International Journal of Energy
ISSN: 2957-9473 | Vol. 3, No. 1, 2023

Study on High-resolution Remote Sensing Image Scene Classification Using Transfer Learning

Qian Ouyang *

School of Electronics and Information Engineering, Jiujiang University, Jiangxi 332000, China
*** Corresponding author Email:** 812916448@qq.com

Abstract: Remote sensing image classification (RSIC) has been increasingly concerned and becoming a challenging task. Recently, deep convolutional neural networks (DCNN) offer the effective classification method include the capacity to handle high-dimensional data and to distinguish classes with very complex characteristics on the remote sensing community. However, the focus of these methods is on publicly available data sets in the field of remote sensing, there are few studies on RSIC composed of different benchmark datasets, which the complexity, diversity, and similarity of data greatly increase the difficulty of classification. In this paper, we reconstructed and selected one new dataset from two standard benchmark remote sensing datasets: UC Merged Land-Use and NWPU-RESISC45. We utilize three transfer learning frameworks to extract the high-level feature map and feed feature information into the proposed model for partial and full fine-tuning. Data augmentation technology is used to increase the number of training samples and dropout strategies to prevent overfitting. The experimental results demonstrate that the proposed methodology achieved remarkable performance in scene classification of overall accuracy: 90.1%,91.0%,93.3% with VggNet, DesNet, InceptionNet, respectively.

Keywords: Convolutional Neural Network; Deep Learning; Transfer Learning; Remote Sensing; Classification Accuracy; Feature Extraction.

1. Introduction

High-resolution RSIC now become a major hot research topic, it has attracted considerable attention in the field of remote sensing because of its significance for a wide range of applications, such as ecological environment [1], Optimizing the consumption of agriculture water [2], Road information extraction [3], Land Use/Land Cover [4],etc. The best-known methodologies about image classification tasks are that DCNN ultimately become mainstream because it made remarkable achievements. However, many state-of-the-art algorithms have been proposed by different researchers, most of them typically utilize a small number of labeled images that are typically available in the remote sensing image from the single standard datasets, very few researchers' study RSIC from different remote sensor benchmark images. High intra-class diversities and low inter-class dissimilarities between the different image categories, render such complexity of spatial and structural patterns of the datasets highly challenging.

Since the earliest benchmark of remote sensing data sets was published, several machine learning (ML) algorithms have been created to support scene classification. Many studies have generally found that the selection of classifier is the key factor to determine the result of classification, these methods tend to produce higher accuracy compared to traditional parametric classifiers, especially for complex data with a high dimensional feature space. The development of high-performance classifiers represents an important step in improving the accuracy of RSIC in the era of high spatial resolution. Belgiu et al. [5] applied a randomly selected subset of training samples and variables produces multiple a random forest (RF) ensemble classifier with the remote sensing community. Pal et al. [6] used decision trees (DT), in which the speed of calculating categorical data is extremely rapid and can handle data that are presented on different measurement scales. Boosted DT classification [7] are adapted in an attempt to minimize the errors of the previous trees. Pu et al. [8] integrated geographic information into the k-nearest neighbors (KNN)classifier, demonstrated better performance in misclassification, and further improves computational efficiency. Foody et al. [9] investigated the effect of four factors on the accuracy with which agricultural crops may be classified by an artificial neural network.

With the rapid development of deep learning (DL) theories, DL models that are composed of different layers can learn more abstract and discriminative powerful features. Particularly, DCNN architectures can model complex function mappings between inputs and outputs, and they produce competitive results in a wide range of areas, including face recognition [10], person re-identification (ReID) [11], natural language processing [12], medical image detection and diagnosis [13], etc. Since DL methods are recognized as the dominant method in computer vision classification and recognition tasks, it is feasible to fully train a deep network structure composed of convolution, pooling and fully-connected layers in the field of remote sensing images [18]. However, such a large multi-layer network architecture usually contains millions of parameters to be learned, thus, it will consume enormous resources to train an efficacious DCNN model from the High-resolution remote sensing datasets. On the other hand, fully train them to need the extreme volume of images, but for remote sensing images only contain thousands or even hundreds of data, it will easily occur over-fitting phenomenon and can't convergence to global optimal.

