Integrating Data and Model Optimization for Improved Multi-Class Land Cover Classification Using Multispectral Remote Sensing Data

Yichuan Li[1], Huanying Sun[2], Junchuan Yu[1,3], Minying Xie[1], Dingjian Jin[1], Ming Wang[1], Guobin Xia[4]

[1] China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, China – {lyichuan, yujunchuan, xieminying, jindingjian, wangming} @mail.cgs.gov.cn
[2] PIESAT Information Technology Co., Ltd, China – sweetshy@126.com
[3] Technology Innovation Center for Geohazard Identification and Monitoring with Earth Observation System, Ministry of Natural Resources, China
[4] Hexi gold mine in Zhaoyuan city – dhtlsse@163.com

Keywords: Multispectral data, Classification, OUNet, Worldview3, Unbalance, Data mining.

Abstract

With the rapid development of artificial intelligence, significant progress has been made in land cover classification using deep learning methods. However, in existing research, most studies focus more on improving classification accuracy by optimizing the model structure and less on mining the value of the data itself. In this paper, experiments on remote sensing multi-class land cover classification were conducted based on Worldview3 data, and strategies to improve classification accuracy were proposed in terms of sampling methods, band combination, loss function, and model optimization. Experiment results show that the proposed improvement strategies are effective for multi-class land cover classification, with recall, F1, and IoU improved by 29%, 17%, and 19%, respectively. The significant improvement in classification accuracy for less-represented targets confirms that enhancing data richness and balance leads to greater improvement than just optimizing the model.

1. Introduction

High-resolution remote sensing images can provide rich spatial information, which is of great significance for remote sensing classification, precision agriculture, and natural resource supervision. In recent years, with the rapid development of artificial intelligence, machine learning methods represented by deep learning have made significant progress in the field of remote sensing classification (Zhou et al., 2019). Benefiting from the powerful learning abilities of a convolutional neural network (CNN), efficiency has an advantage over traditional methods in the case of sufficient samples (Yann et al., 2015). Since the first end-to-end semantic segmentation model, FCN, was proposed in 2015, semantic segmentation models have been rapidly developed in recent years. UNet (Olaf et al., 2015), Deeplabv3+, SegFormer, etc. have propelled the field of semantic segmentation into a new era of development.

Leveraging the powerful learning capabilities of CNNs, it is possible to effectively capture contextual information and achieve recognition of land cover features within complex scenes [23]. Despite these successes, challenges remain in applying deep learning to multi-class target identification in remote sensing. First, there are large differences in scale and texture among different land cover types in high-resolution imagery (Deng et al., 2018). Second, unlike natural images, the background in remotely sensed images is more complex and tends to account for a larger proportion, leading to an imbalance between background and foreground information (He et al., 2016). Third, the imbalance of inter-class samples in multi-classification scenarios can lead to a reduction in the overall classification accuracy (Zheng et al., 2020). At present, in the research on multi-class land cover classification based on high-resolution remote sensing data, many researchers try to solve the problems of scale and morphological differences through multi-scale feature fusion and attention mechanisms (Chen et al., 2020). However, there is relatively less emphasis on addressing the issue of imbalance problems. In addition, a trend can be observed where most studies focus more on improving classification accuracy by optimizing the model structure and less on mining the value of the data itself.

In this paper, experiments on remote sensing multi-class land cover classification were conducted based on Worldview3 data, and strategies to improve classification accuracy were proposed in terms of sampling methods, band combination, loss function, and model optimization. The purpose of our study is to verify that a reasonable optimization approach for training data is more important for remote sensing multi-classification than model optimization alone.

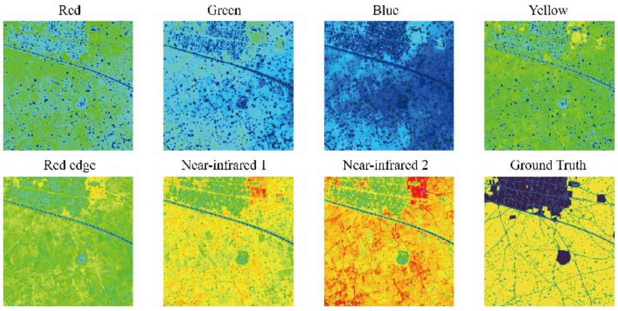

Figure 1. Composition of multispectral bands and classification labels.

