Advances in Deep Learning for Remote Sensing Image Classification (Postprint)

Authors: Fu Weifeng, Zou Weibao

Date: 2018-05-20T00:00:00+00:00

Abstract

With the continuous development of remote sensing technology and computer technology, traditional remote sensing image classification methods have become inadequate for meeting the demands of modern remote sensing image classification. In recent years, the continuous emergence of research achievements in deep learning has provided novel approaches and methodologies for remote sensing image classification. This paper first outlines the evolution of remote sensing image classification and the fundamental concepts of deep learning, then emphatically reviews the research progress of deep learning models—including Deep Belief Networks, Convolutional Neural Networks, and Stacked Auto-Encoders—in remote sensing image classification, and finally proposes the existing problems in current research and the development trends of remote sensing image classification.

Full Text

Review of Remote Sensing Image Classification Based on Deep Learning

Fu Weifeng, Zou Weibao
(College of Geology Engineering & Geomatics, Chang' an University, Xi' an 710054, China)

Abstract: With the continuous development of remote sensing technology and computer technology, traditional remote sensing image classification methods can no longer meet current demands. In recent years, the continuous emergence of deep learning research achievements has provided a new approach for remote sensing image classification. This paper first outlines the development of remote sensing image classification and the basic concepts of deep learning, then focuses on research progress in classification based on deep learning models such as Deep Belief Networks (DBN), Convolutional Neural Networks (CNN), and

Stacked Auto-Encoders (SAE). Finally, it summarizes existing problems and future development trends in remote sensing image classification based on deep learning.

Keywords: deep belief network; convolutional neural network; stacked auto-encoder; remote sensing image classification; deep learning

0 Introduction

Remote sensing image classification is a complex data processing task. Traditional visual interpretation methods, though simple, flexible, and easy to operate, require interpreters to possess extensive experience and professional knowledge. Moreover, they are time-consuming and yield poor interpretation accuracy, making them unsuitable for today's massive volumes of remote sensing imagery. Subsequently, researchers began utilizing spectral, texture, and other features of remote sensing images to achieve interpretation through statistical pattern recognition methods such as maximum likelihood, minimum distance, and K-means clustering.

However, as a vital earth observation technology, remote sensing collects electromagnetic radiation information from ground targets via satellites, aircraft, or other platforms. With the advancement of remote imaging technology, imagery has gradually exhibited characteristics of high spectral, high spatial, and high temporal resolution [1]. Currently, as remote sensing images are increasingly applied in mineral exploration, precision agriculture, urban planning, forestry measurement, military target recognition, and disaster assessment, demands for classification accuracy and efficiency continue to rise. Yet remote sensing image classification is influenced by multiple factors, including band selection and complex surface information, making classification difficult. In high-resolution remote sensing images, the abundance of detail information and the complexity of ground object spectra further degrade classification accuracy. Consequently, new methods such as artificial neural networks [2], support vector machines [3], genetic algorithms [4], and object-oriented approaches [5] have been applied to remote sensing image interpretation. Although these methods have achieved good classification results, their degree of automation remains low.

In recent years, the emergence of deep learning [6] has provided a novel approach for remote sensing image interpretation. As a neural network with deep hierarchical structure, it can better extract remote sensing image features than shallow models such as artificial neural networks and support vector machines. It can achieve dimensionality reduction for hyperspectral remote sensing images and has attained higher classification accuracy than previous methods, better promoting the development of automated and intelligent remote sensing image interpretation. The advantage of deep learning lies in its superior capability for feature selection and extraction. The foundation of remote sensing image application is achieving good classification accuracy, which primarily depends on feature extraction and selection. Applying deep learning technology can effectively extract remote sensing image features, thereby improving classification accuracy for better practical performance.

1 Deep Belief Network (DBN) Models in Remote Sensing Image Classification

Typical deep learning models mainly include Deep Belief Networks (DBN) [6], Stacked Auto-Encoder networks (SAE) [7], and Convolutional Neural Networks (CNN) [8]. This section introduces the applications and research progress of these three models in remote sensing image classification, followed by comparative analysis of their applications in this field, identification of current problems, and prospects for future development directions.

