ISSN: 2664-9640

High-resolution Remote Sensing Image Scene Classification Using Transfer Neural Network Parameters

Qian Ouyang, Young Hoon Joo

School of Electronic and Information Engineering, KunSan National University, KunSan 54150, Korea;

812916448@qq.com

Abstract

The remote sensing image classification (RSIC) has been increasingly concerned and becomes a challenging task. Recently, deep convolutional neural networks (DCNN) offer the effective classification method which includes the capacity to handle high-dimensional data and to distinguish classes with very complex characteristics on the remote sensing community. However, these methods focus on publicly available data sets in the field of remote sensing. There are also few studies on RSIC composed of different benchmark datasets and the complexity, diversity, and similarity of data greatly increase the difficulty of classification. In this paper, we propose and reconstruct one novel dataset from two standard remote sensing datasets: UC Merged Land-Use and NWPU-RESISC45. Moreover, we utilize three transfer learning frameworks to extract the high-level feature map, and feed feature information into the proposed model for the partial and full fine-tuning. Data augmentation technology is used to increase the number of training samples and dropout strategies to prevent overfitting. Finally, we demonstrate that the proposed methodology has remarkable performance through some experimental results.

Keywords

Convolutional neural network; deep learning; transfer learning; remote sensing; classification accuracy; feature extraction.

1. Introduction

High-resolution RSIC now becomes 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 sensing 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

ISSN: 2664-9640

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, the extreme volume of images are needed to fully train them, 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 has 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, spatial information and a multi-layer perceptron (MLP) to conduct the classification task. Nogueira et al. [17] utilized off-the-shelf CNN to extract features of datasets and fine-tuned the model with a linear classifier to obtains high accuracy. Qin et al. [25] selected features that are more reconstructible as the discriminative features with deep belief network. Fotso et al. [19] used a pre-trained DCNN model for extracting heterogeneous feature 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 transferred off-the-shelf pre-trained DCNN for RSIC and achieved better representations for image scenes.

As mentioned previously, the activations from high-level layers of pre-trained DCNN have proved to be powerful generic feature representations with state-of-the-art performance, absolutely outperforms conventional handcrafted feature extraction.

In order to solve the problems of low classification accuracy of traditional ML methods and large computational resources based on the fully-trained network, this paper proposed RSIC with 20 categories based on deep TL model framework. The main contributions of this paper are summarized as follows:

1. A novel dataset was restructured, which cautiously considered the inter-class variance and intra-class deviation between the two standard datasets.

2. Used the pre-trained DCNN model to extract the high-level features with data augmentation, and describe the global features of deep learning.

ISSN: 2664-9640

3. We comprehensively evaluated the impact of fine-tuning different scale of network migration parameters on our proposed datasets. The results have shown satisfactory classification accuracy compared to the traditional ML methods and fully-trained CNN model.

The other organizational structures of this paper are as follows. Datasets are described in detail in Section 2. In section 3, we briefly review the CNN and deep TL. In Section 4, the proposed classification models are introduced. In Section 5, experiments and results analysis are presented, and finally the conclusion in Section 6.

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 these datasets 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, denseresidential, forest, freeway, golfcourse, harbor, river, storagetanks, 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. The classification of the Optimistic-RSD20 is challenging because of the high inter-class similarity and intra-class diversity, e.g., basketball_court and tennis_court have borders in the middle, freeway and bridge have cars on the road. Some examples of each class are shown in Figure 1.

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

ISSN: 2664-9640

Table 2. Optimistic-RSD20 datasets.

datasets	classes	Images per class	Image size	Spatial resolution	Total images
Optimistic-Data20	20	100	256*256	0.2-30m	2000