2. Data

The Worldview3 data used in this study consists of a total of 15 scenes, each with a coverage of 1km × 1km. Each scene was divided into two groups according to different band combinations: one group consisted of three bands for true color, while the other group consisted of seven bands (Figure 1), including yellow, red edge, and two near-infrared bands in addition to true color, covering a wider spectral range. Five land cover types were manually interpreted and labeled: buildings, roads, trees, cultivated land, and water bodies. In this study, 14 scenes were selected for model training and accuracy evaluation.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-1-2024-359-2024

The remaining data was reserved for visual evaluation. All the data was cropped into slices of size 256 × 256, and the 4400 slices of data obtained were divided into training samples. All data has been normalized and divided into two groups, with 70% and 30% used for training and validation, respectively.

3. Methods

3.1 The Baseline Model

The UNet was proposed in 2015, and the network consists of an encoder and a decoder (Figure 2). The encoder includes several convolutional and pooling layers for fea-ture extraction. In order to improve computational efficiency and expand the receptive field, four downsampling operations were performed on the encoder part. On the other hand, the decoder is composed of several convolutional and up-sampling layers, which are used to restore the downsampled features and fuse them with same-scale features in the encoder to extract high-level semantic features, ultimately forming an end-to-end semantic segmentation network. The UNet's encoder is identical to VGG16 network, with the basic convolutional unit consisting of a 3 × 3 convolutional layer and an activation layer. The specific architecture of UNet is as shown in Figure 2.

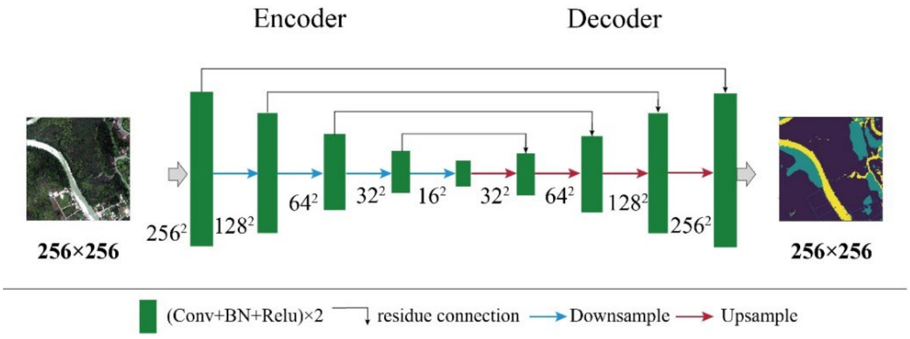

Figure 2. The architecture of Unet

3.2 The Optimized UNet

The UNet model has the advantages of fewer parameters and a simpler structure, which has been widely used in computer vision, medicine and other fields. However, there are still many aspects to be improved in the resolution of land cover classification based on high-resolution remote sensing. Therefore, we propose OUnet, which has made improvements in the following five aspects based on UNet: First, while retaining more shallow spatial information, the downsampling operation is reduced to improve the edge accuracy of remote sensing land classification results. Second, depthwise separable convolution is used to replace ordinary convolution to further reduce model parameters with-out significantly reducing model performance. Third, an attention mechanism is introduced in the decoder to better integrate features at different scales. Finally, a Dropout layer is added to prevent overfitting of the model. While increasing the number of convolutions in the encoder effectively improves model performance, this article did not optimize OUnet in this aspect to facilitate better comparison with the baseline model. The baseline model is optimized using the above method to obtain OUnet, and its architecture is shown in Figure 3.

