SUPERPIXEL CLASSIFICATION OF HIGH SPATIAL RESOLUTION REMOTE SENSING IMAGE BASED ON MULTI-SCALE CNN AND SCALE PARAMETER ESTIMATION

Yangyang Chen [1], Dongping Ming [1, *]

[1] School of Information Engineering, China University of Geosciences (Beijing), 100083, Beijing, China-
jimmyxiyangyang@hotmail.com (Y. Chen), mingdp@cugb.edu.cn (D. Ming)

KEY WORDS: Deep Learning, Spatial Statistics, High Spatial Resolution Remote Sensing Image, Image Segmentation, OBIA

ABSTRACT:

In recent years, considerable attention has been paid to integrate convolutional neural network (CNN) with land cover classification of high spatial resolution remote sensing image. Per-pixel classification method based on CNN (Per-pixel CNN) achieved higher accuracy with the help of high-level features, however, this method still has limitations. Even though per-superpixel classification method based on CNN (Per-superpixel CNN) overcome the limitations of per-pixel CNN, classification accuracy of complex urban is easily influenced by scale effect. To solve this issue, superpixel classification method combining multi-scale CNN (Per-superpixel MCNN) method is proposed. Besides, this paper proposes a novel spatial statistics based method to estimate applicable scale parameter of per-superpixel CNN. Experiments using proposed method were performed on Digital Orthophoto Quarer Quad (DOQQ) images in urban and suburban area. Classification results show that per-superpixel MCNN can effectively avoid misclassification in complex urban area compared with per-superpixel classification method combining single-scale CNN (Per-superpixel SCNN). Series of classification results also show that using the pre-estimated scale parameter can guarantee high classification accuracy, thus arbitrary nature of scale estimation can be avoided to some extent.

1. INTRODUCTION

Spatial details of high spatial resolution remote sensing images have improved dramatically compared with low spatial resolution remote sensing images (Benediktsson et al., 2013). High heterogeneity and low homogeneity of land cover bring difficulties and challenges into image interpretation(Myint et al., 2011; Zhao and Du, 2016a). Pixel-based analysis method and object-based image analysis (OBIA) methods (Blaschke et al., 2008), which use low-level features, leads to misclassification of complex land cover and have failed to gain satisfactory classification results of high spatial resolution remote sensing images. The difficulty in high-level features extraction is a bottleneck that restricts the improvement of image classification accuracy (Pan et al., 2019).

With its strong robustness, generalization capability and feature extraction ability, convolutional neural network (CNN) (Fukushima and Miyake, 1982) is applicable to extract high-level features from remote sensing images (Nogueira et al., 2017). In the last few years, considerable attention has been paid to solve hotspot issues in remote sensing field (e.g., object detection, scene classification and land cover classification). As for CNN based land cover classification, per-pixel classification method based on convolutional neural network (CNN) (Per-pixel CNN) (Längkvist et al., 2016; Zhao and Du, 2016b) achieved higher accuracy with the help of high-level features. However, salt-pepper effect and heavy computation burden, have not yet been entirely overcome. Per-superpixel classification method based on CNN (Per-superpixel CNN) (Audebert et al., 2016; Gonzalo-Martin et al., 2016; Zhao et al., 2017) overcome the limitations of per-pixel CNN by integrating the theory of OBIA, there are still some scale related issues in per-superpixel CNN needed to be explored and studied urgently. Classification accuracy of complex urban is easily influenced by scale effect. Considering the scale of land cover objects is inconsistent, per-superpixel classification method combining

single-scale CNN (Per-superpixel SCNN) can not extract and utilize high level feature of land cover objects. Besides, estimation of scale parameter of CNN directly influences the accuracy of classification result. But in previous studies, there is blindness in scale parameter estimation. To solve above issues, this paper proposes a superpixel classification method combining multi-scale CNN (Per-superpixel MCNN) and a spatial statistics based scale parameter estimation strategy. By carrying out experiments on Digital Orthophoto Quarer Quad (DOQQ) in urban and suburban areas, feasibility and effectiveness of proposed are analysed and discussed.

