Semi-supervised Classification Method for Remote Sensing scene Based on Deep Learning

Yaochang Han, Jie Wang, Li Lu
Air Force Engineering University, Xi’an, 710051, China
Yaochang Han’s e-mail: 271233758@qq.com

Abstract. Aiming at the classification problem of remote sensing scenes, this paper proposes a semi-supervised deep learning method combining the generative generation network with the convolutional neural network. The method improves the loss function of the deep convolutional generation network, and designs the even convolutional network EM-10 and its transposed convolution network as discriminators and generators respectively, which improves its ability to extract remote sensing features. The trained discriminator is used for the classification of remote sensing scenes with an accuracy of over 90%.

1. Introduction
In recent years, with the rapid development of remote sensing satellites and drone technology, remote sensing technology has been widely used in many fields [1]. High-efficiency and high-precision recognition of remote sensing image scenes has become an important issue at present. The traditional manual design based remote sensing scene classification method has limited description ability [2]; the machine learning based method is shallow learning and cannot adapt to the scene recognition task of complex remote sensing image [3].

Convolutional neural network based on the principle of vision, linear and non-linear transformation of image features, with self-learning ability of image features [4], but the classic convolutional neural network network framework, such as AlexNet [5], VGG [6], ResNet [7], etc., the number of layers Deeper and numerous parameters, it is easy to fall into over-fitting when training remote sensing images, which leads to the model not being able to fully exert its advantages.

The idea of generating a counter-neural network adopts the zero-sum game [8]. Through the continuous confrontation between the generator and the discriminator, the Nash equilibrium is achieved, and the false data with high consistency with the real data distribution is generated, which is a better unsupervised learning method. This paper proposes a supervised learning method that combines unsupervised learning of GAN with supervised learning of CNN. The unsupervised learning of remote sensing images is realized by training improved GAN. Based on the convolutional neural network discriminator of remote sensing image features, supervised remote sensing image classification is carried out, which makes full use of the characteristics of remote sensing image dataset and improves the characteristics. Deep learning ability to identify remote sensing scenes.

2. Unsupervised learning by Imaproed DCGAN

2.1. Imaproed DCGAN
The size of the key objects used to determine the scene category in the remote sensing scene picture is quite different, the various features are interlaced, and the probability distribution is complex.
Therefore, the discriminator structure that is originally symmetric with the generator is improved in three ways:

(1) Adding a multi-layer feature fusion link based on the convolution module, combining the features of the middle convolution layer with the high-level semantic information of the high-level convolution layer to enhance its reasoning ability. The result of the multi-layer feature fusion discriminator is shown in Fig3. The discriminator combines the 4th, 5th, and 6th-layer convolution feature maps, which are the transitions of image features from intermediate features to high-level semantic features, and the high-level semantic information. The two stages are combined to improve the readability of the semantic information of its generated data.

(2) The use of $4 \times 4$ convolution kernel: This improvement measure mainly refers to the theoretical analysis results of the literature [9]. The use of even convolution kernels instead of odd convolution kernels can improve the recognition accuracy of objects and effectively reduce the amount of parameters, but even convolution core has no center point and therefore causes a shift in the target point. For the classification task in this paper, it is not required to obtain the object in the image for positioning, but only the classification task.

(3) Introducing the feature matching error guiding generator to improve the feature matching degree between the generated data and the real data, that is, to narrow the probability distribution difference between the two. Assuming that $f(x)$, $f(G(z))$ the feature fusion layer of the discriminator activates the real data and the generated data (0 or 1), the following feature matching error representation can be used:

$$
g_{\text{loss}}_2 = \left\| E_{x \sim p_{data}} [f(x)] - E_{z \sim p_{z}} [f(G(z))] \right\|_2^2$$

(1)

After the correction, the $g_{\text{loss}}$ consists of $g_{\text{loss}}_1$ and $g_{\text{loss}}_2$. Since the optimization target of the generator loss is the maximum value, the feature matching item $g_{\text{loss}}_1$ can be inverted to match the optimization direction of the generator as a whole, and the loss function can be expressed as:

$$
\min_D \max_G V(D,G) = E_{x \sim p_{data}} [\log D(x)] + E_{z \sim p_z} [\log(1-D(G(z)))]$$

$$
\left\| E_{x \sim p_{data}} [f(x)] - E_{z \sim p_z} [f(G(z))] \right\|_2^2$$

(2)

2.2. Improved DCGAN effect

In order to verify the effectiveness of the improvement measures, the GAN and DCGAN [10] were compared and improved respectively. For the improved GAN method, the Adam optimization algorithm was used, and the learning rate attenuation factor was 0.9, the batch size was 64, and the GTX 1080Ti GPU was used. The partial diagram generated by the experiment is shown in Fig 1.

![Fig 1(a) Pictures in AID dataset](image1)

![Fig 1(b) DCGAN generated on the AID dataset](image2)

![Fig 1(c) Remote sensing image generated by improved GAN](image3)
Fig 1(b) is a partial result of DCGAN generated on the AID dataset. Only the underlying features such as texture and color are learned, and no reliable results are generated. As can be seen from Fig 1(c), the improved GAN generated image basically has the characteristics of the remote sensing scene, and can distinguish different types of scenes, such as beaches, residential areas, factories, and oil tanks and forests. The experimental results show that the generator fully learns the characteristics of remote sensing images and associates them with high-level semantic information.

