GaoFen-1 Remote Sensing Image Forest Extraction

Using Object-based CNN

Jie Huang 1, Xiangxiang Zheng 2,* , Dongping Ming 1,* , Yangyang Chen 1 and Keqi Zhou 1

1. School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China; 2104170043@cugb.edu.cn (J.H.); cyy@cugb.edu.cn (Y.C.); 2004170025@cugb.edu.cn (K.Z.)
2. China Aero Geophysical Survey & Remote Sensing Center for Natural Resources, Beijing 100083, China;
* Correspondence: zhxagrs@163.com; Tel.: +86-10-152-1067-6081 (X.Z.) and mingdp@cugb.edu.cn; Tel.: +86-10-135-2090-7831 (D.M.)

Abstract: As an important natural resource, forest plays a vital role in regulating regional climatic conditions and maintaining the balance of the Earth's ecosystem. At the same time, the major changes in China's natural resource management system have put forward new requirements for forest surveys. Efficient and accurate grasp of the spatial distribution of forest resources and the type of forest in China have become an important part of all-weather remote sensing survey and monitoring of natural resources. In recent years, with the strong promotion of the national high-scoring project, the extraction of forest based on high-resolution remote sensing image (HSRI) has become one of the main technical means of forest resource survey in China. However, HSRI contains abundant detailed information and spatial characteristic information, and spectral heterogeneity within the object is increased, which significantly causes difficulties in HSRI processing and information extraction. Therefore, aiming at the problem of extracting forest types from China's domestic GF-1 satellite image, this paper proposes a classification method based on feature learning using an object-oriented convolution neural network (CNN). First, the study area was under-segmented using the Multi-Resolution Segmentation (MRS) algorithm and referred Estimation of Scale Parameter algorithm to determining the segmented scale. Then, a CNN model was obtained for identifying forest type by parameter adjustment. Finally, the multi-layer feature learning of the CNN and the regional Majority Voting algorithm were combined to classify the forest types within the study region. The overall extraction accuracy of the proposed method in this paper was 0.88. The object-oriented method can avoid the “salt and pepper” phenomenon that exists in the pixel-based classification. CNN can extract the deep features contained in HSRI, which...
is effectively conductive to category identification. In addition, the method proposed in this paper has a greater applicability to forest extraction from HSRI.

**Keywords**: high spatial resolution images; forest extraction; multi-resolution segmentation; CNN.

1. **Introduction**

Forest resources are the mainstay of renewable natural resources and ecosystems on Earth. Furthermore, they are also an indispensable part that significantly influences the living and development of the human race. Since the distribution of forest resources is rapidly changing along with the advancement of urbanization, timely and accurate acquisition of forest information is important for forest policies and forest production [1,2]. Nowadays, the main means to acquire forest information is artificial interpretation using remote sensing image and field samplings. However, artificial interpretation is limited in timeliness and objectiveness [3,4]. With the development of remote sensing technology, HSRI, which is characterized by its fast acquisition speed, high spatial resolution, has become an ideal data source and is widely used to acquire and extract forest information [5-7] based on low-level features.

Different objects show diverse spectral features. Object information contained in HSRI is highly detailed, and the object’s spectrum distribution is more of variable and diverse. The difference between spectrums of similar objects become larger and the spectrums of different objects mutually overlap to increase the intraclass variance and decrease the interclass variance, therefore making the spectral domain of HSRI less statistically accurate [8].

Since the HSRI contains inadequate spectral information, Combining spectrum and texture features improves the accuracy of forest information extraction [9,10]. Wang et al. and Hu et al. synthetically utilized spectral and texture information and acquired the eigenvector of feature differences between forest and non-forest areas [11,12]. Hu et al. and Zhou et al. combined shape, spatial relation, and other features to perform a bottom-up cluster-based iterative algorithm to extract forest information according to seed primitives [13,14].

Though the HSRI provides lower spectrum bands, it provides rich spatial information, which enriches the classification and spectral statistics of objects. Various details and complex texture changes of HSRI lead to tremendous difficulties in the analysis and extraction of information. Although some simple features can be selected for classification, deep features of HSRI that are unexcavated lead to the underutilization of tremendous information.