2. METHODS

The outline of proposed method is divided into four main stages: segmentation, scale parameter estimation, CNN training and CNN classification, as illustrated in Figure 1. Firstly, the high spatial resolution remote sensing image is segmented by using superpixel segmentation algorithms. In the meanwhile, the semivariance and the increase of synthetic semivarance of image are calculated to estimate appropriate scale parameter. Using estimated scale parameters, training data and test data are generated based on segmentation result. After that, training data on single-scale / multi-scale are used to train single-scale CNN / multi-scale CNN. In the end, by using the trained single-scale CNN / multi-scale CNN, the VHR remote sensing image are classified.

Figure 1. Framework of per-superpixel CNN method combining scale parameter estimation

2.1 Superpixel Classification

To take full advantage of CNN and OBIA, a segmented object is designated as the basic analysis unit of CNN's training and classification procedure in per-superpixel CNN. Limited by the structure and theory of CNN, data fed into the input layer must be a fixed size image. segmented objects generated by traditional segmentation methods (such as split and merge (Gonzalez and Woods, 1992) or multiresolution segmentation algorithm (Darwish et al., 2003)) vary greatly in size and shape. Uncertainty about objects' locations and boundaries has negative impacts on the classification of superpixel and causes misclassification. For its ability to generate compact and uniform superpixel-level over-segmented objects (Neubert and Protzel, 2012; Ren and Malik, 2003), superpixel segmentation has emerged as candidate segmentation algorithm in proposed method.

2.2 Per-superpixel Classification Using Pre-trained CNN

After superpixel segmentation, the generated superpixel-level segmented objects can be classified by using high-level features extracted by CNN.

For per-superpixel SCNN, the squared image patch sharing the same geometric center is extracted. The land cover of the superpixel and its image patch constitutes the superpixel dataset as label and image data, respectively. The new established superpixel dataset is fed into CNN for network training and superpixel classification as shown in Figure 2.
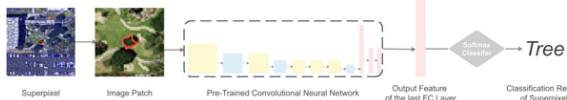

Figure 2. Framework of per-superpixel SCNN method (taking the recognition of tree as example)

The scale effect of remote sensing images has great influence on the classification accuracy, as image objects in different scales have different expressions of feature information (Fauvel et al., 2012; Ming et al., 2015). The feature expression of the image patches extracted from the same superpixel is varies from size to size. Therefore, in order to improve the accuracy of superpixel classification, multi-scale features of superpixels can be extracted by multi-scale convolutional neural network (MCNN).

The dataset for multi-scale superpixel-based image patches is built using the same strategy as per-superpixel SCNN as shown in Figure 3. An overview of proposed Per-superpixel MCNN is illustrated in Figure 4.
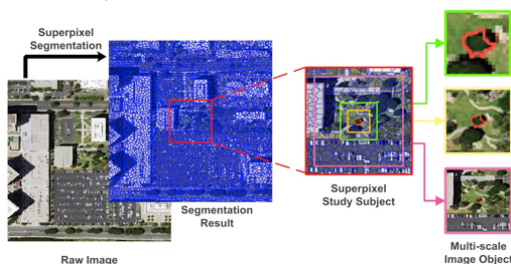

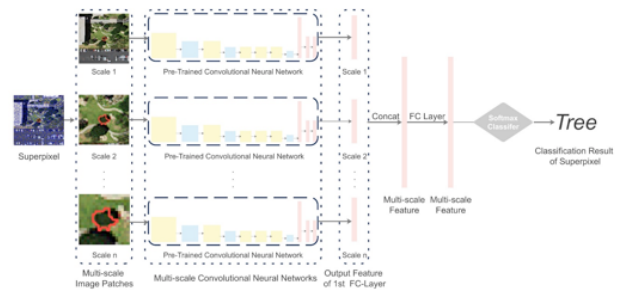Figure 3. Production of superpixel dataset in multiple scales

Figure 4. Framework of proposed per-superpixel MCNN method (taking the recognition of tree as example)

