REMOTE SENSING SCENE CLASSIFICATION USING MULTIPLE PYRAMID POOLING

Y. Yao[1,2], H. Zhao[1,2*], D. Huang[1,2], Q. Tan[1,2]

[1]Department of Civil Engineering, Tsinghua University, Beijing 10084, China
- (yaoyi18, zhr, hdm, tqf17@mails.tsinghua.edu.cn)
[2]3S Center, Tsinghua University, Beijing 10084, China

KEY WORDS: Remote Sensing, Scene Classification, Deep Learning, Convolutional Neural Network, Multiple Pyramid Pooling

ABSTRACT:

Remote sensing image scene classification has gained remarkable attention, due to its versatile use in different applications like geospatial object detection, ground object information extraction, environment monitoring and etc. The scene not only contains the information of the ground objects, but also includes the spatial relationship between the ground objects and the environment. With rapid growth of the amount of remote sensing image data, the need for automatic annotation methods for image scenes is more urgent. This paper proposes a new framework for high resolution remote sensing images scene classification based on convolutional neural network. To eliminate the requirement of fixed-size input image, multiple pyramid pooling strategy is equipped between convolutional layers and fully connected layers. Then, the fixed-size features generated by multiple pyramid pooling layer was extended to one-dimension fixed-length vector and fed into fully connected layers. Our method could generate a fixed-length representation regardless of image size, at the same time get higher classification accuracy. On UC-Merced and NWPU-RESISC45 datasets, our framework achieved satisfying accuracies, which is 93.24% and 88.62% respectively.

1. INTRODUCTION

Remote sensing is a non-contact technology which enables rapid and large-scale acquisition of information. With the development of satellite sensors, large amounts of high-resolution remote sensing (HRRS) images are readily available. Remote sensing image scene classification, which mainly focuses on labelling remote sensing image with a specific semantic category, has gained remarkable attention and has been widely applied in versatile scenarios including geospatial object detection, ground object information extraction, environment monitoring and etc. (Hu et al. 2013, Zhang et al. 2016) As spatial resolution of remote sensing images improves continuously, people want to mine a higher level of semantic information from HRRS images. And ground objects form different semantic scene categories through different spatial distribute pattern, (Bratasanu et al. 2011, Lienou et al. 2010) via using detailed information provided by HRRS images. The scene not only contains the information of the ground objects, but also includes the spatial relationship between the ground objects and the environment. With rapid growth of the amount of remote sensing image data, the need for automatic annotation methods for remote sensing images is more urgent.

In past decades, many methods have been proposed for remote-sensing scene classification. In general, these methods could be divided into three groups, according to features representing level. Firstly, methods using low-level features. These methods usually utilize hand-crafted features to classify remote sensing images, which rely heavily on the experience and domain knowledge of experts, such as spectral, colour, texture and shape information or their combination. One of the most popularly used features is scale invariant feature transform (SIFT) (Lowe et al. 2004, Shao et al. 2013). SIFT is a kind of local feature descriptor in image processing area, which could describe local variations of structures in images and would not be affected by image size and rotation. (Yang et al. 2008) compared SIFT and Gabor texture features for classifying remote sensing images and showed that SIFT performs better. Other low-level features, like histograms of oriented gradients (HOGs) (Dalal, Triggs, 2005) and global color descriptors, could depict the spatial arrangements of images and break the limitation that local descriptors cannot show global distributions of spatial cues. However, due to the lack of consideration for the details of actual data, it is difficult with these low-level features to attain a balance between discriminability and robustness (Chen et al. 2017). And these low-level features often fail to characterize the complex remote sensing scenes in HRRS images.

Secondly, mid-level features. These methods mainly focus on developing a set of basis functions used for feature encoding. One of the most popularly used mid-level models is the bag-of-visual-words (BoVW) (Yang et al. 2010). The BoVW-based models represent an image with the frequency of a collections of 'visual words', which is constructed by encoding local features extracted from local image patches, such as SIFT and HOGs. The original BoVW model ignored the spatial order of local features so that it severely limited the descriptive capability of image representation. Therefore, many BoVW extensions have been proposed to depict spatial relationships of local features. For instance, (Lazebnik et al. 2006) used spatial pyramid matching to avoid this issue. Although these models increased the capability of feature relationship representing, they still demand prior knowledge for hand-crafted feature extraction, lacking flexibility in discovering high-level semantic meaning.

