MULTISCALE CONVOLUTIONAL NEURAL NETWORK FOR THE DETECTION OF BUILT-UP AREAS IN HIGH-RESOLUTION SAR IMAGES

Jingge Li, Rong Zhang and Yue Li
(Department of Electronic Engineering and Information Science, USTC, Hefei, 230027 China)
(Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei, 230027)
Email: lijingg@mail.ustc.edu.cn, zrong@ ustc.edu.cn, lyue@mail.ustc.edu.cn

ABSTRACT

This paper focuses on the problem of built-up areas detection in single high-resolution SAR images. In consideration of the rich structure information of built-up areas in high-resolution SAR images, we put forward a multiscale CNN model to extract multiscale trained features directly from image patches to detect built-up areas. By processing features extraction and classification as a whole, we overcome the difficulty of feature definition of SAR images. Experiments on TerraSAR-X high-resolution SAR images over Beijing show that our approach outperforms the traditional methods and single scale CNNs method.

Index Terms—high-resolution SAR images, built-up areas, multiscale Convolutional neural networks.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is the only imaging system that can generate high resolution imagery anytime - even in inclement weather or darkness. Thus, SAR images are widely used to observe the land, e.g. disaster management, land cover mapping, etc. Built-up areas are main components of urban areas, in urban mapping and planning, the detection of built-up areas is importantly applied to extract urban structures. Especially with the increasing spatial resolution of SAR images, built-up areas present more details in SAR images. We can extract more information to detect built-up areas.

Feature extraction is a key problem of SAR object detection and suitable features can greatly reinforce the detection effects. Generally, features used in built-up areas detection are hand-crafted, and are same as those used for optical images or its improvement, for instance optical texture features [1] and Labeled Co-occurrence Matrix improved from GLCM [2]. These features only take advantage of texture features in SAR images, but the semantic information in SAR images is not used. In addition, hand-crafted features are generally incomplete and frequently out of work because of speckle noise. Therefore, it is tempting to seek for features that have great ability in expressing the distinctive characters of the objects of interest in SAR images.

Recently, deep learning using CNN has gained much success in classic visual recognition tasks [3-5], such as image classification and object detection. Since the generic features extracted from the networks have achieved a superior performance over hand-crafted feature extractor, we can detect built-up areas using CNNs to train features directly from image patches.

In this paper, a multiscale CNN model was proposed to solve the problem of built-up areas detection in high-resolution SAR images. The work is particularly challenging since the built-up areas are diverse and SAR images are suffered strong speckle noise, thus we extract robust multiscale and hierarchical features directly from image patches to detect built-up areas. Experimental results show that our model performs better than traditional features and single scale CNNs.

The paper is organized as follows. Section 2 briefly introduces CNNs and describes the proposed multiscale CNN model in detail. Section 3 presents our experiments and results. Conclusion is made in section 4.

2. MULTISCALE CONVOLUTIONAL NEURAL NETWORKS

2.1. Convolutional neural networks

Convolutional neural networks are multi-layered neural networks，which are specifically designed to extract features directly from the images and do not require any specific feature extractor choice [6]. Unlike norm neural networks, CNNs have three special characteristics (local receptive fields, shared weights and spatial sub-sampling) to make it high degree of invariance of translation, scaling, skewing and other forms of distortion. A typical CNN is shown in Figure 1. Here, Conv-1, Conv-2 and Conv-3 are convolutional layers. Pool-1, Pool-2 and Pool-3 are subsampling layers, fully-connected layers include Ip-1 and Ip-2, respectively.

Convolutional layer is the most characteristic structure of CNN. A convolutional layer feature map is calculated as (1), for instance of Conv-2, each feature map x_j^l in Conv-2 is the result of a sum of convolution of Pool-1's feature maps x_i^{l-1} by their respective kernel k_{ij}^l, a bias b_j^l is added

