SFPN: Semantic Feature Pyramid Network for Object Detection



Yi Gan, Wei Xu, Jianbo Su 627780002ganyi@sjtu.edu.cn, xuweivg@sjtu.edu.cn, jbsu@sjtu.edu.cn

Problem

Feature Pyramid Network(FPN) employs a topdown path to enhance low level feature by utilizing high level feature. However, the information imbalance issue existed between high level feature and low level feature still remain unsolved.



Model Architecture and Detection Result

Semantic segmentation prediction is generated under large receptive field which already separates different categories and still remains extensive localization details of the whole image. Therefore, it composes of both semantic information and spatial information. These properties make it a perfect hint for object detection.

The overall structure of SFPN is illustrated in Figure below



Motivation

we argue that semantic segmentation can become an effective and sufficient element to compensate for both high and low level features. Because the goal of semantic segmentation is to classify each pixel in the input image, it consists of both high level semantic information and low level localization details. Its classification results can largely prevent the low level features suffering from misclassification problem due to lack of semantic information. Moreover, as the segmentation result is pixel-level, fine edge and shape details of the whole picture will be highly kept. Thus it can effectively alleviate the inaccurate regression problem of high level features, which is caused by lack of sufficient localization information.



Following the backbone and FPN, multi-scale features are sent to both Semantic Pyramid Module(SPM) and Semantic Feature Fusion Module(SFFM). The semantic segmentation prediction of whole input image will be generated in SPM at first. Then the predictions of the foreground instance will be selected out as the basic element to construct semantic pyramid. SPM is responsible for semantic segmentation prediction and semantic pyramid construction.



Obviously, the information contained in semantic segmentation is exactly the missing piece of FPN. And by merging semantic segmentation into FPN, the information imbalance problem can be effectively solved and thus further improvement can be made. Furthermore, the scale of the objects will not affect the segmentation result. So when the semantic segmentation result is merged into FPN, dilution issue will not occur during the feature fusion process.



In the SFFM, semantic pyramid will merge with FPN to construct semantic feature pyramid for detection.



The final detection result of Faster with SFPN is the figure below. And according to the result, it is obvious to see that there are less misclassifying issue on small objects and inaccurate localization issue on large objects.



Acknowledgements

This work was partially financially supported by the National Natural Science Foundation of China under grants 61533012, 91748120 and 52041502. The supports are gratefully acknowledged.