2.3 Scale Parameter Estimation of Per-superpixel CNN

Estimating the optimal size of square image patch m is crucial important to fully utilize the performance of per-superpixel CNN and improve the accuracy of final classification result. However, in previous studies, m was normally selected subjectively by human. Therefore, developing a novel method to estimate the size square image patch m is of great significant. For per-superpixel SCNN, a simple and feasible way to estimate the scale parameter is to calculate the semivariance with various sampling intervals on multiple directions, and the optimal scale parameter m can be estimated using the increase of synthetic semivariance. The variation of the semivariogram is calculated as:

$$\gamma_s(h)_i = [\gamma_h(h)_i + \gamma_v(h)_i] \tag{1}$$

$$\Delta\gamma_s(h)_i = \gamma_s(h)_i - \gamma_s(h)_{i-1} \tag{2}$$

where $\gamma_s(h)_i$ represents the synthetic semivariance with computation interval $(h)_i$, i.e., the average value of horizontal semivariance $\gamma_h(h)_i$ and vertical semivariance $\gamma_v(h)_i$. $\Delta\gamma_s(h)_i$ represents the variation of synthetic semivariance with computation interval $(h)_i$. According to (Ming et al., 2012), the optimal size of the computation window W_s (size of medium scale square image patch in per-superpixel SCNN, which is supposed to approximate to the object size) can be calculated by:

$$m = W_s = 2 \times h_s + 1 \tag{3}$$

where h_s is the value of the lag when the $\Delta\gamma_s(h)_i$ is first less than or equal to zero with the increasing h.

For per-superpixel MCNN, the combination of three scales of square image patch (Small-scale m_s, medium-scale m_m and large-scale m_l) is enough for to fully extract the high level features at different scale (Hu et al., 2015). The choices of small-scale image patch size and large-scale image patch size are usually processed manually. Extracting medium-scale features is equally important, for they contain the information about relationship between the superpixel and its environment. Given the above, medium-scale parameter m_m should be identical to the optimal size of single-scale square image patch m of per-superpixel SCNN. Therefore, the proposed semivariogram-based method is also suitable for the estimation of medium-scale m_m in per-superpixel MCNN.

3. EXPERIMENTAL RESULTS AND DISCUSSION

This paper takes two Digital Orthophoto Quarer Quad (DOQQ) images (1.0m per pixel) in the experiments as shown in Figure 5. Experiments were carried out on a Mac OS with single CPU (2.7 GHZ core i7), and TensorFlow was chosen as the deep learning platform.
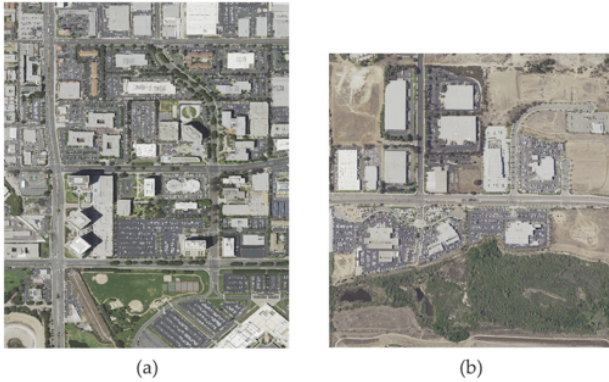

Figure 5. Experimental data. (a) Urban scene; (b) Suburban scene

There are 7 categories of land cover considered in the Urban scene: building, parking space, road, shadow, soil, tree and other vegetation. The suburban scene is composed of 7 land cover categories: building, parking space, road, shadow, soil, vegetation and water. Training and validation datasets were manually labeled through careful visual interpretation. To avoid data redundancy, location of each labeled pixel is the centroid of superpixel. Forty percent of the labeled dataset was randomly selected as the training dataset, while 10 % was selected to be the validation dataset. Validation points for classification accuracy evaluation were randomly selected considering the actual area distribution and complexity of each category of land cover. More than 1000 and 700 validation points were selected in urban scene and suburban scene respectively.

Two state-of-the-art superpixel segmentation algorithms: zero parameter version of SLIC (SLICO) (Achanta et al., 2012) and superpixels extracted via energy-driven sampling (SEEDS) (Van den Bergh et al., 2015) were applied. Two superpixel segmentation algorithm, superpixels extracted via energy-driven sampling(SEEDS) and zero parameter version of SLIC (SLICO) are imployed to classify urban and suburban land covers. The average size of superpixel objects is set as 5 * 5 for SLICO and 6 * 6 for SEEDS through repeated experiments of trial and error.