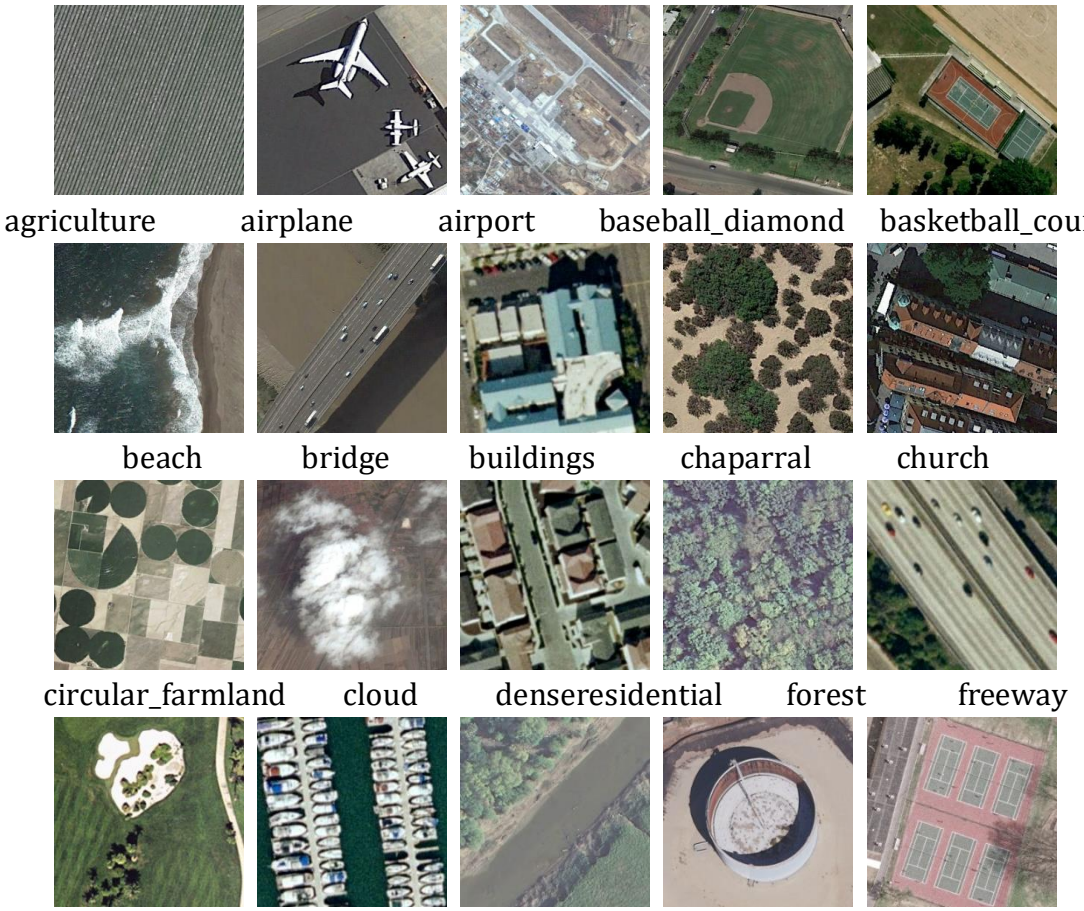

Figure 1. Examples of each classes from the Optimistic-RSD20 datasets.

3. Convolutional Neural Network

Convolutional layer (CONV), is the most important layer in feedforward CNN, which perform 2D/3D convolution operation on the input image and the convolution kernel of network weight. The value of each eigenelement is obtained by the dot product of the convolution kernel with its corresponding local region. Each hidden layer contains low-level features of the original image in the shallow neural unit while deeper neurons can express higher-level features with more combinations of lower-level features. In the case of three-dimension datasets, the input tensor of CONV is $x^l \in \mathbb{R}^{H^l \times W^l \times D^l}$, convolution kernel is $k^l \in \mathbb{R}^{H \times W \times D^l}$, the calculation equation is as follows:

$$y_{i^{l+1},\ j^{l+1},d} = \sum_{i=0}^{H} \sum_{j=0}^{W} \sum_{d^l=0}^{D^l} f_{i,j,d^l,d} \times x^l_{i^{l+1}+i,\, j^{l+1}+j,\, d^l}, \quad (1)$$

Among,

ISSN: 2664-9640

$$0 \le i^{l+1} < H^{l+1} \tag{2}$$

$$0 \le j^{l+1} < W^{l+1} \tag{3}$$

Where H,W represents high and wide of the image respectively, (i^{l+1}, j^{l+1}) is the positional coordinates of the convolution result, M represents the number of convolution kernel, $\boldsymbol{f}_{i,j,d^l,d}$ presents weights learned.

Pooling layer **(PL)**, is another substantial concept in CNN, which reduces the size of input layer by local non-linear functions. The value of each eigenelement is obtained by down sampling and is only related to the size of the receptive field of the pooling nucleus. However, unlike the convolution layer, there are no parameters to learn in the pooling layer. It most important hyperparameters are the size of the pooling window and the step size. Although pooling layer reduces the dimension of features map, it has strong robustness for feature transmission, such as translation, rotation and scaling. The following equation expresses average-pooling and max-pooling configuration:

Average-pooling:

$$y_{i^{l+1},\ j^{l+1},d} = \frac{1}{HW} \sum_{0 \le i < H, 0 \le j < W} \boldsymbol{x}^{\boldsymbol{l}}_{i^{l+1} \times H + i, j^{l+1} \times w + j, d^l} \tag{4}$$