1.1 DBN Model Introduction

A DBN is a probabilistic generative model composed of multiple Restricted Boltzmann Machines (RBMs) stacked together. A typical DBN model with three hidden layers is shown in Figure 1 [Figure 1: see original paper]. The bottom layer of the DBN receives the training dataset, where the first RBM model is a Gaussian-Bernoulli RBM and the remaining RBM models are Bernoulli-Bernoulli RBMs. The DBN learning process consists of two main stages: pre-training and fine-tuning. During pre-training, each layer' s RBM is trained individually using a layer-by-layer approach, with weights and biases between layers fixed and saved for further analysis. In the fine-tuning stage, the DBN model' s weights and biases are updated using the backpropagation algorithm with labeled input data. Specific details can be found in reference [9].

1.2 Research Progress on DBN-Based Remote Sensing Image Classification

With the continuous development of deep learning in recent years, DBN models have been increasingly applied to speech and image datasets, achieving promising results [10,11]. In 2010, Mnih and Hinton first applied DBN models to road detection in airborne remote sensing images, demonstrating that DBN models can effectively extract image features and achieve good classification results [12]. However, as data from hyperspectral, high-resolution remote sensing images, and Synthetic Aperture Radar (SAR) images continue to grow, new problems have emerged in remote sensing image classification. The following sections introduce the application and research progress of DBN models in hyperspectral and high-resolution remote sensing imagery.

1.2.1 DBN Models in Hyperspectral Remote Sensing Image Classification Hyperspectral remote sensing image classification methods are generally based on mixed pixel decomposition, feature space classification, and spectral matching. However, due to the Hughes phenomenon, limited training samples, and spatial variability of spectral features, image classification remains challenging. In 2014, Li et al. introduced DBN into hyperspectral remote sensing imagery for feature extraction and image classification, which could also be used for spectral-spatial classification, though a further balance between learning rate and classification accuracy was needed [13]. Subsequently, Chen et al. proposed a new feature extraction and image classification method based on DBN for hyperspectral data, considering the data characteristics under spectral and spatial variations. This method effectively addressed problems caused by data complexity and limited sample size [14]. Zhou et al. [15] proposed a Grouped Deep Belief Network that achieved better results in the spectral-spatial classification of hyperspectral remote sensing images. This network combines grouped feature knowledge of hyperspectral data with deep neural networks, reducing weights between redundant feature groups connected to the neural network, thereby removing numerous redundant spectral bands for spectral-spatial classification. Recent research by Zhong et al. proposed a diversified DBN model that first uses unsupervised learning to pre-train unlabeled samples, then employs supervised learning fine-tuning to address the small sample problem. By introducing diversified processing to handle the persistent reflection and non-response situations of hidden unit parameters when using DBN, they achieved better accuracy than the original DBN model and other methods for hyperspectral remote sensing image classification [16].

1.2.2 DBN Models in High-Resolution Remote Sensing Image Classification Although high-resolution remote sensing imagery provides accurate information for understanding the world, it also poses challenges for intelligent interpretation in the remote sensing field. Previous pixel-level image classification methods are no longer applicable as remote sensing image resolution increases, because a single pixel may contain different types of ground objects. In 2013, Chen et al. used DBN models to extract aircraft from high-resolution remote sensing images, leveraging DBN' s ability to learn features and adjust parameters via backpropagation, demonstrating better performance than traditional feature classifiers [17]. In 2014, Zou et al. proposed converting the image feature selection problem into a reconstruction problem in DBN for high-resolution remote sensing image classification, achieving an overall classification accuracy of 77% on the RSSCN7 dataset, though the running time was longer than direct DBN-based image classification [18]. In 2016, Diao et al. combined DBN' s unsupervised feature learning with visual saliency advantages to effectively avoid exhaustive searches on images and generate fewer bounding boxes, enabling fast and accurate target localization. However, this method cannot effectively detect multi-scale objects [19].

