ALEXNET FEATURE EXTRACTION AND MULTI-KERNEL LEARNING FOR OBJECT-ORIENTED CLASSIFICATION

Ling Ding [1], Hongyi Li [2,]*, Changmiao Hu [2], Wei Zhang [2], Shumin Wang [1]

[1]Institute of Earthquake Forecasting, China Earthquake Administration, Beijing, China
[2]Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China - lihy_2003@163.com

Commission III , WG III/1

KEY WORDS: AlexNet, GLCM texture, Multi-kernel learning, Object-oriented classification, feature extraction ,SVM

ABSTRACT:

In view of the fact that the deep convolutional neural network has stronger ability of feature learning and feature expression, an exploratory research is done on feature extraction and classification for high resolution remote sensing images. Taking the Google image with 0.3 meter spatial resolution in Ludian area of Yunnan Province as an example, the image segmentation object was taken as the basic unit, and the pre-trained AlexNet deep convolution neural network model was used for feature extraction. And the spectral features,AlexNet features and GLCM texture features are combined with multi-kernel learning and SVM classifier, finally the classification results were compared and analyzed. The results show that the deep convolution neural network can extract more accurate remote sensing image features, and significantly improve the overall accuracy of classification, and provide a reference value for earthquake disaster investigation and remote sensing disaster evaluation**.**

1. INTRODUCTION

High resolution satellite imagery is an image to obtain information of object detailed on the earth’s surface.High resolution image can be used to classify land use from the process of visual interpretation and classification image(Chen etal.,2015).Visual interpretation and classification image of high resolution image can produce good classification accuracy(Gong etal.,2013;Chen etal.,2007;Chen etal.,2014). Visual interpretation and classification image have the disadvantage related to time and energy efficiency involved in satellite image classification so that classification can be done automatically to make the process of image classification faster.With the increase of scales and objects of classification, the computational complexity of object-oriented image processing method is increasing rapidly, and the classification accuracy has decreased. Recently， remote sensing image classification is mainly based on overlay spectral feature classification, researchers have proposed adding texture feature classification, although the classification accuracy is improved, but the texture information is still very limited to improve the classification accuracy (Chen etal.,2007;Chen etal.,2014),

In recent years, Convolutional Neural Networks (CNN) has made a series of breakthroughs in image classification, target detection, semantic segmentation and face recognition. The convolution neural network combines feature extraction and classification as a whole.The local connection weights, sharing and pooling operation and other characteristics can effectively reduce the number of training parameters, reduce the complexity of the network and make the model invariant to image translation, zoom, with a certain degree of distortion, and has strong robustness and fault tolerance(Zhou etal.,2017),Compared with the machine learning method, it has more powerful ability of feature learning and feature expression (Lu etal.,2016).The convolution neural network has achieved successful application in high-resolution remote sensing image scene recognition(Hu etal.,2015;Zhong etal.,2016; Marmanis etal.,2016).those experiments have proved the powerful feature extraction capability of convolution neural networks.Remote sensing image recognition is very similar to land use classification. It is necessary to build classification and recognition related to scene semantics, which are very different from land cover classification.The classification of landsat cover is mainly based on the spectrum, texture, and so on.The main difficulties are the difficulty of the characteristic expression caused by the diversity of the internal spectrum and the problem of the characteristic expression of the mixed spectrum is caused by the mixed pixel.Can the powerful feature extraction capability of the convolution neural network be used to improve the accuracy of surface coverage classification for medium/high resolution remote sensing images? The related research on the improvement of the classification accuracy for high resolution remote sensing images is very few at present.so, the convolution neural network is used to study the feature extraction and classification of high resolution remote sensing images.

This paper studies the convolutional neural network for feature extraction and classification of high resolution remote sensing image, taking Ludian post-earthquake Google image with 0.3 meter spatial resolution of images as experimental data, the object of the segmentation image is the basic unit, AlexNet convolution neural network(Krizhevsky etal.,2012) is used for deep feature extraction, which respectively combines with spectral feature and GLCM texture. and then, the multi- kernel learning is used for fusion feature,SVM classifier is used for classification, The experimental results show that the deep feature can extract more accurate object features and get higher classification accuracy.it also shows the shortage of AlexNet in

* Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-277-2018 | © Authors 2018. CC BY 4.0 License.

the classification of highlighting and lowlighting feature extraction.

