Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

MMUU-Net: A Robust and Effective Network for Farmland Segmentation of Satellite Imagery

Journal of Physics: Conference Series **1651** (2020) 012189 doi:10.1088/1742-6596/1651/1/012189

MMUU-Net: A Robust and Effective Network for Farmland Segmentation of Satellite Imagery

Xumin Gao[1 a *], Long Liu[2] and Huaze Gong[3]

[1]Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan, Hubei, CHXX0138, China

[2]Institute of Geographical Sciences and resources, University of Chinese Academy of Sciences, Beijing, CHXX0008, China

[3]Institute of Geographical Sciences and resources, University of Chinese Academy of Sciences, Beijing, CHXX0008, China

[a]gaoxumin@mcfly.com.cn [*]Corresponding author's e-mail: 15071460998@163.com

Abstract. Aiming at the multi-scale characteristics of satellite imagery, and the adhesion phenomenon in the farmland segmentation results which is caused by the close distance between different farmland blocks, this paper proposes a robust and effective network based on U-Net for farmland segmentation of satellite imagery, which is called MMUU-Net. On the basis of adopting the encoder with higher classification accuracy network, adding ASPP (Atrous Spatial Pyramid Pooling) layer in the middle, and designing the multi-scale feature fusion module in the decoder, so that the multi-scale feature information is fully utilized; in order to better fuse multi-scale information, a more robust loss function is designed; finally, we propose a segmentation strategy of the coarse and refined two-stage to eliminate the adhesion phenomenon. Through the comparative experiments, it is verified that MMUU-Net is better than other segmentation networks, and can be effectively applied to the task of farmland segmentation of satellite imagery.

1. Introduction

Farmland, as the most important place of food sources for human beings, plays an important role in human survival and development. At present, the whole world is in the stage of accelerated urbanization. How to reasonably and efficiently monitor and plan farmland has put forward higher requirements for human capacity. With the development of satellite remote sensing technology, some technologies are used to monitor and manage farmland through satellite imagery, early application technologies [1-5] mainly use the image processing method which combines edge extraction [6] and adaptive thresholding [7] to realize the segmentation of farmland. Because edge extraction and adaptive thresholding are easily disturbed by segmentation threshold, it will lead to the problem of low segmentation accuracy, and the segmentation effect is very poor for the more complex farmland. In recent years, with the developing of deep learning, some segmentation networks including FCN [8], Segnet [9], and U-Net [10], Pspnet [11], Refinenet [12], Deeplab series [13-16] have been effectively applied in the field of image segmentation. Because the U-Net network has simple structure and small amount of parameters, at the same time, it fuses the features of low-level feature map and high-level feature map,which makes it perform well in the field of satellite imagery segmentation. Zhou et al. [17]