Due to their all-weather, all-day, atmospheric propagation and climate independence, and strong penetration capabilities, increasing numbers of SAR images are being used for land use and land cover monitoring, military reconnaissance satellites, and other fields. However, feature selection is difficult during image classification. In 2014, Dou et al. first applied DBN models to urban map production from SAR images [20]. Subsequently, Radu et al. proposed a convolutional DBN model for feature extraction from polarimetric SAR images. DBN models can extract higher-level features layer by layer, then extend them to polarimetric SAR images using the feature invariance property of convolution, effectively extracting features from polarimetric SAR images. This approach achieved 88.7% accuracy in classification verification on fully polarimetric multi-look F-SAR images of the German Kaufbeuren airport [21].

In recent years, due to DBN' s unsupervised learning characteristics and ability to better extract features from images, an increasing number of DBN-based methods have been proposed for hyperspectral and high-resolution remote sensing image classification. Problems such as data complexity and limited sample sizes in hyperspectral images, fine classification and target detection in high-resolution images, and feature extraction from SAR images will all achieve better results with continuous improvements to deep learning models and their integration with spatial information (texture, size, shape, location, etc.).

2 Convolutional Neural Network (CNN) Models in Remote Sensing Image Classification

2.1 CNN Model Introduction

Convolutional Neural Networks are a type of feedforward neural network inspired by the biological mechanism of receptive fields. In the 1960s, Hubel et al. proposed the concept of receptive fields through studies of cat visual cortex cells [22]. In 1980, Fukushima proposed the first convolutional neural network, the Neocognitron, based on the receptive field concept [23]. In 1998, LeCun et al. successfully trained a CNN architecture using backpropagation algorithms and applied it to image classification for the first time [24]. CNN is essentially a multilayer perceptron, but its characteristics of local connectivity, weight sharing, and spatial or temporal subsampling give convolutional neural networks a certain degree of translation, scaling, and distortion invariance when processing two-dimensional graphics.

A typical CNN structure consists of input layers, convolutional layers, pooling layers, and fully connected layers. Each layer contains multiple feature maps, with each feature map extracting one type of feature from the input through a convolutional filter, as shown in Figure 2 [Figure 2: see original paper]. CNN training is conducted through backpropagation and stochastic gradient descent, updating the weights and biases of the convolutional neural network layer by layer. Detailed procedures can be found in reference [25].

Although CNN models were proposed many years ago, they did not receive much attention until 2012, when Krizhevsky et al. trained CNN models for image classification in the ImageNet competition, achieving results nearly 10 percentage points higher than the second-place team [26]. Since then, CNNs have been widely applied in image classification, face recognition, and target detection [27-30]. The weight sharing characteristic of CNN can reduce the number of parameters requiring training, offering obvious advantages for processing high-dimensional data without requiring manual feature selection. Once weights are well-trained, features can be effectively extracted. However, CNN also has many disadvantages: it is prone to overfitting, requires a large number of training samples with known categories to improve model generalization, the neural network itself is a "black box model" with unclear physical meaning, and it involves heavy computation.

2.2 Research Progress on CNN-Based Remote Sensing Image Classification

2.2.1 CNN Models in Hyperspectral Remote Sensing Image Classification Hyperspectral remote sensing images contain rich spectral and spatial information. Due to CNN' s characteristics of local connectivity and weight sharing, some scholars have attempted to use CNN models for hyperspectral image classification in recent years, achieving certain results. In 2015, Adriana proposed an unsupervised neural network for remote sensing image analysis. This method used greedy layer-by-layer unsupervised pre-training to form a deep CNN model that obtained the spectral vector of each pixel, then used CNN convolutional layers to extract local spectral information on the spectral vector, and used the feature maps generated by convolution operations as input to fully connected layers to finally complete hyperspectral remote sensing image classification. This avoided the Hughes phenomenon and overfitting problems caused by small samples when using supervised training methods, though it could not effectively extract image features compared to supervised classification [31]. Subsequently, Chen et al. proposed a supervised 3D-CNN model that could simultaneously extract spectral and spatial features of images, combined with regularization methods and dropout strategies to prevent overfitting and improve model generalization capability, further enhancing classification accuracy [32].