Max-pooling:

$$y_{i^{l+1},\ j^{l+1},d} = \max_{0 \le i < H, 0 \le j < W} \boldsymbol{x}^{\boldsymbol{l}}_{i^{l+1} \times H + i, j^{l+1} \times w + j, d^l} \tag{5}$$

Fully connected layer **(FC)**, which follows the convolutional and pooling layers to form the last layers of the network structure. Increasing the number of **FC** neurons can effectively improve the learning capacity of the DCNN model. Through the feature information acquired from the convolution layer and pooling layer, the high-level feature information is finally fed into the classifier. The deep feature vectors associated with each FC layer can be calculated as follows:

$$y_j^l = \phi\left(\sum_{n=1}^{k} x_i^{l-1} \times w_i^l + b_l^j\right) \tag{6}$$

Where y_j^l is the output of feature map, ϕ represents the activation function, x_i^{l-1} is the feature map of ***l-1*** layer, w_i^l and b_l^j are the weight and bias trainable parameters.

Figure 2. Non-linear activation function. (a) sigmoid(x)=$\frac{1}{1+exp(-x)}$; (b)tanh(x)=$\frac{exp(x)-exp(-x)}{exp(x)+exp(-x)}$; (c) Relu(x)=$\begin{cases} x\ if\ x > 0 \\ 0\ if\ x < 0 \end{cases}$;(d) Leaky Relu(x) = $\begin{cases} x & if\ x > 0 \\ \alpha \cdot x & if\ x < 0 \end{cases}$.

4. Proposed Method

In this paper, a classification method associated with 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 feature map. It is noteworthy that all parameters trained in the previous

ISSN: 2664-9640

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 requiring training to the classifier for training and conduct partial and full fine-tuning, and output layer would provide a score for each class. Its highest score will be identified as the class to be distinguished.

This paper uses 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 does not care about the input size, in fact, 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 3.

Table 3. 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 connected layer. The first fully 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):

$$\mathrm{Z} = \{Z_1, Z_2, Z_3, \cdots, Z_K\}, \mathrm{P} = \{P_1, P_2, P_3, \cdots, P_K\} \tag{9}$$

$$P_j^k = \frac{\exp(Z_j)}{\sum_{j=1}^{K} \exp(Z_j)} \ (0 < P_j^k < 1, \sum_{j=1}^{K} P_j^k = 1) \tag{10}$$

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

$$L_{soft\max} = \sum_{j=1}^{n} \sum_{k=1}^{c} y_j^k \log P_j^k \tag{11}$$

Where $L_{softmax}$ is the 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 needs to be re-trained in the target domain and better adapted to the remote sensing fields.

ISSN: 2664-9640

5. Experiments and Results Analysis

There exist six widely used, standard evaluation metrics in remote image classification: overall accuracy, average accuracy, precision, recall and 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.

Figs.2,3 and 4 present the classification performance of different class through a confusion matrix, where the diagonals of the matrix shows the percentage correctly distinguished for each class. Limited by the space, we here just show the confusion matrix with the best result by fully fine-tuning the three DCNN model. As shown, the Desnet121 method exceeded a classification rate of 90% for 15 classes, while Inceptionv3 achieved 100% classification for 9 classes. More obviously, the classification accuracy of the previously mentioned indistinguishable (baseball_diamond and basketball_court) is more than 70% after fully fine-tuning. e.g., the accuracy of two class achieved 95% with Desnet121 DCNN. However, the scores of three networks on the river classification are very low, for the InceptionV3 with an overall accuracy of 65%, which are misclassified as beach or bridge, because the images of three categories are basically water.

Finally, we compared the training from scratch on the LeNet5 [45] and Alexnet[22] architecture to the our proposed DCNN classification methods. Just the overall accuracy metrics were used to evaluate. It can be seen that in Table.9, by fine-tuning the three DCNN models was further boosted by about 30~40% percentage points, achieve a significant performance enhancement in comparison with the aforementioned two neural networks. In Table.10, compared to the ML classifier, the value of the overall accuracy was very low and all the results were less than 30%, because these traditional ML algorithms are not trained to extract effective feature information.