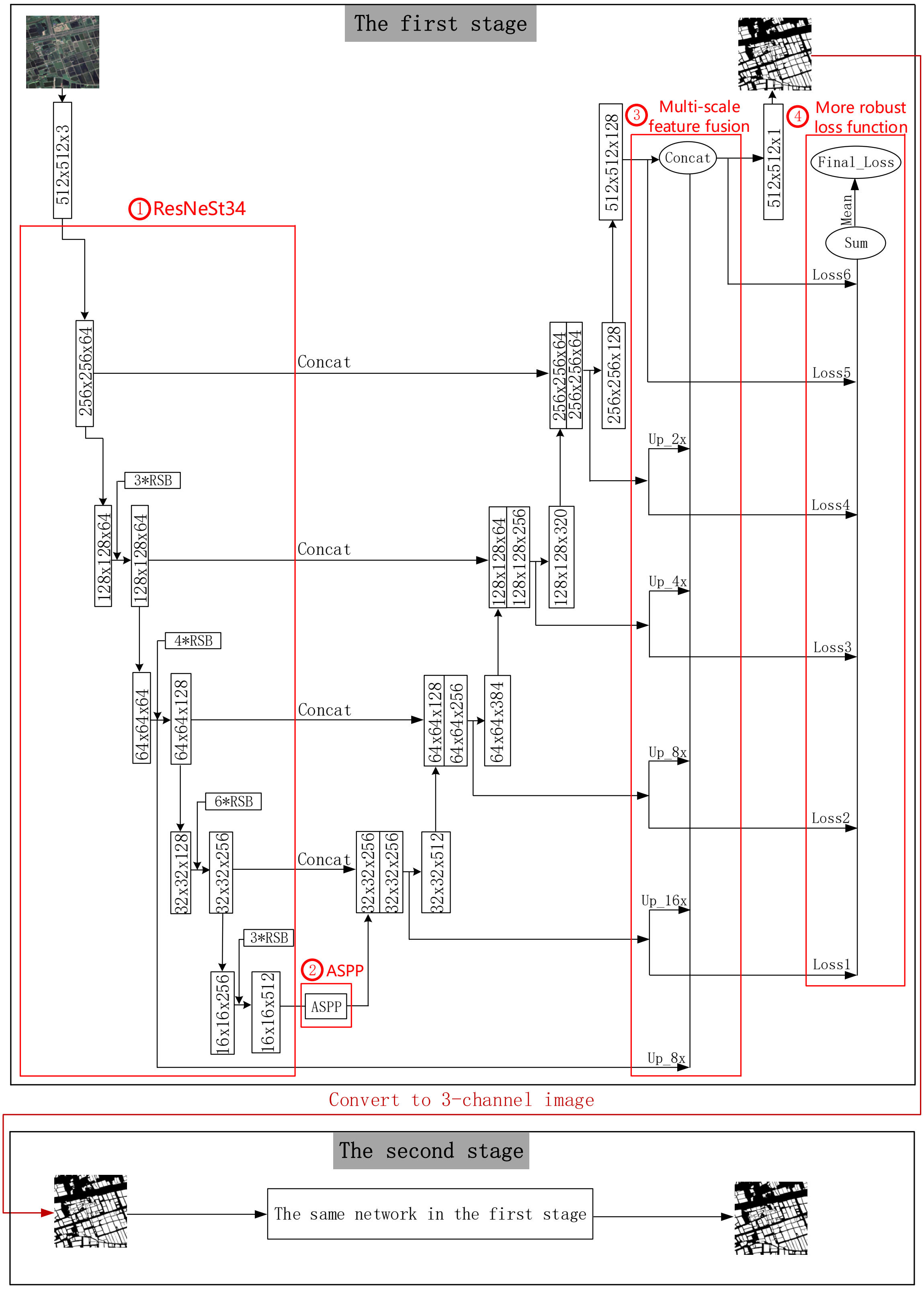

Figure 1. Network architecture of MMUU-Net.

proposed the D-Linknet network based on linknet [18] which is a variant network of U-Net, it was applied to road extraction of satellite imagery, and won the first score in the CVPR DeepGlobe 2018

Road Extraction Challenge; Sun et al. [19] uses the U-Net network to extract the roads of satellite imagery firstly, and then uses the RCF (Rich Convolutional Features for Edge Detection) [20] to extract the individual farmland block. Although U-Net and the variant segmentation network of U-Net which were mentioned in the above article, have made some achievements in the segmentation of satellite imagery, the multi-scale characteristics of satellite imagery are not further considered. At the same time, the characteristic of the close distance between different farmland blocks will lead to adhesion phenomenon in the segmentation result.