Recently, methods based on deep learning technology, especially the convolutional neural network (CNN), have made great

* Corresponding author should be addressed to Hongrui Zhao, Email: zhr@tsinghua.edu.cn.

breakthroughs in image classification, video surveillance, object detection and many other computer vision fields. The typical convolutional neural networks, including Alexnet, VGGNet, GoogleNet, have been successfully applied in image classification tasks (Szegedy et al. 2014, Krizhevsky et al. 2017, Simonyan et al. 2014). These models are mostly composed of multiple convolutional layers, pooling layers and fully connected layers. These outstanding CNNs for object classification tasks can be pre-trained on large natural image datasets such as ImageNet (Deng et al. 2009). As discriminative high-level feature extractor, these pretrained CNN models can be transferable to many other domains. Due to its good feature extraction and classification ability, CNN has gradually attracted the attention of remote sensing communities. (Zeng et al. 2018) integrated global-context and local-object features from remote sensing images to address the issue that it is hard for vanilla CNNs to focus on both global context and small local objects. (Qi et al. 2018) used concentric circle pooling in CNNs to alleviate the problem that conventional CNNs are sensitive to rotation of remote sensing image. These methods showed us it is transferable for CNNs to address remote sensing scene classification.

However, remote sensing images are much larger, both in memory size and image size, than traditional images inputted in CNNs, whereas CNNs often requires a fixed-size smaller image, like 256*256 pixels. In order to satisfy the requirement of conventional CNNs, researchers in remote sensing area usually resized or cropped the large remote sensing image to fixed-size patches. Unfortunately, the down-sampling of the original remote sensing image makes the objects smaller and it is harder to extract corresponding features from image. Besides, the crop operation would change the characteristics of the data. Inspired by the success of fast R-CNN in object detection area (Girshick et al. 2015), to solve the above issue, our model, supporting the input of any size, is designed by adding multiple pyramid pooling layer in traditional CNNs. In general, our model consists of two parts: (1) fine-tuning the VGG16 model, pre-trained on the large-scale natural dataset ImageNet; (2) using multiple pyramid pooling layer to get the fixed-sized feature so that the model supports inputting of any size images.

The remainder of this paper is organized as follows. In Section2, the proposed architecture would be illustrated in detail. In Section3, the different datasets we used and experiment conducting details would be introduced. Experiment result and related discussion would be explained in Section4. Section5 presents the conclusion of this paper.

2. PROPOSED METHOD

In this section, we first introduce the overall architecture of our proposed method; then the VGG-Base and the multiple pyramid pooling layer would be explained in detail.

2.1 Overall Architecture

As illustrated in Figure 1, our proposed architecture can be mainly divided into two parts: VGG-Base (composed of convolutional layers, fully connected layers and softmax layer), multiple pyramid pooling layer. The processing flow of our architecture is as follows. Firstly, an image randomly selected from the dataset is fed in the VGG-Base without cropping and resizing operation. Compared with methods requiring fixed-size input images, usually achieving by resizing images to a certain scale or cropping fixed-size patches, our method can extract more applicable features from original-scale images without resizing or cropping operations.

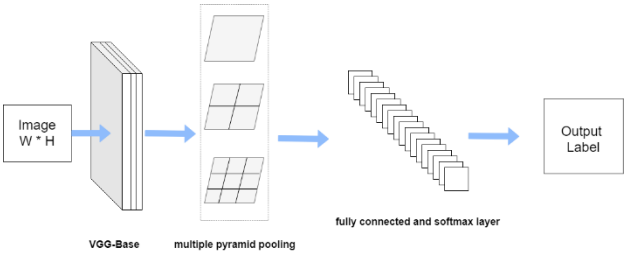


Figure 1 Overall Architecture

Taking an image as input, the VGG-Base network, which is the backbone of work framework, processes multiple convolutional operations to extract high-level features from the image. After this process, we got the feature map. Compared with original image size, the feature map would be down-sampled by 32 times after 5 max-pooling operations. Then we fed feature map into multiple pyramid pooling layer, in which the feature map would be divided into different size grids, according to the given pooling kernels. Features with different size would be produced by pooling kernels in multiple pyramid pooling layer, then these features with different size would be concatenated and expanded to one-dimension fixed-length vector. Then, the fixed-length one-dimension vector would be fed into fully connected layers and softmax classification layer to calculated probability of classification for every class.