It has been proved that via using the pre-trained DCNN models have facilitated considerable improvements to global features of a wide variety of remote sensing images [14,15]. Makantasis et al. [16] exploited a CNN to encode pixels' spectral and spatial information and a Multi-Layer Perceptron to conduct the classification task. Nogueira et al. [17] utilized

the off-the-shelf CNN to extract features of datasets and fine-tuned the model with a linear classifier that obtains high accuracy. Qin et al. [25] selected features that are more reconstructible as the discriminative features with deep belief networks. Fotso et al. [19] used a pre-trained DCNN model for extracting heterogeneous features on three benchmark satellite datasets via transfer learning (TL). Flores et al. [20] also utilized TL technology and the Gaussian mixture model to generate dictionaries of deep features, and obtained excellent results on UC Merged and Brazilian Cerrado-Savana public datasets. In [29], Hu et al. successfully transferring off-the-shelf pre-trained DCNN for RSIC and achieved better representations for image scenes.

2. Dataset Acquisition

Recently, more and more standard benchmark datasets are proposed to support scene classification algorithms. Table 1. summarizes some common benchmark datasets in the past decade, great progress has been made [5,41,42]. In the field of remote sensing image classification, sensor data are usually obtained from a specific angle and direction. However, actual scenes have more complex spatial characteristics than remote sensing images. Because remote sensing images are collected by different sensors often have different characteristic distributions and contained more scene information of a certain kind, enabling people to know more about the state and environment changes of the region.

In order to better illustrate the effectiveness of our algorithm model, In Table 1., UC Merged Land-Use contained a total of 2100 manually selected and uniformly labeled into 21 classes, and we randomly selected 10 classes named Dataset1 from UC Merged Land-Use. It can be seen that there are 45 categories in NWPU-RESISC45, and each category of the dataset contains 700 while the UC Merged Land-Use each category contains 100 images. Hence, we select 10 classes from NWPU-RESISC45 named Dataset2, just only 1/7 data in each class is used to ensure the same training and test samples as Dataset1.

Dataset1 and Dataset2 were merged to constitute new dataset named Optimistic-RSD20 (Table 2.): agricultural, airplane, airport, baseball_diamond, basketball_court, basketball_court, beach, bridge, buildings, chaparral, church, cloud, circular farmland, dense residential, forest, freeway, golf course, harbor, river, storage tanks, tennis court. Table 2. describe the dataset for the detail, where is available for free download on the website [43]. It should be noted that there are duplicate classes in the two datasets, e.g., airplane, baseball_diamond, we try to avoid selecting the same category from the above two datasets.

Table 1. Key information on common benchmarks of remote sensing datasets

datasets	classes	Images per class	Image size	Spatial resolution	Total images
RSC11[21]	11	~100	512*512	0.2m	1232
RSSCN7[25]	7	400	400*400	--	2800
SIRI-WHU[22]	12	200	200*200	2m	2400
UC Merged Land-Use[23]	21	100	256*256	0.3m	2100
NWPU-RESISC45[24]	45	700	256*256	0.2-30m	31500
AID[26]	30	220~420	600*600	0.5-8m	10000

3. Proposed Method

In this paper, a classification method associated with the extract feature map from Optimistic-RSD20 imageries using transfer neural network and fine-tune parameters is proposed. Figure 4. illustrates the schematic diagram of the proposed approach. First, after preprocessing and data augmentation, all training data are sent into pre-trained DCNN (see section 4.1 for detail) model to extract the feature map. It is noteworthy that all parameters trained in the previous source domain will be frozen. Next, the extracted three-dimensional feature vectors are flattened to one-dimension without any data dimension reduction method, where mapping a high-dimensional data projection to a low-dimensional space often loses some important information. Then, transfer different proportional parameters of previously learned to remote sensing classification target domain, and freeze the parameters that do not need to be adjusted. Finally, we feed the parameters of required training to the classifier for training and conduct partial and full fine-tuning, and the output layer would provide a score for each class. Its highest score will be identified as the class to be distinguished.