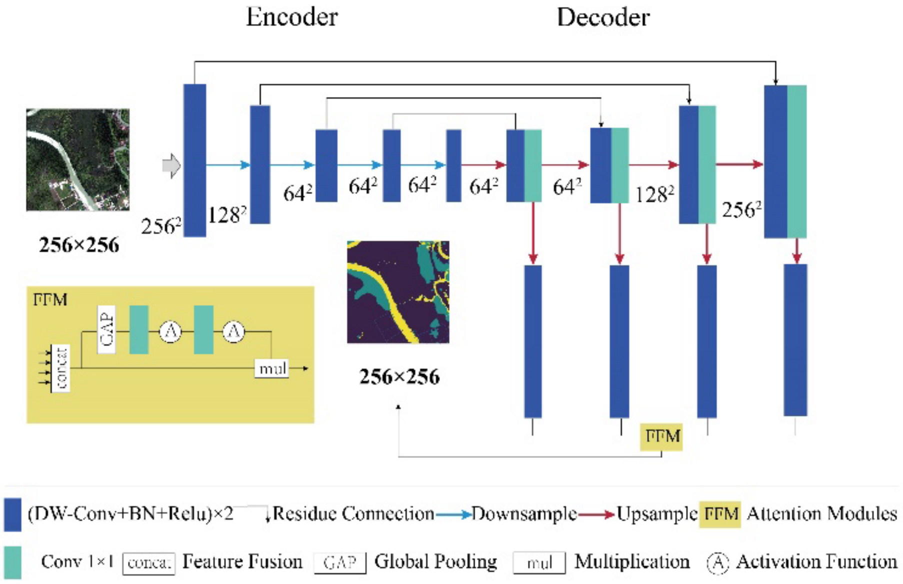

Figure 3. The architecture of OUNet.

Therefore, we propose OUnet, which has made improvements in the following five aspects based on UNet: First, while retaining more shallow spatial information, the downsampling operation is reduced to improve the edge accuracy of remote sensing land classification results. Second, depthwise separable convolution is used instead of common convolution to further degrade model parameters without significantly reducing model performance. Third, an attention mechanism is introduced in the decoder to better integrate features at different scales. Finally, a dropout layer is added to prevent overfitting of the model. The structure of the optimized OUnet model is shown in Figure 1.

3.3 Evaluation Metrics

In this paper, we use Intersection over Union (IoU), Mean Intersection over Union (mIoU) and F1 score as comprehensive evaluation indexes, while referring to precision and recall for evaluating classification results. The calculation formulas of each parameter are as follows:

$$\mathrm{IoU} = \frac{\mathrm{tp}}{\mathrm{fp} + \mathrm{fn} + \mathrm{tp}} \tag{1}$$

$$\mathrm{mIoU} = \frac{1}{\mathrm{k}+1} \sum_{\mathrm{i}=0}^{\mathrm{k}} \frac{\mathrm{tp}}{\mathrm{fp} + \mathrm{fn} + \mathrm{tp}} \tag{2}$$

$$\mathrm{F}_1 = 2 \,/\, \left(\frac{1}{\mathrm{recall}} + \frac{1}{\mathrm{precision}}\right) \tag{3}$$

$$\mathrm{precision} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}} \tag{4}$$

$$\mathrm{recall} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{pn}} \tag{5}$$

where k = feature class
tp = positive sample predicted to be positive
fn = positive sample predicted to be negative
fp = negative sample predicted to be positive
tn = negative sample predicted to be negative

In this study, overall accuracy, recall, F1-score , and Intersection over Union (IoU) are employed as evaluation metrics. Two different loss functions—multiclass cross entropy J_{ce} and weighted multiclass cross entropy J_{wce} —were used for comparison. The formulas for these losses are as follows:

$$J_{ce} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} \, log \, \hat{y}_{ij} \tag{6}$$

where n = the number of samples
m = number of classes
y_{ij} = true label indicating if the i-th sample belongs to the j-th class
$\hat{y}_{ij}$ = the predicted probability of the i-th sample belonging to the j-th class

$$J_{wce} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} w_j y_{ij} \, log \, \hat{y}_{ij} \tag{7}$$

where w_j = the weight of the j-th class

4. Experiment and Results

4.1 Experimental Setup

Different data combinations, sampling methods, models, and loss choices may affect the classification results, and the goal of this study is to quantify such differences experimentally. We set up five experiments, as shown in Table 1. Two combinations of three-band true-color data and seven-band multispectral data were provided in the experiments. The two sampling methods are sequential sampling, where slices are croped along the image length and width in fixed steps, and balanced sampling, where slices are randomly generated within the image and the sampling balance is adjusted by limiting the proportion of each class in the labels. The loss function is compared using multi-class cross-entropy (CE) or weighted multi-class cross-entropy (WCE).

Name	Model	Data	Sampling	Loss
Baseline	UNet	Three-band	sequential	CE
Opt_1	OUNet	Three-band	sequential	CE
Opt_2	UNet	Seven-band	balanced	CE
Opt_3	UNet	Seven-band	balanced	WCE
Opt_4	**OUNet**	**Seven-band**	**balanced**	**WCE**

Table 1. Experimental setup

The experiment was conducted in a Windows 10 environment with a CPU of Gold 5218@2.3GHz (×2), 256GB of memory, and an NVIDIA Tesla A100 GPU. The deep learning framework used was TensorFlow (2.6.0). During the training process, the adaptive learning rate optimization algorithm was used as the optimizer, with an initial learning rate of 0.0001 for optimization. All models were trained for 80 epochs, and the best model among them was selected for comparison.

