Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-5 (2020) 397–402

Semantic Segmentation for Remote Sensing based on RGB Images and Lidar Data using Model-Agnostic Meta-Learning and Partical Swarm Optimization

Kai Zhang*, Yu Han, Jian Chen*, Zichao Zhang*, Shubo Wang*, *****

** College of Engineering, China Agricultural University, Beijing 100083, China (Tel: 86-18810922501; e-mail: jchen@cau.edu.cn, chenjian@buaa.edu.cn).*

*** College of Water Resources & Civil Engineering, China Agricultural University, Beijing 100083, China*

**** State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China*

Abstract: Urban remote sensing has the problems that land cover categories are usually highly unbalanced and annotated samples are scarce, which brings great limitations to monitoring the change of urban coverage and periodically evaluating urban ecological conditions. Semantic segmentation is one of the main applications in urban remote sensing image analysis. Because the ground objects in remote sensing images have the characteristics of disordered distribution and irregular morphology. The classical semantic segmentation model based on deep learning U-Net cannot achieve high semantic segmentation accuracy for urban ground objects. This paper proposes to optimize the neural network structure and introduce lidar data to solve the above problems. In this paper, the Model-Agnostic Meta-Learning and fully convolutional neural networks are fused which be trained and tested by remote sensing images. It makes the training process into inner loop and outer loop. And Partical Swarm Optimization (PSO) is used to optimize the parameter updating process of neural network to further improve the test accuracy. This paper fuses Lidar data and remote sensing images, and comprehensively use the position and elevation information of Lidar data and the spectrum and texture information of remote sensing images to classify the ground features. The datasets used in this paper are RGB remote sensing images and Digital Elevation Model (DEM) images. The test accuracy of U-Net network optimized by MAML can be improved by 6%-7% under the same network parameters and training data sets. With the introduction of Lidar data, the accuracy of the test is increased by 3-5%. The experimental results show that the precision before and after PSO optimization is improved by 6%-9%, which verifies the idea in the paper.

Keywords: Urban remote sensing, semantic segmentation, Model-Agnostic Meta-Learning (MAML), Partical Swarm Optimization (PSO), Lidar data, RGB remote sensing images, Digital Elevation Model (DEM)

1. INTRODUCTION

In recent years, the remote sensing images obtained contain more spatial and spectral information. The various geographic information contained in them has become more complex. But in the face of massive data, the technology of data ex-traction appears to be some backward. Segmenting remote sensing images and acquiring geographic information through semantic segmentation have become one of the most popular qualitative analysis research contents in the field of qualitative remote sensing. However, the training of the traditional neural network requires large number of training sets as support. The calibration and processing of training sets are very time-consuming. This greatly limits the speed and efficiency of semantic segmentation of remote sensing data (Li, 2019; Su, 2019; Zhao, 2019).

Semantic segmentation of remote sensing images is a qualitative classification problem at pixel level, which is more complex than usual picture classification. This requires deeper neural networks to achieve higher precision segmentation. at the same time, it also needs to train neural networks with more data sets. However, remote sensing image data sets are difficult to calibrate, resulting in the scarcity of training sets. In order to solve the problem of the scarcity of training samples when classifying remote sensing images based on deep learning. Rao et al. select multiple supporting samples for each class to ensure that the extracted features are sufficiently stable to avoid HSI intraclass similarity and interclass dissimilarity problems. a spatial-spectral relationship network (SS-RN) of limited training samples is proposed for hyperspectral classification (Rao, 2019). Liu et al. proposed a deepseated small sample learning method for the scarce sample problem in remote sensing image classification. the core idea is that the designed network can learn the metric space from the training dataset (Liu, 2019). Tang et al. apply twin networks to SAR target recognition of limited data, not only take advantage of the advantages of metric learning, improve the accuracy of SAR target recognition in the case of limited data, but also greatly reduce the prediction time consumption based on metric learning model (Tang, 2019). Using different methods such as Meta-Learning to optimize the traditional neural network can effectively reduce the data set needed for training, and the information feature utilization of the data set is more efficient. A particle swarm optimization is used to optimize the training process of the MAML, so that the neural network

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2021.04.117

can learn the information of the image faster to achieve higher accuracy.