2.2 VGG-Base

The VGG-Base network composed of three parts: (1) convolutional layers (2)ReLU activation function (3)pooling layers (4) fully connected and softmax classification layers. The first three parts were shown in Figure 2 and the last part worked behind multiple pyramid layer, which could be seen in Figure 1.

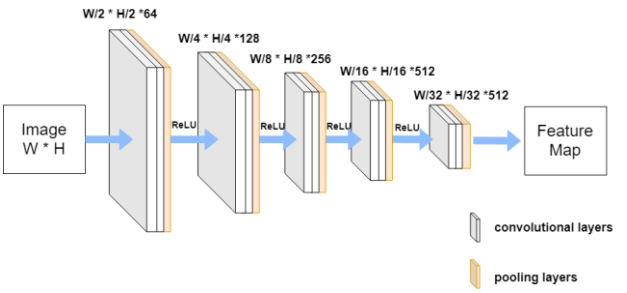


Figure 2 VGG-Base

Comparing to traditional low-level and mid-level feature extraction methods, such as SIFT and BoVW, convolutional layers can automatically extract high-level semantic features from image without hand-crafted feature selection processes. Convolutional layers contain multiple convolutional kernels, which are two-dimension matrixes. These convolutional kernels would slide cross the whole input image based on sliding rules. Through summing up the multiplication result of input image and convolutional kernel, the input image could generate feature map.

In general, an activation function should be added after each convolutional layer to increase the nonlinear fitting capability of neural network, such as the Sigmoid function, Rectified Linear

Unit (ReLU). In our model, we used the ReLU as the activation function. The formula for the ReLU function is :

$$f(x) = \max(0, x) \quad (1)$$

And ReLU function image was shown in Figure 3.

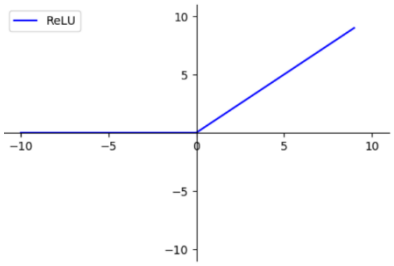


Figure 3 Image of the ReLU function.

The pooling layer is to down-sample the image features so that high-level abstractive semantic information could be extracted from the image. Here we used max-pooling layers in our model, which will return the max value from each sub-area. And the images would be down-sampled by max-pooling layers, causing 1/2 reduction in each images' height and width. Figure 4 shows the illustration of max-pooling layer.

The fully connected layers and softmax layers are located behind multiple pyramid pooling layers, they worked as a classifier to generate final classification probabilities. After multiple pyramid pooling layers, a fixed-size feature map would be produced and we expand it to a one-dimension vector and then fed it into fully connected layers.

In this paper we utilize the VGG-16 model as our baseline. The architecture of VGG-16 contains five convolution parts to generate feature map, two fully connected layers and one softmax classifier to output classification results. Each of convolution part has two or three convolutional layers and one max-pooling layer in it. The parameters of convolution parts are shown in Table1.

Stage	Parameters
Stage1	conv 3⋆3-64 conv 3*3-64 max-pool 2*2
Stage2	conv 3*3-128 conv 3*3-128 max-pool 2*2
Stage3	conv 3*3-256 conv 3*3-256 conv 3*3-256 max-pool 2*2
Stage4	conv 3*3-512 conv 3*3-512 conv 3*3-512 max-pool 2*2
Stage5	conv 3*3-512 conv 3*3-512 conv 3*3-512 max-pool 2*2

Table 1. The parameters of VGG16 model.

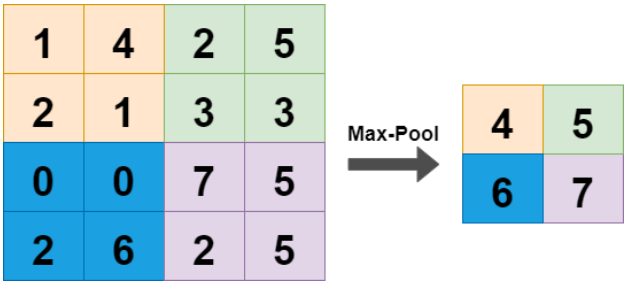


Figure 4. Pooling layer, with filters size 2x2 and stride 2

2.3 Multiple Pyramid Pooling Layer

The convolutional neural network described above demands a fixed image size, due to the requirement that fully connected layers demand a fixed-size input. However, convolutional layers accept inputs of arbitrary sizes and use sliding filters to generate feature maps, which involves not only the strength of the responses, but also their spatial relationship. Therefore, the only part that limits the input image size is fully connected layers. To address issue, a multiple pyramid pooling hierarch is used here.