4.2 Results of the Baseline Experiments

In the baseline experiments, the UNet model is trained on a three-band dataset that was obtained using sequential sampling methods. Cross-entropy is used as the loss function during training. The validation results of the test dataset (Table 2) show that the classification accuracy is not high, with an mIoU of 0.52. There are significant differences in the classification accuracy of various land cover types. The accuracy for farmland and buildings is relatively high, with IoUs both reaching above 0.75 and F1 scores both reaching above 0.85, while the accuracy for roads and water bodies is poor at 0.256 and 0.218, respectively. From the above data, it can be inferred that the strategy used in the baseline experiment has limited ability to extract multi-class land features.

Types	Accuracy	Recall	F1	IoU
Buildings	0.850	0.895	0.872	0.773
Roads	0.716	0.285	0.408	0.256
Trees	0.755	0.685	0.718	0.560
Farmland	0.863	0.904	0.883	0.791
Water	0.466	0.291	0.358	0.218
Average	**0.730**	**0.612**	**0.648**	**0.520**

Table 2. Performance of the baseline experiments

Analysis of the proportion of each land type in the samples shows that there is a serious imbalance in the number of samples in each category. The water and road samples were the least represented (Figure 4), which is one of the main reasons for the overall low classification accuracy.

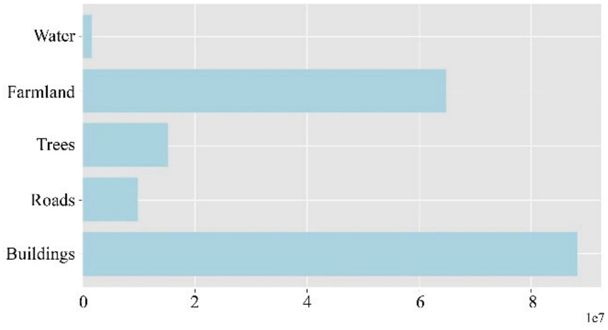

Figure 4. The distribution of various land cover types in the training data.

4.3 Optimization Experiments and Result Analysis

Two optimizations were implemented to address the issue of sample im-balance. First, random sampling was used to obtain image slices for training, and the proportion of each class in the labels was adjusted to control the overall sample size. Second, a weighted multi-class cross-entropy loss function was used. This function gives samples with lower proportions of each category high-er loss penalties based on their actual proportions. This forces the model to learn more features from these types of samples.

