The Detection of Built-up Areas in High-Resolution SAR Images Based on Deep Neural Networks

Yunfei Wu[1,2], Rong Zhang[1,2(✉)], and Yue Li[1,2]

[1] Department of Electronic Engineering and Information Science, USTC, Hefei 230027, China
{wuyunfei,lyue}@mail.ustc.edu.cn, zrong@ustc.edu.cn
[2] Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei 230027, China

Abstract. The detection of built-up areas is an important task for high-resolution Synthetic Aperture Radar (SAR) applications, such as urban planning and environment evaluation. In this paper, we proposed a deep neural network based on convolutional neural networks for the detection of built-up areas in SAR images. Since lables of neighboring pixels have strong correlation in SAR images, informations on labels of neighboring pixels could help making better prediction. In addition, built-up areas in SAR images possess various scales, multiscale representations is critical for the detection of built-up areas. Based on above observations, we introduce the structured prediction into our network, where a network classifies multiple pixels simultaneously. Meanwhile, we attempt to adopt multi-level features in our network. Experiments on TerraSAR-X high resolution SAR images over Beijing show that our method outperforms traditional methods and CNNs methods.

Keywords: High-resolution SAR images · Detection of built-up areas · Structured prediction · Multi-level · Deep neural networks

1 Introduction

In recent years, the urban area is developing rapidly. Consequently, the monitoring and planning of urban areas become an important research field. Different from optical sensors, Synthetic Aperture Radar (SAR) is independent from sun illumination and weather conditions, which makes the information in SAR very useful for cities. In that case, the utilization of SAR data for the monitoring of urban areas has become the topic of recent discussions. Built-up area is the most obvious sign of urban areas. Detection of built-up areas in SAR images promises several applications, such as urban planning, disaster assessment, environmental monitoring. Therefore, the detection of built-up areas is of great importance.

Different techniques for built-up areas detection have been presented in literature. Borghys et al. [1] proposed an automatic detection method of built-up areas in high-resolution polarimetric SAR images in which most features are based on statistical properties of built-up areas. Yang et al. [2] developed a method for the land-over classification of TerraSAR-X imagery over urban areas used texture features. Li et al. [3]

Y. Zhao et al. (Eds.): ICIG 2017, Part III, LNCS 10668, pp. 646–655, 2017.
https://doi.org/10.1007/978-3-319-71598-8_57

employed Labeled Co-occurrence Matrix for the detection of built-up areas in high-resolution SAR images. Generally speaking, the most challenge problem of the detection of built-up areas in SAR images is feature extraction. The features used in all the aforementioned works are hand designed with domain knowledge and can significantly impact the classification accuracy.

Recently, deep learning, especially convolutional neural networks (CNNs) [4, 5], has achieved much success in visual recognition tasks, for instance, object detection and image classification. Experiments showed that features extracted from CNNs are effective and powerful [6, 7]. Lately, CNNs has been applied to the detection of the built-up areas in SAR images [8]. With the help of powerful features extracted by CNNs, Li et al. [8] achived state-of-art result. However, such a method classifies pixels separarely and ignore the strong correlation on labels between neighboring pixels. As we know, pixels belongs to background is more likely adjacent with background pixels than pixels belong to built-up areas in SAR images. We can obtain better result if we could make use of the informations on labels of neighboring pixels.

In this paper, we proposed a deep neural network based on CNNs for the detection of built-up areas in SAR images. To make use of the informations on labels of neighboring pixels, our network is designed to be able to obtain multiple lables for pixles at the same time. In addition, since the built-up areas in SAR images possess various scales, we try to adopt multiscale features in ournetwork. Features extracted from different *conv* layers possess different receptive field sizes, making full use of them could help to detect built-up areas in various scales. We observed the results getting from all *conv* layers in our networks, and discovered that they can be complementary for the detection of built-up areas in SAR images. Based on the above observation, we adopt multi-level features in our network.

