Optik 168 (2018) 127–133

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Scene classification of remote sensing image based on deep network grading transferring

Zhou Yang*, Xiao-dong Mu, Feng-an Zhao

Xi'an High Technology Research Institute, Xi'an, Shaanxi 710025, China

ARTICLE INFO

Article history:
Received 15 March 2018
Accepted 12 April 2018

Keywords:
Remote sensing image
Scene classification
Deep convolutional neural network
Feature fusion
Multi-kernel support vector machine

ABSTRACT

Aiming at low precision of remote sensing image scene classification, a classification method DCNN_GT based on deep convolutional neural network (DCNN) and grading transfer (GT) is proposed. First, the DCNN model pre-trained on a large dataset (such as ImageNet) was transferred to a relatively smaller target dataset to fine-tune and the first-level classification features of images are extracted. Then, the similar images in the target dataset are clustered into several high-similarity categories, and the DCNN model of the first-level transfer is fine-tuned again on them and the second- level classification features are extracted. Then, the two-level classification features are encoded and fused, and the multi-kernel support vector machine (MKSVM) is used for scene classification. The experimental results in the common remote sensing datasets show that the average classification accuracy of the proposed method is improved by at least 1.83% compared with the one-time transfer and fine-tuning, especially the classification of confusable images is increased by at least 4%. In this paper, the DCNN is transferred gradually and is advanced to enhance the representation ability of the extracted image features, which makes the fusion features more recognizable. At the same time, the MKSVM is used to improve the generalization ability of the fusion features, so the classification result is better.

1. Introduction

Scene classification of remote sensing images plays an important role in many fields, such as land management, urban planning, environmental exploration and monitoring, and natural disaster detection. Over the past decades, researchers have done a great deal of experiments in the scene classification for satellites and aerial photographs, and have developed many taxonomies [1–8,32,33]. However, with the continuous improvement of science and technology, the spatial resolution of remote sensing images is getting higher and higher, and their spatial and structural patterns are becoming more and more abundant. The phenomenon of the same objects with different spectrum and the foreign ones with common spectrum are more widespread. However, the most of the classical methods are based on artificial or shallow learning algorithms, and the low middle-level semantic features extracted are limited in the description ability, which makes it difficult to improve the classification accuracy further.

In recent years, the method of deep learning, as the most advanced technique in computer vision recognition [9,10] has been successfully applied to many recognition problems and obtained great improvements over the past, such as object detection [11], face and speech recognition [12,13], behavior recognition [14], semantic segmentation [15], natural image

* Corresponding author.
E-mail addresses: yzmailbox2015@163.com, 114616802@qq.com (Z. Yang).

https://doi.org/10.1016/j.ijleo.2018.04.092
0030-4026/

classification [16] and remote sensing scene classification [6,17–19], etc. Deep Convolutional Neural Network (DCNN), as a prominent branch of deep learning, needs a large dataset to train its model parameters. And the dataset size is one of the determinants affecting its performance. Through the idea of Transfer Learning, a DCNN model pre-trained on a large dataset (such as ImageNet) can be applied to the remote sensing image classification with a relatively small target dataset. And many recent papers [20–23] demonstrated it, and made good classification results.

However, the classification accuracies of the confusion images are not high with the above method. And inspired by the idea of transfer, we use the hierarchical transfer method. That is, after a transfer and adjustment of the deep network model, the confusing images are clustered into a new type of dataset, then the deep network model is transferred and fine-tuned again on it to obtain more specific classification features than the one-time transfer method, thus it can improve the classification accuracies of the confusable images.

2. Remote sensing image classification based on DCNN_GT

By clustering images that are easy to be confused and using two-stage transfer deep network models to extract image features and then classifying them by multi-kernel support vector machines, both common and special features can be taken into account, which not only enhances the represent ability of the classification features, but also can't result in a significant increase in time complexity and error accumulation.

2.1. One grading transferring and fine-tuning deep network model

Using the method in Sharif et al. [20]; Oquab et al. [21]; Donahue et al. [22]; Zeiler et al. [23], the DCNN model is pre-trained on a large dataset, and then it is fine-tuned in the target remote sensing dataset and the first-level eigenvector of the image is extracted, which focuses more on Common features of the image.

2.2. Two grading transferring and fine-tuning deep network model

After extracting the first level eigenvectors of the images, in order to enhance the represent ability of the classification features, the high similar images are clustered into p categories, then the fine-tuned DCNN model is transferred again and the second-level eigenvectors of the images are extracted which focus more on the particularity. In this paper, the method of spectral clustering is used to cluster the similar images. Firstly, the confusion matrix of remote sensing images is obtained by transferring and fine-tuning the deep network for the first time. Then the similarity matrix between each image category is obtained by using Formula (1), and categories are clustered with spectral clustering.

$$D = \frac{1}{2}\left[(1-C)+(1-C)^T\right] \tag{1}$$

2.3. Multi-kernel support vector machine

After fusing the features extracted from the two-stage transferred and fine-tuned deep network, MKSVM classifier [24] is used to replace Softmax layer of DCNN to improve the accuracy of image classification. In this paper, the k kernel function [25] is used as the base function of multi-kernel support vector machine because it avoids the complex exponential operation of radial basis function (RBF), and has the advantages of less computation of polynomial kernel function and high approximation accuracy and good generalization ability of RBF, and it has superior performance. The k kernel function and the multi-kernel function are shown in Eqs. (2) and (3), respectively:

$$k(x_i, x_j) = \prod_{t}^{n} \frac{1}{1 + l^2 \left(x_i^{(t)} - x_j^{(t)}\right)^2} \tag{2}$$

$$\left.\begin{aligned} K(x_i, x_j) &= \sum_{m=1}^{M} d_m K_m(x_i - x_j) \\ \sum_{m=1}^{M} d_m &= 1 \quad d_m \geq 0 \end{aligned}\right\} \tag{3}$$

In Eq. (2), the width of the k kernel function is expressed with $l > 0$. In Formula (3), M is the number of the base kernel functions, and K_m is the base kernel function that is the k kernel function represented by the Formula (2). The optimization of MKSVM can be performed by two steps of alternative optimization, that is, the kernel function weights are fixed firstly, and the basic SVM problem is solved, then the hyperplane weights are fixed, and the objective function with d_m is solved by the gradient descent method.