3.1. Pre-trained Model

A Pre-training model is widely used in computer vision classification tasks include VggNet [30], InceptionNet [31,32,33,34], DesNet [35] and so on. Although these network structures and depth are different, all extracted feature information is convoluted and pooled. All the network parameters have been trained on the same ImageNet datasets[36], which contains about 1,200,000 images divided into 1000 classes.

VggNet [30] is developed based on AlexNet [37], which main contribution is to use very small 3x3 convolution kernel and 2x2 pooling kernel for network architecture design. Despite increase the network architecture depth to 16-19 layers, multiple 3x3 convolution kernel can greatly reduce the number of parameters and accelerate the convergence speed. Meanwhile, the alternating structure of multiple convolutional layers and nonlinear activation layers can extract deeper features than the structure of single convolutional layer.

InceptionV1 [31] proposed that the convolution kernel of 1x1 is added before 3x3 and 5x5 and after Max pooling respectively to reduce the thickness of feature map. InceptionV2 [32] network adds BN layer, which reduces Internal Covariate Shift (changes in the data distribution of neuron) and normalizes the output of each layer to a N (0, 1) gaussian distribution. One of the most important improvements to InceptionV3 [33] is Factorization. The decomposition of two one-dimensional convolution can accelerate the calculation, not only increasing the depth of the network but also increasing the non-linearity of the network. InceptionV4 [34] proposed the method of combining inception module with residual connection module. The experimental results show that this model can greatly accelerate training and improve performance.

Desnet [35] alleviates the problem of gradient disappearance, enhances feature propagation and feature reuse, and significantly reduces the number of parameters. The main contribution is that it establishes a direct connection between all feature map, so the L layer has a direct connection of L feature map, the first L layer has L(L+1)/2 connections in total.

3.2. Data Pre-processing

To extract semantic information expressing deep features, improve the generalization ability of neural network during training, a variety of quality training data are required as input image to effectively improve classification accuracy [39]. For the limited remote sensing data set, expanding data through data enhancement strategy and reducing complexity of pre-trained DCNN model can effectively prevent overfitting and more robustness in deep network training. However, it is difficult to modify the structure of a mature pre-trained model. Hence, we adopted data augmentation strategy to generate more diversified training data samples to achieve better results [40,44].

The essence of image transformation is that the matrix of the original image pixels is multiplied by the transformation matrix to form a new image. Furthermore, these operations help the deep model learn intrinsic feature of remote sensing data. In this work, we used two types of methods for generating new augmented data, are described as follows:

Translation**:** Change the x and y coordinates of the image by certain size. The translated points [X, Y] can be calculated under the following Equation:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

Where [x, y] is image original coordinates, [dx, dy] represent offsets for the image original coordinates, [X, Y] represent new coordinates for the translated image.

Rotate: randomly rotate the image at a certain angle to change the orientation of image content. The rotated points [X, Y] can be calculated under the following Equation:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

Where [x, y] is image original coordinates, β represent the angle of counterclockwise rotation, [X, Y] represent new coordinates for the rotated image.

Although we exploit translations and rotates as the data augmentation strategy because they do not change the spatial distribution, while learned more high-level feature map information of remote sensing data. Fig.5. visualize some examples of augmentation methods and skills with rotation and translation by OpenCV.

3.3. Classification Method

This paper use three different state-of-the-art pre-training architecture model to extract the high-level feature maps. Although different models have special size requirements for input data (Vgg16 need to (224,224,3)), a pre-trained network can be used as a depth feature extractor of any image, the convolution and pooling layer do not care about the input size, different input size have little influence on the classification accuracy results for the same model. In order to make a fair comparison and accelerate convergence, we reshaped the size of all input images to 150×150×3 on three transfer learning frameworks. The feature dimension vectors extraction through multiple stages of convolution, max-pooling are shown in Table 2.