Our multiple pyramid pooling module works between convolutional layers and fully connected layers. The multiple pyramid pooling module fuses feature under different scales. To illustrate it in detail, the input feature map, generated by convolutional layers, was divided into different size grids by adopting varying-size pooling kernels with varying strides. Our pyramid pooling module adopts a three-level pyramid pooling structure, with bin sizes of 1×1, 2×2, 3×3, after performing experiments to test classification accuracy under different pyramid pooling level. In each spatial bin, we found the maximum number in feature map patches to produce fixed-size responses. With multiple pyramid pooling, any size input image would be divided according to given spatial bins and generate a fixed-size feature map. And then the fixed-size feature map would be expanded to one-dimension vector which is the input of fully connected layer.

Interestingly, under the extreme circumstance, the pyramid structure could be one single bin and the operation would find max number in feature map. This is in fact a 'global pooling' operation, which is also investigated in several concurrent works. (Lin, Chen, 2013) used a global average pooling to reduce the number of parameters and also reduce overfitting. (Oquab et al. 2014) used a global max pooling for weakly supervised object recognition.

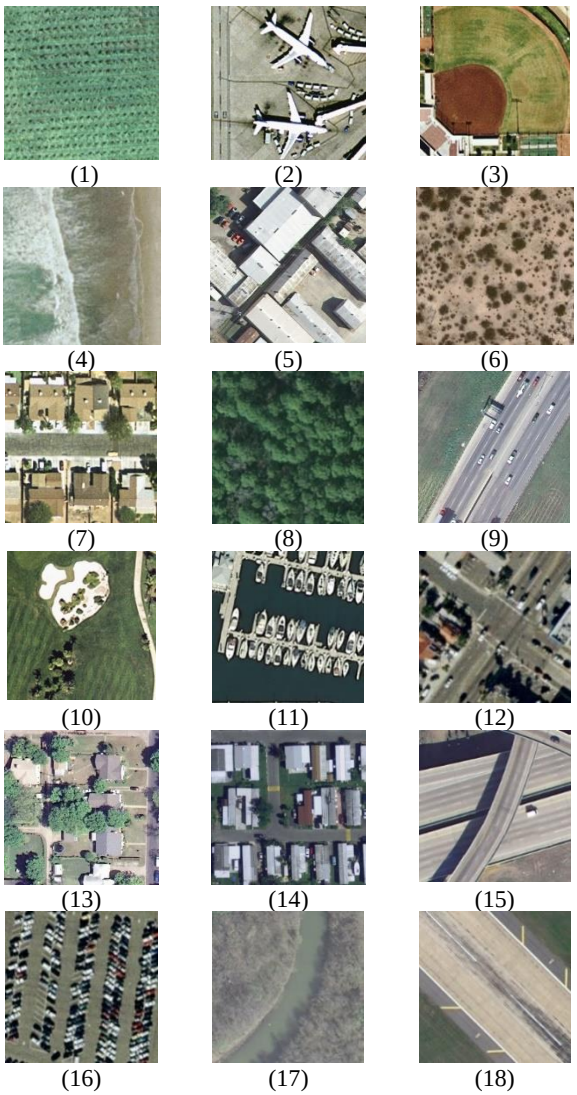
3. MATERIALS AND EXPERIMENT

3.1 Datasets

In this paper we selected two datasets for experiment, UC-Merced dataset (Yang, Newsam, 2010) and NWPU-RESISC45 dataset (Cheng et al. 2017). The UC-Merced dataset was provided by United States Geological Survey (USGS). There are 21 classes of scene and 100 images in each scene category, with the spatial resolution of 0.3 meter. And every image is composed of 256×256 pixels in red green blue (RGB) color space. The 21 categories, as shown is Figure 5, include agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbour, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tank and tennis courts. This dataset has several highly overlapping classes, such as sparse

residential, medium residential and dense residential, which only differ in the density of buildings so that it is a challenging work to classify this dataset.