2.2.2 CNN Models in High-Resolution Remote Sensing Image Classification Although deep learning CNN models have achieved excellent performance in many fields, particularly in hyperspectral image research, SAR image classification remains difficult due to complex scattering mechanisms and random noise. Chen et al. proposed using fully convolutional networks (without fully connected layers, only sparse connection layers) for automatic target recognition and classification of SAR images, effectively alleviating overfitting problems caused by limited samples and achieving 99% accuracy on the MSTAR dataset, significantly outperforming traditional classification methods [33]. Subsequently, Zhou et al. studied the applicability of deep CNN models in supervised classification of polarimetric SAR images. They converted polarimetric SAR images into 6-dimensional feature vectors for input into CNN models for feature extraction and classification, achieving good classification accuracy on airborne SAR images of the San Francisco and Flevoland regions, providing a new approach for future polarimetric SAR image classification [34]. Recently, Hu et al. combined hyperspectral data and polarimetric SAR data for urban image classification, proposing a two-stream convolutional neural network to process them. This approach can utilize both the spectral information from hyperspectral data and the scattering mechanisms from polarimetric SAR data to better distinguish ground objects in urban areas [35].

Recent studies show that Grant et al. proposed combining transfer learning and data augmentation methods with deep CNN to effectively overcome the problem of limited samples in remote sensing images [36]. Weng et al. combined CNN' s feature learning capability with the generalization performance of extreme learning machines [37] for land use classification in high spatial resolution images, reducing computation compared to traditional methods while achieving good classification accuracy [38]. With the development of deep learning, CNN has also been continuously applied to remote sensing image applications such as road and building extraction, target detection, and land use classification [39-42].

---

3 Stacked Auto-Encoder (SAE) Models in Remote Sensing Image Classification

3.1 SAE Model Introduction

A Stacked Auto-Encoder is a deep neural network composed of several sparse auto-encoder structural units, with network architecture shown in Figure 3 [Figure 3: see original paper]. Stacked auto-encoders are typical deep neural networks widely used for feature learning and representation. Parameters are determined layer by layer through greedy learning, then backpropagated from the top layer to adjust the entire network parameters. If the topmost layer has no label information, the learning process is unsupervised; if sample label information is added to the top layer to adjust network parameters in reverse, the learning process becomes supervised. Detailed details can be found in reference [43].

3.2 Research Progress on SAE-Based Remote Sensing Image Classification

3.2.1 SAE Models in Hyperspectral Remote Sensing Image Classification Hyperspectral remote sensing images have high-dimensional feature spaces with strong data correlation and redundancy, posing significant challenges for image classification. Generally, dimensionality reduction is required to remove redundant band information before feature extraction. Traditional dimensionality reduction methods for hyperspectral remote sensing images mainly use algorithms such as Principal Component Analysis, Independent Component Analysis, or Minimum Noise Fraction transformation to extract representative bands for classification. Although these algorithms effectively avoid the “Hughes phenomenon” and improve classification efficiency, they lose rich information contained in the original data to some extent, leading to decreased classification accuracy.

In recent years, more researchers have adopted deep learning-based methods for hyperspectral remote sensing image classification. In 2014, Chen applied stacked auto-encoder networks to hyperspectral remote sensing image classification for nonlinear feature extraction. Although this method achieved significant breakthroughs compared to traditional methods, the fully connected layers of SAE networks generate many parameters, requiring numerous samples for training [44]. Subsequently, Tao et al. improved the learning mechanism based on this approach using a stacked sparse encoder model, and utilized deep learning' s ability to process large amounts of data simultaneously and learn feature representations common to multiple images, verifying that spectral-spatial features learned from one image could be transferred to other related images [45]. Lv et al. proposed combining extreme learning machines with SAE to address the problems of data complexity and limited sample sizes in hyperspectral image classification [46]. Meanwhile, remote sensing image classification based on stacked auto-encoders has gradually been applied to land use classification [47,48].