Multi-spectral or hyperspectral images obtained by a sensor have the disadvantage of information redundancy between different bands, which often cannot meet the requirements of large range and high precision. Wenqing Feng et al. proposed a target-level change detection method based on multi-scale segmentation fusion, using fine to coarse segmentation to obtain initial objects of different sizes. According to the characteristics of the object, the change vector analysis method was used to obtain the change detection results of various scales. In order to improve the accuracy of change detection, fuzzy fusion and two decision-level fusion methods were also introduced to obtain the results of multi-scale fusion (Feng, 2016). Dalponte, M. et al. proposed a joint effect analysis of hyperspectral and lidar (LIDAR) data fusion for classification of complex forest areas, illustrating the validity and potential of joint use of hyperspectral and LIDAR data and the accuracy of different classification techniques analyzed in the proposed system (Dalponte, 2008). Khodadadzadeh et al. proposed a new effective strategy for the fusion and classification of hyperspectral and LiDAR data. The proposed method is designed to integrate multiple types of features extracted from these data. An important feature of the proposed method is that it does not require any regularization parameters to effectively develop and integrate different types of features in a collaborative and flexible manner (Khodadadzadeh, 2015). The remote sensing task can be accomplished more effectively by fusion with the unpassed data, but it is necessary to select reasonable data for different tasks and extract useful features to improve the accuracy of the results.

In this paper, our main contribution is the innovative application of MAML to the U-Net network and the use of RGB remote sensing images for training.

1. MAML is used to optimize neural network training process.

2. In this paper, the bionic optimization algorithm is to optimize neural network parameter updating process.

3. We introduce the lidar data to improve semantic segmentation accuracy.

The test accuracy of U-Net network optimized by MAML can be improved by 6%-7% under the same network parameters and training data sets. With the introduction of Lidar data, the accuracy of the test is increased by 3-5%. The experimental results show that the precision before and after PSO optimization is improved by 6%-9%. MAML can get a better segmentation effect when the semantic segmentation has the same calibrated training set. And through PSO optimization of network parameters and introduction of Lidar data can effectively improve the accuracy of testing. The rest of the article is organized as follows: in the section 2, the specific structure of MAML network are introduced; in the section 3, MAML and PSO are fused; in the section 4, MAML and U-Net are fused; in the section 5, the experimental conditions are set and the comparative experiment results are summarized and demonstrated; in the section 6, this article conclusion is summarized and analyzed.

2. Model-Agnostic Meta-Learning (MAML)

The main objective of MAML is to find a better initial parameter through a series of training iterations. By finding new initial parameters, the model used can converge with less training times. When neural networks perform classification tasks, the network is trained by initializing random weights and minimizing losses. It uses gradient descent to minimize loss and gradient descent to find the optimal weight, which will give us the smallest loss.

Suppose there are three related tasks: T_1, T_2 and T_3. The initial parameters θ of the neural network are firstly set randomly. For task T_1, the neural network receives the training data first and then produces a loss, which is minimized T_1 by stochastic gradient descent or other gradient descent methods according to the loss. The optimized model parameters $\theta_1^{'}$ are obtained by several steps of iterative training. Similarly, for tasks T_2 and T_3, it will start with a randomly initialized model parameter θ and find the optimal parameters $\theta_2^{'}$ and $\theta_3^{'}$ for tasks T_2 and T_3 by gradient descent respectively. As shown in the Fig. 1, optimal parameters $\theta_1^{'}$, $\theta_2^{'}$ and $\theta_3^{'}$ are found for each task T_1, T_2 and T_3.

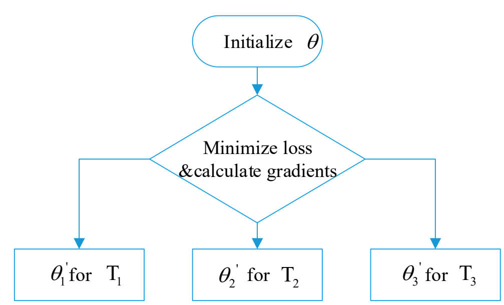

Fig. 1. The process of optimizing parameters for different tasks.