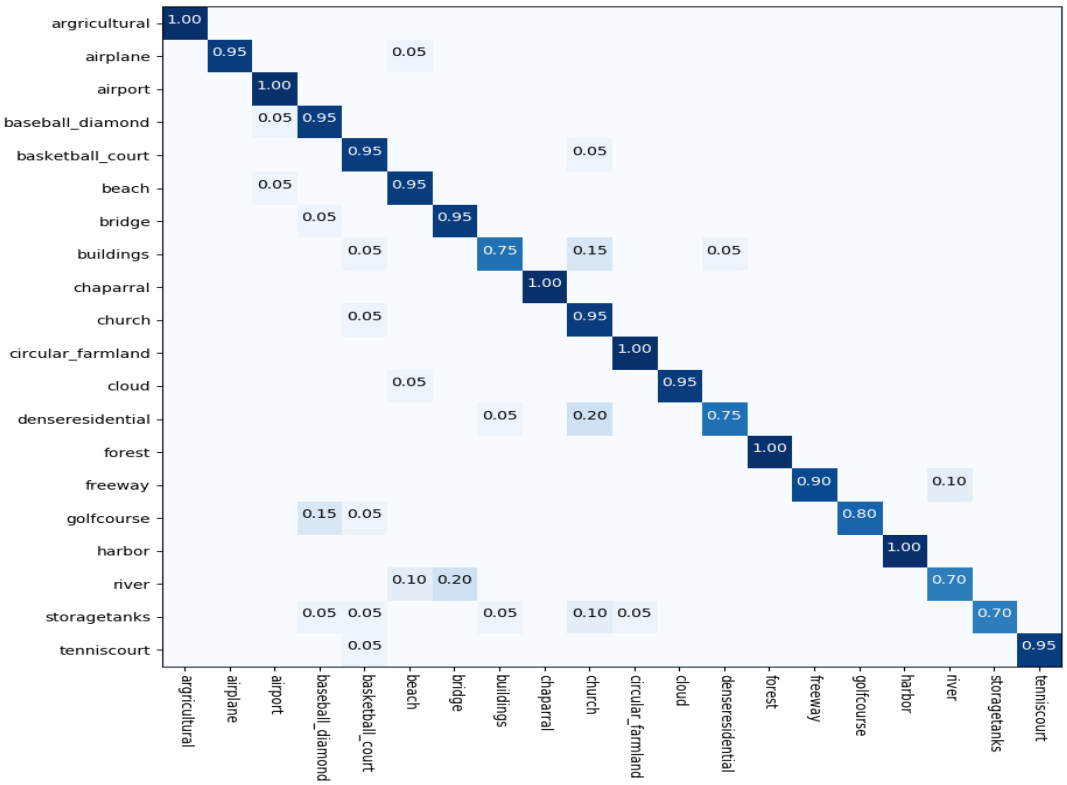

Figure 2. The confusion matrices showing the classification accuracies (%) for the Vgg16