Another dataset is NWPU-RESISC45, which is one of the largest remote sensing image scene datasets. This dataset was provided by North-western Polytechnical University (NWPU). The dataset contains 31,500 images in total with 45 scene classes and each class contains 700 images. And the spatial resolution of images varies from 30m to 0.2m. The different spatial resolution and rich variations, such as shooting angle, illumination, background, contained in each category extremely increased difficulty for classification. The scene categories in NWPU-RESISC45 dataset are: airplane, airport, beach, bridge, chaparral, runway, sea ice, ship, commercial area, tennis court, sparse residential, dense residential, desert, forest, mobile home park, mountain, freeway, intersection, church, circular farmland, cloud, baseball diamond, basketball court, golf course, ground track field, harbor, industrial area, island, lake, meadow, medium residential, overpass, palace, parking lot, wetland, railway, railway station, rectangular farmland, river, roundabout, snow berg, storage tank, terrace, thermal power station and stadium.



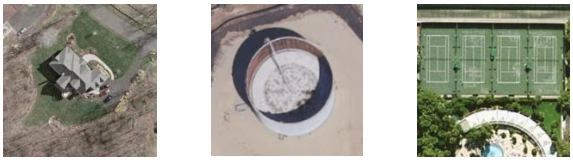


Figure 5. Categories in UC-Merced dataset: (1) agriculture; (2) airplane; (3) baseball diamond; (4) beach; (5) buildings; (6) chaparral; (7) dense residential; (8) forest; (9) freeway; (10) golf course; (11) harbor; (12) intersection; (13) medium residential; (14) mobile home park; (15) overpass; (16) parking lot; (17) river; (18) runway; (19) sparse residential; (20) storage tanks; and (21) tennis court.

3.2 Experiment and Evaluation Protocol

We used open-source TensorFlow (https://www.tensorflow.org/) framework to implement our proposed architecture. For UC-Merced and NWPU-RESISC45 datasets, 20% of the samples are randomly selected for testing. Data augmentation was used to generate sufficient data to train an effective model. Our augmentation operations mainly included rotating original images in four different orientations(0°, 90°, 180°, 270°) and randomly adding the White Gaussian Noise. Back propagation algorithm was used to train the convolutional layers and the fully connected layers. We used stochastic gradient descent algorithm based on mini-batch to optimize parameters, and the batch size for each iteration in the training is 32. Other hyper-parameters used for training were set as follows. The learning rate was set to 0.001. The maximum number of iterations was 20000. It is worth noticing that our architecture is fine-tuned from the pre-trained VGG16 model on ImageNet, while the multiple pyramid pooling module was trained from scratch. In all experiments, all parameters trained from scratch were initialized by Gaussian distribution with zero mean and unit variance. Our program was run on a PC with 2 3.2GHz 8-core CPUs, 32GB memory and a NVIDA TITAN X GPU for acceleration.

Overall accuracy and confusion matrix were reported to compare results with original VGG16 network. The overall accuracy is defined as the number of correctly classified images divided by the total number of images. The confusion matrix is an informative table used for analysing the classification errors and confusions between different categories. We obtained confusion matrix through counting correct and incorrect classifications of test images in each class and summing up the results.

4. RESULTS AND DISCUSSION

4.1 Experiments on UC-Merced

Table 2 shows the performance comparison between the original VGG16 and our model on UCM dataset and NWPU dataset. As can be seen from Table 2, our classification method, by adding multiple pyramid pooling layer, achieved better overall accuracy of 93.24% and 88.62%, both on UCM dataset and NWPU-RESISC45 dataset, respectively 1.39%, 4.06% higher than original VGG16 model. In original VGG16 network and our model, the convolutional layers have the same structures, whereas the multiple pyramid pooling layer was added between convolutional layers and fully connected layers in our model. For results in Table 2, a 3-level pyramid pooling structure was used. The pyramid structure is {1×1, 2×2, 3×3} (totally 14 bins). Worthy of mention is that the overall accuracy improvement not simply due to more parameters; rather, it is because the multiple level pooling operation is robust to the variance in object

deformations and spatial layout (Lazebnik et al. 2006). In remote sensing image scene, object deformation is easy to occur as shooting gesture changing. And spatial layout of objects is one of the most important features for remote sensing image classification, because most natural and artificial ground objects showed clustering feature in aerial images. The multiple spatial pyramid pooling could reduce the negative influences of spatial layout and object deformation so that the overall accuracy gained improvement, at the same time the input image size limitation was eliminated.