978-1-5090-3332-4/16/$31.00 ©2016 IEEE

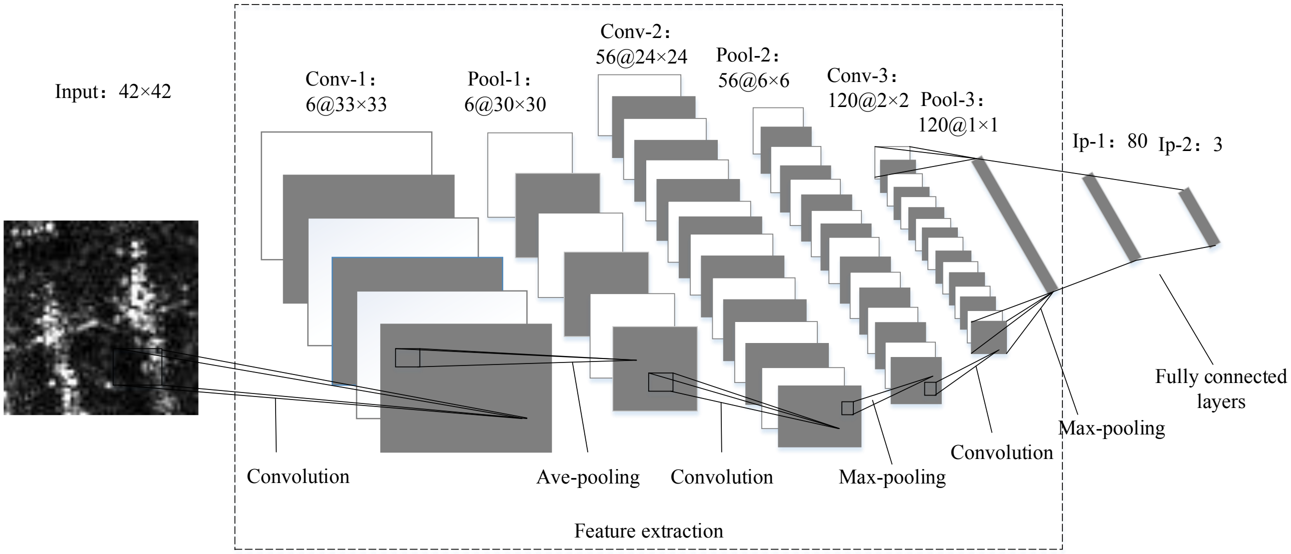


Figure 1. Architecture of CNN$_{42}$

and the result is passed through an activation function. The CNNs's special characteristic of local receptive fields is formed by every pixel in Conv-2's maps being connected with a small patch in Pool-1's feature maps because of convolutional process, and a feature map of pool-1 sharing the same convolutional kernel constitutes the shared weights characteristic, which decreases the complexity of network to avoid the overfitting problem.

$$x_j^l = f\Big(\sum_{i\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\Big) \qquad (1)$$

The subsampling layers which always follow the convolutional layers subsample the feature maps from the convolutional layers. The subsampling process decreases the sensitivity of displacement and deformation and increases the robustness of features. Meanwhile, the subsampling layer is also the process of aggregating low-level features to high-level semantics features.

Fully-connected layer is a classifier fed with features after convoluted and sub-sampled, and fully-connected layers generally contain two or three layers. The last layer, in the case of supervised learning, contains as many neurons as the number of classes. In our model, the last layer contains 3 neurons, and usually a softmax activation function is used to turn outputs into probabilities.

2.2. Multiscale CNN models

CNNs were directly fed with SAR image patches, and using a single scale image patch to detect built-up areas is problematic because of the diversity of buildings in urban areas. To extract enough semantics information and make a more precise detection, a multiscale CNN model is proposed which uses multiscale patches to get multiscale CNN features in the detection of built-up areas. It is well admitted that human vision is a multiscale process, then we use a large image patch to feed into CNNs to gather more robust semantics information and a small image patch to get more precise location information, thus we can get more complete features to detect built-up areas.

Our architecture of multiscale CNN is shown in Figure 2. Features trained from CNNs of different scales are fused into multiscale features to make a decision. Considering the size of buildings and the structure of network，we use image patch sizes of 14×14, 42×42 and 84×84 to extract different scales CNN features to detect the built-up areas. CNNs models of different input scales are named CNN$_{14}$，CNN$_{42}$ and CNN$_{84}$ according to its input image sizes. Features after convoluted and sub-sampled by CNN net of each scale are fused to feed into the full-connected layers for classification. In our work, SAR images are grouped into three categories: the built-up areas, the ground areas and the water areas, thus the output of our multiscale CNN is three.