Aiming at the shortcomings of the above segmentation networks, this paper proposes a robust and effective network for farmland segmentation of satellite imagery based on U-Net, which is called MMUU-Net. The network architecture of MMUU-Net is shown in figure 1: encoder part uses ResneSt [21] which has higher classification accuracy; ASPP layer is inserted into the middle; the multi-scale features of satellite imagery are further explored from two aspects. Firstly, the feature maps of each scale in decoder and some part of encoder are concatenated as the final feature layer, so as to realize the fusion of multi-scale features. Secondly, we compare the feature map with the downsized GT (Ground Truth) from different scales in decoder to calculate the loss; finally, we average the losses from different scales as the final loss, so as to design a more robust loss function. For the adhesion phenomenon, we use the segmentation strategy of the coarse and refine two-stage to solve it.

2. MMUU-Net

2.1. Network architecture

The network architecture of MMUU-Net proposed in this paper is shown in figure 1. It can be seen from figure 1 that compared with U-Net, the main differences of MMUU-Net are four parts which are marked by serial numbers in figure 1. Next, we will detailly explain these improvements.

2.2. Encoder design

The encoder of MMUU-Net uses ResneSt [21] which has higher classification accuracy. The basic unit of ResneSt is RSB (ResneSt Block). The structure of RSB is shown in figure 2:

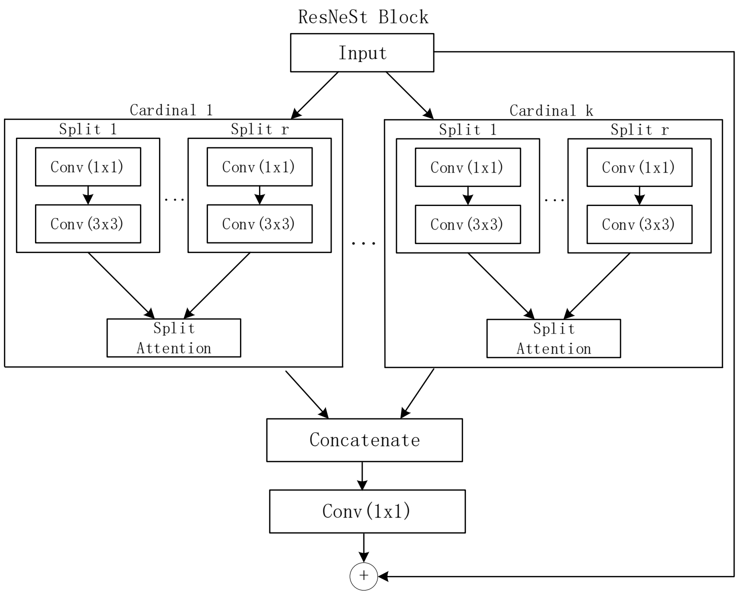

Figure 2. Structure of ResNeSt Block.

It can be seen from figure 2 that the main feature of RSB is that it combines the design of group convolution and the idea of split-attention blocks, which makes the classification performance more superior and surpasses the classical classification networks such as ResNet [22].

2.3. ASPP layer
Inspired by Deeplab [15], we insert the ASSP layer into the middle of MMUU-Net. The design structure of ASPP layer is shown in figure 3:

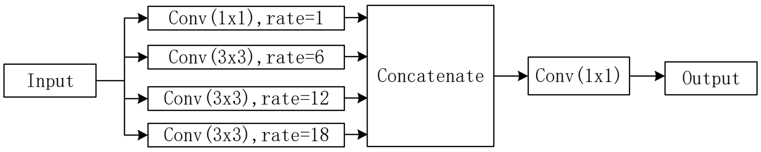

Figure 3. Structure of ASPP.

In the ASSP layer, the feature map extracted from encoder will pass through different dilated convolutions which dilated rates are 1, 6, 12, 18 respectively, then feature fusion will be carried out by concatenate + conv (1x1), which will make the receptive field of network more larger and fuse multi-scale information.

2.4. Multi-scale feature fusion
The feature maps of different scales which are extracted from decoder and some parts of encoder, are resized to the original image size by upsample. Finally, these feature maps are concatenated together as the final output feature map in MMUU-Net, which further realizes the fusion of multi-scale features.

