Joint COCO and Mapillary Workshop at ICCV 2019: Detection Challenge Track

Technical Report: Instance Segmentation mask

Shuchun Liu, Feiyun Zhang, Li Long, Weiyuan Shao, Jiajun Wang
The AI Lab of ELEME Inc, China
{shuchun.liu, feiyun.zhang, li.long02, weiyuan.shao, jiajun.wang}@ele.me

Abstract

Abstract: This paper introduces our solution for the full track of the COCO Detection Challenge (Segmentation Mask). Our detection system is based on Hybrid Task Cascade network. Guided Anchor Region proposal network is adopted to improve the performance. Some ticks is used to make our model has better robust performance, including data augment, method of keeping balance of positive and negative sample, multi-scale training and so on. We choose two backbones to extract features. Our submission is an ensemble of two models, it achieved mAP 0.46 on the test-challenge 2019 set.

1. Introduction

Instance segmentation has drawn more researcher's attention in recent years, which is a fundamental and hot research topics in the computer vision. With the rapid process of deep neural network and computing capacity, the Instance Segmentation task achieves high performance, which focus on the detection and segmentation of things. It should identify the category of object, then the bound box, coordinate of object is predicted. In addition, if an object is detected, a pixel mask should be predicted. The main purpose of instance segmentation is to distinguish different instances of things classes. This is depicted in Fig. 1.

Currently, one of the best state-of-the-art methods is based on Mask R-CNN[1], which extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. To achieve higher performance, Cascade Mask R-CNN[2] is published. The multi-stage object detection architecture reduced the overfitting problem, which are trained stage by stage, leveraging the observation that the output of a detector is a good distribution for training the next higher quality detector. Hybrid Task Cascade(HTC)[3], is based on Cascade R-CNN detector and Mask R-CNN, combining cascade and instance segmentation, it can learn more discriminative features progressively while integrating complementary features together in each stage. HTC has a strong improvement in Instance segmentation.

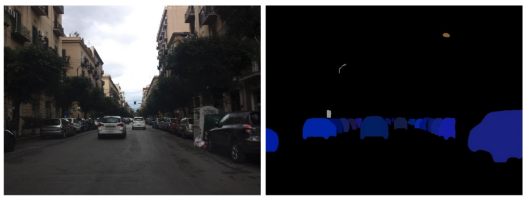

Figure 1: Predictions by the network for an image from the Mapillary Vistas validation set. Left: original image. Right: instance segmentation

2. Method

In this section, we will introduce our methods and details, including our backbone, obtained score and so on.

A. Train Data. We only use COCO Dataset to train our model with no other extra dataset. Some data augment methods are applied, such as random jitter, random horizontal flip, random contrast and so on. In addition, we apply InstaBoost in model training[4].

B. Network architecture. We adopt HTC network to predict instance segmentation outputs. Two backbones are used to extracted feature, including resnext-32x8d-101, resnext-64x4d-101[5]. Deformable convolution is used in resnext layer[2-4]. A Region Proposal Network called GA-RPN[6] is used to generate region proposals.

C. Process of model training. Multi-scale training is used to improve performance. The long side is range from 400 to 1400 with keep length-width ratio. OHEM[7] is adopted on training.

D. Inference. Multi-scale testing is used for each model. Firstly, region proposal is predicted on each single scale. Then map all region proposals' size to origin image size,

AP	AP^{50}	AP^{75}	AP^{L}	AP^{M}	AP^{S}	AR^{L}	AR^{M}	AR^{S}	AR^{1}	AR^{10}	AR^{100}
0.46	0.70	0.51	0.63	0.49	0.26	0.76	0.64	0.44	0.36	0.57	0.61

Table 1: Results on COCO test-challenge2019 dataset

and combine it to process Soft-NMS[8]. Region proposals remain are through the RCNN network on respective scale to predict the coordinate of bounding box. The mean of coordinate of bounding box on each scale after NMS is the detector result. Each bounding box has a segmentation mask. Also the mean of pixel mask on every scale is the final segmentation result that we need.

3. Result

We achieve 0.46 mAP on COCO Detection Challenge 2019 (segmentation mask). In Tab 1, we show the results on test-challenge2019 (segm) dataset.

References

[1] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1

[2] 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*, pages 6154–6162, 2018. 1

[3] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, et al. Hybrid task cascade for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4974–4983, 2019. 1

[4] Hao-Shu Fang, Jianhua Sun, Runzhong Wang, Minghao Gou, Yong-Lu Li, and Cewu Lu. Instaboost: Boosting instance segmentation via probability map guided copy-pasting. *arXiv preprint arXiv:1908.07801*, 2019. 1

[5] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017. 1

[6] Jiaqi Wang, Kai Chen, Shuo Yang, Chen Change Loy, and Dahua Lin. Region proposal by guided anchoring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2965–2974, 2019. 1

[7] Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 761–769, 2016. 1

[8] Navaneeth Bodla, Bharat Singh, Rama Chellappa, and Larry S Davis. Soft-nms–improving object detection with one line of code. In *Proceedings of the IEEE international conference on computer vision*, pages 5561–5569, 2017. 2