Recognition and location of typical scenes in large hyperspectral remote sensing image based on deep transfer learning

Tong Shi*, Jie Wang1, Pengfei Wang1, Qihang Cai1 and Yaochang Han1

1 Air and Missile Defense College, Air Force Engineering University, Xi’an 710038, China
* m18202715025@163.com

Abstract. The recognition and location of military scenes in hostile battlefield are of great strategic significance. Such scenes are the main targets of our long-range reconnaissance and direction strike. Deep transfer learning algorithm is always adopted to improve the accuracy of image recognition based on DCNN model. And on this basis, this paper mainly studied the application of deep transfer learning algorithm to recognize and locate typical scenes in large hyperspectral remote sensing image. Nichetargeting and impeccable DCNN model was accomplished after the training by typical scenes dataset. In the face of a large hyperspectral remote sensing image, the method of grid cutting, recognizing one by one and marking distinctively could pinpoint the location of typical scenes within. Experimental results showed that deep transfer learning algorithm could get a good application in the fast recognition and accurate location of typical scenes in large hyperspectral remote sensing image.

1. Introduction
In modernized and informationized warfare, some typical scenes in hostile battlefield are of great strategic significance, such as missile base, port, airport, oil storage, bridge and so on. Such scenes are the main targets of our long-range reconnaissance and directional strike. The fast recognition and accurate location of these typical scenes will help us draw more detailed military maps, carry out more comprehensive threat analysis and formulate more reasonable military strategies[1,2]. In view of the particularity of battlefield environment, long-range ground scene could be identified mainly by its hyperspectral remote sensing image.

With the development of remote sensing platform, sensor technology and communication technology, aerospace remote sensing technology has been greatly improved[3]. Data amount of remote sensing image is larger, update speed is faster, detection range is wider, and hidden information is more valuable. Remote sensing image data has been presented with typical 4V features of big-data -- Volume, Velocity, Variety and Value. However, in the face of remote sensing image big-data, traditional image processing technology has become inadequate, which is low in automation and intelligence[4].

In recent years, deep learning, reinforcement learning, transfer learning and other new intelligent algorithms have come to the fore. Deep learning is an algorithm of machine learning, attempting to use multi-layer neural network with complex structure to perform high-level abstraction of data features. And for the deep learning’s limitation of only focusing on features, reinforcement learning and transfer learning can provide reward and adaptation respectively.
These new intelligent algorithms have made remarkable achievements in the field of image recognition, speech recognition, big-data analysis, robot control and other fields. Furthermore, they have been rapidly extended to other smaller branches, and aroused profound changes in all trades and fields\(^\text{[5]}\), including the field of remote sensing image recognition.

2. DCNN

As a typical algorithm of deep learning, convolution neural network (CNN) is particularly pronounced in the direction of image recognition and classification\(^\text{[6]}\). With the number of CNN layers increasing and the structure of CNN being more complex, deep convolutional neural network (DCNN) arises. DCNN avoids complex image preprocessing, and extracts features directly from original image by the way of local connection and weight sharing. Feature representation is more and more abstract and specific from the shallow layer to the deep layer. The final advanced feature can be classified well by a simple classifier.

The basic network structures of DCNN include:

- Convolution layer, also known as feature extraction layer. Each neuron of this layer is connected with the receptive field of previous level, and the features of receptive field are extracted by filter and nonlinear transformation.
- Pooling layer, also known as feature mapping layer. Features extracted in convolution layer is likely to be redundant, and the feature vector can be reduced by sub-sampling to increase model's resistance of distortion at the same time.
- Full connection layer. There is generally at least one full connection layer in a DCNN model, and the neurons in this layer are all connected with all input units of previous level to synthesize the features extracted before for classification.

With the continuous improvement of computer performance, especially the development of GPU and TPU, and the support of big-data, DCNN grows rapidly. In recent years, many excellent DCNN models have been proposed in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)\(^\text{[7]}\). The accuracy record has been constantly refreshed. Table 1 shows the comparison of several network models.

| Name     | AlexNet | VGGNet | GoogLeNet | ResNet |
|----------|---------|--------|-----------|--------|
| Year     | 2012    | 2014   | 2014      | 2015   |
| Number of layers | 8       | 19     | 22        | 152    |
| Number of convolution layers | 5       | 16     | 21        | 151    |
| Convolution kernels size | 11,5,3  | 3      | 7,1,3,5   | 7,1,3,5|
| Number of full connection layers | 3       | 3      | 1         | 1      |
| Full connection layers size | 4096,4096,1000 | 4096,4096,1000 | 1000 | 1000 |
| TOP-5 error rate | 16.4%   | 7.3%   | 6.7%      | 3.57%  |
| Inception | No      | No     | Yes       | No     |

3. Deep transfer learning

Transfer learning is to use the knowledge learned from an environment to help the learning task in a new environment, which can be defined as: given a source domain \(D_s\) and a learning task \(T_s\), a target domain \(D_t\) and a learning task \(T_t\), where \(D_s \neq D_t\), \(T_s \neq T_t\), using existing knowledge in \(D_s\) and \(T_s\) to help improve the target predictive function \(f_t(\cdot)\) in the target domain \(D_t\).