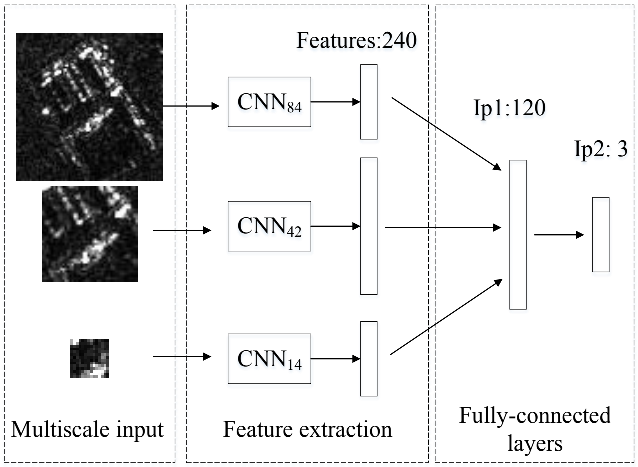
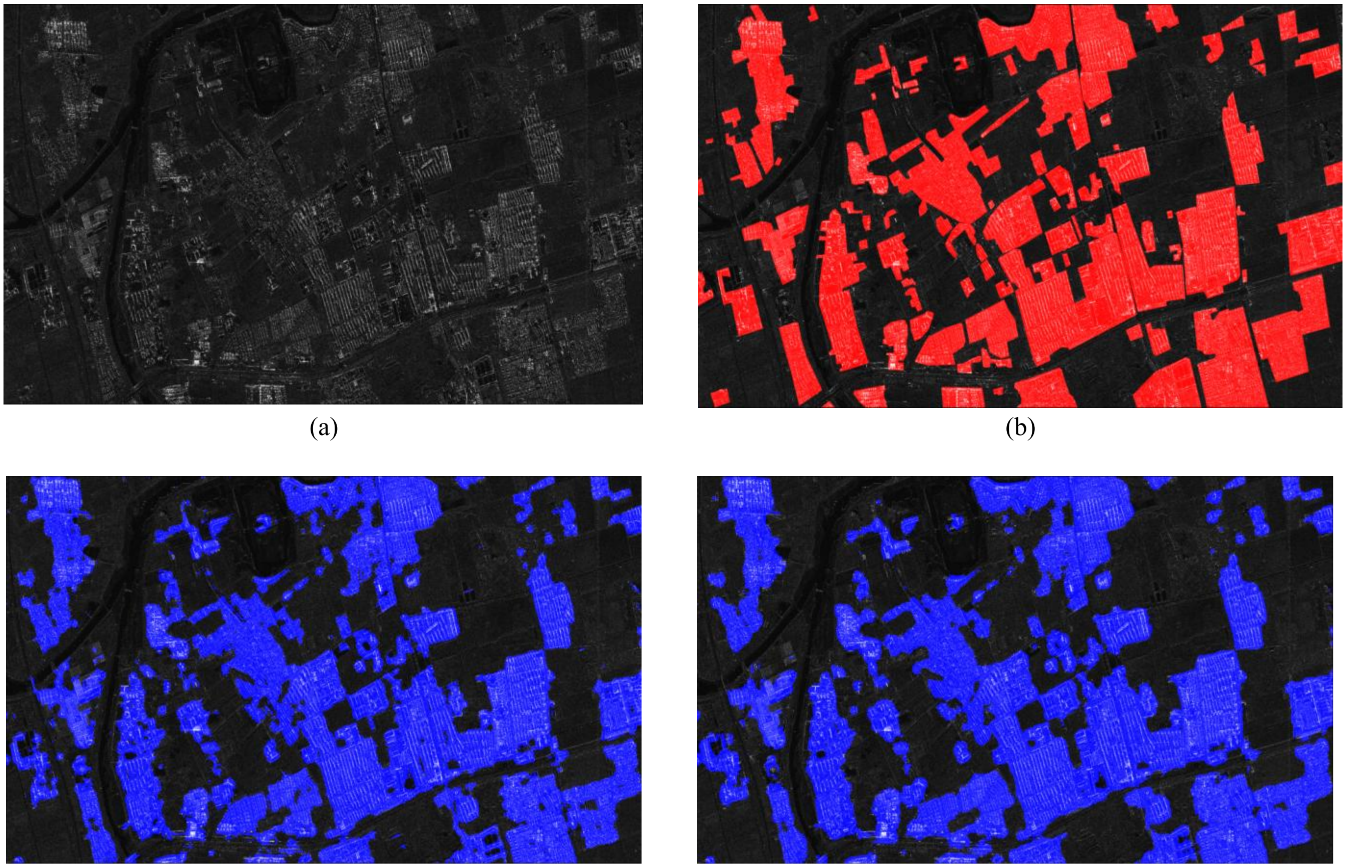


Figure 2. Architecture of multiscale CNNs



(a) (b)

(c) (d)

Figure 3. Experimental results (a) SAR Image of northern areas of Beijing. (b) Manually Labeled Image.
(c) Detection Result of CNN42. (d) Detection Result of Proposed Multiscale CNN.

Table 1 CNN architecture of each scale
Top: kernel size, Bottom: kernel number

Layer	CNN$_{14}$	CNN$_{42}$	CNN$_{84}$
Conv-1	8×8 60	10×10 6	7×7 6
Pool-1	4×4 60	4×4 6	4×4 6
Conv-2		7×7 56	7×7 60
Pool-2		4×4 56	3×3 60
Conv-3		5×5 120	
Pool-3		2×2 120	
Ip1	45	80	45
Ip2	3	3	3

Figure 1 is the architecture of CNN$_{42}$. Experiments indicate that we have to increase the kernel size in the front of the network because of the strong speckle noise, thus prevent the noise to Spread into the deep layers of network. ReLU was used in every convolutional layer and fully-connected layer to avoid gradient vanishing and to speed up learning. In addition, the drop-out technique was used to help generate more robust features by learning on different random subsets [4]. The architectures of the other CNNs are summarized in Table 1.There are some blanks because the net does not have that layer. Ip1 and Ip2 layer were equipped to train CNNs of each scale.