Table 2. Extracted feature dimension vectors on three transfer learning frameworks

Pre-trained model	Training set Feature dimension vectors	Validation set Feature dimension vectors
Vgg16	(4,4,512)	(4,4,512)
InceptionV3	(3,3,2048)	(3,3,2048)
Desnet121	(4,4,1024)	(4,4,1024)

After flattened three-dimension feature vectors, one-dimension feature vector finally sent to fully connection layer. The first full connection layer contains 256, the second contains 20 neural units that corresponding to the number need classification categories. These neurons are fully connected by matrix parameters to be learned. It can be regarded as multi classification problem, high-resolution remote sensing image classification task can be calculated by SoftMax classifier, defined as Equation (10):

$$\mathrm{Z} = \{Z_1, Z_2, Z_3, \cdots, Z_K\},\ \mathrm{P} = \{P_1, P_2, P_3, \cdots, P_K\} \quad (3)$$

$$P_j^k = \frac{\exp(Z_j)}{\sum_{j=1}^{K} \exp(Z_j)} \quad (0 < P_j^k < 1, \sum_{j=1}^{K} P_j^k = 1) \quad (4)$$

Where Z represents the original output of the last fully layer, k represents the number of classes, P_j^k is probability that sample j belongs to class k. The loss function associated with SoftMax is defined as follows:

$$L_{soft\max} = \sum_{j=1}^{n} \sum_{k=1}^{c} y_j^k \log P_j^k \quad (5)$$

Where $L_{softmax}$ is cross entropy loss function, n is the total number of samples, y_j^k represents true probability that sample j belongs to class k, P_j^k represents predicted probability that sample j belongs to class k.

In this paper, two strategies for parameter sharing and fine-tuning are adopted in the task of RSIC. Parameter sharing proposes that the parameters are completely transferable, which can directly copy the parameters learned in the source domain to the target domain. Fine-tuning proposes that the parameters in the source domain are fully or partially adaptive, but it need to be re-trained in the target domain and better adapted to the remote sensing fields.

4. Experiments and Results Analysis

There exist six widely used, standard evaluation metrics in remote image classification: overall accuracy, average accuracy, precision, recall, F1-score. Because Optimistic-RSD20 have the same image number per class, so the value of overall accuracy equals to the value of average accuracy.

Table 3. Shows the experimental results of three fine-tuned CNN features, respectively, under the transfer learning parameters ratio of 25%, 50% and 100%. We can observe that

full fine-tuning parameters give better results compared to those partial fine-tuning parameters. Only by training to adjust the parameters of the last FC layers have the lowest classification accuracy (detail in 5.3.1). The overall accuracy has improved 8.8%, 9.7%, 10%, the precision has improved 10.2%, 6.4%, 10.3%, the recall has improved 10.2%, 11.1%, 12.4% and the F1-score has improved 8.9%,10.1%,12.2% from the pre-trained models with full-fining, respectively. The F1-score of the transfer learning based on Vgg16, Desnet121, Inceptionv3 model reached 90.3%, 91.5 and 93,6%, respectively. Experimental results demonstrated that the InceptionV3 model was the best pre-trained framework.

Table 3. Compared the three results of deep transfer learning under the fine-tuning different parameters ratios on the Optimistic-RSD20. Vgg16, Desnet121, Inceptionv3 from up to down, respectively.

