Training strategy of CNN for remote sensing image classification with active learning

Liu Xiao Lei¹, Liu Peng¹
1. School of information science and engineering, Yanshan University

Emails: 2084595068@qq.com

Abstract. With the rapid development of remote sensing image classification technology, remote sensing image classification is more and more widely used in forest vegetation detection, land use and land cover, environmental monitoring, geographic information system and so on. However, because of the large amount of remote sensing image data, it is extremely difficult to select a large number of effective annotated data from the massive data. Especially when the classifier is oriented to deep learning, the sample labeling strategy of remote sensing image is particularly important. In the training process of remote sensing image classification based on CNN, the labeling strategy needs to consider both the uncertainty of samples and the representativeness of samples. In this paper, an active deep learning sample labeling strategy for remote sensing image classification is proposed. In the training process of VGG-16 convolutional neural network model, a density-based spatial clustering of application with noise is added, and information entropy is combined to improve the labeling strategy. This method focuses on the analysis of sample information, mining the intrinsic features contained in the sample, taking into account the uncertainty and correlation of the sample, and ultimately achieves the selection of the most informative sample for model training. Experiments show that this method effectively improves the labeling efficiency of samples and achieves higher classification accuracy under the same number of conditions.

Keywords: active learning, deep convolution neural network, deep learning, remote sensing images scene classification
1. Introduction
The remote sensing image is the image of the reflected electromagnetic wave information of a long-distance target radiated by the electromagnetic wave emitted by various sensing instruments, it is a spectral space composed of luminance characteristics \[1\]. With the development of the remote sensing technology, the remote sensing images already became the main approaches to obtain essential factor information, such as geomorphic feature, geomorphic interpretation and so on \[2\]. In the remote sensing images, there are rich in information, we can mine this information for disaster monitoring, farmland detection and so on. Classification methods of remote sensing images have always been the focus of research in the field of remote sensing, which has developed from the traditional remote sensing image classification method at the beginning to the remote sensing classification method based on the deep learning. As a result of the development of remote sensing technology and the increasing amount of remote sensing data, image classification based on deep learning has been favored by more and more users.

In the term of remote sensing image classification, mining of sample information is an important task, however, most of the current classification models focus on using the information of the labeled samples, unmarked sample information is not fully and reasonably utilized, such as support vector data description(SVDD) \[3\], Mixture of Gaussians \[4\], OCSVM \[5\], these methods usually use only labeled training samples \[6\]. In the field of remote sensing image classification, the labeling of samples is a very tedious task. In the view of labeled sample, the number of labeled samples is usually small, Contrary to the former, the number of unlabeled samples is very large. In the process of calculating the uncertainty of sample, MS is a common active learning method in sampling strategy, but MS only calculates two maximum probability labels, which ignoring the other information. In the process of image classification training, not only the uncertain of samples, but also the representativeness of samples should be considered to ensure that the information of samples is comprehensive. However, many sampling strategies still have the problem of incomplete mining information and lack of extraction and analysis of the information of sample representativeness and diversity.

2. Convolutional Neural Network

2.1 Convolutional network
Convolutional network, also known as convolutional neural network (CNN), is a kind of neural network specially used to process data with similar grid structure \[7\]. Convolution neural network belongs to a multilayer depth feedforward artificial neural network, it refers to those neural networks that use convolution operation to replace the general matrix multiplication operation in at least one layer of the network, which is also a typical representative of the influence of the principles of neuroscience on deep learning \[7\]. At present, thanks to the characteristics of less training parameters, simple structure and easy to adapt, convolutional neural network has become one of the most representative algorithms in deep learning algorithm, which plays an important role in the history of deep learning. In addition, convolutional neural network is widely used in various fields, it was first used to read cheques, and then the Microsoft deployed handwriting recognition system. In recent years,
convolutional neural network is frequently used in face recognition, target detection, gesture recognition and other fields.

Convolutional neural network has developed rapidly, from the simplest lenet-5 network to the later AlexNet, ImageNet, VGGNet, GoogleNet series networks, as well as ResNet series and SeNet series networks in recent years.