2.5. More robust loss function
In order to fuse the multi-scale feature into the final model better, we design a more robust loss function for MMUU-Net. According to the size of different feature maps in decoder, it carries out the corresponding downsample for ground truth, then calculates the loss respectively at different scales. Finally, the losses calculated on different scale of decoder and the final feature output layer are averaged as the final loss. The specific calculation process is as follows:

$$Final\_Loss = \frac{\sum_{i=1}^{N} Loss_i}{N} \tag{1}$$

It can be seen from figure 1 that N is equal to 6 in equation (1).

For each $Loss_i$, it is composed of two parts: DC Loss (Dice Coefficient Loss) and BCE Loss (Binary Cross Entropy Loss), the calculation method is as in equation (2):

$$Loss = 1\text{-}\sum_{i=1}^{N} DCLoss(P_i, GT_i) + \sum_{i=1}^{N} BCELoss(P_i, GT_i) \tag{2}$$

In equation (2), P denotes predicted image， GT denotes annotation image, N is batch_size.

The calculation of DC Loss and BCE Loss are as in equation (3) and equation (4) respectively:

$$DCLoss(P, GT) = \sum_{i=1}^{W}\sum_{j=1}^{H} \frac{2\left|gt_{ij} \cap p_{ij}\right|}{\left|gt_{ij}\right| + \left|p_{ij}\right|} \tag{3}$$

$$BCELoss(P, GT) = -\sum_{i=1}^{W}\sum_{j=1}^{H}\left[gt_{ij}\cdot\log p_{ij} + (1-gt_{ij})\cdot\log(1-p_{ij})\right] \tag{4}$$

In equation (3) and equation (4), W and H are width and height of the image respectively, gt_{ij} denotes every pixel of GT, p_{ij} denotes every pixel of P.

2.6. The coarse and refined two-stage segmentation
For the task of farmland segmentation of satellite imagery, expect to the multi-scale characteristics, it is another key point to separate each farmland block completely. We found that there are a lot of adhesions between different farmland blocks in the segmentation results through previous experiments. In order to solve this problem, morphological processing is generally used to solve this problem.

Journal of Physics: Conference Series **1651** (2020) 012189 doi:10.1088/1742-6596/1651/1/012189

However, morphological processing is aimed at the whole image, so that it will lead to abnormal changes in the corrected segmentation region, which will affect the overall segmentation accuracy. Therefore, this paper proposes a segmentation strategy combining coarse and refined two-stage to solve adhesion problem: the first stage network is used to roughly segment farmland, then the output of the first stage is transformed into a three channel image and is input to the second stage network to eliminate adhesions, so that it realizes the refined segmentation for farmland.

3. Experiment

3.1. Dataset

In this paper, we use google map loader to randomly sample satellite imagery from typical farming areas with different resolutions and time periods, and 3000 satellite imagery with the size of 1024x1024 are selected as the dataset of farmland segmentation. After labeling the dataset with labelme tool, the dataset is divided into training dataset and testing dataset according to the ratio of 8:2.

3.2. Training

We use the MMUU-Net to train the training dataset. The GPU of the training machine is GXT 1080, the training parameters are set as follows: batch size is two images; we use Adam optimizer, which the basic learning rate is 1e-3 and weight_decay is 5e-4; the maximum number of training iterations is 6000.

3.3. Comparative experiment

In order to verify the effectiveness of MMUU-Net proposed in this paper, we also use U-Net, D-LinkNet [17], and RCF [19] to train the dataset under the same conditions. Then we compare the results of different segmentation network by MIoU (Mean Intersection over Union) and FPS (Frames Per Second) on the testing dataset. The comparison results are shown in table 1:

Table 1. The comparison results of different segmentation network on testing dataset

Method	MIoU(%)	FPS
U-Net	68.39	81
D-LinkNet [17]	76.88	46
RCF [19]	75.35	7
MMUU-Net (The first stage)	78.18	43
MMUU-Net (The second stage)	79.30	43

It can be seen from table 1 that the MIoU of MMUU-Net (The second stage) is the highest which get 79.30%. The FPS of MMUU-Net is 43, which is close to D-LinkNet [17], and both are higher than RCF [19], this segmentation speed has basically met the requirement of most segmentation tasks of satellite imagery. Therefore, considering the segmentation accuracy and speed, the MMUU-Net proposed in this paper is better than other segmentation network.

