Research on Unbalanced Sample Segmentation of Remote Sensing Image

Yanping Li¹, Lu Wang¹*, Lei Zhang², Shiyong Wang¹, Haiwen Chen¹ and Xiaolan Zhu¹

¹College of Computer Science, Qinghai University, Xining, Qinghai, China
²College of Computer Science, Sichuan University, Chengdu, Sichuan, China
Email: 465232991@qq.com

Abstract. This article proposes a framework for the unbalanced sample segmentation of remote sensing images on an open data set. For the few sample categories, improved by converting it into binary segmentation and combining with weighted cross-entropy loss, and then merged with the segmentation result of the sufficient sample categories. Finally, the mIoU (mean Intersection of Union) of the 8 categories is increased by 4.5% compared to the results before the improvement, especially, the results of the few sample category road and grassland are increased by 10.2% and 9.7%. The experiments show that the framework can greatly improve the segmentation performance of the few sample categories, and have a good guiding significance for the problem of multiclass unbalanced sample segmentation.

Keywords. Unbalanced sample segmentation; binary segmentation; loss compensation; model fusion.

1. Introduction
Remote sensing image is widely used in land use investigation [1-3], vegetation classification [4-7], natural disaster warning [8], agriculture census [9-10] and other fields. As the basis of intelligent interpretation of remote sensing images, ground object segmentation has great research and application value in related industries such as remote sensing and artificial intelligence. The remote sensing image usually shows the characteristics of low resolution and small size due to the distance, atmosphere, which makes the performance of traditional image segmentation methods on remote sensing images greatly reduced. In addition, when the research scope is limited to a certain area on the surface, depending on the actual surface conditions, it often shows the characteristics of the unbalanced sample. While the current usage of AI methods needs to rely on a large number of samples for training, the lack of samples makes the interpretation of the few sample object more difficult, resulting in the much lower segmentation accuracy of the few sample categories than that of the sufficient sample categories, and it is also easy overfitting. Therefore, unbalanced sample segmentation, especially few sample segmentation, is a very important research direction in the intelligent interpretation of remote sensing images, and it has a wide range of significance.

2. Related Work

2.1. Unbalanced Learning
Unbalanced learning refers to the large gap in the number of samples in different categories, which is a relatively common phenomenon in real research. In recent years, with the development of unbalanced
learning research, some methods to solve the problem have been proposed from the perspective of data, classification method, and evaluation index.

The data-based method resamples the entire data to adjust the sample ratio of each category in the data set, by oversampling increases the number of samples of the few sample categories and undersampling reduces the number of samples of the sufficient sample categories. The popular oversampling methods include random oversampling [11], synthetic minority oversampling [12] and adaptive synthetic oversampling [13] and so on. The purpose of under-sampling is to delete some samples according to certain rules. The popular under-sampling methods based on different deleting rules include random under-sampling, nearest neighbour compression [14], nearest neighbour elimination [15] and so on. In the resampled data, the gap between the number of the few sample categories and the sufficient sample categories is narrowed, thereby alleviating the imbalance from the source of the data.

Methods based on classification algorithms focus on improving the performance of the few sample categories, such as ensemble learning and few-shot learning. Ensemble learning is suitable when a single classifier cannot solve the specific classification problem by combining multiple classifiers with better performance than that of any single classifier. The integrated classifiers can be the same or different. Few-shot learning studies machine learning problems on limited data, and can strengthen the learning of the few sample category in unbalanced learning. The mainstream method is meta-learning, studying the problem of how to learn to learn.

The method based on the evaluation index no longer uses the average accuracy for both the sufficient sample categories and the few sample categories instead of measurement designed to consider the imbalance and compensate for the few sample categories, such as F-measure, G-mean.

2.2. Semantic Segmentation

Semantic segmentation is done by dividing each pixel in the image into different categories. The classic semantic segmentation network FCN [16] (Fully Convolutional Networks) is the pioneering work of image semantic segmentation, since then, mainstream semantic segmentation networks have been mainly constructed based on deep convolutional neural networks. This article only introduce some of the deep learning networks.

