Land-use Scene Classification in High-Resolution Remote Sensing Images by Concentric Circle Pooling Networks

Kunlun Qi 1, Feifei Peng2, Qingfeng Guan 1, \*

1 Faculty of Information Engineering, China University of Geosciences, Wuhan, China

2 College of Urban and Environmental Sciences, Central China Normal University

\*Corresponding author, e-mail: [guanqf@cug.edu.cn](mailto:guanqf@cug.edu.cn)

Keywords: deep learning, rotation-invariation, scene classification, convolutional neural network

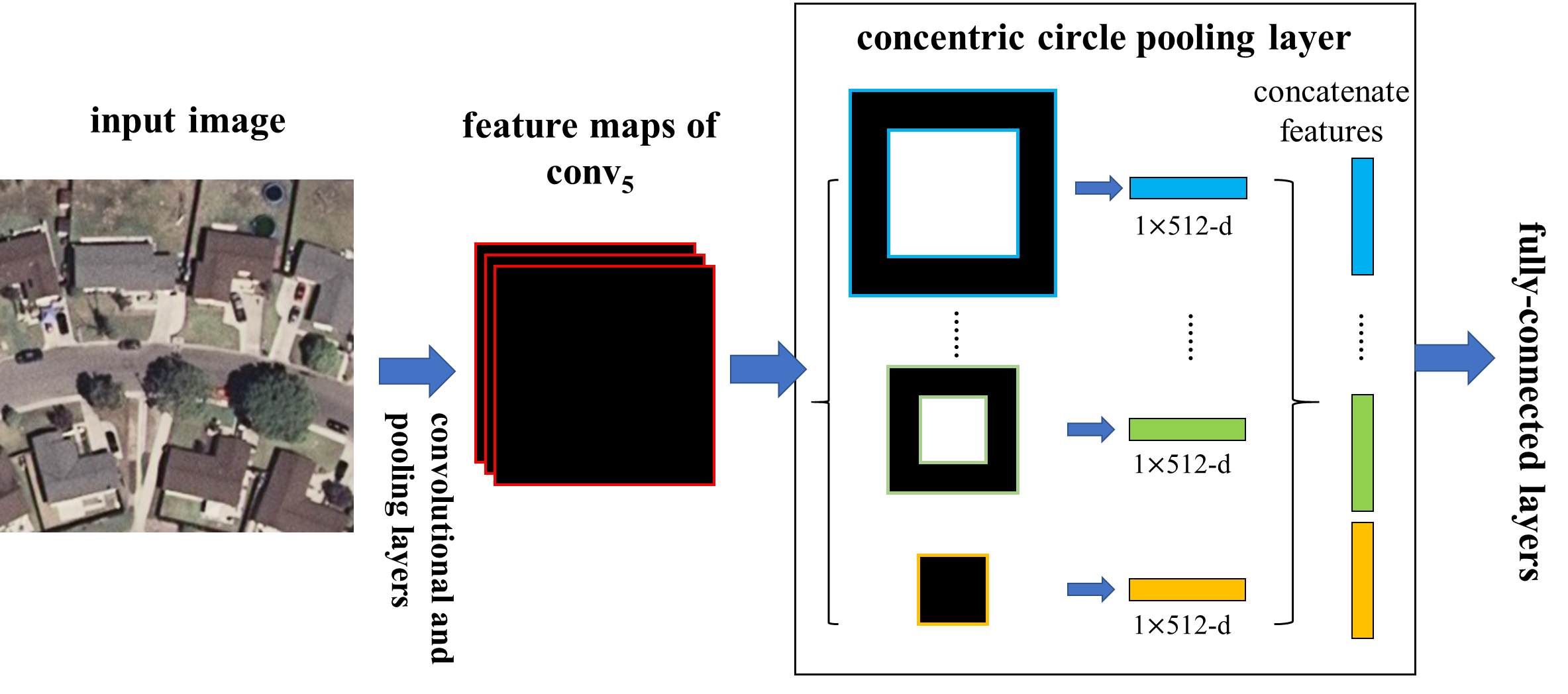
*Objective*: With the popular use of High-Resolution Satellite (HRS) images, more and more research efforts have been placed on land-use scene classification. However, it makes the task difficult with HRS images for the complex background and multiple land-cover classes or objects. This article presents a “concentric circle pooling” for land-use scene classification to alleviate the sensitive to the rotation of image scene in existing convolutional neural network (CNN). The new network structure, called CCP -net can generate a concentric circle-based spatial-rotation-invariant representation for describing remote sensing scene classification. Experiments on two public scene classification datasets demonstrate that the proposed model is efficient for land-use scene classification and achieves competitive classification results compared with the state-of-art methods.

*Background*: HRS images are increasingly available and playing an ever-more important role land-use classification [1]. HRS images proved more of the appearance and spatial arrangement information helpful in land-use scene category recognition [2]. However, it is usually difficult to recognize land-use scene categories because they usually cover multiple land covers or ground objects, such as airports with airplanes, runways, and grass[3]-[5]. Land-use scene categories are largely affected and determined by human and social activities, thus the recognition of land-use scene image is based on a priori knowledge. These characteristics make the traditional pixel-based [6] and low-level feature-based image classification techniques [7], [8] inadequate for land-use scene classification.