The rest of this paper is organized as follows. In Sect. 2, we describe the method we proposed in detail. Section 3 shows the experiments and results. We present our conclusion in Sect. 4.

2 Structured Prediction

By automatically learning hierarchies of features from massive training data, CNNs obtained state-of–art results in most visual tasks of natural images, such as classification [9] and object detection [10]. Inspired of the great success made in natural images, several reseachers have attempt to adopt CNNs to proccess SAR data [11, 12]. Since built-up areas in SAR images are rich of structure informations, Li et al. [8] proposed a multiscale CNN model to extract the features of built-up areas to detect the built-up areas in SAR images. By densely predicting patches in SAR images, Li obtained good detection result compared with traditional methods. However, the multiscale CNN model classifies individual pixels separarely. As a result, the strong correlation on labels between neighboring pixels in SAR images would be ignored. It is well known that pixels in background is more likely to be adjacented with background pixels than pixels in built-up areas in SAR images. The information on labels of neighboring pixels could help making better decision.

As pointed by Liskowski et al. [13], we could make use of the information if labels for all pixels are available at the same time. In their work, Liskowski posed blood vessels segementation task as multilable inference problem on a set of binary predictions subject to a joint loss. This is a special case of structured prediction [14].

The structured prediction (SP) networks is designed to obtain information on multiple labels of pixels at the same time (Fig. 1). It can be achived by slight modification on existing deep architectures: we only need to set the number of units in final fully connected layer as m^2, which indicate that if the centural m^2 pixels of input patch are belong to built-up areas in SAR images. The loss function of SP network employs cross entropy (CE) loss:

$$J_{CE}(\hat{y}, y) = -\sum_i (y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)) \tag{1}$$

where $\hat{y}_i$ and y_i are the prediction and the target for i th output node.

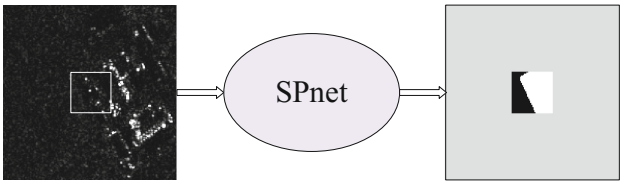

Fig. 1. An example for structured prediction: the n*n patch processed by CNN and get m*m labels of central m*m pixels of input patch.

In order to better analyse our method, i.e. the improved SP network, in the following experiment, we explore two kinds of SP networks: the plain SP networks and the

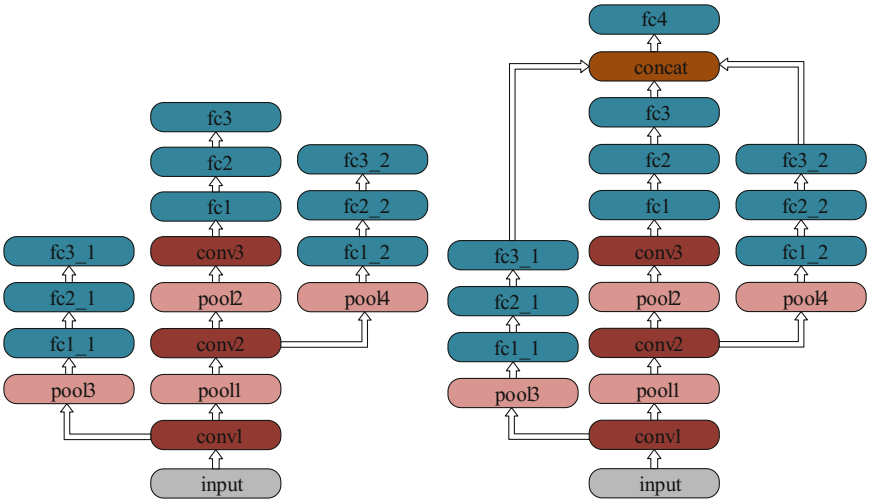

Fig. 2. Models of two SP networks: left is the plain SP networks, right is the improved SP networks

improved SP networks. The models of these two networks are shown in Fig. 2, and the architectures are shown in Table 1. Pointed by Peng et al. [15], large kenel helps obtaining better performance. And Li et al. [8] indicate that large kenel help to reduce the effect caused by strong speckle noise in SAR images. To such consideration, we employ large kernel in convolutional layers in these two kinds of SP networks.

