

Joint COCO and Mapillary Workshop at ICCV 2019:
COCO Instance Segmentation Challenge Track
Technical Report: enhanced hybrid task cascade

Luo Linfeng

Ping An International Smart City Technology Co., Ltd

LUOLINFENG619@pingan.com.cn

Xie Yuyang

Ping An International Smart City Technology Co., Ltd

xieyuyang273@pingan.com.cn

Jiang Zhengkai

Institute of Automation, Chinese Academy of Science, Beijing

zhengkai.jiang@nlpr.ia.ac.cn

Zhang Yuqi

Ping An International Smart City Technology Co., Ltd

ZHANGYUQI286@pingan.com.cn

Jin Qizhao

Institute of Automation, Chinese Academy of Science, Beijing

qizhao.jin@nlpr.ia.ac.cn

Guo Chaoxu

Institute of Automation, Chinese Academy of Science, Beijing

chaoxu.guo@nlpr.ia.ac.cn

Wang Hongbin

Ping An International Smart City Technology Co., Ltd

WANGHONGBIN800@pingan.com.cn

Xiang Shiming

Institute of Automation, Chinese Academy of Science, Beijing

smxiang@nlpr.ia.ac.cn

Abstract

The Hybrid Task Cascade and Cascade-mask-rcnn are used as our basic architectures. Several techniques are made in the network, such as deformable convolution V2, GCNet, global average pooling FPN, all level roi extractor, extra convolution layers in the mask head. The network uses resnet-101, resnet-50 as backbone, and the final model use resnet-101. We train the network with coco-2017 dataset. The model is trained with multiscale schedule and random

flipping and the final prediction is made with the multiscale testing and flipping.

1. Method

We use the [4] to build our model. Starting with the [1], with resnet-50, and using the COCO2017 dataset only, we obtain the result of segmentation mAP of 35.6. The global average pooling operation and [5] are applied, and the result of mAP is 36.4. The [6] is then applied

on top, with the result of 38.8. The mAP of 39.2 is achieved with help of [2]. After that, the cascade mask head, heavier mask heads with more convolution layers and all level roi extractor are applied to obtain the mAP of 40.7. Finally, we use all the tested techniques which worked with Cascade-mask-rcnn in the [3]. The baseline of hybrid task cascade with resnet-50 and FPN has the mAP of 36. Then the HTC model is applied with all the techniques mentioned above along with multiscale training and semantic segmentation head, and the model achieves the mAP of 41. The final submitted prediction result is from the single model which stemmed from the HTC model mentioned before, except changing the backbone from resnet-50 to resnet-101 and adding multiscale testing at test time, which achieves the final test mAP of 42.4.

References

- [1] Zhaowei Cai and Nuno Vasconcelos. “Cascade r-cnn: Delving into high quality object detection”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 6154–6162.
- [2] Yue Cao et al. “GCNet: Non-local Networks Meet Squeeze-Excitation Networks and Beyond”. In: *arXiv preprint arXiv:1904.11492* (2019).
- [3] Kai Chen et al. “Hybrid task cascade for instance segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4974–4983.
- [4] Kai Chen et al. “MMDetection: Open MMLab Detection Toolbox and Benchmark”. In: *arXiv preprint arXiv:1906.07155* (2019).
- [5] Tsung-Yi Lin et al. “Feature pyramid networks for object detection”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2117–2125.
- [6] Xizhou Zhu et al. “Deformable convnets v2: More deformable, better results”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 9308–9316.