In this paper, we introduce a Concentric Circle Pooling (CCP) layer to incorporate rotation-invariant spatial layout information of remote sensing scene images. The concentric circle-based partition strategy of an image has been proven effective for rotation-invariant spatial information representation in color and texture feature extraction[10], [11], and Bag-Of-Visual-Words (BOVW) and FV representation [11], [12]. Specifically, we add a CCP layer on top of the last convolutional layer. The CCP layer pools the features and then fed into the fully-connected layers. Thus, for CCP layer uses annular spatial bins, we can pool the convolutional features to achieve a rotation invariant spatial representation. Experiments were conducted based on two public ground truth image datasets, manually extracted from publicly available high-resolution overhead imagery. Experimental results show that the CCP layer helps CNNs to represent the remote sensing scene images and achieves high classification accuracies.

*Data*: Experiments were conducted based on a ground truth image data set of 21 land-use classes UC Merced dataset (UCM), manually extracted from publicly available high-resolution overhead imagery [4]. The experimental results show that the deep correlaton model using CNN is a simple yet effective way to represent the land-use scene image and achieve good performances in terms of classification accuracies.

*Methodology*: To adopt the deep network for the spatial information of rotation-invariance, we replace the last pooling layer (the pool5 after the last convolutional layer conv5 in VGG-VD16) with a CCP layer. Figure 3 illustrates our method. In each annular subregion, we pool the response of each filter with max pooling. The output size of last convolutional layer may not divide exactly by the number of subregions, so the outputs of the CCP layer are K-dimensional vectors where R denotes the circle number, which is number of subregions, and K denotes the number of filters in the last convolutional layer. In the next subsection, we interpret the output size of CCP layer in detail.



1. A network structure with a concentric circle pooling layer. The convolutional and pooling layers except the last pooling layer in VGG-VD16 are transformed to this network and 512 is the filter number of the conv5 layer, which is the last convolutional layer.

*Result and Discussion*: We randomly select samples of each class for training CCP-net and the rest for testing. The sampling setting as for the UCM dataset are: 80 training samples per class. The dataset is divided 10 times (each run with randomly selected training and testing samples) to obtain reliable results, and all the results, as well as the classification accuracy rate for categories were recorded as the mean and standard deviation of these 10 runs. The results in Table 1 show that the deep correlaton can get a better performance that some other state-of-art methods.

1. erformance Comparison

| Methods | Accuracy（%） |
| --- | --- |
| BoVW | 72 |
| SPM | 74 |
| CaffeNet | 93.41.0 |
| OverFeat | 90.91.2 |
| CCP-net | 95.51.4 |

*Conclusion and Future Work*: This paper presents a CCP-net for HRS image scenes classification. Experiments on the UCM datasets indicate that our method performs competitively to several representative state-of-the-art approaches. In the future work, we will consider multi-scale feature learning method for HRS image scene classification.

##### Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant No. 41701410, Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing under Grant No. 16R02, and China Postdoctoral Science Foundation under Grant No 2017M622550.

##### References

1. Zhou, W.; Troy, A. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *Int. J. Remote Sens.* **2008**, *29*, 3119-3135.
2. Zhang, H.; Lin, H.; Li, Y.; Zhang, Y. Feature extraction for high-resolution imagery based on human visual perception. *Int. J. Remote Sens.* **2013**, *34*, 1146-1163.
3. Rogan, J.; Chen, D. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Prog. Plann.* **2004**, *61*, 301-325.
4. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classiﬁcation. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 270–279.
5. Xia, G.S.; Yang, W.; Delon, J.; Gousseau, Y.; Sun, H.; Maitre, H. Structrual High-Resolution Satellite Image Indexing. In *Processdings of the ISPRS, TC VII Symposium Part A: 100 Years ISPRS - Advancing Remote Sensing Science*, Vienna, Austria, 5–7 July 2010.
6. Chehdi, K.; Soltani, M.; Cariou, C. Pixel classification of large-size hyperspectral images by affinity propagation. *J. Appl. Remote Sens.* **2014**, *8*, 083567-083567.
7. Yu, Q. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 799-811.
8. Zhao, Y.; Zhang, L.; Li, P.; Huang, B. Classification of high spatial resolution imagery using improved gaussian markov random-field-based texture features. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1458-1468.
9. Lu, X.; Zheng, X.; Yuan, Y. Remote sensing scene classification by unsupervised representation learning. *IEEE Trans. Geosci. Remote Sens.* **2017**, *PP*, 1-10.
10. Battiato, S.; Farinella, G.M.; Gallo, G.; Ravi, D. Spatial hierarchy of textons distributions for scene classification, In Conference on Multimedia Modeling, Berlin, Heidelberg, 2009; pp 333-343.
11. Zhao, L.J.; Tang, P.; Huo, L.Z. Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4620-4631.
12. Zhou, L.; Zhou, Z.; Hu, D. Scene classification using multi-resolution low-level feature combination. *Neurocomputing* **2013**, *122*, 284-297.