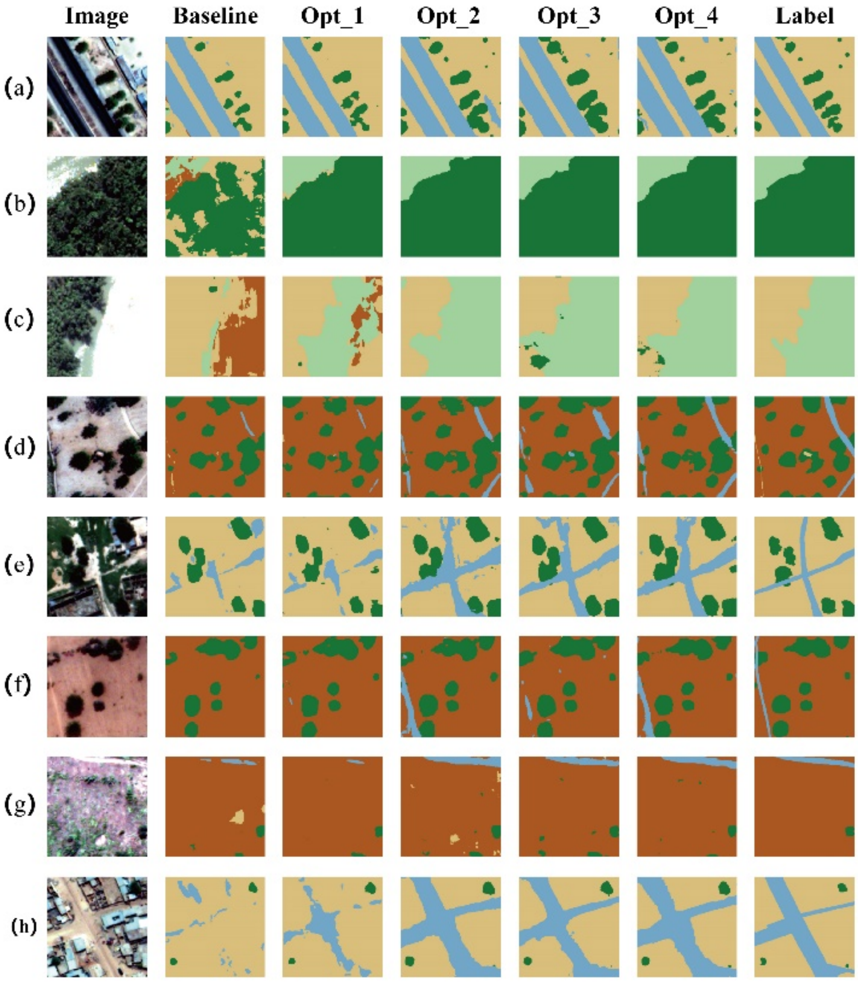

Figure 5. Classification results of different models.

To verify the effectiveness of the proposed optimization strategy, five sets of comparative experiments were performed for validation (Table 1). Figure 5 shows the prediction results of different experiments. It can be seen that in the prediction results of the baseline experiment, water, trees, and roads, which account for a relatively small proportion of the data, have more obvious misidentifications. However, with other optimized solutions, this situation has improved. Figure 5 (b, c) shows that switching the input data to seven-band multispectral data made the Opt_2 group much better at classifying trees, water, and roads, but it was still had a significant gap with the ground truth. The Opt_3 group added a sample balance sampling strategy and a weighted information entropy loss function based on Opt_2, which further improved the recognition accuracy of small sample categories such as water and roads. This indicates that enhancing the richness of the input data and sample balance are very effective in improving the multi-class classification accuracy of remote sensing. The overall prediction results of three experimental groups, Opt_3, Opt_1, and Opt_4, are relatively close to the ground truth. Among them, the Opt_4 group, which adopted all optimizing methods, performed the best.

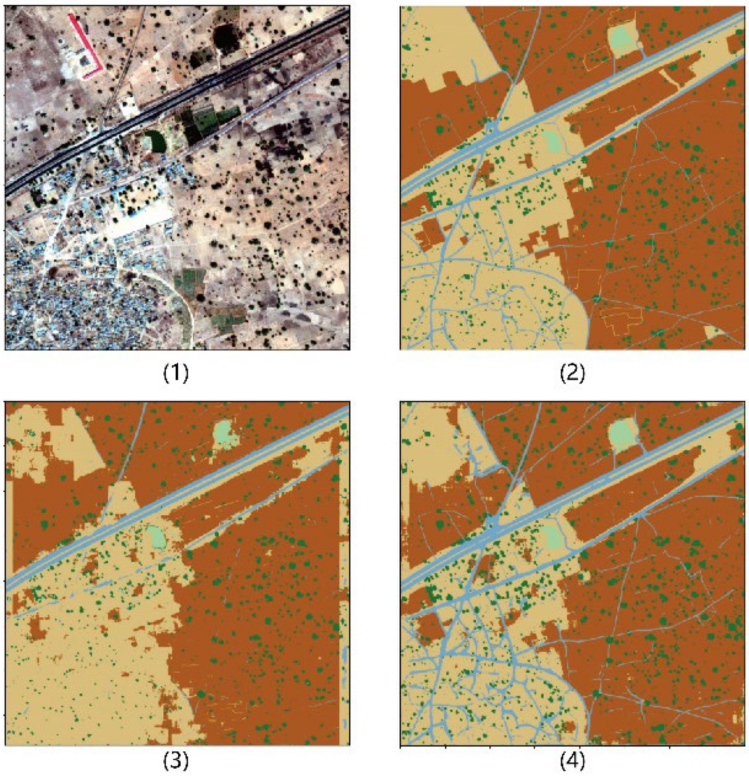

Figure 6. Performance of baseline and improved experimental results on test data. (a) RGB, (b) Opt_4, (c) baseline, (d) GT.

As shown in Table 3, the accuracy of the optimal experimental group is significantly better than the baseline group. The recall of roads and water is improved by 66% and 70%, respectively, and the IoU is improved by 30% and 58%, respectively. Tests on the whole scene image (Figure 6) also confirm the previous conclusions that the improved method is closer to the ground truth, while the baseline method has obvious misclassification and omission. The quantitative analysis results in Table 4 also confirm that the proposed improvement strategies are effective for multi-class classification applications, with recall, F1, and IoU improved by 29%, 17%, and 19%, respectively.