There are a model f_θ parameterized by θ and a distribution $\mathrm{P}(\mathrm{T})$ on the task. The initial parameters θ of the neural network are firstly set randomly. Next, some tasks T_i are extracted from the above tasks, that is, $\mathrm{T}_i \sim \mathrm{P}(\mathrm{T})$. MAML does the above training process by minimizing the loss. For each task T_i, updating the parameters by following:

$$\theta_i^{'} = \theta - \alpha \nabla_\theta L_{T_i}\left(f_\theta\right) \tag{1}$$

where $\theta_i^{'}$ is the optimal parameter for a task $\mathrm{T_i}$, θ is the initial parameter, α is the hyperparameter, $\nabla_\theta L_{T_i}\left(f_\theta\right)$ is the gradient of a task $\mathrm{T_i}$.

By updating the initial parameters, different update parameters are obtained for each task:

$$\theta^{'} = \left\{\theta_1^{'}, \theta_2^{'}, \theta_3^{'}, \theta_4^{'}, \theta_5^{'}\right\} \tag{2}$$

Different parameters $\theta_i^{'}$ obtained by updating different task training are distributed differently. For new tasks $\mathrm{T_i}$ that need to be learned, you can train faster and update the random initialization parameter θ according to the parameters that have been learned. This allows random initialization parameter θ to move to a relatively optimal position to get a better initial parameter. The method of the entire algorithm can be implemented with the following formula:

$$\theta = \theta - \beta\nabla_\theta \sum_{T_i \sim p(T)} L_{T_i}\left(f_{\theta_i^{'}}\right) \tag{3}$$

where θ is the initial parameter, β is the hyperparameter, $\nabla_\theta \sum_{T_i \sim p(T)} L_{T_i}\left(f_{\theta_i^{'}}\right)$ is the gradient value of each different task $\mathrm{T_i}$ with respect to parameter $\theta_i^{'}$.

MAML training process can be divided into inner and outer loop. The whole process is shown in Fig. 2:

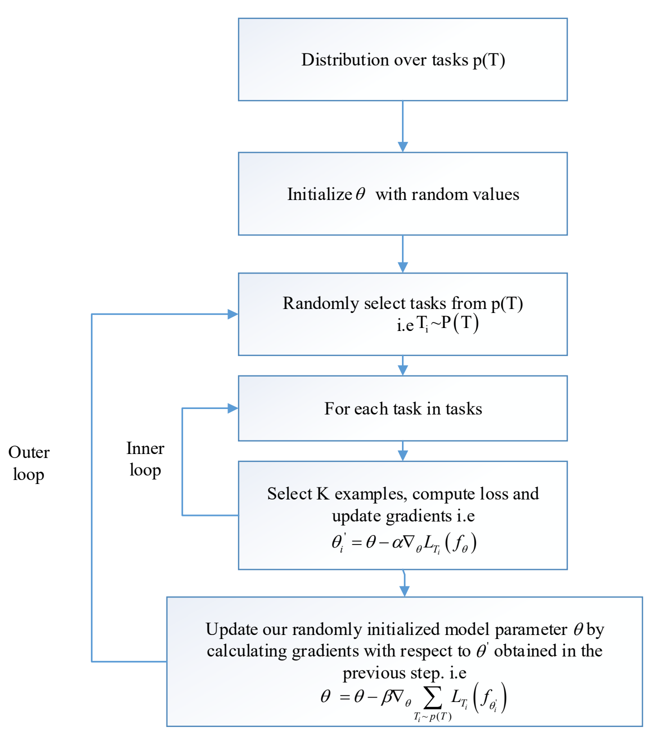

Fig. 2. Application process of MAML algorithm.

2. Particle Swarm Optimization to optimize MAML

Particle swarm optimization (PSO) belongs to a swarm intelligence algorithm (Li, 2002). The specific implementation of the algorithm is to use a particle to simulate an individual bird. Each particle can be regarded as a search individual in a N dimensional search space. The current position of the particle is a candidate solution corresponding to the optimization problem. The flight process of the particle is the search process of the individual. The flying speed of particles can be dynamically adjusted according to the optimal position of particle history and the optimal position of population history. Particles have only two attributes: speed and position, speed represents the speed of movement, position represents the direction of movement. The optimal solution searched by each particle alone is called the individual extremum, and the optimal individual extremum in the particle swarm is taken as the current global optimal solution. Iteratively continuously, updating speed and position. finally, the optimal solution satisfying the termination condition is obtained.