3.2.2 SAE Models in High-Resolution Remote Sensing Image Classification As high-resolution remote sensing images are increasingly used in practical applications, new classification methods continue to emerge. Ding et al. proposed a new land use classification method for remote sensing images based on SAE and validated it using GF-1 remote sensing data. Experimental results showed that compared with support vector machines and BP neural networks, the SAE-based remote sensing image land use classification method could achieve better results, with an overall classification accuracy of 95.5% [49]. However, auto-encoders cannot effectively solve pooling problems in image data, and forcing a large number of redundant parameters to participate in computation makes operational efficiency low. Therefore, convolutional auto-encoders were proposed for feature extraction from high-resolution remote sensing images. Convolutional auto-encoders utilize the unsupervised learning approach of traditional auto-encoders combined with the convolution and pooling operations of convolutional neural networks to achieve feature extraction, finally building a deep neural network through stacking. In 2015, Geng et al. proposed convolutional auto-encoders for classification of high spatial resolution SAR images, which could effectively solve the problems of lacking effective feature representation and speckle noise in SAR images, though model parameter selection was a factor limiting accuracy [50]. Recently, Gong et al. proposed combining sparse auto-encoders, CNN, and unsupervised classification to comprehensively utilize the advantages of all three to solve the ternary change detection problem in SAR images. SAE is used to convert different images into appropriate feature spaces to suppress noise and extract key change information, and the feature mapping learned by SAE is used as input for CNN to provide reliable pseudo-labels for training CNN as a change feature classifier. Experiments verified the effectiveness and superiority of the proposed framework [51].

4 Comparative Analysis of Deep Learning Models in Remote Sensing Image Classification

Since the proposal of deep learning, numerous domestic and international scholars have continuously improved deep learning models, achieving great success in speech recognition, image classification, face recognition, and other fields. As deep learning applications have made breakthrough progress in various domains, some scholars have also applied deep learning to the remote sensing field. This paper provides an overview of the applications of three main deep learning models in remote sensing image classification and analyzes the research status of each model in image classification. The application of Deep Belief Networks, Convolutional Neural Networks, and Stacked Auto-Encoder networks in remote sensing image classification is comparatively analyzed and summarized as follows:

a) Although CNN models can be used for image classification, hyperspectral remote sensing image classification generally requires dimensionality reduction before CNN can be applied due to the integration of spectral and spatial information, high dimensionality, and nonlinear components in the data. SAE can reduce the dimensionality of high-dimensional images and can be combined with extreme learning machines to solve problems of data complexity and limited sample sizes. Since DBN models can better extract deep features from images, DBN is also introduced into hyperspectral remote sensing imagery for feature extraction and image classification to improve classification accuracy.

b) High-resolution remote sensing images have high precision and can more accurately identify ground object information. In practical applications, high-resolution images are generally used for target recognition and land cover change detection. CNN can effectively obtain image texture information and capture spatial features between pixels, giving a well-trained CNN model significant advantages in ground object extraction. Most image recognition tasks use convolutional neural networks. Additionally, convolutional auto-encoders have potential advantages in SAR image feature extraction and image classification because they can improve the limited feature representation and speckle noise problems in SAR images.

5 Problems and Future Research Directions in Deep Learning-Based Remote Sensing Image Classification

In remote sensing applications, interpreting and identifying various ground objects through remote sensing images is an important objective of remote sensing image processing, and classification is one of the most fundamental problems in remote sensing image information extraction. Although new deep learning theories and methods have achieved certain results in remote sensing image classification research, several prominent issues remain:

a) Although DBN models can effectively extract image features, they cannot directly extract features from high-dimensional data images. Moreover, the network structure and parameter selection of the model need to be optimized based on experience or experimental methods, resulting in long computation times.

b) Due to limited remote sensing image samples, CNN models are prone to overfitting during image classification, and small sample sizes also limit the generalization capability of CNN models.

c) Although SAE models can achieve dimensionality reduction for high-dimensional images, the layer-by-layer training approach makes existing pre-training algorithms and system parameter optimization strategies strongly dependent on labeled datasets, which is not truly "unsupervised classification." Moreover, as the number of hidden layers and neurons increases, gradient sparsity becomes more severe, still leading to local optima.

Based on these problems in current deep learning-based remote sensing image classification, several directions deserve further research when applying deep learning to remote sensing image classification in the future:

a) How to more quickly and accurately determine the optimal deep learning model structure and number of neural units to achieve higher remote sensing image classification accuracy while avoiding overfitting and local optima problems in deep models.

b) To address the problem of limited remote sensing data samples, methods such as transfer learning, extreme learning machines, and data augmentation can be combined to expand sample size, effectively solving overfitting problems caused by limited sample volume.

c) Deep learning models can be combined for remote sensing image classification. For example, models such as convolutional DBN and convolutional auto-encoders can fully leverage their respective advantages and compensate for their shortcomings.