To implement the per-superpixel SCNN and per-superpixel MCNN, the appropriate scale parameters were estimated by using proposed spatial statistics method, as shown in Figure 6 and Figure 7. The optimal single-scale parameter m in per-superpixel SCNN was estimated as 15, the small-scale parameter m_s, medium-scale parameter m_m and large-scale parameter m_l were selected as 15, 25, and 45 respectively in per-superpixel MCNN.
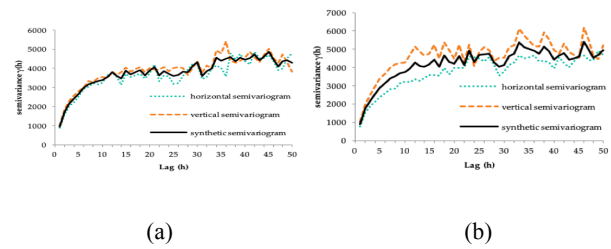

Figure 6. Semivariogram of the experimental images. (a) Urban Scene; (b) Suburban scene
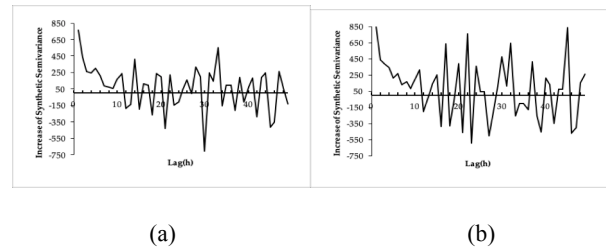

Figure 7. Computed results of the increase of the synthetic semivariance. (a) Urban scene; (b) Suburban scene.

To prove the superiority of per-superpixel CNN, multiple existing methods were used in two study area. the OBIA method based on support vector machine (SVM) classifier (OBIA-SVM) was implemented via eCognition Developer 9.0 with multiresolution segmentation algorithm (MRS) (Darwish et al., 2003). Series of experimental results (Figure 8, Figure 9, Figure 10 and Table 1) show that per-superpixel MCNN can effectively avoid misclassification in complex urban area compared with Per-superpixel SCNN. Classification results also show that using the pre-estimated scale parameter can guarantee high classification accuracy, thus arbitrary nature of scale estimation can be avoided to some extent.
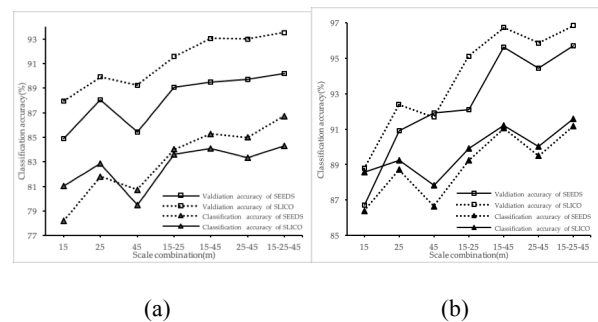

Figure 8. Validation accuracy and classification accuracy with different m.

	BD	PS	RD	SD	SO	VG	WT	OA
OBIA-SVM	63.41	41.90	71.85	54.17	93.70	96.14	100.00	81.20
Per-pixel SCNN	84.14	78.38	82.22	62.50	93.70	96.14	100.00	88.98
Per-pixel MCNN	95.12	82.43	83.70	70.83	94.54	95.65	100.00	91.18
SLICO-SCNN	92.68	79.38	79.26	79.17	92.86	94.69	100.00	89.23
SLICO-MCNN	95.12	82.43	86.67	83.33	93.70	94.69	100.00	91.57
SEEDS-SCNN	84.15	77.03	81.48	62.50	93.28	96.62	100.00	88.72
SEEDS-MCNN	92.68	75.68	82.96	83.33	94.96	97.58	100.00	91.18