Figure 6 shows the per-class classification accuracies of our model on UC-Merced dataset. From the data shown in Figure 5, we can see that chaparral, forest, parking lot and runway got highest classification accuracy, whereas intersection got lowest accuracy 69%, 6% of which were classified into storage tanks. Through manual interpreting for images in intersection class, at the same time comparing them with the images of storage tanks category, we see that these two totally unrelated categories in literal, share a common feature that most of them had huge shadows area around the object. This result is possibly explained by the fact that shadow area in image is sharp color transition area and the network memorized this color transition feature incorrectly. Particularly, for similar and easily confused scenes like medium residential and dense residential, only small portion of dense residential were misidentified as medium residential, which partly showed the classification capability of our model.

Figure 7 showed us the classification performance of our model in each category on the NWPU-RESISC45 dataset. For categories which has clear feature to distinguish, such as lake, cloud and sea ice, higher classification accuracy was obtained. It is worth noticing that 12% medium residential was classified as dense residential, similar to the experiment result on UCM dataset.

Dataset	Method	Overall Accuracy(%)
UC-Merced	VGG16	91.85
	Our model	**93.24**
NWPU-RESISC45	VGG16	84.56
	Our model	**88.62**

Table2. Overall Accuracy Comparison

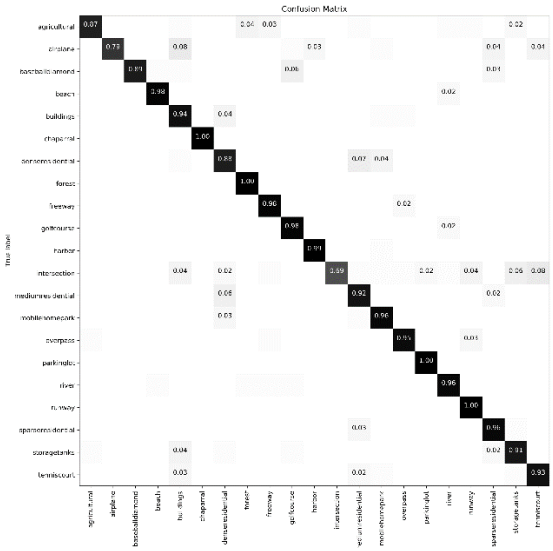


Figure 6. Confusion matrix of our model on UCM dataset

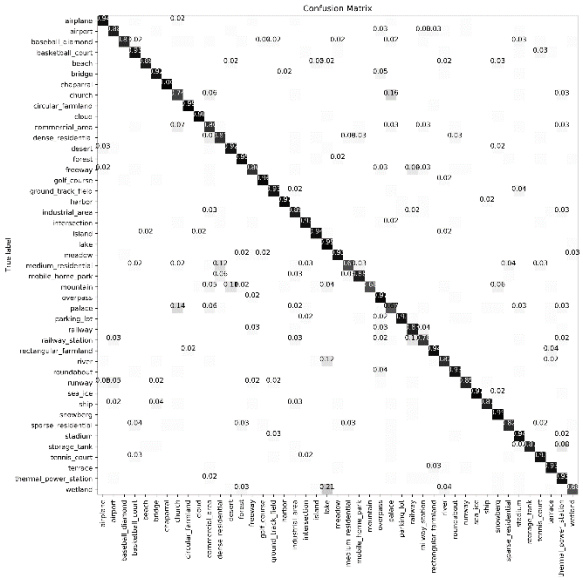


Figure 7. Confusion matrix of our model on NWPU dataset

To analyse the impact of multiple pyramid pooling structure (bin size and level), multiple experiments were conducted with different bin size and pyramid level on UCM dataset. As shown in Table 3, it can be seen that single bin one level pyramid got worse result than original VGG16 network, whereas 3×3 bin structure got highest accuracy in one-level pyramid structure. As the pyramid level increased, the accuracy also increased. But three-level pyramid pooling structure worked better than four-level one.

Dataset	Pyramid Structure	Overall Accuracy(%)
UC-Merced	None	91.85
	1×1	90.07
	3×3	91.89
	5×5	90.58
	1×1, 3×3	92.78
	1×1, 2×2, 3×3	**93.24**
	1×1, 2×2, 3×3, 5×5	93.01

Table 3. Pyramid Structure Impact Comparison

Summarizing the above result and discussion, we find our architecture could not only support arbitrary size of input image, but also got improvement in accuracy. According to the result that it got a little more accuracy improvement on NWPU-RESISC45 dataset which has more categories than UCM, it is possible that multiple pyramid pooling module works better for dataset with more categories. Through analysing different structure of pyramid pooling, we see that multi-level pyramid structure gained better accuracy than single-level pyramid, but accuracy is not always increasing as pyramid level goes up.