Fine-tune parameter	Overall accuracy	Precision	Recall	F1-score
Last 2 fully layer	81.3±0.04	81.1±0.02	80.2±0.06	80.4±0.07
Parameters (25%)	83.0±0.05	86.1±0.05	86.3±0.01	86.2±0.08
Parameters (50%)	86.5±0.02	88.2±0.06	86.4±0.02	87.7±0.04
Parameters (100%)	90.1±0.25	91.3±0.03	90.4±0.02	90.3±0.05
Last 2 fully layer	80.3±0.01	86.0±0.06	80.2±0.03	81.4±0.05
Parameters (25%)	81.7±0.05	87.4±0.05	82.9±0.06	83.2±0.04
Parameters (50%)	83.3±0.02	88.2±0.02	83.4±0.04	83.5±0.02
Parameters (100%)	91.0±0.07	92.4±0.04	91.3±0.06	91.5±0.07
Last 2 fully layer	80.7±0.04	84.0±0.05	81.2±0.04	81.4±0.08
Parameters (25%)	84.0±0.02	88.4±0.04	84.9±0.07	84.3±0.05
Parameters (50%)	90.5±0.02	91.5±0.04	91.4±0.03	91.6±0.09
Parameters (100%)	93.3±0.03	94.3±0.01	93.6±0.07	93.6±0.03

5. Conclusion

We used the pre-trained DCNN to extract high-level features from different remote sensing data, further improved the accuracy of classification by fine-tuning different proportions of transfer learning parameters, and comprehensively evaluated our proposed model include overall classification accuracy, precision, recall, F1 score. On the Optimal-data20, we achieved 90.1%,91.0%,93.3% overall classification from the VggNet, DesNet, InceptionNet DCNN, respectively. As we can see from the confusion matrix by full fine-tuning parameters, many of the classes achieved 100% classification accuracy. Finally, compared to the two classical deep learning architecture and different classifiers, the experiment proves that based on transfer learning and fine-tuning methodology achieved stat-of-the-art results and obvious advantage.

Consequently, in the future work, we will continue to study imperfect, complex feature space, the similarity between classes training data set classification task. All the uncertain information is in fact very difficult for the classification and recognition of remote sensing images because it will become more and more difficult to learn powerful feature representations. Although deep learning algorithms offer a powerful set of tools for remote sensing image scene classification, we would also explore new developments and new algorithms for the combination of different sensor data, such as reduce the complexity of the model or use the fewer features to achieve satisfactory results.

References

[1] Qiu, B.; Li, W.; Tang, Z.; Chen, C.; Qi, W. Mapping paddy rice areas based on vegetation phenology and surface moisture conditions. Ecological Indicators. 2015, 56, 79–86.

[2] Mokhtrai, A.; Noory, H.; Vazifedoust, M.; Palouj, M. Evaluation of single crop coefficient curves derived from Landsat satellite images for major crops in Iran. Agricultural Water Management. 2019, 218,234-249.

[3] Zhou, T.; Sun, C.; Fu, H. Road Information Extraction from High-Resolution Remote Sensing Images Based on Road Reconstruction. Remote Sens. 2019, 11, 78-98.

[4] Cai, G.; Ren, H.; Yang, L.; Zhang, N.; Du, M.; Wu, C. Detailed Urban Land Use Land Cover Classification at the Metropolitan Scale Using a Three-Layer Classification Scheme. Sensors. 2019, 19, 3119-3143.

[5] Belgiu, M.; Dragut, L.Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 2016, 114, 24-31.

[6] Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment. 2003, 86, 554-565.

[7] Chan, J.C.; Paelinckx, D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment. 2008, 112, 2999-3011.

[8] Pu, Y.; Zhao, X.; Chi, G.; Zhao, S.; Wang, J.; Jin, Z.; Yinf, J. Design and implementation of a parallel geographically weighted k-nearest neighbor classifier. Computers & Geosciences. 2019, 127, 111-122.

[9] Foody, G.M.; Arora, M. K.; An evaluation of some factors affecting the accuracy of classification by an artificial neural network. International Journal of Remote Sensing. 1997, 18, 799-810.

[10] Kute, R.S.; Vyas, V.; Anuse, A. Component-based face recognition under transfer learning for forensic applications. Information Sciences. 2019, 476, 176-191.