Table 2. Classification accuracy (In percentage) comparison of Suburban scene (BD: Building, PS: Parking space, RD: Road, SD: Shadow, SO: Soil, VG: Vegetation, WT: Water, OA: Overall accuracy)
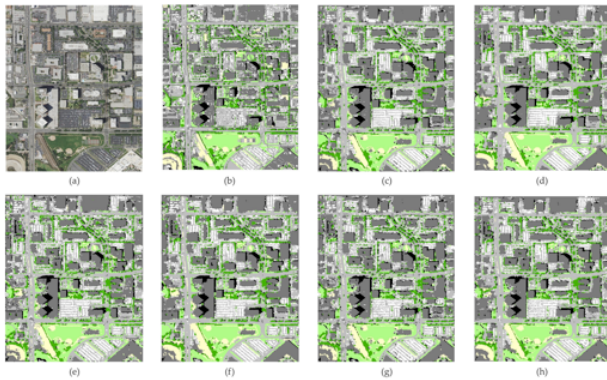
Figure 9. Classification result comparison of different methods for Urban scene. (a) Original image (b) OBIA-SVM; (c) Per-pixel SCNN; (d) Per-pixel MCNN; (e) SLICO based SCNN; (f) SLICO based MCNN; (g) SEEDS based SCNN; (h)SEEDS based MCNN
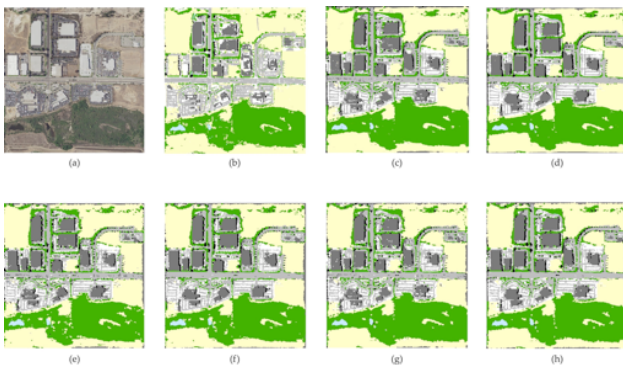

Figure 10. Classification result comparison of different methods for Suburban scene. (a) Original image (b) OBIA-SVM; (c) Per-pixel SCNN; (d) Per-pixel MCNN; (e) SLICO based SCNN; (f) SLICO based MCNN; (g) SEEDS based SCNN; (h)SEEDS based MCNN

	BD	PS	RD	SD	SO	TR	OV	OA
OBIA-SVM	81.92	42.62	78.72	90.56	54.38	84.61	81.82	74.32
Per-pixel SCNN	95.38	79.51	78.97	79.25	70.97	71.15	79.09	82.17
Per-pixel MCNN	94.23	80.33	79.26	77.35	87.10	85.58	79.09	83.91
SLICO-SCNN	90.38	69.97	80.68	84.91	77.42	86.54	83.64	82.49
SLICO-MCNN	94.23	75.41	80.97	83.02	83.87	82.69	83.64	84.30
SEEDS-SCNN	93.46	71.31	77.56	90.57	80.65	73.08	83.64	81.78
SEEDS-MCNN	94.62	77.05	84.66	96.77	90.57	82.69	84.55	86.72

Table 1. Classification accuracy (In percentage) comparison of Urban scene (BD: Building, PS: Parking space, RD: Road, SD: Shadow, SO: Soil, TR: Tree, OV: Other vegetation, OA: Overall accuracy)

Through making full use of the multi-scale high level features and estimating the applicable scale parameter estimated by the proposed spatial statistics based method, the classification accuracy of the per-superpixel MCNN is well ahead of per-superpixel SCNN in two study areas. Classification results of urban and suburban Vista scene show that applying multi-scale strategy to per-superpixel CNN is of great significance to improve the classification accuracy of high spatial resolution remote sensing images, especially for complex urban area. Additionally, the per-superpixel CNN achieved higher accuracy than the OBIA-SVM method, which proved the superiority of using high-level features in classification.

The influence of scale parameter m on the final classification results is evaluated by using SCNN. With the help of the proposed spatial statistics based scale parameter estimation method, classification results with medium-scale parameters are superior for striking the balance between the features extracted from superpixels and their surrounding environment. As can be seen in Figure 8, the classification results with scale parameter m=25 achieved higher accuracy than the classification results with other scale parameters. Given the above, proposed scale parameter estimation method provides a new and feasible way for scale parameter estimation of per-superpixel SCNN, which is is of great significance to improve the classification accuracy.

The proposed per-superpixel MCNN method made comprehensive use of the features in three different scale. Using the proposed scale parameter estimation method to estimate the medium one of three scale parameter that is statistically suitable to the objects that appear most frequently, which is helpful for overcoming the blindness in parameter configuration. Even though the selection of other two scale parameters relies on human intervention, the proposed scale parameter estimation strategy is still instructively meaningful to further improve the final classification accuracy.