The traditional convolution neural network is generally divided into the input layer, the convolution layer, the pooling layer, the full connection layer and the output layer, using weight sharing. The input layer is mainly used to preprocess the original data; the convolution layer is composed of several convolution kernels, among which the convolution kernels are used to calculate different characteristic graphs; the role of the pooling layer is to reduce the output characteristic vectors of the convolution layer; the full connection layer is mainly used to connect all the neuron weights, and transmit signals to other full connection layers\(^8\); the role of the output layer is the final output result, generally it will follow the full connection layer.

The traditional convolution neural network is not perfect. With the development of deep learning, the convolution neural network model is also improving. For example, in 2015, the full convolution structure proposed by springenberg JT et al.\(^9\) replaced the full connection layer in traditional CNN with convolution layer, which improves the loss of two-dimensional spatial position information caused by mapping the feature map to a fixed size one-dimensional feature vector in the full connection layer, reduce error rate by up to 10% compared with traditional CNN; in the view of the problem that the pooling layer will cause the loss of local feature information, the hole convolution proposed by vladlen koltun et al.\(^10\) replaces the pooling layer with the hole convolution, so as to increase the receptive field without loss of information.

### 2.2 VGGNet network

VGGNet series proposed by researchers from Oxford University and Google DeepMind company, the most commonly used VGGNet series is vgg-16 network. The VGG-16 network is composed of 13 convolutions and 3 fully connected layers, including three fully connected layers and five convolutions. Compared with the traditional CNN, VGG-16 greatly deepens the network, improves the accuracy, and simplifies the neural network structure, which adopts convolution kernel of 3 * 3, step 1 and same convolution. In the VGG-16 network, with the deepening of the network, the proportion of image reduction and the proportion of channel number increase are regular. The following figure shows the networks with different structures of VGGNet series.

The structure of VGGNet is as shown in Table 1. In this table, from A to E, the depth of network configuration gradually increases. Among them, D is the vgg-16 network structure diagram, which is constructed by repeatedly using 3 * 3 small convolution kernel and 2 * 2 maximum pooling layer. Because the convolution kernel is very small, the network depth can be increased steadily by adding the number of convolution layers, and the convolution series is used in VGG-16, compared with only using a larger convolution kernel, it will have less parameter amount and more nonlinear transformation.
Table 1 the network structure of VGGNet

| ConvNet Configuration | A       | A-LRN  | B       | C       | D       | E       |
|-----------------------|---------|--------|---------|---------|---------|---------|
| 11 weight layers      | 11 weight layers | 13 weight layers | 16 weight layers | 16 weight layers | 19 weight layers |

Input(224*224 RGB image)

|               | Conv3-64 | Conv3-64 | Conv3-64 | Conv3-64 | Conv3-64 | Conv3-64 |
|---------------|----------|----------|----------|----------|----------|----------|
|               | LRN      |          |          |          |          |          |
| maxpool       |          |          |          |          |          |          |
|               | Conv3-128| Conv3-128| Conv3-128| Conv3-128| Conv3-128| Conv3-128|
| maxpool       |          |          |          |          |          |          |
|               | Conv3-256| Conv3-256| Conv3-256| Conv3-256| Conv3-256| Conv3-256|
| maxpool       |          |          |          |          |          |          |
|               | Conv3-512| Conv3-512| Conv3-512| Conv3-512| Conv3-512| Conv3-512|
| maxpool       |          |          |          |          |          |          |
|               | Conv3-512| Conv3-512| Conv3-512| Conv3-512| Conv3-512| Conv3-512|
| maxpool       |          |          |          |          |          |          |
|               | FC-4096  | FC-4096  | FC-1000  | Soft-max |          |          |

3. Active Learning Algorithm

The active learning method adopts information entropy and density-based spatial clustering of application with noise \(^{[11]}\).