ISSN: 2664-9640

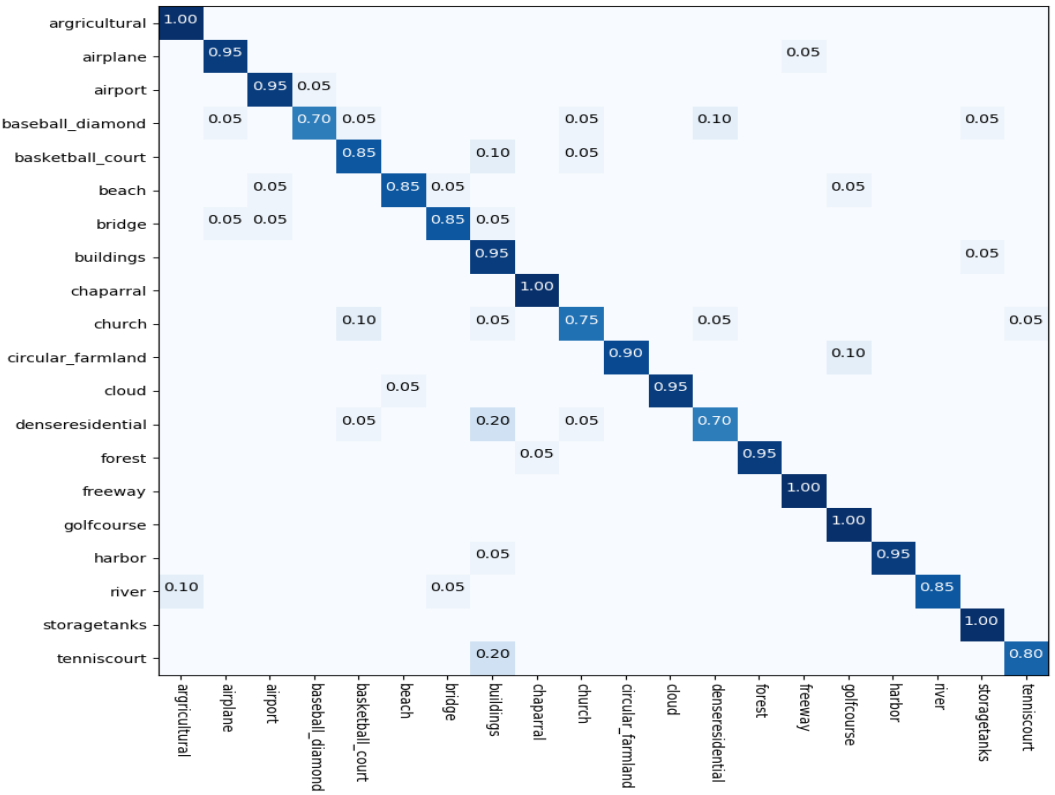

Figure 3. The confusion matrices showing the classification accuracies (%) for the Desnet121.

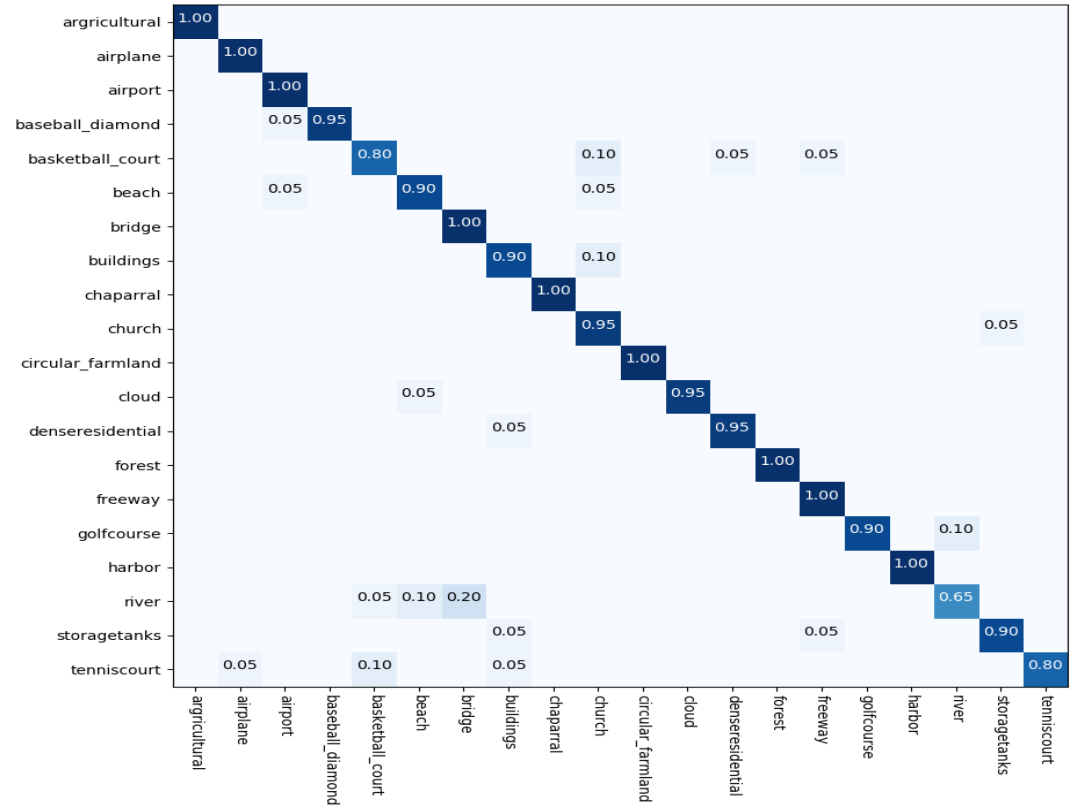

Figure 4. The confusion matrices showing the classification accuracies (%) for the Inceptionv3.

ISSN: 2664-9640

6. Conclusions

We used the pre-trained DCNN to extract high-level feature from different remote sensing data, further improved the accuracy of classification by fine-tuning different proportion of transfer learning parameters, and comprehensively evaluated our proposed model including overall classification accuracy, precision, recall and 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 classifier, 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, and 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 model or use fewer feature 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.

ISSN: 2664-9640

[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] Pan, S. J.; Yang, Q. A survey on transfer learning. IEEE Transactions on knowledge and data engineering. 2010, 22, 1345–1359.

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

[30] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv: 1409.1556.