2. ALEXNET FEATURE EXTRACTION AND MULTI-KERNEL LEARNING CLASSIFICATION

The classification process is shown Figure 1 based on the convolution neural network AlexNet and SVM, It mainly consists of three steps：

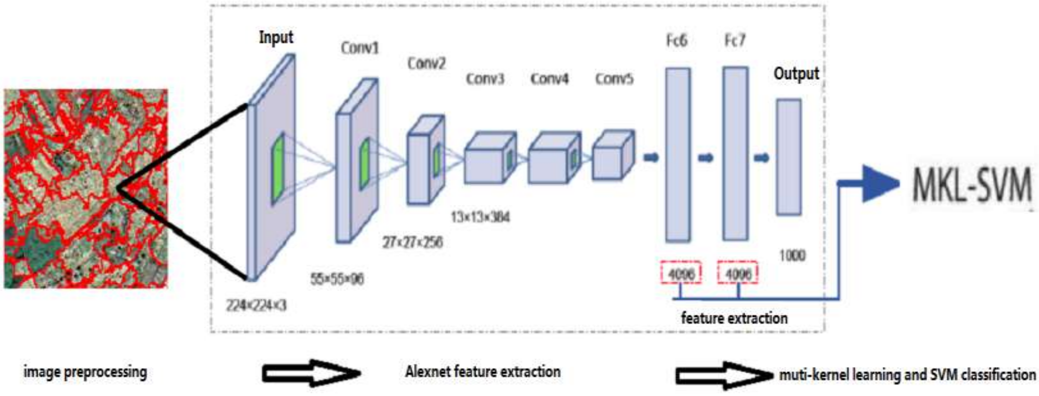


Figure 1. The flow chart of classification with deep features extraction by deep CNN

1) Image preprocessing: since the pre-trained AlexNet network model on the ImageNet data is as a feature extractor, the objects of remote sensing image segmentation are as the input of AlexNet model.

2) Deep feature extraction: the objects of image segmentation are as the feature extraction unit. According to the AlexNet input requirements, each object is normalized and neighborhood interpolation sampling(Krizhevsky etal.,2012). In this paper. The feature extraction is completed on the MatConvNet(Vedaldi etal.,2015)platform.

3) Multi-kernel learning SVM classification: Different image features correspond to a kernel function. These kernel functions form a unified kernel function through a combination of weights and then, the kernel function is used to classify. In this paper, the selected kernel function is linear kernel function, and the optimal solution is selected by using the most commonly used grid selection method. It is used to fuse spectral features, deep features, and GLCM texture features, and to extend to object-t oriented classification. the compound kernel method is used to fuse two features (spectral features and deep features), and its weighted kernel method is very attractive, because the method balances the spatial information and spectral characteristics. The weighted core is in(1):

$$K\,(x_i,\ x_j)=\mu K\,(x_i^s,\ x_j^s)+\,(1-\mu)\,K_t(x_i^t, x_j^t) \qquad (1)$$

where μ is a positive parameter that needs to be regulated in the training process. A pixel X_i is redefined by using spectral features in the spectral feature domain $x_i^s \in R^{N_s}$,and redefined in spatial feature domain using spatial features, $x_i^t \in R^{N_t}$, N_s and N_t relatively represent the number of bands of the vectors of spectral and spatial characteristics, respectively. K_s and K_t are the kernel matrices that describe their spectral information and spatial information. K is the overall kernel matrix. In the original paper, it is used for pixel based classification. In this study, the complex kernel method is used to fuse spectral features and complex deep features, and is extended to the object-oriented classification method.

3. RESULTS AND ANALYSIS

The experimental data are the image of Google post-earthquake in Ludian, Yunnan. The date of acquisition is August 20, 2017, the spatial resolution is 0.3 meter. The image size is 2400*2400. After segmentation, there are 1085 objects (see Figure 2). The categories include woodlands, intact buildings, cultivated land, bare land and collapsed buildings. The training sample and the test sample are randomly selected.

