Summary

In this contest, considering that the semantic information of the small-sized objects will be weakened or even disappear in the deeper layers of neural network, we design a novel detection framework called Multibranched Feature Pyramid Network (MFPN) to enhance the information extraction ability of neural network, which is based on Feature Pyramid Network. In addition, due to the large size of remote sensing images, image cutting is a common process for saving the computational cost in training. However, a roughly cutting images into small pieces will undermine the integrity of some objects and make the network hard to recognize. Thus, to solve the problem, an adaptive resize rate for dividing the images into smaller sub-images is designed to keep the large object perfectly after cutting down. Finally, we propose a novel method to get the adaptive aspect ratio of Anchors and further improve the detection performance.

Method 1-- Design Multibranched Feature Pyramid Network (MFPN)

The designed Multibranched Feature Pyramid Network (MFPN) is based on Feature Pyramid Network (FPN). The construction of the FPN involves a bottom-up pathway, a top-down pathway, and lateral connections. In this paper, we make two improvements to the FPN.

First, for the top-down pathway and lateral connections, the proposed MFPN employs a low complexity Multi-branched dilated bottleneck (MBDB) module(Figure 1) to improve semantic information, which is added at the lateral connection of feature pyramid networks to achieve the feature maps with more details at all scales. The details of MBDB module is shown in the Figure 1. It can be seen that the MBDB module combines multiple branches with different dilated convolution layers. We have experimented with more dilated convolution layers and observed marginally better results. Thus, in order to achieve approximate optimal effect without introducing too many parameters, we choose to introduce the MBDB module. At each level of the feature pyramids, MFPN not only maintains high spatial resolution feature maps but also keeps large receptive field due to the added MBDB module, thus it has stronger semantic information extraction capability.

Second, for the bottom-up pathway, we adopt ResNeXt-101 as our backbone, which is widely used as the backbone network in a lot of object detectors due to its superior performance in many fields of computer vision. There are often many layers of the backbone producing output maps of the same size and we say these layers are in the same network stage. In the ResNeXt-101, it has five stages, which are denoted by stage1, stage2, stage3, stage4 and stage5. We only use the stage1, stage2, stage3, stage4 of the ResNeXt-101 as our backbone and keep these stages as the same as original form. The reasons that we don’t use the stage5 for the MFPN are as follows. On the one hand, traditional backbone networks use large down-sampling factor to bring large valid receptive field, which is good for image classification but compromises the object location ability. Thus the stage5 with a scaling step of 32 is of little use in pinpointing larger objects and adding semantic information of the smaller objects which may have disappeared. On the other hand, with the proposed MBDB module, we have enlarged the receptive field to get richer semantic information based on stage4.

Method 2-- Design the adaptive a resize rate

Some remote sensing images are too large for training the convolutional neural network. Thus, resizing the large image to a smaller size image is a common process for save computational and memory cost in training. However, resizing process may lead more small objects missed in the deeper layers. To solve this problem, the general solution is to simply cut large images into small chunks, but if the cut images including relatively large size objects, such as ground track field, the object will be broken up into small pieces(can be called sub-objects) which are more likely to become difficult samples. When a certain category contains many difficult samples in the training dataset, our neural network model is difficult to learn the characteristics of this category and accurately identify it. Therefore, to solve this problem, we propose an adaptive resize rate of the original image for dividing the images into smaller sub-images to protect the integrity of large objects in the remote sensing images. Using the proposed adaptive resize rate to resize the original image before dividing the images into smaller sub-images, can ensure most of the sub-objects are simple samples and improve the recognition ability of deep neural network.

Method 3-- Design the adaptive aspect ratio of Anchors

The aspect ratio of Anchors (the initialization of candidate boxes of Faster R-CNN) for object detection is generally set artificially and empirically to several initial values. Unlike natural images, some objects in remote sensing images are also of very different shapes and with large aspect ratios, such as bridges and harbor. Improper prior aspect ratio setting generally affects the accuracy of positioning in detection. Therefore, it may not be appropriate to directly use the prior aspect ratio of natural images for remote sensing image detection. In this contest, to solve this problem, we propose a novel method to get the adaptive aspect ratio of Anchors. We analyze the training data through a special clustering method, and obtain the appropriate aspect ratio of Anchors.

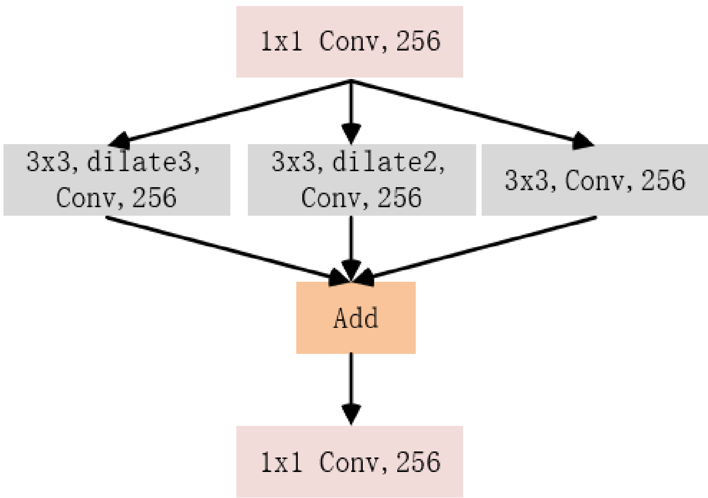

Figure 1: Multi-branched dilated bottleneck (MBDB) module