In order to intuitively reflect the superiority of MMUU-Net in the task of farmland segmentation of satellite imagery, we use different segmentation networks to segment three images from testing dataset, and get the visual comparison results, as shown in figure 4:

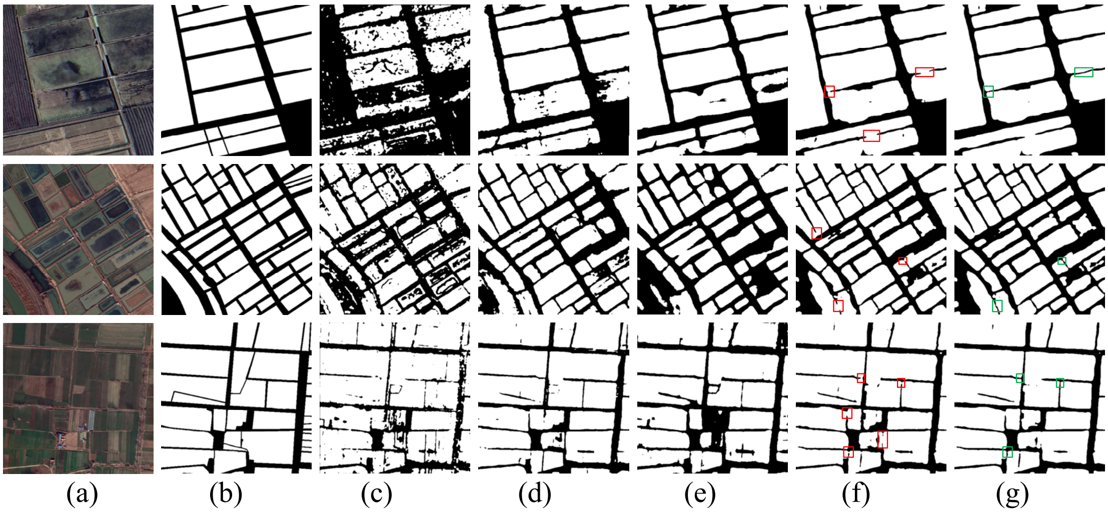

Figure 4. The visual comparison results of different segmention network. From left to right column: (a) Original image, (b) Ground truth, (c) U-Net, (d) D-LinkNet [17], (e) RCF [19], (f) MMUU-Net (The first stage), (g) MMUU-Net (The second stage).

It can be seen from figure 4 that the segmentation result of U-Net is very poor, and many pixels of the images are incorrectly classified. The segmentation result of D-Linknet is not complete in segmentation boundary of some farmland blocks, and there are a small amount of noised points. Because RCF is designed based on the edge detection principle, it leads to be excessively attention to edge features, and the overall segmentation result shows corrosion effect. The segmentation result of MMUU-Net is best, but there are some adhesions between different farmland blocks in MMUU-Net(The first stage), we use red rectangular boxes to mark the specific position of adhesions in figure 4(f). In order to solve this problem, we put the segmentation result of MMUU-Net(The first stage) into MMUU-Net(The second stage) to realize the refined segmentation, we can see that most of adhesions are eliminated in the segmentation result of MMUU-Net(The second stage), we use green rectangular boxes to mark the specific position which adhesions have been removed in figure 4(g). Through the visual comparison results, the adaptability and superiority of MMUU-Net in the task of farmland segmentation of satellite imagery are more intuitively verified.

