Algorithm technology solution

Yang Lihe

Nanjing

University lihe.yang.cs@gmail.com

1. Name of participating team

Leo

2. Competition items

Change detection

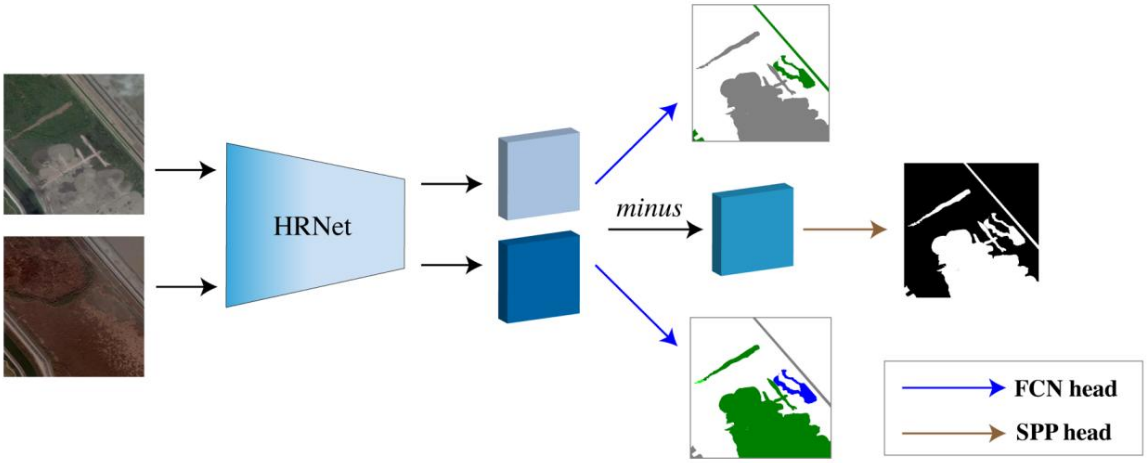
3. Algorithm description

This problem can be regarded as a multi-task semantic segmentation problem. There are two tasks, namely: 1) binary classification segmentation of whether there is a change 2) multi-category semantic segmentation of changed areas

3.1 Network structure

In this task, the performance of a series of backbones such as ResNet[2], ResNeXt[5], ResNeSt[6], and HRNet[3] was compared. Finally, it was found that with similar parameter amounts, HRNet performed better than ResNet, ResNeXt, ResNeSt, so the backbone used in all network structures mentioned later is the HRNet series.

According to the above analysis, in order to solve two different segmentation tasks, a structure is adopted that shares the backbone and uses two segmentation heads to handle their respective segmentation tasks. The rough outline is as follows:



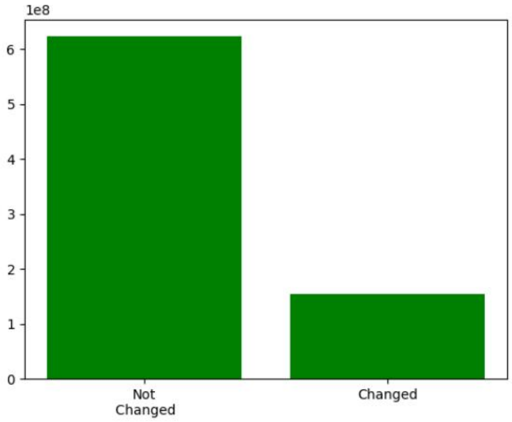
1) For two images in the same area but at different times, input the backbone of shared parameters, and then use the segmentation head of shared parameters for semantic segmentation. Specifically, the classic FCN head was used in the experiment, and other

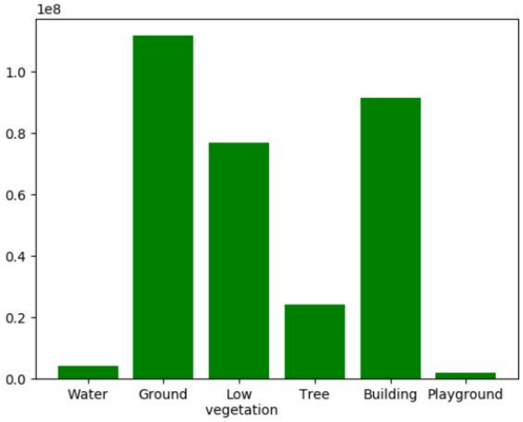
Some heads, such as Spatial Pyramid Pooling head (PSPNet [7]), ASPP head (DeepLabv3 [1]), but the effect is not as good as FCN head. For areas that have not changed, since there is no specific category in the annotation, the calculation of loss is ignored in these areas.