6 Conclusion

Currently, traditional classification methods are still more commonly used in remote sensing image classification. In recent years, with the proposal and continuous development of deep learning, this approach has been increasingly applied to remote sensing image classification. Although classification accuracy has significantly improved, some shortcomings remain. To further improve classification, comprehensive utilization of various information for remote sensing image classification is essential. This paper mainly introduced three main deep learning models and their specific applications in remote sensing image classification, summarized the reasons limiting remote sensing image classification accuracy and the advantages of each model, and presented research by numerous domestic and international scholars on deep learning and remote sensing image classification, along with existing problems. In the future, with the continuous development of remote sensing and computer technologies, intelligent image interpretation will be an important trend.

References

[1] Li Deren, Tong Qingxi, Li Rongxing, et al. Some frontier scientific issues in high-resolution earth observation [J]. Science China: Earth Sciences, 2012, 42 (6): 805-813.
[2] Canty M J. Boosting a fast neural network for supervised land cover classification. [M]. [S. l.]: Pergamon Press, Inc. 2009.
[3] Huang C, Davis L S, Townshend J R G. An assessment of support vector machines for land cover classification [J]. International Journal of Remote Sensing, 2002, 23 (4): 725-749.
[4] Tseng M H, Chen S J, Hwang G H, et al. A genetic algorithm rule-based approach for land-cover classification [J]. Isprs Journal of Photogrammetry and Remote Sensing, 2008, 63 (2): 202-212.
[5] Huan Y U, Zhang S Q, Kong B, et al. Optimal segmentation scale selection for object-oriented remote sensing image classification [J]. Journal of Image and Graphics, 2010, 15 (2): 352-360.
[6] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. [J]. Neural Computation, 2006, 18 (7): 1527.
[7] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786): 504.
[8] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.
[9] Bengio Y, Lamblin P, Dan P, et al. Greedy layer-wise training of deep networks [C]// Proc of International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2006: 153-160.
[10] Mohamed A, Dahl G, Hinton G. Deep belief networks for phone recognition [C]// Proc of NIPS Workshop on Deep Learning for Speech Recognition and Related Applications. 2009, 1 (9): 39.

[11] Farabet C, Couprie C, Najman L, et al. Learning Hierarchical features for scene labeling [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (8): 1915-1929.

[12] Mnih V, Hinton G E. Learning to detect roads in high-resolution aerial images [C]// Proc of European Conference on Computer Vision. 2010: 210-223.

[13] Li T, Zhang J, Zhang Y. Classification of hyperspectral image based on deep belief networks [C]// Proc of IEEE International Conference on Image Processing. 2015: 5132-5136.

[14] Chen Y, Zhao X, Jia X. Spectral–spatial classification of hyperspectral data based on deep belief network [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8 (6): 2381-2392.

[15] Zhou X, Li S, Tang F, et al. Deep learning with grouped features for spatial spectral classification of hyperspectral images [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 14 (1): 97-101.

[16] Zhong P, Gong Z, Li S, et al. Learning to diversify deep belief networks for hyperspectral image classification [J]. IEEE Trans on Geoscience and Remote Sensing, 2017, 55 (6): 3516-3530.

[17] Chen X, Xiang S, Liu C L, et al. Aircraft detection by deep belief nets [C]// Proc of the 2nd IAPR Asian Conference on Pattern Recognition. 2013: 54-58.

[18] Zou Q, Ni L, Zhang T, et al. Deep learning based feature selection for remote sensing scene classification [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12 (11): 2321-2325.

[19] Diao W, Sun X, Zheng X, et al. Efficient saliency-based object detection in remote sensing images using deep belief networks [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13 (2): 137-141.

[20] Lv Q, Dou Y, Niu X, et al. Classification of land cover based on deep belief networks using polarimetric RADARSAT-2 data [C]// Proc of Geoscience and Remote Sensing Symposium. 2014: 4679-4682.

[21] Tanase R, Datcu M, Dan R. A convolutional deep belief network for polarimetric SAR data feature extraction [C]// Proc of Geoscience and Remote Sensing Symposium. 2016: 7545-7548.