4. CONCLUSIONS

This paper proposes a superpixel-based land cover classification method combining multi-scale CNN and a novel scale parameter estimation strategy. Through making full use of the multi-scale high level features the classification accuracy of the per-superpixel MCNN is well ahead of per-superpixel SCNN which only utilizes single-scale features. Besides, proposed scale parameter estimation method provides a new and feasible way for scale parameter estimation of per-superpixel SCNN and

per-superpixel MCNN, which is is of great significance to improve the classification accuracy. Moreover, proposed scale parameter estimation strategy solves the problem that the estimation of scale parameter mainly depends on human's subjective judgement and experience to a certain extent. Even though the selection of other two scale parameters of per-superpixel MCNN relies on human intervention, the proposed scale parameter estimation strategy is still instructively meaningful to further improve the classification accuracy of land cover.

However, the size of two study areas is far from real application. Multiple type of remote sensing images and larger scale study area should be considered in the future research, in order to further demonstrate the validity and feasibility of proposed methods and superpixel segmentation algorithms,

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence, 34, 2274-2282.

Audebert, N., Le Saux, B., Lefevre, S., 2016. How useful is region-based classification of remote sensing images in a deep learning framework?, Proceedings of 2016 IEEE International Geoscience & Remote Sensing Symposium (IGARSS'16). IEEE, pp. 5091-5094.

Benediktsson, J.A., Chanussot, J., Moon, W.M., 2013. Advances in Very-high-resolution Remote Sensins, Proceedings of the IEEE. IEEE, pp. 566-569.

Blaschke, T., Lang, S., Hay, G., 2008. Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications. IEEE Transactions on Geoscience and Remote Sensing, 65, 2-16.

Darwish, A., Leukert, K., Reinhardt, W., 2003. Image segmentation for the purpose of object-based classification, Proceedings of 2003 IEEE International Geoscience & Remote Sensing Symposium (IGARSS'03) IEEE, pp. 2039-2041.

Fauvel, M., Chanussot, J., Benediktsson, J.A., 2012. A spatial–spectral kernel-based approach for the classification of remote-sensing images. Pattern Recognition, 45, 381-392.

Fukushima, K., Miyake, S., 1982. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition, Competition and cooperation in neural nets. Springer, pp. 267-285.

Gonzalez, R.C., Woods, R.E., 1992. Digital image processing. Addison-wesley, Boston.

Gonzalo-Martin, C., Garcia-Pedrero, A., Lillo-Saavedra, M., Menasalvas, E., 2016. Deep learning for superpixel-based classification of remote sensing images, GEOBIA 2016.

Hu, F., Xia, G.-S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7, 14680-14707.

Längkvist, M., Kiselev, A., Alirezaie, M., Loutfi, A., 2016. Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sensing, 8, 329.

Ming, D., Ci, T., Cai, H., Li, L., Qiao, C., Du, J., 2012. Semivariogram-based spatial bandwidth selection for remote sensing image segmentation with mean-shift algorithm. IEEE Geoscience and Remote Sensing Letters, 9, 813-817.

Ming, D., Li, J., Wang, J., Zhang, M., 2015. Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. ISPRS Journal of Photogrammetry and Remote Sensing, 106, 28-41.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote sensing of environment, 115, 1145-1161.

Neubert, P., Protzel, P., 2012. Superpixel benchmark and comparison, Proceedings of Forum Bildverarbeitung, pp. 1-12.

Nogueira, K., Penatti, O.A., dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 61, 539-556.

Pan, X., Zhao, J., Xu, J., 2019. An object-based and heterogeneous segment filter convolutional neural network for high-resolution remote sensing image classification. International Journal of Remote Sensing, 1-25.

Ren, X., Malik, J., 2003. Learning a classification model for segmentation, Proceedings of 2003 IEEE International Conference on Computer Vision. IEEE, pp. 10-17.

Van den Bergh, M., Boix, X., Roig, G., Van Gool, L., 2015. Seeds: Superpixels extracted via energy-driven sampling. International Journal of Computer Vision, 111, 298-314.

Zhao, W., Du, S., 2016a. Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 113, 155-165.

Zhao, W., Du, S., 2016b. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing, 54, 4544-4554.

Zhao, W., Jiao, L., Ma, W., Zhao, J., Zhao, J., Liu, H., Cao, X., Yang, S., 2017. Superpixel-based multiple local CNN for panchromatic and multispectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 55, 4141-4156.