2) In order to perform change detection, the features extracted from the above two images are differenced and the absolute values are taken to obtain feature maps of the same size. On this feature map, the Spatial Pyramid Pooling (SPP) head is used for the final two-class segmentation.

3.2 The difficulty of segmentation of different categories is inconsistent

The above scheme has built the most basic network structure, but in the experiment it was found that the segmentation effect is not very ideal, the difficulty of segmentation of different categories is quite different, and the training data has a serious category imbalance problem. There are two aspects of category imbalance problem. First, there is a large difference in the area between the unchanged area and the changed area, while the area of the changed area is smaller. Secondly, there is a large difference in the area of various landform types, such as "water body", "tree", The three categories of "Sports Ground" have smaller areas in the annotation. The schematic diagram is as follows:





In order to solve the above-mentioned problems of large differences in segmentation difficulty between different categories and the problem of category imbalance, a relatively simple method of weighting each category of cross-entropy loss is adopted. For categories with poor segmentation accuracy, the weight is set to 2 (In the semantic segmentation branch it is "water body", "low vegetation", "tree", in the two-category change detection it is "changed area"), and the weights of other categories are set to 1.

Through this improvement, there can be an improvement of about 1% on the public test set.

3.3 Pseudo labels alleviate the problem of small data volume

After alleviating the problems of inconsistent segmentation difficulty and unbalanced number of categories, the experiment found that the visualization results of segmentation are still not ideal, because a variety of backbones (ResNet, ResNeXt, ResNeSt, HRNet) and some commonly used in semantic segmentation have been tried. multi-scale[7], attention[4] mechanism, and the cutting-edge work in change detection has not made any significant modifications to the network structure, so we hope to continue to improve the performance of the model from the perspective of data.

Considering that the amount of data in this dataset is very small, and more seriously, the unchanged areas in a set of images do not have specific semantic category annotations, and the pixels with semantic category annotations only account for a small part of each image. Therefore, we try to use pseudo-labels to predict semantic category labels for unchanged areas, thereby increasing the number of labeled pixels and alleviating the problem of small data volume.

Specifically, a change detection model is first trained according to the given groundtruth annotation. This model can perform two tasks:

The first task is change detection and the second is semantic segmentation. For unchanged areas marked in groundtruth, use semantic segmentation branch prediction

Specific semantic categories, in other areas, use the specific categories already given by groundtruth. Combined with pseudo-labels of unchanged regions and