The OCRNet proposed by Microsoft Asia Research Institute is based on the context information of the object region and only uses the information of the category of the object to convert the pixel classification problem into the object region classification problem, puts the similarity between the object region feature representation and the pixel feature representation together as input, has won the first place in the 2020 ECCV Cityscapes semantic segmentation task [17]. PSPNet obtains context information through the pyramid pooling module and designs a deep-supervised loss function for the backbone network ResNet to assist in training, reaches 85.4% mIoU on PASCAL VOC2012 [18]. DANET uses a dual attention network to capture global feature dependencies and significantly improves the segmentation results by modeling rich context dependencies on local features, reaches 82.6% mIoU on PASCAL VOC2012 [19]. DeepLab V3 is the third version of the DeepLab series, it uses hole convolution to obtain a larger receptive field and improves ASPP by cascading or paralleling hole convolutions and batchnorm layers of different sampling rates, and reaches 86.9% mIoU on Pascal VOC2012 [20]. DeepLab V3+ designs an encoder-decoder structure on the basis of DeepLab V3 by using DeepLab V3 as the encoder, Xception as the backbone network and adding a new decoder, reaches 89.0% mIoU on PASCAL VOC2012 [21].

3. Unbalanced Sample Segmentation Framework

The unbalanced sample segmentation framework proposed in this study is shown in figure 1. By converting the few samples categories into two binary segmentation tasks after data augmentation and then compensating the few samples categories by assigning higher weight in the loss function. The segmentation result of the sufficient sample categories are obtained by fusion of multiple models, and
the results of the binary segmentation of the few sample categories and the sufficient sample categories are merged again to get the final result.

![Unbalanced sample segmentation framework.](image)

**Figure 1.** Unbalanced sample segmentation framework.

First, data augmentation is performed on the few sample category of road and grassland. We perform four geometric transformations by rotating 90 degrees, rotating 270 degrees, flipping up and down and flipping left and right, then perform four colour transformations by adjusting brightness, contrast, saturation and adding Gaussian noise, and perform one of the above four geometric transformations and then one of the above four-color transformations at last. Then take the data set with the best result, convert the multiclass task into two binary segmentation tasks of grassland/non-grassland and road/non-road by using all other categories as background. In the binary segmentation, the few sample category of grassland and road are compensated by weighted cross-entropy loss to make the whole training inclining to the few sample category. The segmentation results of the sufficient sample categories are obtained by voting by multiple models and then corrected with the binary segmentation results.

4. **Experiment**

The research data set comes from the CCF Big Data and Computational Intelligence Competition, contains about 145,000 images with categories of building, farmland, forest, water, road, grassland, other and unmarked area. In the experiments, the data set is divided into training set of 120,000 images, while verification set and test set about 13,000 images. The segmentation result of each category is measured by mIoU. The performance of the benchmark model is shown in table 1.

| Network    | Building | Farmland | Forest | Water | Road | Grassland | Other | Unmarked | mIoU |
|------------|----------|----------|--------|-------|------|-----------|-------|----------|------|
| OCRNet     | 77.7     | 88.9     | 85.2   | 91.1  | 35.2 | 22.5      | 67.2  | 98.7     | 70.8 |
| PSPNet     | 76.9     | 88.5     | 84.5   | 90.5  | 38.3 | 21.4      | 65.6  | 98.0     | 70.5 |
| DANet      | 77.9     | 88.9     | 85.0   | 91.1  | 31.6 | 24.7      | 68.2  | 98.6     | 70.7 |
| DeeplabV3+| 73.5     | 87.4     | 83.7   | 90.1  | 24.9 | 16.0      | 63.3  | 97.6     | 67.1 |
| DeeplabV3 | 72.9     | 84.9     | 82.6   | 88.5  | 17.0 | 17.4      | 60.0  | 6.8      | 53.8 |
The main difference in the results of each network is the category road and grassland. According to statistics, these two categories only account for 0.4% and 2% of the overall sample pixels. Therefore, the key to improving the segmentation performance lies in the few sample categories - road and grassland.