Table 1. Architecture of SP networks.

Layer Name	Kenel size*channel	Stride
conv1	9*9*50	1
pool1	4*4	4
conv2	8*8*100	1
pool2	4*4	4
conv3	3*3*300	1
fc1/fc1_1/fc1_2	1000	
fc2/fc2_1/fc2_2	1000	
fc3/fc3_1/fc3_2/fc4	100	
pool3	4*4	4
pool4	2*2	2

In both two SP networks, Rectification non-linearity was used in used in all convolutional layers and fully connected layers to accelerate the convergence of stochastic gradient decent. In adidtion, drop out layer is employed in first two fully connected layers.

2.1 The Plain SP Network

The plain SP network is a sequential combination of convolutional layers, maxpooling layers and fully connected layers. However, since we introduce large kenel in networks, SP networks would be hard to train and easily encounter the problem of overfitting. In consideration of such circumstances, we add extra supervision to the plain SP networks. As pointed out by [16], extra supervision using hidden layer feature maps leads to reduction in testing error. In prediction stage, the result obtained by extra supervison will be abandoned.

2.2 The Improved SP Network

The model of the improved SP network is shown in Fig. 2. Each *conv* layer in the improved SP network is connected to a stack of fully connected layers. And the results are concatenated and processed by a fully connected layer to obtain final classifier output.

The motivation behind this is that we would like to introduce multiscale features to the final classifier. As we know, the neurons in different levels have different receptive field sizes, they can be seen as representations of multiscale. Considering the dynamics

of the sizes of built-up areas in SAR images, we hypothesize that the hierarchical information could help to make better decision.

Thus, we combine hierarchical features from all the *conv* layers. Instead of combining hierarchical features directly, we choose to combine them after they are proccessed by serveral fully connected layers so that features to be concatenated are trained to be more discriminative.

Since receptive field sizes in different *conv* layers in our network are different, our network could learn multiscale features. Such information is helpful for the detection fo built-up areas in SAR images. SAR images are corrupted by speckle noise, which could significantly impact the detection result. And the sizes of built-up areas in SAR images are so dynamic. By introducing multi-level features, our network could have the ability to suppress the effect of speckle noise and obtain good detection result at the same time.

We show the intermediate results of the plain SP network in Fig. 3. From left to right, the receptive field sizes decrease. We can see that under the complicated environment condition in SAR images, network with single receptive filed size can not always obtain satisfactory detection result. By embedding the multi-level features into classifier, our network is expected to achive better detection result.