In the field of image recognition, DCNN models in Table 1 have achieved unprecedented success using Imagenet natural color image dataset, not only because network model is being optimized, but also because millions of images provide data support for network training. But there are problems when transferring above DCNN models to a new specific filed without big-data, because a DCNN model, directly trained by a small image dataset, is easy to cause a phenomenon of overfitting.
Learning from the idea of transfer learning, some scholars have put forward the deep transfer learning algorithm based on DCNN[8].

A DCNN model, trained by Imagenet natural color image dataset, has amounts of network parameters, whose shallow layers already have feature extraction function. When this trained DCNN model applied in a new recognition task, network parameters of shallow layers could be fixed, and parameters of deep layers should be randomly initialized. Then the small image dataset is used for network training with a rapid learning rate to realize the fine tuning of the whole network. In addition, appropriate network optimization methods are necessary, for example, adjusting some layers’ settings based on input image size and the number of categories, or replacing full connection layer with better classification methods after advanced feature is extracted. Finally a nichetargeting and impeccable DCNN model is completed for the specific recognition task.

However, transfer learning is not omnipotent. It is required that target domain dataset is close enough to source domain dataset. For example, natural color image and remote sensing color image can transfer between each other, because their imaging principle and image feature are similar, and there are few color differences between photo and object. Images like infrared image, grayscale image, SAR image are not suitable for transfer learning. Figure 1 displays different types of images[9].

Figure 1. Different types of images.

4. Recognition and location
For a large hyperspectral remote sensing image, there are many different kinds of scenes within. Some are of great military significance, like port, airport, oil storage, dam and bridge, while others are not, like forest, playground, residential area, pond and park.

At present, most of research on the recognition and location of remote sensing target is locating before recognizing, which means that it first find rough target area through image detection, then identify the target[10,11]. Different from the above method, the means of our method are recognition and classification, and the purposes of our method are mark and location.

Firstly, in the face of a large hyperspectral remote sensing image, proper grid cutting strategy is adopted to cut original large image into numerous small images with numbers in order. Secondly, these small images are identified and classified by the nichetargeting DCNN model one by one. Thirdly, classification results of each category are fed back to the original image, where the positions of different typical scenes are marked with different colors. Finally, if we know the longitude and latitude of a zero point in original image, typical scenes could be pinpointed.

4.1. Recognition
In order to achieve a very precise positioning result, we must ensure high accuracy of recognition. That is to say, we need to train a nichetargeting and impeccable DCNN model with remote sensing typical scenes image dataset based on deep transfer learning.

The new remote sensing image dataset is made up of 1000 port images, 1000 airport images, 1000 oil storage images, 1000 dam images, 1000 bridge images, and 2000 other scenes images including playground images, residential images, park images and so on. Figure 2 shows the categories contained in the new dataset.
AlexNet, VGGNet, GoogLeNet and ResNet are selected for experiments. The full connection layers are replaced by support vector machine (SVM) to improve classification accuracy\cite{12,13}. The performance is defined by the accuracy rate of DCNN model during testing. A correct answer means that the network classifies an image as the class it belongs to. Table 2 shows the testing results.

Table 2. Testing results.

| Category     | Network  | AlexNet | VGGNet | GoogLeNet | ResNet |
|--------------|----------|---------|--------|-----------|--------|
| port         |          | 0.76    | 0.83   | 0.91      | 0.90   |
| airport      |          | 0.78    | 0.86   | 0.92      | 0.93   |
| oil storage  |          | 0.81    | 0.87   | 0.95      | 0.95   |
| dam          |          | 0.75    | 0.82   | 0.89      | 0.90   |
| bridge       |          | 0.79    | 0.85   | 0.94      | 0.93   |
| others       |          | 0.75    | 0.83   | 0.91      | 0.92   |

According to the comprehensive performance, AlexNet < VGGNet < GoogLeNet ≈ ResNet. So on the basis of laboratory equipment, the selected network is GoogLeNet, which is quite niche-targeting and impeccable enough to meet our needs for typical scenes recognition after training.