4.1. Augmentation on the Few Sample Categories
The data set was expanded to about 230,000 through augmentation. By evaluating the result of different data augmentation methods on the segmentation accuracy, it is found that only geometric transformation has the best segmentation performance. The performance after geometric transformation on the test set is shown in Table 2.

| Network  | Building | Farmland | Forest | Water | Road | Grassland | Other | Unmarked | mIoU |
|----------|----------|----------|--------|-------|------|-----------|-------|----------|------|
| pspnet   | 78.7     | 89.8     | 85.9   | 91.6  | 31.0 | 33.8      | 73.2  | 98.5     | 72.8 |
| ocrnet   | 78.9     | 89.6     | 85.9   | 92.1  | 42.3 | 29.4      | 69.8  | 99.5     | 73.4 |
| danet    | 77.8     | 89.6     | 85.7   | 90.4  | 36.7 | 29.0      | 70.7  | 98.2     | 72.3 |
| Deeplab v3 | 72.0   | 87.1     | 83.6   | 85.4  | 28.8 | 25.4      | 64.6  | 4.6      | 56.4 |
| Deeplab v3+ | 75.6  | 88.4     | 84.5   | 90.9  | 33.0 | 25.3      | 65.4  | 99.0     | 70.3 |

After data augmentation, the highest accuracy for road and grassland reaches 42.3% and 33.8%. Comparing the performance of all the networks in the few sample categories, PSPNet has a stronger learning ability on category grassland, and OCRNet has a stronger learning ability on category road.

4.2. Converting into Binary Segmentation Task
Convert the category road and grassland into two binary segmentation tasks on the geometrically transformed data set with OCRNet and PSPNet and choose cross-entropy as loss function. The performance is shown in Table 3.

| Network  | Road | Non-road | Grassland | Non-grassland |
|----------|------|----------|-----------|---------------|
| PSPNet   | 32.4 | 99.7     | 31.2      | 98.3          |
| OCRNet   | 48.1 | 99.8     | 21.9      | 97.3          |

After binary segmentation, the segmentation accuracy of OCRNet of the category road is increased to 48.1% and the segmentation accuracy of PSPNet of the category grass is slightly reduced to 31.2%.

4.3. Compensation for the Few Sample Categories
The weighted cross-entropy is used as the loss function to compensate the category grassland and road in binary segmentation. The optimal weight is determined through experiments to be 0.7/0.3. Under the optimal weight, the performance of PSPNet and OCRNet is shown in Table 4.

| Network  | Road | Non-road | Grassland | Non-grassland |
|----------|------|----------|-----------|---------------|
| PSPNet   | 33.0 | 99.7     | 34.1      | 98.3          |
| OCRNet   | 49.6 | 99.7     | 25.0      | 97.9          |
After using weighted cross-entropy, the segmentation performance of the few sample category is improved on both PSPNet and OCRNet. Compared with no weight, the mIoU of the category road and grassland is increased by 1.5% and 2.9%.

4.4. Model Fusion
Firstly, the trained PSPNet, DeepLab V3, DeepLab V3+, DANet and OCRNet models are integrated through voting, with OCRNet adopted as the main model and the other four models as the supplemented model. Only when the prediction of the four models is the same then correct OCRNet’s prediction, the performance after fusion is shown in Table 5.

| Building | Farmland | Forest | Water | Road | Grassland | Other | Unmarked | mIoU |
|---------|----------|--------|-------|------|-----------|-------|----------|------|
| 79.8    | 90.1     | 86.5   | 92.4  | 41.1 | 30.9      | 71.3  | 99.4     | 73.9 |

The result after the fusion is 0.7% higher than OCRNet before the fusion, then fuses with the binary segmentation result. During the fusion, the result of the few sample categories is used to correct the multiple model segmentation result. The performance of the final result is shown in Table 6.

| Building | Farmland | Forest | Water | Road | Grassland | Other | Unmarked | mIoU |
|---------|----------|--------|-------|------|-----------|-------|----------|------|
| 79.9    | 90.0     | 86.5   | 92.4  | 48.5 | 34.4      | 71.2  | 99.4     | 75.3 |

The overall mIoU in the final segmentation result is 1.4% higher than above and the mIoU of category road and grassland reaches 48.5% and 34.4%.

