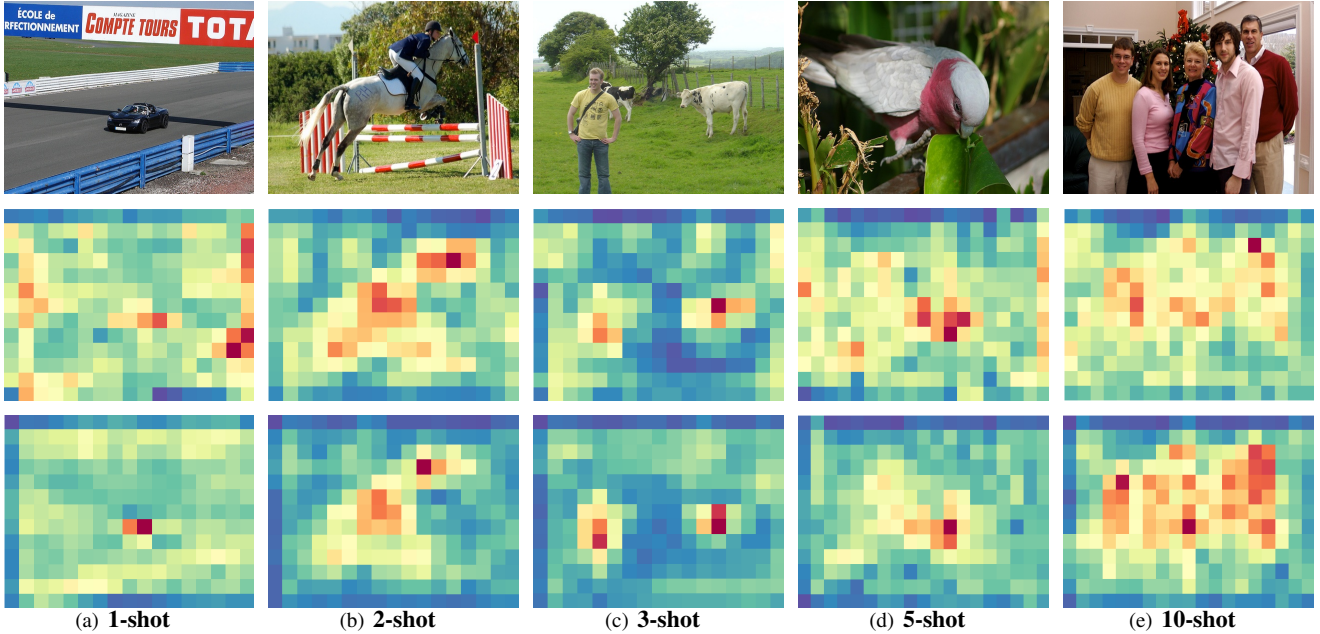


Figure 2: Visualization of the feature map used for RPN based on different shots. The second and third row separately indicate F and the output of Ψ (see Eq. (2)). For each feature map, the channels corresponding to the maximum value are selected for visualization.

few-shot object detection.

2. Visualization Analysis

Based on VOC Novel Set 1, we further make an ablation analysis of the proposed universal-prototype enhancement.

In this paper, we fuse descriptors that are generated based on universal prototypes into the current features F . To further demonstrate the effectiveness of this operation, in Fig. 2, we show many visualization results of the original features F and the output of Ψ (see Eq. (2) in the submitted paper). We can see that for cases with different shots, fusing descriptors promotes the features to contain more object-related information. Taking the first image as an example, the original features F contain much object-irrelevant information, which affects the accuracy of detection. Through fusing descriptors, the output of Ψ contains much object-related information, which is beneficial for improving detection performance.