With the development of science and technology, deep-learning technology has alleviated the lack of learning and application of high-level features in traditional research. In scene classification [15-18], target detection [19-21], semantic segmentation [22,23], thematic feature extraction [24-26], and so on, CNN's multi-layer feature learning proved its effectiveness [27]. In recent years, scholars have also applied CNN in the extraction of forest information. Some researchers apply multi-spectral data, hyperspectral data, LiDAR data, and other multi-source data as input data for the CNN model [28-31]. Multiple data is able to obtain rich feature information [32]. However, the acquisition of LiDAR data is difficult, and the integration with other data is more complicated. Meanwhile, the same problem exists with multi-source data. These methods are generally classified as the pixel-level classification, that is, all pixels in the
image are classified by a sliding window. However, the limitations of this pixel-level classification method are very large because there are ‘similar objects with different spectrums’ and ‘different objects with similar spectrums’. Therefore, extracted forest maps are usually significantly influenced by ‘salt and pepper’ noise.

Based on the Geographic Object-based Image Analysis (GeOBIA) [33], this paper proposes an HSRI forest extraction method. In the data-processing stage, a false color fusion image was obtained from preprocessed GF-1 image and served as the input of CNN. Then, a model obtained for identifying forest type by parameter adjustment. At last, the majority voting strategy was adopted to perform forest types classification on segmentation results to obtain fine extraction result of forest categories.

2. Data and Methods

As shown in Figure 1, the workflow of the experiment was divided into three stages: (1) Sample set construction, (2) Model training and testing, and (3) Classification and Evaluation.

In the data set construction stage, a raw image generated by GF-1, which was acquired in 2019, was firstly synthesized to serve as the input data of the neural network by carrying out false color fusion. After that, the false color image was under-segmented by MRS using its texture information. The multi-resolution under-segmentation algorithm can not only ensure high homogeneity inside segmented objects, but it also generates multi-scale segmented objects according to the scale characteristics of different objects. Then, training data points and validation data points were selected in the study area, while test data points were generated from the segmentation result. In the next stage, a neural network was trained by the training dataset. Through adjustment and optimization of parameters, the trained neural network had the ability to identify the vegetation category of test data points. In the final step, classified test data points were spatially correlated with the segmentation result, and the regional majority voting algorithm was applied to obtain the classification results.

![Figure 1. Workflow of forest types identification using object-based CNN.](image)

2.1. Data Sources and Study Area

The GF-1 satellite image used in the experiment was provided by the China Aero Geophysical Survey & Remote Sensing Center for Natural Resources. The GF-1 image
consists of four spectral bands: red, green, blue, and near infrared. The experimental data is obtained by image fusion of 2 meters and 8 meters image. The study area is Yakeshi City, which is located in the middle of Hulun Buir and the western slope of the middle section of the Greater Khingan Range of China, and is illustrated in Figure 2. The geographical coordinate center of the study area is 120°46'15"E, 49°02'30"N, and the image size is 14796 pixels × 15380 pixels. Forest coverage in the Greater Khingan Range is relatively high. Considering the near-infrared bands are sensitive to vegetation reflection and three-band data are required by Alexnet network, near-infrared, red, and green bands were selected to synthesize the false color image as the input data for the CNN network. The study area was classified into 7 ground object categories based on field investigation. The study area was classified into arbor, shrub, sparse wood, grass, farmland, construction and water. The image of the study area is shown in Figure 2. The field photographs of the ground object categories and the corresponding visual features of the false color high spatial resolution image are shown in Figure 3, which also served as the visual interpretation base for selecting the training sample.

Figure 2. The raw image data.
Figure 3. Field data and class features: (a) arbor; (b) shrub; (c) sparse wood; (d) grass; (e) farmland; (f) image visual features.

2.2. Multi-resolution Image Segmentation

MRS can be understood as a bottom-up region-growing process, starting from seed pixels, continuously merging under specific criteria to generate non-overlapping segmented objects [34], which is the key basic work of transforming remote sensing images from discrete regular pixels into homogeneous image object primitives [35,36], is the foundation for GeOBIA; however, confirming the best segmentation scale is still a challenge. According to the Estimation of Scale Parameter (ESP) algorithm and the Local Variance (LV) of the object heterogeneity in the scene, the variation of heterogeneity at different scales can be estimated [37]. The ideal segmentation scale can be obtained according to change rate of LV [38]. The curves of LV and ROC-LV at different scales presented in Figure 4 show that ROC-LV tended to be stable from the scale of 30. Therefore, the scale parameter trend graph and the boundary of typical ground object categories were combined to obtain the 30 scale as the final segmentation scale of the study area. At the same time, the segmentation results of several typical ground object categories under scale 30 are shown in Figure 5, which show that the segmented objects not only ensured consistency within the categories, but also the segmented boundaries distinguished different types of ground objects very well.