First, we set the maximum number of iterations, the number of independent variables of the objective function, the maximum velocity of the particle, and the position information for the whole search space. We randomly initialize the velocity and position in the velocity interval and the search space, and set the particle swarm size to initialize a flying velocity randomly M. Then the fitness function is defined, and the individual extremum is the optimal solution found by each particle find a global value from these optimal solutions, which is called this global optimal solution. Compared with the historical global optimum for updating. Updated formula for speed and position:

$$\begin{aligned} V_{id} &= \omega V_{id} + C_1 random(0,1)\left(P_{id} - X_{id}\right) \\ &+ C_2 random(0,1)\left(P_{gd} - X_{id}\right) X_{id} = X_{id} + V_{id} \end{aligned} \tag{4}$$

$$X_{id} = X_{id} + V_{id} \tag{5}$$

where, ω is inertial factor, C_1 and C_2 are called acceleration constant, it is generally taken $C_1 = C_2 \in [0,4]$. $\mathrm{random}(0,1)$ represents the random number on the interval $[0,1]$. P_{id} represents the d first dimension of the individual extremum of the i variable. P_{gd} represents the d dimension of the global optimal solution. The termination condition is the maximum number of iterations.

This paper uses the PSO to optimize the learning rate of the MAML, and obtains different learning rates through optimization to speed up the learning speed and improve the performance of the network model. The MAML is divided into two learning rates: inter loop and outer loop. The fitness function is as follows:

$$\min f\left(r_1, r_2\right) = \min L\left(f_\theta\right) \tag{6}$$

where r_1 and r_2 are the learning rate of inner loop and outer loop in the MAML.

3. The Fusion of MAML and U-Net

MAML can achieve the purpose of fast training by quickly learning the initial parameters. And the model obtained by MAML training has the advantage of good generalization. This paper attempts to use MAML to optimize U-Net neural networks, and to find the optimal initial parameters which can reduce the time period required to train the neural network and increase the generalization of the model. All these optimized initial parameters refer to the weights of different layers of the neural network and it determine the final classification effect.

This paper presents a classification task for semantic segmentation, so the loss function of the neural network can be expressed as follows (Yi, 2017):

$$L_{T_i}\left(f_\theta\right)=\sum_{x_j,y_j\sim T_i} y_i \log f_\theta\left(x_j\right)+ \left(1-y_j\right)\log\left(1-f_\theta\left(x_j\right)\right) \tag{7}$$

This is how U-Net is applied to MAML:

a) Suppose there is a parameterized model f_θ by parameter θ and a distribution on the task $\mathrm{P}(\mathrm{T})$. First, the model parameter θ is randomly initialized.

b) Some different batches of tasks $\mathrm{T_i}$ are randomly selected from task T, that is, $\mathrm{T_i}\sim\mathrm{P}(\mathrm{T})$. For example, there are selected three tasks $\mathrm{T}=\{\mathrm{T_1},\mathrm{T_2},\mathrm{T_3}\}$.

c) Inner loop: For each tasks $\mathrm{T_i}$ taken from T, this paper sample k data points and prepare our training and test data sets:

$$D_i^{train}=\left(x_1,y_1\right)\ldots\left(x_k,y_k\right),\ D_i^{test}=\left(x_1,y_1\right)\ldots\left(x_k,y_k\right)$$

Now U-Net network is used to train on D_{train} to calculate the loss. Different gradient minimization methods are used to reduce losses to obtain the best parameter $\theta_i^{'}$, that is, $\theta_i^{'}=\theta-\alpha\nabla_\theta L_{T_i}\left(f_\theta\right)$. Performing the same calculations for the other tasks to get the parameters $\theta_1^{'}$, $\theta_2^{'}$ and $\theta_3^{'}$.

d) Outer loop: The meta-test set D_{test} are used to minimize the loss based on the parameters obtained in the previous step, and obtain an optimal initialization parameter θ.