Types	Accuracy	Recall	F1	IoU
Buildings	0.966	0.773	0.858	0.752
Roads	0.569	0.947	0.710	0.551
Trees	0.645	0.909	0.755	0.606
Farmland	0.910	0.918	0.914	0.842
Water	0.797	0.996	0.885	0.794
Average	**0.777**	**0.909**	**0.825**	**0.709**

Table 3. Performance of the Opt_4 experiment.

In order to observe the impact of data imbalance on the classification accuracy in multi-class classification scenarios more carefully, we analyzed the recall rate, F1-score, and IoU index of the prediction results for the two least represented land cover types (roads and water bodies) in five sets of experiments. From Figure 7, it can be seen that all metrics in the baseline experimental group are the lowest values, while the Opt_3 experi-mental group has better results relative to Opt_2, suggesting that the WCE loss function is very effective for the data imbalance case for multiple landcover classification. By observ-ing all the experimental data, it can be found that although the sample size of the water body category is small, its features are more distinguishable from the background com-pared to roads. Roads, as a typical linear shallow feature, spatial information is very im-portant to improve the recognition accuracy, which is taken into account in the design of OUNet. Therefore, the road accuracy improvement in the Opt_1 group is very obvious. When compared with Opt_3, it has limited room for improvement in sample conditions compared to data optimization.

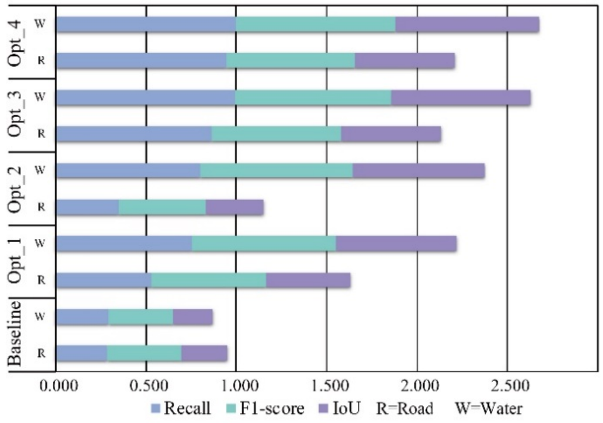

Figure 7. Comparison table of accuracy for roads and water in all experiments.

Name	Precision	Recall	F1-score	IoU
Baseline	0.730	0.612	0.648	0.520
Opt_1	0.843	0.763	0.797	0.675
Opt_2	0.851	0.770	0.793	0.685

Name	Precision	Recall	F1-score	IoU
Opt_3	0.767	0.889	0.814	0.692
Opt_4	**0.777**	**0.909**	**0.825**	**0.709**

Table 4. Quantitative comparison results of all groups

5. Conclusion

In this paper, the optimization method for multi-class land cover classification of high-resolution remote sensing imagery is investigated from different perspectives, such as sampling strategy, band combination, loss function, and model optimization. We propose OUnet by optimizing it in four aspects to effectively improve its classification ability. In a multi-classification scenario, the balance of data and the effective number of samples are crucial to the classification results. The proposed optimization strategies for data combination, sampling strategy, and loss function have greater improvement in multi-classification accuracy compared to model optimization.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2021YFC3000400) and Major Project on High-Resolution Earth Observation System of Systems (04-H30G01-9001-20/22).

References

Zhou, H.Q., Huang, L.L., Wang, Y.T., 2019: Deep learning algorithm and its application in optics. *Infrared and Laser Engineering,* 48(12), 1226004.

Yann, L.C., Yoshua, B., Geoffrey, H., 2015: Deep learning . *Nature*, 521(7553), 436-444.

Olaf, R., Philipp, F., Thomas, B., 2015: U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention, Springer,* 234-241.

Deng, Z.P., Lei, L., Sun, Hao., Zhao, J.P., Zhou, S.L., Zou, H.X.,2018: Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 3.

He K., Zhang X., Ren S., Sun J., 2016: Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 770-778.

Zheng Z., Zhong, Y.F., Wang, J.J., Ma A., 2020: Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4096-4105.

Chen, L.C., Yang, Y., Wang, J., Xu, W., Yuille, A.L., 2016: Attention to scale: Scale-aware semantic image segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3640-3649.