5. Conclusion
Taking an open unbalanced data set as an example, this paper designs a framework consisting of the sufficient sample categories segmentation network and the few sample categories segmentation network. By using different optimization methods for the different networks, the overall mIoU of 8 categories is increased by 4.5% compared with before the improvement, especially the few sample category road and grassland have increased by 10.2% and 9.7%. This result has a good guiding significance for the unbalanced segmentation task. In general, the framework proposed in this article can be further extended to other multiclass segmentation task. However, the optimization measures tried in this article are not comprehensive, and the optimization effects of other measures need to be further explored in subsequent research.

Acknowledgments
This paper is supported by the Youth Natural Science Foundation of Qinghai Provincial Department of Science and Technology (No. 2021-ZJ-952Q).

References
[1] Song Y P 2020 Research on Land Use Classification Technology Based on Multi-Source Image Fusion and Object-Oriented (Shihezi University).
[2] Lei M 2019 Research on Urban Land-Use Change Detection Based on High-Resolution Remote Sensing Images (Central China Normal University).
[3] Li Z 2018 Research on High-resolution Remote Sensing Image Land Use Information Extraction Technology (Xi’an University of Science and Technology).
[4] Huang J 2020 Woodland Type Recognition Based on OBIA-CNN from Gaofen-2 Satellite Image (China University of Geosciences (Beijing)).
[5] Cao W N, Wang W G, Wang X, Yu Y J and Liu S J 2021 Study on crop classification method based on Gaofen-2 satellite data Surveying and Geospatial Information 44 (04) 158-161.

[6] Xu Z Y, Zhou Y, Wang Sh X, Wang L T and Wang Z Q 2021 U-Net urban green space classification for GF-2 remote sensing images Journal of Image and Graphics 26 (03) 700-713.

[7] Wang N, Min J, Guo J L and He N 2021 Research on remote sensing classification of tree species based on multi-scale feature transfer learning Journal of Ezhou University 28 (02) 93-97.

[8] Jia J T 2020 Automatic Landslide Recognition Based on Gaofen-1 Satellite Image (Southwest University of Science and Technology).

[9] Wu J F 2021 Research on the monitoring of planting area based on remote sensing technology Journal of Longyan University 39 (02) 50-54.

[10] Hastuya, Chen Z X, Li C and Liu S X 2021 Object-oriented remote sensing representation method of film-covered farmland information Science of Surveying and Mapping 46 (03) 140-146.

[11] Ling C X and Li C 1998 Data mining for direct marketing: Problems and solutions Knowledge Discovery and Data Mining 73-79.

[12] Chawla N V, Bowyer K W, Hall L O, et al. 2002 SMOTE: Synthetic minority over-sampling technique Journal of Artificial Intelligence Research 16 (1) 321-357.

[13] He H, Yang B, Garcia E A, et al. 2008 ADASYN: Adaptive synthetic sampling approach for imbalanced learning IEEE International Joint Conference on Neural Networks pp 1322-1328.

[14] Wilson D L 2007 Asymptotic properties of nearest neighbour rules using edited data IEEE Transactions on Systems Man and Cybernetics 2 (3) 408-421.

[15] Laurikkala J 2001 Improving identification of difficult small classes by balancing class distribution Proceedings of the 8th Conference on AI in Medicine in Europe: Artificial Intelligence Medicine.

[16] Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 3431-3440.

[17] Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y 2019 Object-contextual representations for semantic segmentation European Conference on Computer Vision pp 2881-2890.

[18] Chen L C, Papandreou G, Schroff F, et al. 2016 Pyramid scene parsing network Preprint arXiv:1612.01105.

[19] Fu J, Liu J, Tian H, et al. 2019 Dual attention network for scene segmentation Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 3146-3154.

[20] Chen L C, Papandreou G, Schroff F, et al. 2017 Rethinking atrous convolution for semantic image segmentation Preprint arXiv/1706.05587.

[21] Chen L C, Zhu Y, Papandreou G, et al. 2018 Encoder-decoder with atrous separable convolution for semantic image segmentation Preprint arXiv/1802.02611.