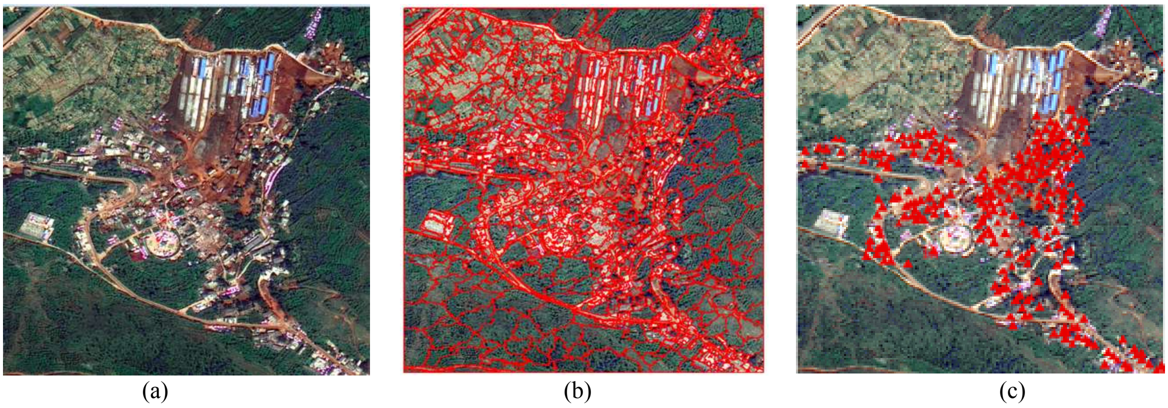


Figure 2. Images of the experimental area: (a) original image;(b)segmented map;(c) sample distribution

The main experiments of this paper are as follows:

1) The effects of different layers features on the classification results;

In order to analyze which feature of the AlexNet has more expressive ability, the last two fully connected layers Fc6, Fc7 and all convolution layers are extracted.in Figure 3, it can be seen that the classification accuracy is on the rise with the increase of the depth of the number of layers.The classification accuracy of the fully connected layer is higher than that of convolution layers.This is because the characteristics of the deep layers are more abstractive and more expressive.But the classification accuracy of Fc6 is higher than Fc7, this is because the pre-trained AlexNet is obtained in ImageNet natural image, while the semantic features of Fc7 is stronger, but it is more in line with the classification attribute of the training set, so the classification accuracy decreases.Therefore, for remote sensing images on the non-training set, the Fc6 is more expressive.

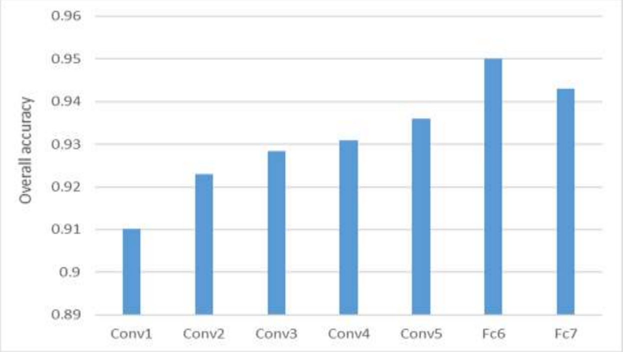


Figure 3. The classification accuracy comparison chart with deep feature layers

2) The influence of the size of different neighborhood windows on the classification results.

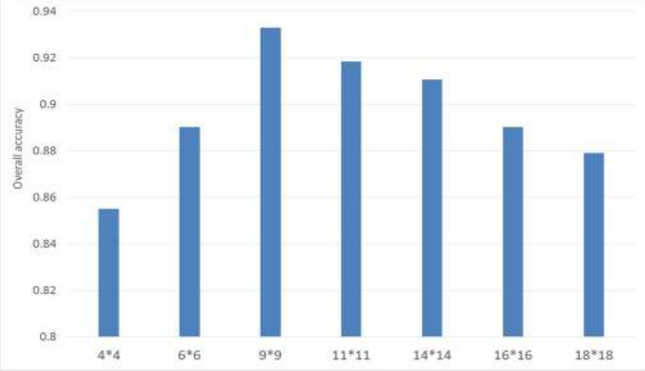


Figure 4. Classification accuracy chart with different window sizes

In order to analyze the effect of window size on the classification results,We select the optimal window size to extract Fc6.We has selected 4 * 4, 6 * 6, 9 * 9, 11 * 11, 14 * 14, 16 16 and 18 * 18 window sizes to analyze the classification results.Figure 4 is a comparison chart of the classification results. As can be seen from Figure 4, as the size of the window increases, the accuracy increases, the maximum value is reached at 9 * 9, and the precision will decrease when the size is increased.This is because the size of the window is small, the neighborhood information is too little, it can not extract the features of the objects very well, so the precision is low.while the size of the window is large, the information contained is too much, and there are many redundant information that will affect the classification results, instead, the classification accuracy is reduced.

3) Comparison and analysis of experimental results

The optimal classification accuracy can be obtained when the size of the neighborhood window is 9 * 9, and the Fc6 features are extracted. In order to analyze the effectiveness of the proposed method, the same training samples and testing samples were used to compare 3 methods of the experiments.