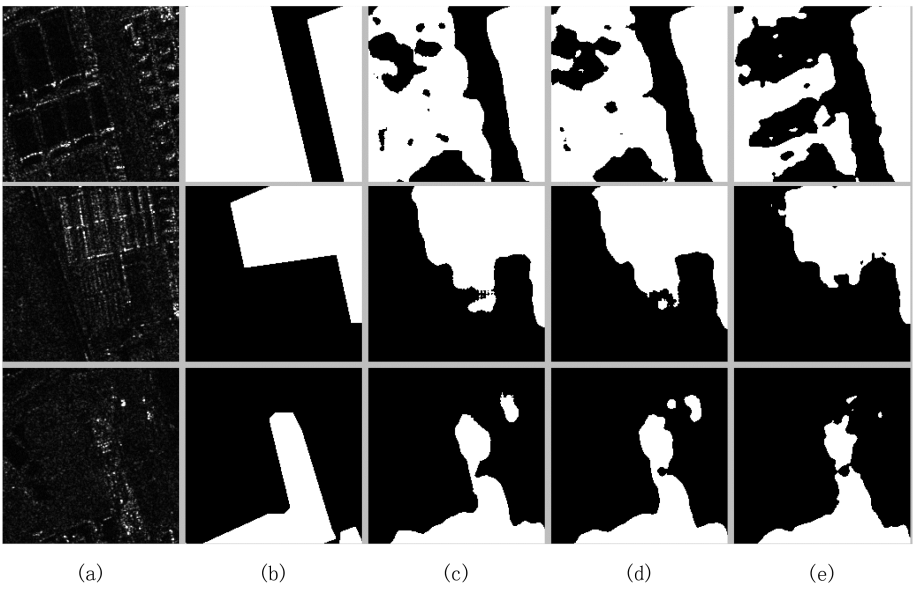

Fig. 3. Example of intermediate results of the plain SP network (a) Original image. (b) Reference. (c) Result of fc3 layer. (d) Result of fc3_2 layer. (e) Result of fc3_1 layer.

3 Experiment and Result

High-resolution TerraSAR-X SAR images of Beijing collected on November 25, 2011 were selected to verify our method. The SAR image is of range resolution of 2.3 m, and azimuth resolution of 3.3 m. The types of building areas in images includes Dot villa district, residential quarter buildings, squatter settlement and etc.

Training: We used caffe [17] to train our networks, and Stochastic Gradient Descent is used for training. The initial learning rate is 0.0001. We use momentum of 0.9 and weight decay of 0.0005.

Dataset: In the following experiments, we set the size of input patches as 84*84, and choose the output of our network as 10*10. We selected 90000 patches as train data, and 24000 patches as validation data. The test data is formed by an SAR image of 2500*4000 pixels.

Qualitative results: Fig. 4 shows the fragments of detection results obtained by multi-scale CNN and two kinds of the SP networks. The first column of Fig. 4 shows the detection result of road area in SAR images. We can see clearly from it that road areas are quite similar with build-up areas in SAR images, and by making use of the

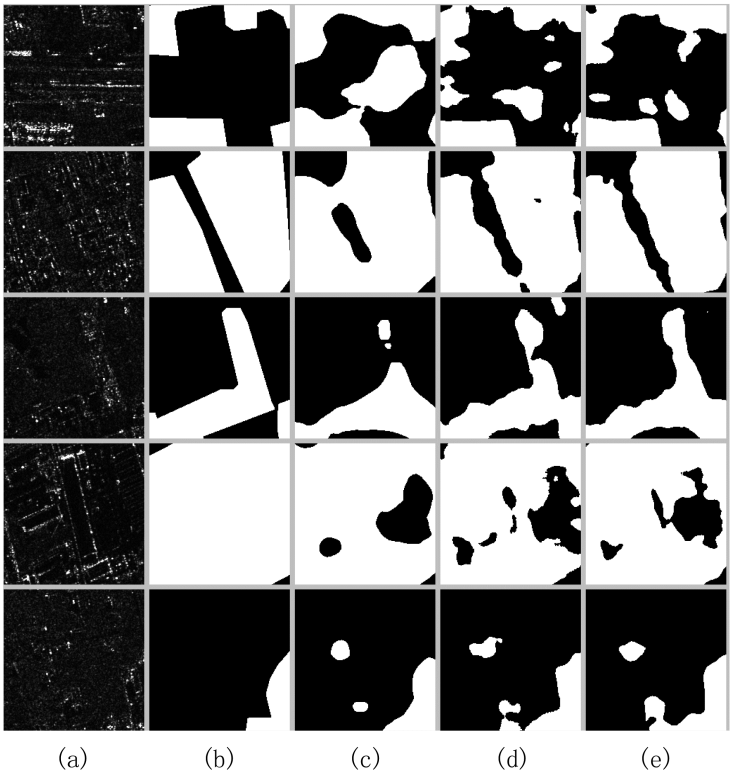

Fig. 4. Fragment of the detection result of SAR images (a) Original image. (b) Reference. (c) Multiscale CNN. (d) Plain SP network. (e) Improved SP network.

