VIPriors Instance Segmentation Challenge: Technical Report

Abstract. This report introduces the approaches of team RalleyTheValley used in the of VIPriors Instance Segmentation challenge. We combine SCNet[1], DCN[2] and Seesaw Loss[3] in the submission and achieves 18.5 mask AP on test set.

Zhenhong Chen Ximin Zheng

chenzhnhn@outlook.com yeyuweimeng@126.com

1. Introduction

The performance of deep neural network is more often limited by the lack of data, with the advanced model architectures easier to access than ever. Obtaining high performance by using a small amount of dataset efficiently is a challenging task and can benefit the projects in real-world scenario where the large-scale training data is often unavailable. The Deep Sport dataset of VIPriors Instance Segmentation Challenge is one of this kind. It provides 184, 62 and 64 images for training, validation and testing respectively. The participants are to segment the human and basketball from the images token from basketball courts. We have tried out serval architectures and data augmentation techniques in our work and describe them in the next section.

2. Our Approach

SCNet. Sample Consistency Network is a cascade architecture that matches the distribution between training and inference. To train the semantic branch, We derived the segmentation masks from the existing instance segmentation annotations without supplementing other annotations.

Seesaw Loss. We adopted Seesaw Loss as the classification loss for the instances of human heavily outnumbers samples of basketball.

DCN. We use Deformable Convolution in the stage 1, 2 and 3 of the backbone.

Data Augmentation. We utilized Instaboost[4] in addition to the standard data augmentation method. Instaboost copy and paste the instance of an image according to the similarity of semantic information.

3. Experimental Results

We kept the hyper-parameters unchanged as the default mmdetection[5] 20e training schedule, except the batch size is 2. In our final submission, we trained models with the backbone ResNet-50-FPN, with both training and validation set. The submission result where we combine SCNet, DCN and Seesaw Loss is shown in the Table 1.

Table 1. Experiment Result on test set of our final submission

AP @ 0.50:0.95	AP @ 0.50	AP @ 0.75	AP @ 0.50:0.95 (small)	AP @ 0.50:0.95 (medium)	AP @ 0.50:0.95 (large)
0.185	0.371	0.170	0.000	0.372	0.376

4. Conclusion and Discussion

As the description of our submission goes, we are the rookie in the task of instance segmentation and the results we achieve is far behind the top ones. We planned to go for the semantic segmentation challenge at first, but it turns out that the semantic segmentation is replaced by instance segmentation this year. The inefficient transition we made to cope with the change accounts for the main reason of the unsatisfactory results. We were unfamiliar with the codebase and annotation format of instance segmentation and the amount of time we spent on them is discouraging. Apart from this, we are aware that we did not work enough on finding the better training strategy. We are looking forward to learning from the top submissions and will move forward with the lessons that we have taken.

Reference

[1] Vu, T. , Kang, H. , & Chang, D. Y. . (2020). Scnet: training inference sample consistency for instance segmentation.

[2] Dai, J. , Qi, H. , Xiong, Y. , Li, Y. , Zhang, G. , & Hu, H. , et al. (2017). Deformable Convolutional Networks. IEEE. IEEE.
[3] Wang, J. , Zhang, W. , Zang, Y. , Cao, Y. , & Lin, D. . (2020). Seesaw loss for long-tailed instance segmentation.
[4] Fang, H. S. , Sun, J. , Wang, R. , Gou, M. , Li, Y. L. , & Lu, C. . (2019). InstaBoost: Boosting Instance Segmentation via Probability Map Guided Copy-Pasting. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE.
[5] Chen, K. , Wang, J. , Pang, J. , Cao, Y. , Xiong, Y. , & Li, X. , et al. (2019). Mmdetection: open mmlab detection toolbox and benchmark.