(1) Method 1:Multi-kernel learning and SVM classification is carried out using spectral features and 8-dimensional GLCM texture features.

(2) Method 2:Multi-kernel learning and SVM classification is carried out using deep features and spectral features.

(3) Method 3: SVM classification is carried out using deep features.

In the quantitative evaluation of classification accuracy (see Figure 5), it can be shown that the overall classification accuracy fused the deep features is higher than that of the other methods.In particular, the deep feature added to the spectral information makes the classification precision highest.

In Figure 4,there are a large number of misclassification of woodlands into cultivated land in Method 1.Compared to Method 1, the overall accuracy of the classification is improved obviously by Method 3,it uses the deep features so the small objects are relatively few(see Figure 6).Method 2 the classification results are the best, and the misclassification phenomenon is very few(see Figure 6). It can keep the continuity of the objects in the classification result, and can save the post-processing operation of the classification.

But careful observation of the details of the classification results in Method 2 (see Figure 6) can be found that the range of collapsed buildings expanded than actual collapsed buildings. Intact buildings is less than actual buildings. In summary, Method 2 makes the extraction area increase or shrink by making highbrighting and lowbrighting objects (such as intact buildings and collapsed buildings). in Table 1, The misclassification and leakage of the bare land and collapsed buildings are relatively serious in Method 2.it may be due to the application of ReLU nonlinear excitation function and maximum pooling function in the AlexNet model, making the highbrighting and lowbrighting objects (such as intact buildings and bare land) more erroneous.

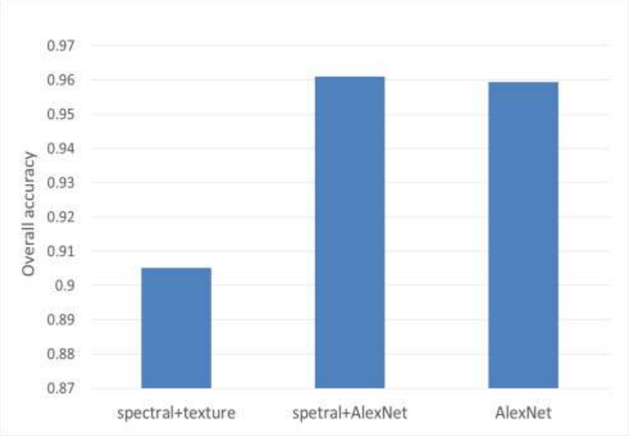
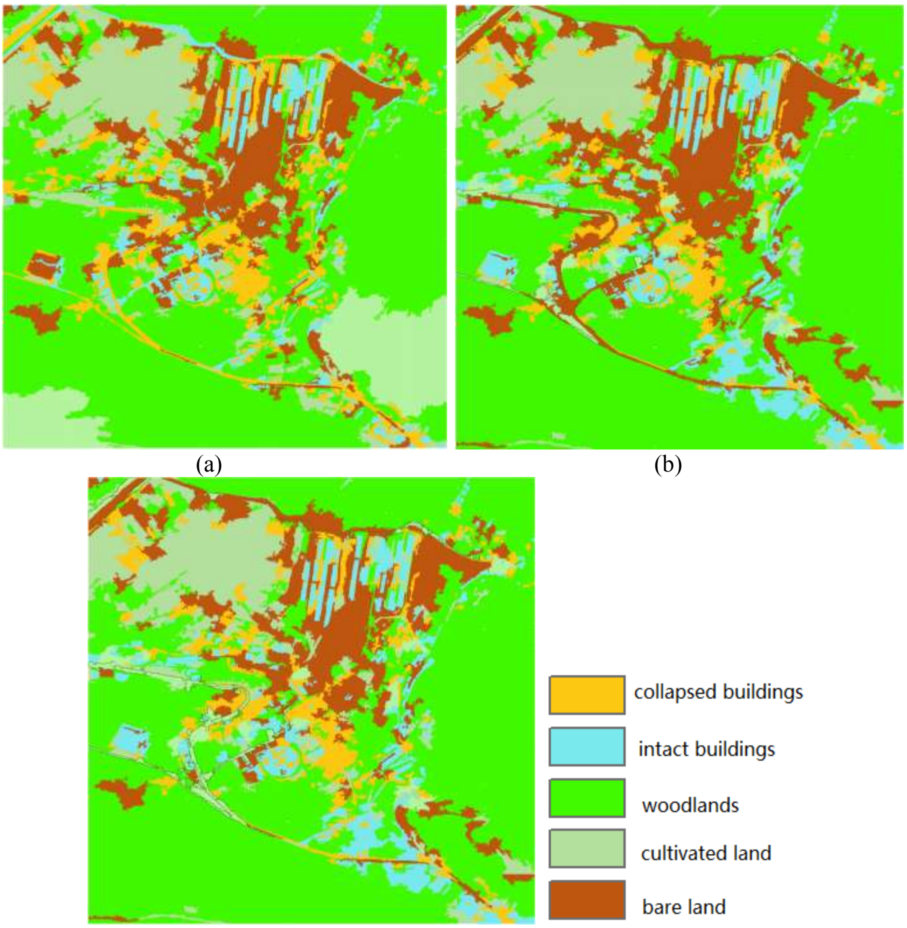


