Joint COCO and Mapillary Workshop at ICCV 2019: COCO Panoptic Segmentation Challenge Track
Technical Report: ZigZagPanoptic

Dingguo Shen Siting Shen Yuanfeng Ji Di Lin
Shenzhen University
Shenzhen University Shenzhen Guangdong PRC
sitingshen1@gmail.com

Abstract

Panoptic segmentation has become a new trend of segmentation, which will be applied for autonomous driving in the future. Therefore, our team is committed to promoting the development of panoptic segmentation and has made the following contributions: First, integrating advanced network module; Second, improving the network module corresponding to the scene; Third, We got the 50.2 PQ performance on COCO test-dev dataset finally.

1. Overview

We implement the panoptic segmentation system based on the end-to-end training the panoptic segmentation model in mmdetection. We use ResNeXt-101-DCN[3][5] as the backbone network. Our model consists of instance segmentation branch and semantic segmentation branch, which are combined by UPSNet[4] to generate the final panoptic segmentation result. For better instance segmentation and semantic segmentation, we improve ZigZagNet[2] to enhance multi-scale feature extraction in FPN by bidirectional information flow. Furthermore, we implement a ZigZag-fasion interaction among bbox, mask and semantic heads to improve the panoptic segmentation performance. Our single panoptic segmentation model achieves 50.2 PQ on test-dev. The whole model overview is illustrated in Figure 1.

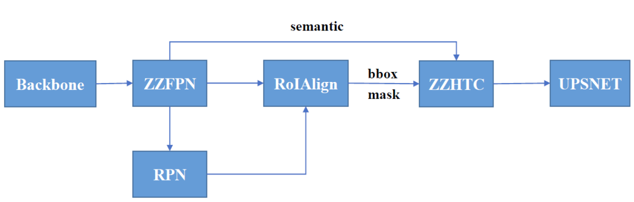

Figure 1: ZigZagPanoptic architecture

2. Dataset

In this competition, we strictly use the images and annotations given by the competition without extended data. We experiment with the about 120k images on the training set from train2017, 5k images on the validation set from val2017 and their corresponding annotations such as instance_train2017, instance_val2017, panoptic_train2017, and panoptic_val2017. The test data is the test-dev2017 and test2017 datasets.

3. Competition Details

This section contains network model adopted in this competition. The network model is mainly divided into the following parts: Backbone Architecture, ZZFPN, ZZHTC and UPSNet.

3.1. Backbone Architecture

Because of our poor hardware, we can't afford to use SE-X152 as the backbone of our network for the best performance. Meanwhile, we think about that DCN can get the better features with spatial deformation, so we adopted X101-DCN[1][2] as our final backbone.

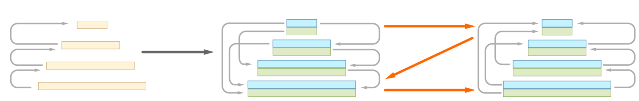

Figure 2: ZigZagFPN Module

3.2. ZigZag Feature Pyramid Network

For better instance segmentation and semantic segmentation, ZigZagNet[3] is adopted to enhance multi-scale feature extraction in FPN by bidirectional information flow. RCE module in ZZFPN also provides regional hybrid information at the same time. In the end of ZZFPN, the outputs of the three iterations in ZZFPN corresponding to the input of three iterations in ZZHTC which provide favorable and diversified evidence for each task branch illustrated as figure 2.

3.3. ZigZag Hybrid Task Cascade

In the original HTC[1], the author fused the feature information of semantic segmentation into the mask and bbox branches. The mask branches and bbox branches of different stages only received the information of their previous stage respectively, and there was no intersection between the feature information of mask and bbox except the acquired box coordinates. Thanks to ZigZag thought, we implement ZigZag-fasion to build feature fusion at each stage of HTC illustrated as figure 3.

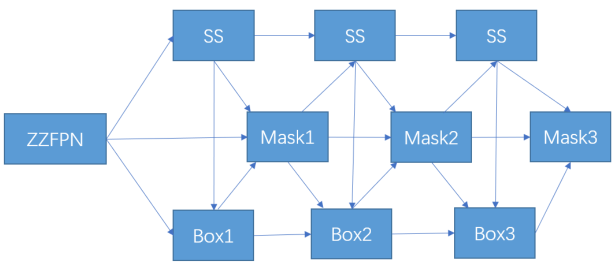

Figure 3: ZZHTC Module

3.4. Unified Panoptic Segmentation Network

In order to obtain panoptic segmentation results by end2end model, we combined semantic segmentation task and instance segmentation task with UPSNet instead of simple heuristic segmentation illustrated as figure 4.

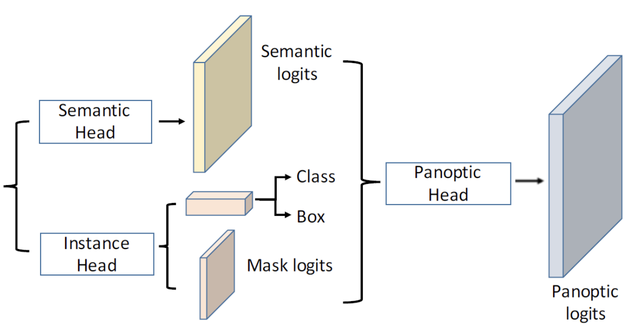

Figure 4: UPSNet Module

4. Ablation Experiments

In this section, we give extensive experiments to verify the efficacy of our main method illustrated as figure 5. First, we build the panoptic model with simple heuristic segmentation, which gets 38.6 PQ on the validatation set. When we implement the UPSNet instead of simple heuristic segmentation, we get 42.1 PQ on the validatation set. Furthermore, we get respectively 43.0 PQ and 44.5 PQ on the validatation set with ZZFPN and ZZHTC. Finally, thanks for the multi scale and large backbone, the model achieve 49.7 PQ on the validatation.

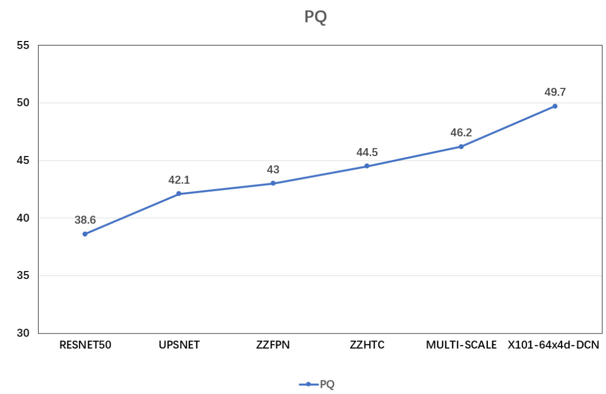

Figure 5: Ablation Experiments

References

[1] 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. 2

[2] Di Lin, Dingguo Shen, Siting Shen, Yuanfeng Ji, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Zigzagnet: Fusing top-down and bottom-up context for object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7490–7499, 2019. 1

[3] 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

[4] Yuwen Xiong, Renjie Liao, Hengshuang Zhao, Rui Hu, Min Bai, Ersin Yumer, and Raquel Urtasun. Upsnet: A unified panoptic segmentation network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8818–8826, 2019. 1

[5] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9308–9316, 2019. 1