The change in the loss value of DCGAN and the improved GAN when training on the AID data set is shown in Fig 2. The ordinates in Fig 2(a) correspond to the \(\log(D(x)) + \log(1 - D(G(z)))\) and \(\log(1 - D(G(z)))\) in the loss function of DCGAN, respectively. Ignoring the abnormal loss value in the initial stage of training, it can be seen from Fig 2(a) that during the training process, the loss value of the discriminator shows a downward trend as a whole, which is consistent with the minimization of the discriminant error in the objective function; the generator loss value is in a small range. After the rise, the fluctuation is basically maintained. The larger the value of \(\text{loss}_g\) is, the closer the distribution of the generated image to the real image is, that is, the performance of the generator is not significantly improved.

It can be seen from Fig 2(b) that as the loss of the improved GAN generator increases, it shows that its ability to simulate real data is constantly increasing, and its oscillating rising form fully reflects its confrontation. In Fig 2(b), the value of the discriminator finally oscillates around 0.5, indicating that the discriminator fully plays a counter-game role.

3. Supervised Learning Using the Trained Discriminator

Semi-supervised learning [11] is a combination of supervised learning and unsupervised learning. The confrontation process between the GAN generator and the discriminator in Section 2 can also be regarded as the unsupervised learning part of the convolutional neural network (discriminator) using the generated data and the real data. This section designs the supervised classification part of the entire semi-supervised classification.

In this section, we use the upper section to generate a parameter discriminator with better performance against the neural network, fix the convolution layer parameters, and change the structure after the convolution layer to facilitate the classification task. The overall structure of the 10-layer convolutional neural network EMNet-10 constructed for the subdivision of remote sensing image scene classification is shown in Fig 3.
4. Trails and analysis
This section uses the designed EMNet-10 convolutional neural network structure to experiment on three sets of remote sensing scene classification datasets of AID, UCM [12] and WHU-19 [13], and it is combined with the ImageNet data set. Comparative experiments were conducted with trained AlexNet, VGG19 and ResNet101. The number of iterations of the experiment is 50,000 times, the learning rate is Adam optimization algorithm, the attenuation factor is 0.9, the batch size is set to 100, and the experimental result format is “training precision/test accuracy”. The experimental results are shown in Table 1, and the left data is training accuracy, the right side of the data is the test accuracy.

According to the test results, the EMNet-10 has a 10-layer depth network structure, and its test accuracy and training accuracy significantly exceed AleNet, which is close to the 19-layer VGG and 101-layer ResNet classification accuracy. Among them, the training precision of VGG and ResNet are obviously higher than the test accuracy, which indicates that the classical deep convolutional neural network structure has different degrees of over-fitting in the remote sensing image scene classification task.

The experimental results show that the convolutional layer and feature fusion layer of EMNet-10 have fully learned the characteristics of remote sensing pictures through unsupervised and supervised learning. The structure is relatively simple and the generalization ability is strong. It is a suitable classification task for remote sensing scenes. Lightweight convolutional neural network structure.

Table 1 EMNet-10 and its comparison experiment results

|          | AID        | UCM        | WHU-19     |
|----------|------------|------------|------------|
| EMNet-10 | 93.2% / 91.4% | 91.6% / 89.5% | 90.2% / 91.1% |
| AlexNet  | 79.6% / 62.1% | 74.2% / 68.5% | 73.5% / 64.7% |
| VGG19    | 86.2% / 70.8% | 83.3% / 79.4% | 87.6% / 81.2% |
| ResNet101| 96.2% / 91.6% | 95.3% / 94.2% | 95.1% / 90.8% |

5. Conclusion
In this paper, a semi-supervised classification method combining GAN and CNN is proposed for the classification of remote sensing scenes. The features of the limited-scale remote sensing scene marker data are fully utilized. This paper creatively divides the semi-supervised learning process into two parts. Firstly, the DCGAN is improved. After training, the discriminator that learns the remote sensing features is obtained. Based on this, 10 layers of convolutional nerves for remote sensing scene classification are designed. Network structure of EMNet-10 was used for supervised learning and the effectiveness was verified by experiments.
References

[1] Bing Zhang. A Survey of Optical Remote Sensing Information Technology and Application[J]. Journal of Nanjing University of Information Science and Technology, 2018, 10(01): 1-5.

[2] SWAIN M J, BALLARD D H. Indexing via Color Histograms[M]//Active Perception and Robot Vision. Springer Berlin Heidelberg, 1992: 390-393.

[3] BELGIU M, DRAGUT L. Random forest in remote sensing: A review of applications and future directions[J]. Isprs Journal of Photogrammetry & Remote Sensing, 2016, 114: 24-31.

[4] F Y Zhou. A review of convolutional neural networks[J]. Journal of Computer. 2017, 40(06): 1229-1251.

[5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[8] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

[9] He K, Sun J. Convolutional neural networks at constrained time cost[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 5353-5360.

[10] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[11] E M Tu, J Yang. Overview of Semi-supervised Learning Theory and Its Research Progress. Journal of Shanghai Jiaotong University. 2018, 52(10): 1280-1291.

[12] http://weegee.vision.ucmerced.edu/datasets/landuse.html

[13] http://dsp.whu.edu.cn/cn/staff/yw/HRSscene.html