[22] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat' s visual cortex [J]. Journal of Physiology, 1962, 160 (1): 106.

[23] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position [J]. Biological Cybernetics, 1980, 36 (4): 193-202.

[24] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.

[25] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors [J]. Nature, 1986, 323 (6088): 533-536.

[26] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. 2012: 1097-1105.

[27] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: a convolutional neural-network approach [J]. IEEE Trans on Neural Networks, 1997, 8 (1): 98-

113.
[28] Taigman Y, Yang M, Ranzato M, et al. DeepFace: closing the gap to human-level performance in face verification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 1701-1708.
[29] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 580-587.
[30] Xu W, Xu W, Yang M, et al. 3D Convolutional neural networks for human action recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (1): 221-231.
[31] Romero A, Gatta C, Camps-Valls G. Unsupervised deep feature extraction for remote sensing image classification [J]. IEEE Trans on Geoscience and Remote Sensing, 2016, 54 (3): 1349-1362.
[32] Chen Y, Jiang H, Li C, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks [J]. IEEE Trans on Geoscience and Remote Sensing, 2016, 54 (10): 6232-6251.
[33] Chen S, Wang H, Xu F, et al. Target classification using the deep convolutional networks for SAR images [J]. IEEE Trans on Geoscience and Remote Sensing, 2016, 54 (8): 4806-4817.
[34] Zhou Y, Wang H, Xu F, et al. Polarimetric SAR image classification using deep convolutional neural networks [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13 (12): 1935-1939.
[35] Hu J, Mou L, Schmitt A, et al. FusioNet: a two-stream convolutional neural network for urban scene classification using PolSAR and hyperspectral data [C]// Urban Remote Sensing Event. 2017.
[36] Scott G J, England M R, Starms W A, et al. Training deep convolutional neural networks for land-cover classification of high-resolution imagery [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14 (4): 549-553.
[37] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70 (1-3): 489-501.
[38] Weng Q, Mao Z, Lin J, et al. Land-use classification via extreme learning classifier based on deep convolutional features [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14 (5): 704-708.
[39] Alshehhi R, Marpu P R, Wei L W, et al. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks [J]. Isprs Journal of Photogrammetry and Remote Sensing, 2017, 130: 139-149.
[40] Chen X, Xiang S, Liu C L, et al. Vehicle detection in satellite images by hybrid deep convolutional neural networks [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 11 (10): 1797-1801.
[41] Zhang Yide, Hu Changyu, Hu Chunyu. Aircraft detection in remote sensing images based on convolutional neural networks [J]. Optoelectronic Technology, 2017, 37 (1): 66-71.
[42] Castelluccio M, Poggi G, Sansone C, et al. Land use classification in remote sensing images by convolutional neural networks [J]. Acta Ecologica Sinica, 2015,

28 (2): 627-635.

[43] Bengio Y. Learning Deep Architectures for AI [M]. [S. l.]: Now Publishers, 2009.

[44] Chen Y, Lin Z, Zhao X, et al. Deep learning-based classification of hyperspectral data [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7 (6): 2094-2107.

[45] Tao C, Pan H, Li Y, et al. Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12 (12): 2438-2442.

[46] Lv F, Han M, Qiu T. Remote sensing image classification based on ensemble extreme learning machine with stacked autoencoder [J]. IEEE Access, 2017, 5: 9021-9031.

[47] Ding A, Zhou X. Land-Use Classification with Remote Sensing Image Based on Stacked Autoencoder [C]// Proc of International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration. 2017: 145-149.

[48] Li W, Fu H, Yu L, et al. Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping [J]. International Journal of Remote Sensing, 2016, 37 (23): 5632-5646.

[49] Ding A, Zhou X. Land-use classification with remote sensing image based on stacked autoencoder [C]// Proc of International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration. 2017: 145-149.

[50] Geng J, Fan J, Wang H, et al. High-Resolution SAR image classification via deep convolutional autoencoders [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12 (11): 2351-2355.

[51] Gong M, Yang H, Zhang P. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images [J]. Isprs Journal of Photogrammetry and Remote Sensing, 2017, 129: 1-16.

Note: Figure translations are in progress. See original paper for figures.