Figure 4. Estimation of segmentation scale parameter.
2.3. The Sample Set

Through continuous iterative learning by using training samples, and simultaneously optimizing and adjusting parameters of neural network, CNN acquires multi-layer abstract data features. Test data can be classified by the extracted deep features. Traditional machine learning classification methods utilize manually set features to train the model and get the classification results of data [39]. The selection of sample points categories referred to natural geographical data, natural resource interpretation data from 2016 to 2017, and field investigation data in August 2019. The data set of sample points was established manually by using ArcGIS software. Independency and non-repetition were the principles of the training set for the selection and labeling of the validation set. The distribution of sample points followed the principle of the overall distribution. The number of sample points referred to the proportional area of each ground object category. The input data of CNN was obtained by setting the size manually according to point sample coordinates information. Detailed information about the training and test sample sets is given in Table 1.

| Class       | 2019 train | 2019 validation |
|-------------|------------|-----------------|
| arbor       | 6906       | 2432            |
| construction| 290        | 132             |
| farmland    | 12196      | 3910            |
| grass       | 7733       | 3617            |
| shrub       | 633        | 231             |
| aPARSE      | 1036       | 357             |
| water       | 149        | 118             |

Table 1. The number of ground feature categories in the data set.
2.4. Alexnet Network

Alexnet was proposed by Alex in 2012 for classifying large image datasets, and won the championship of the annual image recognition competition [40]. This made CNN prosper again and gradually became the core algorithm for image classification. The size of the three-channel input image is 227 x 227. Alexnet consists of eight layers with weights, including 5 convolutional layers and 3 fully connected layers. The maximum pooling layer is connected to the first, second, and fifth convolutional layers respectively. The softmax classifier, which is linked by the output of the last layer of the fully connected layers, can be used to classify image into more than 1000 categories.

The structure of the Alexnet network is illustrated in Figure 7.

In order to reduce the impact of artificially selected samples, 80% of the sample points were selected from the training samples randomly before training all scale network models. These sample points were used as training samples and validation samples in a ratio of 4:1 in the model training. The initial learning rate of the network model was set as 0.01. Restricted by hardware conditions, it was impossible to input all data into the network at one time. By setting the batch size to 100, the input of multiple data was achieved. The epoch represents one traversal of all data and was set as 200 according to the validation precision. Meanwhile, to avoid overfitting of the network, the dropout rate was set as 0.5.

![Figure 6. Several types of sample legends.](image)

![Figure 7. The network structure of Alexnet. Conv: convolution layer; Max pooling: maximum pooling layer; FC: fully connected layer. The numbers below the figures are the sizes and dimensions of data or features.](image)
2.5. Regional Majority Voting

Generally, the shapes of the segmented objects generated by MRS are diverse, which results in it being incapable of ensuring the accuracy of object classification when only the central point of the region is used as the category identification benchmark. So, to solve the related problems, Lv et al. (2018)[7] proposed a majority voting strategy for region-based forest cover types classification by using CNN. The ideas and steps of the majority voting algorithm are as follows.

1. Center point generation: For a polygon $V$ with $n$ vertices, the coordinates of the center point can be calculated by Equation 3:

$$ C = \frac{1}{n} \sum_{i=1}^{n} V_i $$

where $V_1 V_2 \ldots V_{n-1} V_n$ are the vertex coordinates of the polygon $V$.

2. Random point generation: A specified number of random points can be generated by the polygon constraint. The random points should be in the coordinate range of the polygon domain. A random function was used to generate random points, by constraining $x$ and $y$ coordinates of the polygon domain. Then, the random points were determined to be in-plane or not using geometry properties of the polygon. If the condition was satisfied, the specified number of random points could be generated by continuously repeating the random points generation process.

3. The center points and random points were combined to obtain the test points for the vegetation region. The trained model was used to predict and assign the categories of the test points, then each test point was assigned to the corresponding category. Through the spatial correlation between test points and segmentation results of the vegetation region, the final classification result can be obtained by using the regional majority voting algorithm.

The implementation process of the majority voting algorithm is detailed in Figure 8.

3. Experiments

3.1 Classification Results

According to the segmentation results of the study region, a center point and six random points were respectively generated for each object, and they were combined into test points afterwards. The test points were then inputted into the trained network model and assigned category attributes. Test points with category information were spatially associated with the segmentation results, and the regional majority voting was performed to obtain category information of each segmented object. At this point, the study area contained a total of seven categories (arbor, shrub, sparse wood, grass, farmland, water and construction). As the
purpose of the study was to extract forest types, the non-forest object attributes were all assigned as the other category (background category). And then adjacent objects of the same category were combined into one object. The final classification result of the study area included four categories: three types of forest categories (arbor, shrub, and sparse wood) and non-forest. The result of forest extraction types is shown respectively in Figure 9.

Figure 9. Classification result.