4. Conclusion

In this paper, a robust and effective network for farmland segmentation of satellite imagery is proposed, we call it MMUU-Net. It uses fully multi-scale feature information, and proposes a segmentation strategy of the coarse and refined two-stage to eliminate the adhesion phenomenon in the preliminary segmentation results. Finally, the effectiveness and superiority of MMUU-Net in the task of farmland segmentation of satellite imagery are verified through comparative experiments.

Acknowledgments

This research was supported by National Natural Science Foundation of China (NO. 51505347) and Natural Science Foundation of Hubei Province of China (NO. 2018CFB626).

References

[1] Mas, J.F., Stéphane Couturier, J.P.G., Skutsch, M. (2016) Comment on gebhardt et al. mad-mex: automatic wall-to-wall land cover monitoring for the mexican redd-mrv program using all landsat data. remote sens, 6: 3923–3943.

[2] Yan, L., Roy, D. P. (2014) Automated crop field extraction from multi-temporal web enabled landsat data. Remote Sensing of Environment, 144(144): 42-64.

[3] Fritz,S., See, L, Mccallum, I. (2015) Mapping global cropland and field size. Global Change Biology, 21(5):32-47.

[4] Waldner, F., Fritz, S., Di Gregorio, A. (2016) A unified cropland layer at 250 m for global agriculture monitoring. 1(1):56-57.
[5] Graesser, J., Ramankutty, N. (2017) Detection of cropland field parcels from landsat imagery. Remote Sensing of Environment, 201: 165-180.
[6] Youfu, W. (2006). A survey of the research on edge detection in image processing. Ship Electronic Engineering. 11: 695-712
[7] Manay, S., Yezzi, A. (2003) Anti-geometric diffusion for adaptive thresholding and fast segmentation. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society. 12(11):1310-23.
[8] Long, J., Shelhamer, E., Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39(4): 640-651.
[9] Badrinarayanan, V., Kendall, A., Cipolla R. (2019) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39(12): 2481-2495.
[10] Ronneberger, O., Fischer, P., Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. 6: 98-107
[11] Zhao, H., Shi, J., Qi, X. (2017) Pyramid scene parsing network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii. pp. 195-209.
[12] Lin, G., Milan, A., Shen, C. (2017) RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation. In: Computer Vision and Pattern Recognition (CVPR). Hawaii. pp. 337-341.
[13] Liangchieh, C., Papandreou, G., Kokkinos, I., Murphy, K. (2014) Semantic image segmentation with deep convolutional nets and fully connected crfs. Computer ence. 4: 357-361.
[14] Chen, L.C., Papandreou, G., Kokkinos, I. (2018) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 834-848.
[15] Chen, L.C., Papandreou, G., Schroff, F. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.
[16] Chen, L.-C. Zhu, Y. Papandreou, G. (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: In Proceedings of the European conference on computer vision (ECCV). Munich. pp. 801-818.
[17] Zhou, L., Zhang, C., Wu, M. (2018) D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Saline Lake. pp. 755-763.
[18] Chaurasia, A., Culurciello, E., (2017) LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation[J]. arXiv preprint arXiv:1707.03718.
[19] Sun, Y., Luo, J., Xia, L., Wu, T., Gao, L. (2019) Geo-parcel-based crop classification in very-high-resolution images via hierarchical perception. International Journal of Remote Sensing. 9: 63-72
[20] Liu, Y., Cheng, M.M., Hu, X.W. (2019) Richer Convolutional Features for Edge Detection. IEEE transactions on pattern analysis and machine intelligence. 41:8-17.
[21] Zhang, H., Wu, C., Zhang, Z. (2020) ResneSt: Split-Attention networks. arXiv preprint arXiv:2004.08955
[22] He, K., Zhang, X., Ren, S., Sun, J. (2015) Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition. Las Vegas. pp. 770–778.