4.2. Location

In the course of location experiment, the selection of large hyperspectral remote sensing image has a slight effect on the experimental results. Because one image may not include all the typical scene types, it is necessary to collect multiple images for every kind of scene to conduct location experiments more than once.

We take a large hyperspectral remote sensing image of part of Shenzhen China as an example. The image is 631 MB, whose size is $15301 \times 14421$. Only one airport is within the image, so as shown in Fig. 3, here is a brief overview of the process from the aspect of airport’s marking and location:

- Grid cut with $256 \times 256$ size for the whole picture without incomplete edge, and get 3304 small figures with numbers in order, so that the position of each small figure in the large one is determined.
• Identify and classify all the 3304 small pictures with the nichetargeting DCNN model one by one, and record the classification results of airport target[14].

• In the corresponding position of a digital image with the same proportion as the original image, mark the airport scene in red, mark the other scene in black.

• If we know the longitude and latitude of a zero point in original image, adopt central point method to pinpoint the location of the airport area.

![Figure 3. Process of the mark and location of airport.](image)

The above is only the location of airport, and other scenes could be located in the same way. After many related experiments, it has been proved that our method can achieve good results. This method can provide a basis for long-range reconnaissance and directional strike. The fast recognition and accurate location of these typical scenes will play a decisive role in grasping military opportunity, analyzing military situation and making military tactics.

5. Conclusion
This paper mainly studied the application of deep transfer learning algorithm to recognize and locate typical scenes in large hyperspectral remote sensing image. Nichetargeting and impeccable DCNN model was accomplished after the training by typical scenes dataset. In the face of a large hyperspectral remote sensing image, the method of grid cutting, recognizing one by one and marking distinctively could pinpoint the location of typical scenes within. Experimental results showed that deep transfer learning algorithm could get a good application in the fast recognition and accurate location of typical scenes in large hyperspectral remote sensing image.

In order to make this method more perfect, some optimizations are still needed to be done later. In particular, the means of our method are recognition and classification, and the purposes of our method are mark and location, which means that the accuracy of location heavily depends on the accuracy of recognition. So the optimization of DCNN model is of great importance. Better network models and
more optimized learning algorithms will bring higher accuracy of recognition. Moreover, the choice of grid cutting strategy and image preprocessing method in the course of location should be also studied in depth.

References

[1] Cheng D. PLA Views on Informationized Warfare, Information Warfare and Information Operations[M]// Chinese Cybersecurity and Defense. John Wiley & Sons, Inc. 2014:55-80.
[2] Metz S. Armed Conflict in the 21st Century: The Information Revolution and Post-Modern Warfare[J]. Carlisle Barracks U.s.army War College Strategic Studies Institute, 2000(1).
[3] Yang Z H, Han J F, Gong D P, et al. The Development and Application of Hyperspectral-Remote-Sensing Technology[J]. Hydrographic Surveying & Charting, 2003.
[4] Lurie J B. Remote Sensing Technology and Applications[J]. Optical Engineering, 2002, 41(9):2075.
[5] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.
[6] Shi X, Fang X, Zhang D, et al. Image classification based on mixed deep learning model transfer learning[J]. Journal of System Simulation, 2016.
[7] Ballester P, Araujo R M. On the performance of GoogLeNet and AlexNet applied to sketches[C]// Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, 2016:1124-1128.
[8] Yang Q. When Deep Learning Meets Transfer Learning[C]// ACM on Conference on Information and Knowledge Management. ACM, 2017:5-5.
[9] Galanti T, Wolf L, Hazan T. A theoretical framework for deep transfer learning[J]. Information & Inference A Journal of the Ima, 2016(2):iaw008.
[10] Shang Q, Qiuci R, Xiaoli L I. Target recognition and location based on binocular stereo vision[J]. Ciai Transactions on Intelligent Systems, 2011.
[11] Luo X, Luo Q, Han B, et al. Special Target Recognition and Location Using Differential Image Detection Technology[C]// Second International Conference on Intelligent Human-Machine Systems and Cybernetics. IEEE Computer Society, 2010:116-119.
[12] Yu J, Weng K, Liang G, et al. A vision-based robotic grasping system using deep learning for 3D object recognition and pose estimation[C]// IEEE International Conference on Robotics and Biomimetics. IEEE, 2014:1175-1180.
[13] Zhang L, Zhang L, Du B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art[J]. IEEE Geoscience & Remote Sensing Magazine, 2016, 4(2):22-40.
[14] Qu J, Sun X, Gao X. Remote sensing image target recognition based on CNN[J]. Foreign Electronic Measurement Technology, 2016.