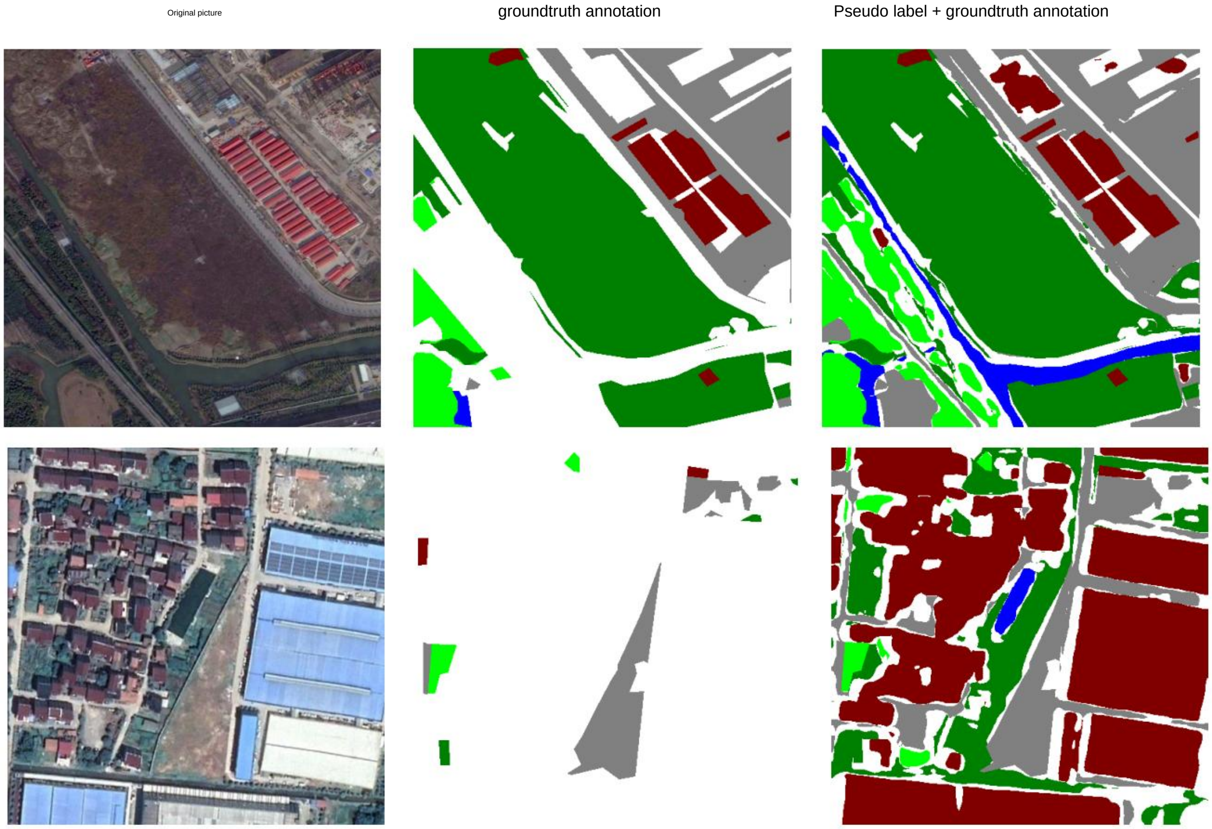
With the true labels of the changed regions, a segmentation model can be retrained, which is ultimately used for testing.

In order to improve the quality of pseudo labels, the voting method of multiple larger models (HRNet_w48, HRNet_w44, HRNet_w40, etc.) is adopted.

Mode. It should be noted that these large models will only be used in the process of pseudo-labeling during training, and these models are not needed during testing.

Through this improvement, there can be an improvement of about 2% on the public test set.

The visualization results of some pseudo-labels are as follows:



Original picture

groundtruth annotation

Pseudo label + groundtruth annotation

3.4 Test-time augmentation (TTA) and model integration

TTA is a prediction result that combines six transformations, namely rotation of 0, 90, 180, and 270 degrees, and horizontal and vertical flipping. Integrated in the model

In terms of performance, considering the limitations of testing time and model size, two backbones were finally integrated, namely HRNet_w40 and HRNet_w18.

model.

4. Experimental environment

1) No external data sets were used, only the pre-trained model on ImageNet provided by HRNet official

2) Hardware resources: 4 Tesla V100 GPUs (experiments were also conducted on 4 2080Tis, their video memory can complete the entire training, and the The accuracy is not affected. V100 is used for faster training). The whole training process takes about 4 to 5 hours. The training framework is PyTorch.

3) Docker run command: sudo docker run --gpus all -it --rm -v /local-val-path:/val -v /local-output-path:/output --shm-size 8G name-of-the-image

4) Code open source address: https://github.com/LiheYoung/SenseEarth2020-ChangeDetection

References

[1]. Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[2]. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[3]. Wang J, Sun K, Cheng T, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2020.

[4]. Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.

[5]. Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.

[6]. Zhang H, Wu C, Zhang Z, et al. Resnest: Split-attention networks[J]. arXiv preprint arXiv:2004.08955, 2020.

[7]. Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.