information on labels of neighboring pixels, the SP networks behave better than multi-scale CNN. The second column and the third column indicate that the SP networks obtain good results in building dense areas and "slender" built-up areas. The last two column of Fig. 4 are the failure examples of the SP networks, but we can see that the SP networks still obtain comparable results in such areas. In general, the improved SP network obtain better result in the examples by introducing multi-level features.

We visualize the entire detection result in Fig. 5. From Fig. 5(c), we can see that multiscale CNN model obtained a good performance, most built-up areas have been detected successfully. However, as mentioned above, we can see that multiscale CNN model performs not so satisfactory in road areas and building dense areas, and then cause

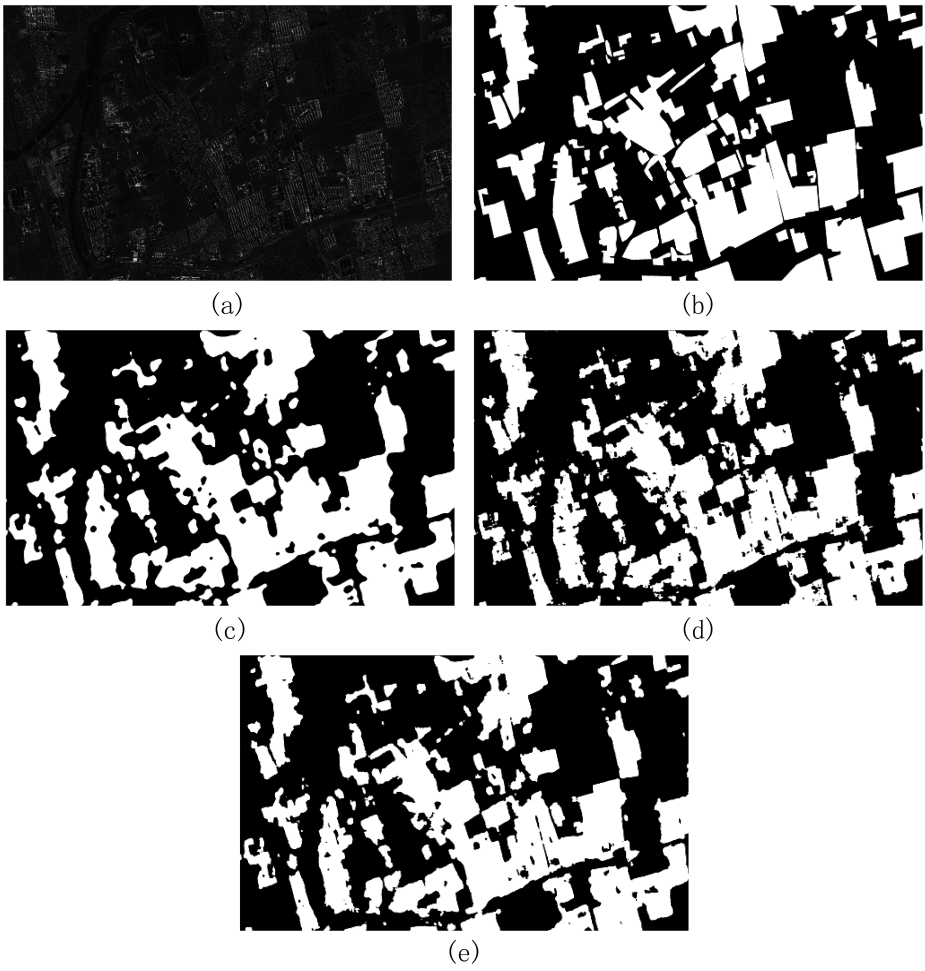

Fig. 5. Experiment results (a) SAR image of northern areas of Beijing. (b) Manually labeled image. (c) Detection result of multiscale CNN. (d) Detection result of the plain SP network. (e) Detection result of the improved SP network.