For n iterations, we repeat steps b) to d). The introduction of the inner loop training process makes the model learn new tasks quickly with a large learning rate, and the learned parameters update the model parameters with a small learning rate. The following diagram gives an overview of MAML in U-Net:

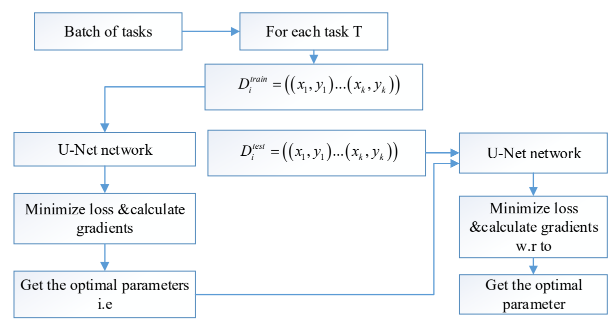

Fig. 3. The overview of MAML in U-Net.

4. Building Detection Experiments

4.1 Dataset

The training images is the data from the 2015 IGRSS Data Fusion Competition. To explore the fusion of RGB and LiDAR data. Including two datasets obtained simultaneously using active and passive sensors, acquired on march 13,2011, at an air-borne platform of 300 m flight altitude in belgian Zeebruges ports (51.33°N,3.20°E). the passive dataset is a 5 cm resolution RGB orthophoto image obtained in the visible wavelength range; the active dataset is LiDAR data collected using repetition rate, angle and frequency of 125 kHz,20° and 49, respectively, and includes a 10 cm spatial resolution digital surface model obtained point cloud data. In this case, the number of buildings is multi-density, which contains tagged large size RGB remote sensing images with a total of one class of objects, buildings (mark 1) and others (mark 0). To better observe the images, we have visualized two of the training pictures as follows: red- buildings. There is a DEM image obtained from Lidar data as shown in the Fig. 4.

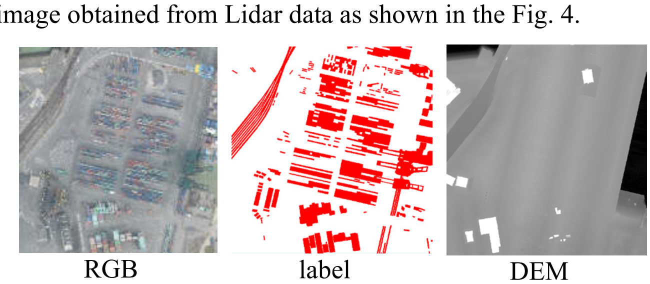

Fig. 4. Remote sensing images and label images.

4.2 Experimental setup

This paper needs to do six groups of comparative experiments. The experimental network to be optimized is U-Net network, which is optimized by using MAML and MAML optimized by PSO. The datasets used in the experiment are RGB remote sensing images and RGB-DEM images.

In terms of parameter settings for U-Net network training, the initial training step is set to 0. 0001, and Adam is applied as the optimizer. MAML is divided into an inner loop and an outer loop. There are two learning rates that need to be set, where the meta- training step is 0.0001, and the training step is 0.001. As the basic network, the U-Net network structure is a ten-layer network. Because it is a symmetric structure, a Batch Normlization layer is used after a convolutional layer. The first five layers are followed by max-pooling, and the next five layers are followed by upsampling layers. The size of the core of each convolutional layer is 3*3, and the output layers of each convolution layer is 32,32,64,64,128,64,64,32,32,5. Since the training image is an RGB image, the first layer input has 3 layers. Because it is a two-class classification problem, the last layer is output two layers. The upsampling and maxpooling filter sizes are 2*2, and other parameters are set by default.

The experimental operating environment is Windows 10 system. It is an end-to-end open source machine learning platform using TensorFlow in the anaconda virtual environment. The programming language is python and the hardware environment is Z7-CT5NS workstation equipped with NVIDA GeForce GTX1660Ti GPU.