5. CONCLUSION

This paper proposes a new framework for high resolution remote sensing (HRRS) images scene classification based on convolutional neural network. To eliminate the requirement of fixed-size input image, multiple pyramid pooling strategy is equipped. It can be concluded that, this new method not only

supports arbitrary size of input image, but also achieves higher accuracy over the original convolutional neural network model. On both UC-Merced and NWPU-RESISC45 datasets, our method achieves good accuracy.

In future, we intend to use more modern neural networks as our baseline to prove the effectiveness of multiple pyramid pooling module. In order to get more accurate remote sensing scene classification result, other features could be introduced such as point of interest (POI), social media data, etc. New technologies to combine these features also should be explored.

ACKNOWLEDGMENTS

This study was supported by the research and development program of Anhui province[Grant number 1704d0802178].

REFERENCES

Hu, Q., Wu, W., Xia, T., Yu, Q., Yang, P., Li, Z., Song, Q., 2013. Exploring the use of Google Earth imagery and object-based m ethods in land use/cover mapping. *Remote Sensing*. 5,6026–6042.

Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensingdata: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine.* 4, 22–40.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision.* 60, 91–110.

Krizhevsky, A., Sutskever, I., Hinton, G., 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A., , 2014. Going Deeper with Convolutions. arXiv:1409.4842.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556.

Shao, W., Yang, W., Xia, G.S., Liu, G., 2013. A Hierarchical Scheme of Multiple Feature Fusion for High-Resolution Satellite SceneCategorization. *In Proceedings of the International Conferenceon Computer Vision Systems*, St. Petersburg, Russia, 16–18 July 2013; Springer: Berlin/Heidelberg, Germany, 324–333.

Bratasanu, D., Nedelcu, I., Datcu, M., 2011. Bridging the SemanticGap for Satellite Image Annotation and Automatic MappingApplications. *IEEE Journal of Selected Topics in Applied EarthObservations & Remote Sensing*, 4, 193-204.

Lienou, M., Maitre, H., Datcu, M., 2010. Semantic Annotation of Satellite Images Using Latent Dirichlet Allocation. *IEEE Geoscience & Remote Sensing Letters*, 7, 28-32.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human d etection. *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 20–25 June 2005, 886–893.

Chen, X., Xiang, S., Liu, C.L., Pan, C.H., 2017. Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience And Remote Sensing Letters*, 11, 1797–1801.

Yang, Y., Newsam, S., 2010. 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, 3–5November 2010, 270–279.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scenecategories. *IEEE Conference on Computer Vision & PatternRecognition (CPRV '06), 2*, 2169-2178.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami,FL, USA, 20–25 June 2009, 248–255.

Zeng, D., Chen, S., Chen, B., Li, S., 2018. Improving Remote Sensing Scene Classification by Integrating Global-Context and Local-Object Features. *Remote Sensing*, 10(5), 734.

Qi, K., Guan, Q., Yang C., Peng, F., Shen, S., Wu, H., 2018. Concentric Circle Pooling in Deep Convolutional Networks for Remote Sensing Scene Classification. *Remote Sensing*, 2018, 10(6), 934.

Girshick, R. Fast r-cnn. arXiv 2015, arXiv:1504.08083.

Lin, M., Chen, Q., Yan, S., 2013. Network in network. arXiv:1312.4400.

Oquab, M., Bottou, L., Laptev, I., Sivic, J., 2013. Learning and transferring mid-level image representations using convolutional neural networks, *IEEE Conference on Computer Visionand Pattern Recognition*, November 2013.

Yi Y., Shawn N., 2010. Bag-Of-Visual-Words andSpatial Extensions for Land-Use Classification, *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS)*.

Cheng, G., Han, J., Lu, X., 2017. Remote Sensing Image Scene C lassification: Benchmark and State of the Art. *Proceedings of the IEEE*.

Lazebnik, S., Schmid, C., Ponce J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *IEEE Conference on Computer Vision and Pattern Recognition*.

Yang, Y., Newsam, S., 2008. Comparing SIFT descriptors and Gabor t exture features for classification of remote sensed imagery. *In Proceedingsofthe15thIEEEInternationalConference on ImageProcessing (ICIP)*, San Diego, CA, USA, 12–15 October 2008, 1852–1855.