high false alarm rate in such areas. The performance of the plain SP network can be seen form Fig. 5(d). In general, the pain SP network achived better detection result than multiscale CNN. But since the plain SP network belongs to single scale network, it is not able to deal with the complex size of built-up areas in SAR images. And then the plain network obtain lower detection rate. The detection result of the improved SP network can be seen in Fig. 5(e), it shows that the improved SP network achived best result.

Pixel level results: Performance of pixel level accuracy is presented in Detection rate (DR), False alarm rate (FA), Accuracy of classification (Acc) [18], they are defined as:

$$DR = \frac{TP}{TP + FN}, \quad FA = \frac{FP}{TP + FP}, \quad Acc = \frac{TP + TN}{TP + TN + FP + TP},$$

where TP, TN, FP and FN are respectively the numbers of true positive, true negative, false positive, and false negative decisions.

In multiscale CNN, the three performance indicators are based on the output of network: positive decision is made if the network judge the input patch belongs to built-up areas, otherwise, negative decision is made. And in SP networks, the three performance indicators are based on the default interpretation of network decisions: positive decision is made if the output of the unit (sigmoid) is greater than 0.5 threshold, otherwise, negative decision is made.

Pixel level accuracy is shown in Table 2, the detection result of multiscale CNN is result of [8]. We can see that the improved SP network obtain best result on Detection rate and Accuracy of classification. On the False alarm rate, our method is a little higher than the plain SP network, we think that it is because network is hard to optimise when introducing multilevel features.

Table 2. Pixel level accuracy.

Method	Detection rate	False alarm rate	Accuracy of classification
GLCM	84.38%	15.82%	88.78%
LCM [3]	89.39%	23.40%	86.16%
CNN_{42}	90.43%	12.77%	90.52%
CNN_{84}	90.38%	17.10%	89.64%
Multiscale CNN [8]	92.14%	10.71%	92.86%
Plain SP network	91.00%	**9.08%**	93.18%
Improved SP network	**92.40%**	9.87%	**93.32%**

As mentioned above, in SP networks, positive decision is made if the output is greater than 0.5 threshold. However, this threshold is not enough to show the advantage of SP networks. Figure 6 shows the change of result when setting different thresholds in the improved SP network. From Fig. 6, we can find that the curve of Accuracy of classification changes slower near the threshold of 0.5. On the contrary, the Detection rate decrease when threshold increase. It indicates that we can slightly change the threshold to obtain different Detection rate and False alarm rate while keeping Accuracy of

classification in a stable state. For example, we can set the judge threshold smaller than 0.5 to get high performance of detection rate or bigger than 0.5 to get lower false alarm rate (Table 3). To some extent, it means that we can control the detection result by setting different thresholds.

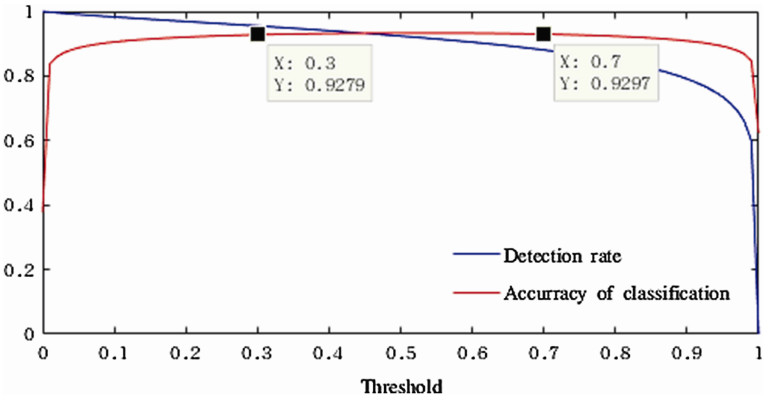

Fig. 6. Experiments results getting from different threshold.