The commonly used methods to select samples are margin sampling (MS), but most of these methods are suitable for shallow models. For more complicated convolutional neural network
structure and more data sets of category labels, it cannot make good use of sample information. Due to in the process of using sample information, the number of labeled samples is usually small, and the number of unlabeled samples is large, for the margin sampling method, MS only considers two maximum probability category labels, and does not consider other category labels, which is very unfavorable for the calculation of sample uncertainty, therefore, information entropy (IE) is selected as the sample selection method. The information entropy is used to calculate all probability distribution of each sample and make statistical measurement, then calculate the information quantity of the sample and judge its importance to the classifier.

The information entropy formula is as fellow:

$$H(x) = H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \ln p_i$$ \hspace{1cm} (1)

Where X represents any sample, $p_1, \ldots, p_n$ represents the probability value of the corresponding label of X. With the increase of information entropy, the uncertainty of samples is also increasing.

The principle of density-based spatial clustering of application with noise is to assume that categories can be determined by the compactness of sample distribution. There are samples of the same category in the area close to the samples, by dividing the closely connected samples into one category, the clustering category is obtained. By classifying all closely connected samples into different categories, the final classification results of cluster categories can be obtained.

Fig.1 The flow diagram of algorithm

The specific steps of the density-based spatial clustering of application with noise algorithm can be summarized as follows: The convolutional neural network is used to train a small number of labeled data sets, then the training parameters of the network are obtained, and the training network is saved; then the training network is used to calculate the probability value of each sample; the cluster analysis method based on density with noise is used to cluster the validation set, and the uncertainty of the validation set samples is calculated by using information entropy to select some samples with high uncertainty value; the clustering analysis method based on density with noise is used to analyze the clustering categories of the selected samples, and the samples with low uncertainty value in each
clustering cluster are removed; the remaining samples are the samples with high amount of information. These samples are added to the training set, and then they are trained by using the training network. The cycle stop condition is that the test accuracy meets the requirements.

4. Experimental Process and Results

4.1 Experimental Process

In this experiment, the high-resolution remote sensing data set NWPU-RESISC45 is used as the data set, VGG-16 pre training model is used as the networks, and the active learning method combining density-based spatial clustering of application with noise with information entropy is used.

![Fig.2 The sample example of NWPU-RESISC45](image)

The specific steps of the experiment, as shown below:

1. Select experimental data. The NWPU-RESISC45 data set includes 45 categories, ten categories of which are selected as experimental data. The ten categories are airplane, beach, chaparral, dense residential, forest, freeway, island, sea ice and terrace respectively. In this experiment, the total number of samples is 7000, the Initial training set is set to 1000 and the validation set is set to 7000.

2. CNN is used to train the training set and get the classifier parameters. In the experiment of this method, CNN network is selected as the vgg-16 network pre training model. The convolution network is set to carry out active learning and selecting samples every 25 epochs, and certain samples are selected for each active learning.

3. The role of density-based spatial clustering of application with noise is clustering the characteristics of samples. The density-based spatial clustering of application method by classifying all closely related samples into different categories, we can get the final results of all cluster categories and get the cluster division.

4. Using CNN classifier to estimate the labeled sample data set and get the prediction results.

5. For each sample, according to the prediction results, using information entropy \( H(X) = H(P_1, ..., P_n) = - \sum_{i=1}^{n} P_i \ln P_i \), where \( X \) represents any sample, \( P_1, ..., P_n \) represents the probability value of the corresponding label of \( X \). It can be seen from the formula that the larger the entropy value is, the higher the uncertainty of the sample is.) to calculate the uncertainty.

6. By sorting the prediction results in ascending order by the uncertainty value, the top samples are selected to get the new dataset.
(7) For each cluster obtained by cluster analysis method, in the same cluster, the samples with higher uncertainty are selected first, and the samples with higher uncertainty value in each cluster are put into the new data set, then in this way, we can get a sample set with high uncertainty and rich information.

(8) Add the new sample set to the training set as the new train dataset, repeat the above steps to cycle, and the cycle termination condition is that the test accuracy meets the requirements.

(9) In addition to method a, which is the main research method proposed in this paper, this paper also uses two other methods, method b and method c, respectively. Method b: VGG16 pre training model was used as training network, after active learning, a hundred randomly selected samples were added to the training set and then put into the network for training; method c: using LeNet network as training network, after active learning, the first 100 selected samples were added to the training set and then put into the network for training.