4.3 Results

The experiment uses the above experimental settings for training under U-Net network, MAML-U-Net and PSO-MAML. The above neural network model is used to train the two sets of data of RGB image and RGB-DEM image respectively to verify the improvement of DEM image for semantic segmentation task. The experiment has one large remote sensing images, and choose one part for training and the other for testing. These images cannot be fully used for neural network training, because workstation cannot bear the memory and their sizes are different. Therefore, it is firstly made them to do random cutting. That is, we random generate x, y coordinates, and then pick out the small graph of 200*200 under this coordinate. A total of 1000 small images are obtained for training and testing, where 75% is used for training model and 25% is used for model testing. MAML makes the training process into internal training and external training. Our dataset is also divided into training set and meta-training set, where the first data set of inner loop is the basic ability to learn to handle this task for each task. The second data set of the external training is the generalization ability of learning multiple tasks. In the course of the experiment, the training set is inputted into the inner loop and outer loop simultaneously. The basic ability and generalization of the model processing tasks obtained from the training can be very good, so the test accuracy is the best. Then use the prepared test data set to test the accuracy of the trained model, the test results as shown in Table 1. The MAML means the U-Net network optimized by MAML and the PSML means the U-Net network optimized by MAML optimized by PSO.

It can be seen from the experimental results that the test accuracy of the neural network with the introduction of Lidar data and the PSO optimization is the highest. U-Net network without any optimization means is very low for complex remote sensing image classification, especially when the spectral similarity of ground objects is very high. Although the high-resolution remote sensing image records continuous spectral information, including important information such as spatial structure, ground object boundary, color attributes and so on. With the introduction of Lidar data, the accuracy of the test is increased by 3%-5% under the same neural network model. The results show that the Lidar data can help the common spectral images to increase the interpretation height information, and thus improve the classification of ground objects in the case of high spectral similarity. The test accuracy of U-Net network optimized by MAML can be improved by 6%-7% under the same network parameters and training data sets, respectively, so MAML has a good effect on the semantic segmentation of remote sensing images of complex tasks. This is because the increase of generalization by increasing the training process of the outer training loop, which has a good effect in segmenting the complex problem of remote sensing image. The learning rate of MAML two learning is optimized by PSO algorithm, so as to speed up the training learning speed and strengthen the extraction of image features. The experimental results show that the precision before and after PSO optimization is improved by 6%-9%, which verifies the idea in the paper.

Table 1. The test results.

Model	pixel acc	F1 score	mIoU
U-Net-RGB	0.57682	0.52965	0.36178
MAML-RGB	0.65350	0.62370	0.41936
PSML-RGB	0.71178	0.78864	0.44259
U-Net-RGBD	0.60363	0.54522	0.39950
MAML-RGBD	0.66606	0.65447	0.46473
PSML-RGBD	0.75441	0.70236	0.52716

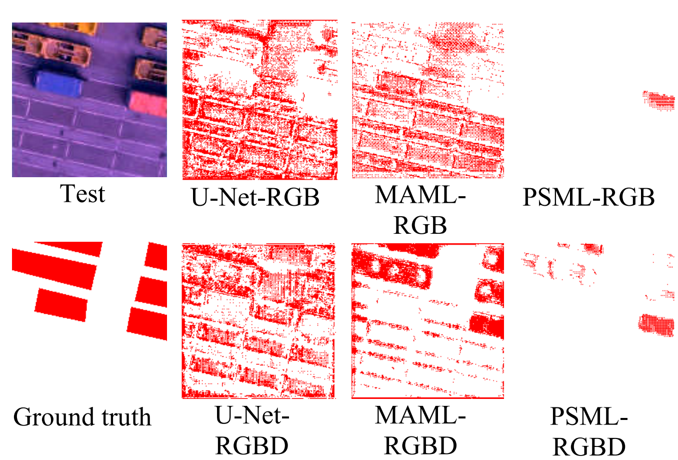

Fig. 5. The predicted results.

The predicted results of the model obtained by our training are shown in Fig. 5. Results that the test accuracy after applying MAML are obviously increased, but the overall accuracy is not very good. By introducing lidar data, the misclassification of objects with obvious height difference can be effectively reduced. The appropriate learning rate can be updated during training by using PSO to optimize the MAML learning rate. Experimental results show that the misclassification rate for features is greatly reduced. Compared with the non-optimized model, the performance of the optimized model is greatly improved. Because of the optimization of MAML training process, the parameters learned from the training are more generalized. It is more effective for the classification accuracy of complex remote sensing semantic segmentation task.

