Supplementary Material for DCNet

Hanzhe Hu[1], Shuai Bai[2], Aoxue Li[1], Jinshi Cui[1]*, Liwei Wang[1]*
[1]Key Laboratory of Machine Perception (MOE), School of EECS, Peking University
[2]Beijing University of Posts and Telecommunications
{huhz, lax}@pku.edu.cn baishuai@bupt.edu.cn {cjs, wanglw}@cis.pku.edu.cn

In this supplementary material, we provide additional visualization results. In section 1, we present additional few-shot detection results on PASCAL VOC dataset. In section 2, qualitative results on PASCAL VOC dataset are presented. In section 3, qualitative results on COCO dataset are presented.

1. Additional Experimental Results

In this section, we provide extra few-shot detection results on PASCAL VOC [1] dataset.

1.1. Run-time Latency

As shown in Table 1, during inference, the baseline method Meta R-CNN takes 0.125s per image while our DCNet takes 0.137s, where the speed is only about 6.3% slower since our method can get rid of the time-consuming class-specific prediction. The results are averaged over three splits on PASCAL VOC.

2. Qualitative Results on PASCAL VOC Dataset

In this section, we provide qualitative results on PASCAL VOC [1] dataset. As shown in Fig. 1, we compare the detection results of our proposed DCNet with the baseline method Meta R-CNN [3] under our own implementation. Specifically, in the first line, due to the occlusion, the baseline method fails to detect the bicycle behind, while DCNet successfully retrieves all the query objects; in the second line, with a relatively small object to detect, though performing the right localization, the baseline method fails to predict the corresponding category, while DCNet is capable of dealing with classification and localization for the small object; in the third line, DCNet serves as increasing the confidence of classification and resolving the ambiguity; in the fourth line, due to severe occlusions, some query object is neglected by the baseline method, and DCNet successfully overcomes the occlusion problem and detects all the objects. Hence, the proposed DCNet can effectively improve the detection performance under occlusions and scale variations.

3. Qualitative Results on COCO Dataset

We provide visualization results on COCO [2] dataset. In Fig. 2, we compare the detection results of our proposed DCNet with the baseline method Meta R-CNN [3] under our own implementation. Specifically, the first line reflects the classification robustness of DCNet. The second and third line demonstrate the superiority of our method under severe occlusions. The last line manifests that our method can effectively overcome the problems of scale variations and appearance changes. Hence, our method exhibits significant improvement over the baseline method.

References

[1] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 1

[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 1

[3] Xiaopeng Yan, Ziliang Chen, Anni Xu, Xiaoxi Wang, Xiaodan Liang, and Liang Lin. Meta r-cnn: Towards general solver for instance-level low-shot learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9577–9586, 2019. 1

* Corresponding authors.

Method	Inference Time per Image (s)
Meta-RCNN	0.125
DCNet	0.137

Table 1. Run-time latency comparisons during testing on PASCAL VOC dataset.

Figure 1. Qualitative results on PASCAL VOC dataset. From left to right: query image, prediction by baseline method, prediction by DCNet and ground truth.

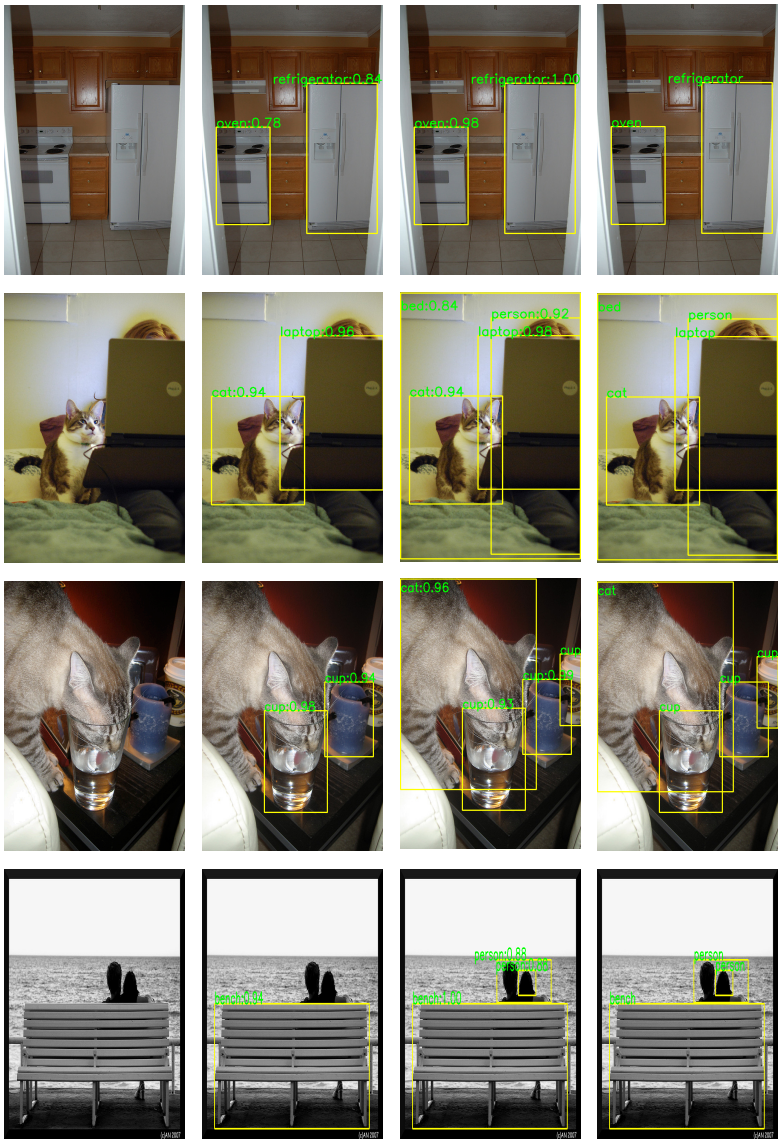

Figure 2. Qualitative results on COCO dataset. From left to right: query image, prediction by baseline method, prediction by DCNet and ground truth.