To fully train the multiscale CNN, we use trained CNNs of each scale to finetune the multiscale model. Firstly, the parameters of the multiscale CNNs are initialized using the parameters of CNNs of each scale trained separately, then the finetuning of the multiscale CNN is operated to finish the train process.

3. EXPERIMENTS AND RESULTS

High-resolution TerraSAR-X SAR images collected on November 25, 2011 of Beijing areas were selected to verify our method. The SAR images have range resolution of 2.3m and azimuth resolution of 3.3m. Building types in images includes Dot villa district, residential quarter buildings, squatter settlement, etc. 54000 multiscale samples were extracted in eastern areas of Beijing according to the high-resolution optical maps to train CNNs. 25% samples separated from the train set were divided as validation set. The test set was a SAR image of 2500×4000 pixels collected from the northern areas of Beijing, which was separated with the train and validation set. Adaptive gradient [7] was used to train every CNNs. The result on validation set got an accuracy of 94.3%. The result proves that our multiscale CNN features are effective.

Our trained multiscale CNN model was tested on the test SAR image. For every pixel we got the local patches around the pixel defined in multiscale CNNs, then, we feed it into multiscale CNN model to get the result of this pixel. Figure 4 shows our detection result on the SAR image. From Figure 4 (a), it can be seen that building types and sizes in this areas are complex, and both large span of built-up areas and scattered building appeared in this areas. Figure 4 (b) shows the built-up areas labels manually labeled according to optical maps, and the regions in red are the built-up areas. Figure 4 (c) shows the result of CNN_{42}, it can found that some missed detection occurred in built-up areas. Figure 4 (d) shows the result of multiscale CNNs, all types of buildings were detected and we obtained a good performance in sparse building areas and boundary portions. This indicates that our multiscale CNN model is more effective to detect built-up areas.

Table 2 Pixel Level Accuracy

method	Detection rate	False alarm rate	Accuracy of classification
GLCM	84.38%	15.82%	88.78%
LCM[2]	89.39%	23.40%	86.16%
CNN_{14}	78.61%	16.62%	86.02%
CNN_{42}	90.43%	12.77%	90.52%
CNN_{84}	90.38%	17.10%	89.64%
Multiscale CNN	**92.14%**	**10.71%**	**92.86%**

Pixel level accuracy is shown in Table 2. We use detection rate, false alarm rate and accuracy of classification to evaluate our result [8]. CNN_{14}, CNN_{42} and CNN_{84} represent the result using corresponding single scale CNN. GLCM is the detection result using GLCM texture features consulting [1], and LCM is the result of [2], results show that our method performs better than other methods. This illustrates that features extract from proposed multiscale CNN model are more efficient than contrast features.

4. CONCLUSION

In this paper, a model based on multiscale CNNs has been proposed to solve the problem of built-up areas detection in high-resolution SAR images. By combining with the great feature extraction ability of CNNs, we use multiscale CNNs to extract multiscale features to make a detection. Experimental results on TerraSAR-X SAR images get a detection rate of 92.14%. This indicates that the multiscale CNN model is effective to detect built-up areas in high-resolution SAR images.

5. ACKNOWLEDGMENT

This work was supported in part by the National Nature Science Foundation of China (No.61331020). The authors would also like to thank Mr. Huaiyu Ruan for his work in data collections.

6. REFERENCES

[1] Yang, Wen, et al. "Supervised land-cover classification of TerraSAR-X imagery over urban areas using extremely randomized clustering forests." Urban Remote Sensing Event, 2009 Joint. IEEE, 2009.

[2] Li, Na, et al. "Labeled co-occurrence matrix for the detection of built-up areas in high-resolution SAR images." SPIE Remote Sensing. International Society for Optics and Photonics, 2013.

[3] LeCun, Yann, Koray Kavukcuoglu, and Clément Farabet. "Convolutional networks and applications in vision." Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. IEEE, pp. 253-256. 2010.

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. pp. 1097-1105. 2012.

[5] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556. 2014.

[6] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 pp. 2278-2324. 1998.

[7] Duchi, John, Elad Hazan, and Yoram Singer. "Adaptive subgradient methods for online learning and stochastic optimization." The Journal of Machine Learning Research pp. 2121-2159. 2011.

[8] Shufelt, Jefferey. "Performance evaluation and analysis of monocular building extraction from aerial imagery." Pattern Analysis and Machine Intelligence, IEEE Transactions on 21.4. pp. 311-326. 1999.