Figure 5. The accuracy comparison chart by different methods



(a) (b)

(c)

Figure 6. Classifition image obtained by different methods (a) Spectral-texture features; (b)Spectral-AlexNet features; (c) AlexNet features

	collapsed buildings	collapsed buildings	collapsed buildings	collapsed buildings	collapsed buildings
collapsed buildings	95.86	0.54	0.00	0.58	3.02
intact buildings	0.42	98.68	0.26	0.00	0.65
woodlands	0.00	0.00	94.65	2.37	0.61
cultivated land	0.31	0.98	2.91	94.91	0.89
bare land	2.65	0.00	0.00	0.43	96.92

Table 4. Confusion matrices obtain by Spectral-AlexNet features by Method 2

4. CONCLUSION

This paper makes use of AlexNet features for multi-kernel learning and classification. The features of fully connected layer of AlexNet are more expressive than the features of the convolutional layer. With the increase of the window size, the overall classification accuracy is increased first and then reduced, the appropriate window size is chosen for feature extraction; Compared to spectral features and texture features, deep convolutional neural network is more expressive, and can obviously improve the classification accuracy. Because ReLU is used as a nonlinear excitation function in the AlexNet model, and the maximum value is used to pool, the phenomenon of misclassification and leakage will occur for highbrighting and lowbrighting objects. In the future work, AlexNet model will be transformed to further improve the classification accuracy by selecting the appropriate nonlinear excitation function, pool operation and characteristic layer number

ACKNOWLEDGEMENTS

This work was supported in part by the National High Technology Research and Development Program of China under Grant 2013AA12A301, and in part by the Fundamental Research Funds of the Institute of Earthquake Forecasting,China Earthquake Administration under Grant 2015IES0203.

REFERENCES

Chen J, Chen J, Liao A P, et al.,2015.Global land cover mapping at 30m resolution: a pok-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing, 103,pp.7-27

Gong P, Wang J, Yu L, et al.,2013. Finer resolution observation and monitoring of global land cover: first mapping results with

Landsat TM and ETM+ data. International Journal of Remote Sensing, 34(7), pp. 2607-2654.

Chen B, Zhang Y J, Chen L.,2007.RS image classification based on SVM method with texture. Engineering of Surveying and Mapping, 16(5),pp.23-27.

Chen G F, Zeng G W, Chen H, et al.,2014. Study of RS image classification method based on texture features and neural network algorithm. Journal of Chinese Agricultural Mechanization, 35(1),pp.270-274.

Zhou F Y, Jin L P, Dong J.,2017.Review of convolutional neural network. Chinese Journal of Computers, 40(7),pp.1-23.

Lu H T, Zhang Q C.,2016. Application of deep convolutional neural network in computer vision. Journal of Data Acquisition and Processing, 31(1),pp.1-17.

Hu F, Xia G S, Hu J W, et al.,2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing. 7(11),pp. 14680-14707.

Zhong Y F, Fei F, Zhang L P.,2016. Large patch convolutional neural networks for the scene classification of high spatialresolution imagery. Journal of Applied Remote Sensing. 10(2),pp. 025006(1-20).

Marmanis D, Datcu M, Esch T, et al.,2016. Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geoscience and Remote Sensing Letter, 13(1): pp.105-109.

Krizhevsky A, Sutskever I, Hinton G E.,2012. ImageNet classification with deep convolutional neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe, USA: NIPS, pp.1106-1114.

Vedaldi A, Lenc K.,2015. MatConvNet-Convolutional Neural Networks for MATLAB[C]//Proceeding of the 23rd ACM International Conference on Multimedia. Brisbane, Australia:ACM,pp. 689-692.