[11] Wu, D.; Zheng, S.J.; Zhao, X.P.; Yuan, C.A.; Cheng, F. Deep learning-based methods for person re-identification: A comprehensive review. Neurocomputing. 2019, 337, 354-371.

[12] Plappert, M.; Mandery, C.; Asfour, T. Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks. Robotics and Autonomous Systems. 2018, 109, 13-26.

[13] Zilong, H.; Jinshan, T.; Ziming, W.; Kai, Z.; Ling, Z.; Qingling S. Deep learning for image-based cancer detection and diagnosis − A survey. Pattern Recognition. 2018, 83, 134-149.

[14] Penatti, A. B.; Nogueira, K.; Jefersson, A. Do Deep Features Generalize From Everyday Objects to Remote Sensing and Aerial Scenes Domains? The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2015, pp. 44-51.

[15] Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. arXiv 2015, arXiv:1508.00092.

[16] Makantasis, K.; Karantzalos, K.; Doulamis, A.; Doulamis, N. Deep supervised learning for hyperspectral data classification through convolutional neural networks. IEEE IGARSS. 2015, 4959-4962.

[17] Nogueira, K.; Penatti, A.B.; Jefersson, A.S. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition. 2017, 16, 539-556.

[18] Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. IEEE Transactions on Geoscience and Remote Sensing. 2017, 55, 645-657.

[19] Guy, F.K.; Akram, T.; Laurent, B.; Naqvi, S.R.; Alex, M.M.; Muhammad, N. A deep heterogeneous feature fusion approach for automatic land-use classification, Information Sciences. 2018, 467, 199-218.

[20] Flores, E.; Zortea, M.; Scharcanski, J. Dictionaries of deep features for land-use scene classification of very high spatial resolution images. Pattern Recognition. 2019,89,32-44.

[21] Zhu, Q.; Zhong, Y.; Zhao, B.; Xia, G.S.; Zhang, L. Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. IEEE Geosci. Remote Sens. Lett. 2016,13,747-751.

[22] Zhao, B.; Zhong, Y.; Xia, G.S.; Zhang, L. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 2016,54, 2108–2123.

[23] Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. Proc. ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. 2010, 270–279.

[24] Cheng, G.; Han, J.; Lu, X. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE. 2017,105,1865-1883.

[25] Zou, Q.; Ni, L.; Zhang, T.; Wang, Q. Deep Learning Based Feature Selection for Remote Sensing Scene Classification. IEEE Geoscience and Remote Sensing Letters. 2015, 12, 2321-2325.

[26] Xia, G.S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. Aid: a benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. 2017, 55, 3965-3981.

[27] Nair, V.; Hinton, G. Rectified linear units improve restricted Boltzmann machines. ICML. 2010, 807-814.

[28] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. arXiv 2015, arXiv:1512.00567.

[29] Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi. A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv 2016, arXiv:1602.07261.

[30] Huang, G.; Liu, Z.; Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. arXiv 2016, arXiv:1608.06993.

[31] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. ImageNet large scale visual recognition challenge. International Journal of Computer Vision. 2015, 115, 211–252.

[32] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Neural Information Processing Systems, 2012,1097–1105.

[33] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 2014, 15, 1929–1958.

[34] Mikołajczyk, A.; Grochowski, M.; Data augmentation for improving deep learning in image classification problem. 2018 International Interdisciplinary PhD Workshop (IIPhDW). 2018, 117-122.

[35] Yu, X.; Wu, X.; Luo, C.; Peng, Ren. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. GIScience & Remote Sensing, 2017, 54, 741-758.

[36] Avramovi´c, A.; Risojevi´c,V. Block-based semantic classification of high-resolution multispectral aerial images. Signal, Image Video Process. 2016, 10, 75–84.

[37] LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998.

[38] Qian, N. On the momentum term in gradient descent learning algorithms. Neural Networks .1999, 12, 145–151.