3.2 Accuracy evaluation

As Figure 9 shows, the voting strategy which the category of an object is determined by multiple test points scattered within the object based MRS made a good classification result. The learning process of the CNN network model is to update the model parameters continuously through the input images, thereby learning the middle- and high-level features of the input data and then classifying the data according to those multi-layer features. The quantitative evaluation of classification results to the forest types can further verify the applicability of CNN in classifying the forest types.

Table 2. The number of ground feature categories in the data set.

| Year | Accuracy | Kappa |
|------|----------|-------|
| 2019 | 0.88     | 0.84  |

4. Conclusions

Compared with medium and low spatial resolution images, HSRI contains a large volume of spatial details of ground objects (texture, shape, size, etc.), and the technical difficulties of spatial analysis and information extraction are significantly increased. The deep learning method does not need to manually design the feature extractor, but it can be obtained by automatic machine learning, which is especially suitable for changeable natural data and has a strong generalization ability and robustness.

From the experimental results, CNN intelligent identification can realize the fine extraction of ground objects by learning the middle- and high-level features of ground objects. Through the learning of domain spatial information, complex sparse wood and grassland with high similarity to shrubs can be extracted completely and accurately.
However, the classification results of the proposed method depend on the segmentation results of MRS. The selection of segmentation scale affects the integrity of the ground features and the accuracy of the boundaries. The selection of an optimal segmentation scale is one of the focuses of future research.

**Funding:** This research was supported by the Remote Sensing Geological Survey of Key Earth Belts (Grant No. DD20190536), the National Natural Science Foundation of China (41671369) and "the Fundamental Research Funds for the Central Universities".

**References**

[1] Ling, C.-L.; Zhu, L.-Y.; Wu, L.-M. The research and realization of the forest information extraction based on the WorldView-2 image. Science of Surveying and Mapping 2010, v.35;No.167, 205-207.

[2] Coops, N.; Wulder, M.; Culvenor, D.; St-Onge, B. Comparison of forest attributes extracted from fine spatial resolution multi-spectral and LiDAR data. Canadian Journal of Remote Sensing 2004, 30, 855-866.

[3] Goodbody, T.; Coops, N.; Hermosilla, T.; Tompalski, P.; Crawford, P. Assessing the status of forest regeneration using digital aerial photogrammetry and unmanned aerial systems. International Journal of Remote Sensing 2017, 1-19.

[4] Fan, G.; Chen, F.; Li, Y.; Liu, B.; Fan, X. Development and Testing of a New Ground Measurement Tool to Assist in Forest GIS Surveys. 2019, 10, 643.

[5] Liu, H.J.; Zhang, T.; Shi, H.; Xu, Y.N.; Wu, W.L.; Yu, P.L. Classification evaluation on high resolution remote sensing image based on RF. Journal of Nanjing Forestry University(Natural Science Edition) 2015, 39, 99-103.

[6] Wulder, M.; Hall, R.J.; Coops, N.; Franklin, S. High spatial resolution remotely-sensed data for the study of forest ecosystems. BioScience 2004, 54, 511-521.

[7] Lv, X.; Ming, D.; Lu, T.; Zhou, K.; Wang, M.; Bao, H. A New Method for Region-Based Majority Voting CNNs for Very High Resolution Image Classification. Remote Sens 2018, 10, 1946.

[8] Bruzzone, L.; Carlin, L. A multilevel context-based system for classification of very high spatial resolution images. Ieee Transactions on Geoscience and Remote Sensing 2006, 44, 2587-2600.

[9] Johansen, K.; Phinn, S. Mapping structural parameters and species composition of riparian vegetation using IKONOS and landsat ETM plus data in Australian tropical savannahs. Photogramm Eng Rem S 2006, 72, 71-80.

[10] Mallinis, G.; Koutsias, N.; Tsakiri-Strati, M.; Karteris, M. Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. Isprs Journal of Photogrammetry & Remote Sensing 2008, 63, 237-250.

[11] Hu, Z.H.; Zhao, Z.H.; Wang, Y.C. Information Extraction Method of the Land Forest Resource through Remote Sensing Image of High Resolution. 2015.

[12] Wang, X.L. Research on Forestland Resource Information Extraction Method Based on High Resolution Remote Sensing Image. Information Recording Materials 2018, 19, 96-98.

[13] Hu, H.L.; Wu, B.; Qin, Z.Y.; Lai, G.L. Forest Boundary Extraction Method Applied to High Resolution Remote Sensing Images. Journal of Surveying and Mapping Science and Technology 2016, 33, 394-399.
[14] Zhou, R.P. Study on GF-1 Remote Sensing Satellite Data Forestry Information Fast Extraction. 2019, 31, 72-76.