Table 3. Pixel level accuracy.

Method	Detection rate	False alarm rate	Accuracy of classification
Multiscale CNN [8]	92.14%	10.71%	92.86%
Plain SP network-0.4	92.41%	10.17%	93.19%
Improved SP network-0.4	**94.04%**	11.44%	93.17%
Plain SP network-0.6	89.41%	**8.03%**	93.06%
Improved SP network-0.6	90.48%	8.43%	**93.27%**

In Table 3, the suffix "-0.4" or "-0.6" means we choose 0.4 or 0.6 as the threshold of SP networks. From Table 3, we can find that we could obtain controllable results by setting different thresholds.

4 Conclusion

In this paper, we proposed an improved structured prediction network for the detection of built-up areas in SAR images. By making use of the information on labels of neighboring pixels and multi-level features, our network achived success in the detection of built-up areas in SAR images. In particularly, we can obtain controllable results by setting different thresholds on the output of the improved SP networks. the experiments carried out on TerraSAR-X SAR image of Beijing confirmed that our method is effective to detect built-up areas in SAR images.

Acknowledgment. This work was supported in part by the National Nature Science Foundation of China (No.61331020).

References

1. Borghys, D., Perneel, C., Acheroy, M.: Automatic detection of built-up areas in high-resolution polarimetric SAR images. Pattern Recogn. Lett. **23**(9), 1085–1093 (2002)
2. Yang, W., Zou, T., Dai, D., et al.: Supervised land-cover classification of TerraSAR-X imagery over urban areas using extremely randomized clustering forests. In: Urban Remote Sensing Event, 2009 Joint, pp. 1–6. IEEE (2009)
3. Li, N., Bruzzone, L., Chen, Z., et al.: Labeled co-occurrence matrix for the detection of built-up areas in high-resolution SAR images. In: SPIE Remote Sensing. International Society for Optics and Photonics, p. 88921A-88921A-12 (2013)
4. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 253–256. IEEE (2010)
5. Krizhevsky, A., Sutskever, I., Hinton, G E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
6. Karpathy, A., Toderici, G., Shetty, S., et al.: Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1725–1732 (2014)
7. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv: 1408.5882 (2014)
8. Li, J., Zhang, R., Li, Y.: Multiscale convolutional neural network for the detection of built-up areas in high-resolution SAR images. In: 2016 IEEE International on Geoscience and Remote Sensing Symposium (IGARSS), pp 910–913. IEEE (2016)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
10. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
11. Chen, S., Wang, H., Xu, F., et al.: Target classification using the deep convolutional networks for SAR images. IEEE Trans. Geosci. Remote Sens. **54**(8), 4806–4817 (2016)
12. Gong, M., Zhao, J., Liu, J., et al.: Change detection in synthetic aperture radar images based on deep neural networks. IEEE Trans. Neural Netw. Learn. Syst. **27**(1), 125–138 (2016)
13. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016)
14. Bakir, G.: Predicting Structured Data. MIT press, Cambridge (2007)
15. Peng, C., Zhang, X., Yu, G., et al.: Large kernel matters–improve semantic segmentation by global convolutional network. arXiv preprint arXiv:1703.02719 (2017)
16. Lee, C Y., Xie, S., Gallagher, P W., et al.: Deeply-supervised nets. AISTATS. 2(3), p. 5 (2015)
17. Jia, Y., Shelhamer, E., Donahue, J., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)
18. Shufelt, J.A.: Performance evaluation and analysis of monocular building extraction from aerial imagery. IEEE Trans. Pattern Anal. Mach. Intell. **21**(4), 311–326 (1999)