(10) Finally, the experimental results of the three methods are compared.

4.2 Experimental Results
The accuracy of different methods in different training samples are shown in the table below.

| The number of the samples | Method a     | Method b     | Method c     |
|--------------------------|--------------|--------------|--------------|
| 1100                     | 77.86%       | 77.60%       | 63.77%       |
| 1300                     | 78.63%       | 77.94%       | 68.20%       |
| 1500                     | 80.73%       | 78.41%       | 69.31%       |
| 1700                     | 81.56%       | 80.69%       | 70.40%       |
| 1900                     | 82.47%       | 81.54%       | 71.47%       |
| 2000                     | 83.11%       | 82.51%       | 72.58%       |

In order to analyze and understand the experimental results more intuitively, the experimental results are presented in the form of line graph, where the horizontal axis of the graph represents the number of labeled samples, and the vertical axis of the graph represents the test accuracy. In this figure, the black line represents the result of method a, the blue line represents the result of method b, and the red line represents the result of method c.
Fig. 3 The final experimental results

From the above table and figure, we can see that the effect of method a and method b is much better than that of method c, and the effect of method a is a little better than that of method b. In active learning, the samples processed by information entropy and clustering are processed in descending order. Therefore, the samples rich in information are relatively top-ranking, and the top-ranking samples are better, so the effect of method a is better than that of method b.

Final experiment demonstrated that this method can improve the test accuracy from 77% to 83%, and improve the uncertainty and representativeness of the sample.

5. Conclusion
In this paper, convolutional neural network is introduced, then the paper introduces the active learning method combining density-based spatial clustering of application with noise with information entropy. Experiments show that this active learning method can make full use of the sample information, improve the uncertainty and representativeness of the sample, and improve the test accuracy. Because the uncertainty of this method only considers the output value of the convolutional neural network in the last layer, the information of the sample can not be fully displayed, and this method needs further improvement.

Reference
[1] CUI Lu, ZHANG Peng, CHE Jin. Outline of Deep Neural Network Based Classification Algorithms for Remote Sensing Images[J].Computer Science, 2018,45(S1):50-53.
[2] JIN Fei, WANG Fan, LIU Zhi, et al. Extraction of residential areas from remote sensing images based on log-polar transformation[J]. Science of surveying and mapping, 2013, 38 (3) :102-105.
[3] Tax D M J,Duin R P W. Data domain description using support vectors[C]. Proceedings of European Symposium on Artificial Neural Networks.Bruges:[s.n.],1999:251-256.
[4] Ilonen, J., Paalanen, P., Kamarainen, J.-K., Kalviainen, H.. Gaussian mixture pdf in one-class classification: computing and utilizing confidence values[P]. Pattern Recognition, 2006. ICPR 2006. 18th International Conference on,2006.
[5] Xuefei Hu, Qihao Weng. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks[J]. Remote Sensing of Environment, 2009, 113(10).

[6] Sun Yi, Li Peijun. Using active learning to improve single class classification of remote sensing images: a case study of positive and unlabeled sample learning methods[J]. Journal of Beijing University (NATURAL SCIENCE EDITION): 1-10 [2019-12-09]. https://doi.org/10.13209/j.0479-8023.2019.035.

[7] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning[M]. MIT Press, 2016.

[8] ZHOU FeiYan, JIN LinPeng, DONG Jun. Review of Convolutional Work[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.

[9] Springenberg JT, Dosovitskiy A, Brox T, et al. Striving for simplicity: The all convolutional net[J]. arXiv preprint arXiv:1412.6806, 2014.

[10] Yu, Fisher, and Vladlen Koltun. “Multi-scale context aggregation by dilated convolutions.” arXiv preprint arXiv:1511.07122, 2015.

[11] BORAH B, BHATTACHARYYA D K. An improved sampling-based DBSCAN for large spatial databases[C]. International Conference on Intelligent Sensing & Information Processing, 2004.