[15] Zhang, Z.; Wang, Y.; Liu, Q.; Li, L.; Ping, W. A CNN based functional zone classification method for aerial images. In Proceedings of Geoscience & Remote Sensing Symposium.

[16] Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Zhang, L. Learning Transferable Deep Models for Land-Use Classification with High-Resolution Remote Sensing Images. 2018.

[17] Nogueira, K.; Penatti, O.A.B.; Santos, J.A. Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification. Pattern Recognition 2017, 61, 539-556.

[18] Gong, C.; Yang, C.; Yao, X.; Lei, G.; Han, J. When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs. IEEE Transactions on Geoscience and Remote Sensing 2018, 56, 1-11.

[19] Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Transactions on Medical Imaging 2016, 35, 1285-1298.

[20] Jiang, H.; Learned-Miller, E. Face Detection with the Faster R-CNN. In Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition.

[21] Wang, X.; Shrivastava, A.; Gupta, A. A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.

[22] Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation; 2014; Vol. 39, pp. 640-651.

[23] Dai, J.; He, K.; Jian, S. Instance-aware Semantic Segmentation via Multi-task Network Cascades. 2015.

[24] Hamaguchi, R.; Nemoto, K.; Imaizumi, T.; Hikosaka, S. Detecting Buildings of any Size Using Integration of Cnn Models. In Proceedings of IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.

[25] Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. IEEE Geoscience Remote Sensing Letters 2017, PP, 1-5.

[26] Shao, Z.; Wang, L.; Wang, Z.; Du, W.; Wu, W. Saliency-Aware Convolution Neural Network for Ship Detection in Surveillance Video. IEEE Transactions on Circuits Systems for Video Technology 2019.

[27] Zhou, W.; Newsam, S.; Li, C.; Shao, Z.J.R.S. Learning Low Dimensional Convolutional Neural Networks for High-Resolution Remote Sensing Image Retrieval. 2016, 9, 489.

[28] Mizoguchi, T.; Ishii, A.; Nakamura, H.; Inoue, T.; Takamatsu, H. Lidar-based Individual Tree Species Classification using Convolutional Neural Network. In Proceedings of SPIE Optocal Metrology 2017, Videometrics, Range Imaging, and Applications.

[29] Zou, X.; Ming, C.; Cheng, W.; Yan, X.; Li, J. Tree Classification in Complex Forest Point Clouds Based on Deep Learning. IEEE Geoscience and Remote Sensing Letters 2017, 14, 2360-2364.

[30] Jamil, A.; Bayram, B. Tree Species Extraction and Land Use/Cover Classification From High-Resolution Digital Orthophoto Maps. IEEE Journal of Selected Topics in Applied
Earth Observations & Remote Sensing 2017, PP, 1-6.

[31] Dian, Y.; Yong, P.; Dong, Y.; Li, Z. Urban Tree Species Mapping Using Airborne LiDAR and Hyperspectral Data. Journal of the Indian Society of Remote Sensing 2016, 44, 595-603.

[32] Shao, Z.; Lei, Z.; Zhou, X.; Lin, D.J.I.G.; Letters, R.S. A Novel Hierarchical Semisupervised SVM for Classification of Hyperspectral Images. 2014, 11, 1609-1613.

[33] Hay, G.J.; Blaschke, T.; Hay, G.J.; Blaschke, T. Special Issue on Geographic Object-Based Image Analysis (GEOBIA). Photogrammetric Engineering & Remote Sensing 2010, 76, 121-122.

[34] Rabiee, H.R.; Kashyap, R.L.; Safavian, S.R. Multiresolution segmentation-based image coding with hierarchical data structures.

[35] Draguț, L.; Tiede, D.; Levick, S.R. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. International Journal of Geographical Information Science 2010, 24, 859-871.

[36] Blaschke, T. Object based image analysis for remote sensing. Isprs Journal of Photogrammetry & Remote Sensing 2010, 65, 2-16.

[37] Denny, C.K.; Nielsen, S.E. Spatial Heterogeneity of the Forest Canopy Scales with the Heterogeneity of an Understory Shrub Based on Fractal Analysis. 2017, 8, 146.

[38] Ming, D.; Li, J.; Wang, J.; Min, Z. Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. Isprs Journal of Photogrammetry Remote Sensing 2015, 106, 28-41.

[39] Lopes Queiroz, G.; McDermid, G.J.; Castilla, G.; Linke, J.; Rahman, M.M. Mapping Coarse Woody Debris with Random Forest Classification of Centimetric Aerial Imagery. 2019, 10, 471.

[40] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of International Conference on Neural Information Processing Systems.