5. Conclusion

In this paper, MAML is proposed in semantic segmentation and combined with U-Net to solve qualitative remote sensing analysis. The test results show that MAML has very good performance on the accuracy of semantic segmentation. The improvement is mainly due to the MAML algorithm trains a set of initialization parameters. Through MAML, we can learn better initial parameters, which has achieved the purpose of fast training and has increased model generalization. This paper proves that MAML can optimize the structure of U-Net. Under the same data set and experimental conditions, it can indeed speed up the training speed and accuracy. The learning rate in the training process of MAML is optimized by PSO algorithm, which obviously improves the testing accuracy and training speed of the model. In view of the low classification accuracy of ordinary spectral images in the case of spectral similarity, Lidar data is introduced to solve the above problems. The experimental results show that adding height information can increase the effective characteristics of classification and thus improve the accuracy of testing. The proposed method can effectively improve the accuracy of feature extraction of remote sensing images.

Acknowledgement

This research was funded by the National Natural Science Foundation of China (Grant No. 51979275), by the National Key R&D Program of China (Grant Nos. 2017YFD0701003 from 2017YFD0701000, 2018YFD0700603 from 2018YFD0700600, and 2016YFD0200702 from 2016YFD0200700), by the Jilin Province Key R&D Plan Project (Grant Nos. 20180201036SF), by the Open Fund of Synergistic Innovation Center of Jiangsu Modern Agricultural Equipment and Technology, Jiangsu University (Grant No. 4091600015), by the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (Grant No. 19R06), by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University (Grant No. ICT20021), and by the Chinese Universities Scientific Fund (Grant Nos. 2020TC033 and 10710301).

REFERENCES

LI, X., TANG, W., and YANG, B. (2019). Semantic Segmentation of High-score Remote Sensing Images Using Deep Residual Network. Journal of Applied Science, 37 (02), 136-144.

SU, J., YANG, L., and JING, W. (2019). A U-Net Method for Semantic Segmentation of Remote Sensing Images. Computer Engineering and Applications, 55 (07), 212-218.

ZHAO, B., ZHANG, W., and YAN, Z. (2019). Semantic Segmentation of Remote Sensing Image Based on Multi-Digraph Pyramid Fusion Depth Network. Journal of Electronics and Information, 41 (10).

Rao, M., Tang, P., and Zhang, Z. (2019). Spatial–Spectral Relation Network for Hyperspectral Image Classification with Limited Training Samples. IEEE Journal of Selected Topics in Ap-plied Earth Observations and Remote Sensing, 12 (12), 5086-5100.

Liu, B., Yu, X., and Yu, A. (2019). Deep Few-Shot Learning for Hyperspectral Image Classification. IEEE Transactions on Geoscience & Remote Sensing, PP (99), 1-15.

Tang, J., Zhang, F., and Zhou, Y. (2019). A Fast Inference Networks for SAR Target Few-Shot Learning Based on Improved Siamese Networks. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 1212-1215.

Feng, W., Sui, H., and Tu, J. (2016). Object-Oriented Change Detection for Remote Sensing Images based on Multi-Scale Fusion. Acta Geodaetica Et Cartographica Sinica, XLI-B7 (10), 483-491.

Dalponte, M., Bruzzone, L., and Gianelle, D. (2008). Fusion of Hyperspectral and LIDAR Re-mote Sensing Data for Classification of Complex Forest Areas. IEEE Transactions on Geoscience and Remote Sensing, 46 (5), 1416-1427.11

Khodadadzadeh, M., Li, J., and Prasad, S. (2015). Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning. Selected Topics in Applied Earth Observations & Remote Sensing IEEE Journal of, 8 (6), 2971-2983.

Li, A., Qin, Z., ang Bao, F. (2002). Particle swarm optimization algorithm. Computer Engineering and Applications, 038 (021), 1-3.

YI, M. and SUI, L. (2017). Semantic Classification of Aerial Images based on Improved Full Convolutional Neural Network. Computer Engineering, 043 (010), 216-221.