High Spatial Resolution Remote Sensing Scene Classification Method using Transfer Learning and Deep Convolutional Neural Network

Wenmei Li, *Member, IEEE*, Ziteng Wang, Yu Wang, Jiaqi Wu, Juan Wang, Yan Jia, and Guan Gui, *Senior member, IEEE*

***Abstract*—Deep convolutional neural network (DeCNN) is considered one of promising techniques for classifying the high spatial resolution remote sensing (HSRRS) scenes, due to its powerful feature extraction capabilities. It is well-known that huge high quality labeled datasets are required for achieving the better classification performances and preventing over-fitting, during the training DeCNN model process. However, the lack of high quality datasets often limits the applications of DeCNN. In order to solve this problem, in this paper, we propose a HSRRS image scene classification method using transfer learning and DeCNN (TL-DeCNN) model in few shot HSRRS scene samples. Specifically, three typical DeCNNs of VGG19, ResNet50 and InceptionV3, trained on the ImageNet2015, the weights of their convolutional layer for that of the TL-DeCNN are transferred, respectively. Then, TL-DeCNN just needs to fine-tune its classification module on the few shot HSRRS scene samples in a few epochs. Experimental results indicate that our proposed TL-DeCNN method provides absolute dominance results without over-fitting, when compared with the VGG19, ResNet50 and InceptionV3, directly trained on the few shot samples.**

***Index Terms*—Transfer learning, deep convolutional neural network (DeCNN), few shot, high spatial resolution remote sensing (HSRRS), scene classification.**

I. Introduction

With the development of satellite remote sensing and computer technology, the spatial resolution and texture information of remote sensing image is improved and the processing approaches have been updated. High spatial resolution remote sensing (HSRRS) image with higher spatial resolution and abundant texture details have been performed well in object identification, classification and information extraction [1]–[3]. In recent years, a lot of HSRRS images have been acquired and significant efforts have been made for land use land cover (LULC) scene classification in the field of pattern recognition [4]–[8]. These approaches extract features firstly from training data and then build a classification model for testing other data. Most of the recognition methods are based on deep learning.

Deep learning has been successfully applied in extraction of abstract and semantic features [9]–[15], and it performs well in target identification, object detection and classification. Convolutional neural network (CNN) is one of typical deep learning algorithms, and many types of algorithms based on CNN (e.g., ResNet, VGG, Inception) have been developed in computer vision, natural language processing, medical and remote sensing image processing [16]. These practical applications indicated that the depth of a network is vital for the model, when adding layers to the network, it can extract more complex features. While the model with a deeper layer will obtain better performance and training CNN model, especially deep CNN (DeCNN) model often requires a lot of labeled data. However, it is hard to obtain a huge amount of labeled data to train the DeCNN model for HSRRS scene classification. In addition, it takes a lot of manpower and resources to label the HSRRS data. When the size of labeled data is not large enough, the trained DeCNN model easily show an over-fitting problem. Several studies have shown that transfer learning get a good performance in classification and recognition for small scale training data [17].

In this paper, we propose a transfer learning and DeCNN model (TL-DeCNN) based classification method to reduce the over-fitting problem and improve the classification accuracy with limited labeled samples. Specifically, three typical deep CNN models, i.e., VGG19, ResNet50 and InceptionV3, are combined with transfer learning, respectively, and these combined algorithms are called TLVGG19, TLResNet50 and TLInceptionV3. To assess the performance of TL-DeCNN for few shot HSRRS scene classification, the retraining and testing accuracy, loss, confusion matrix, overall accuracy (OA) and kappa coefficient (KC) are used. The main contribution of the paper includes three aspects:

- DeCNN based HSRRS image scene classification method is presented in a few shot samples. We train three CNN models, i.e., VGG19, ResNet50 and InceptionV3, in a few shot samples and evaluate their accuracy. Experiment

Manuscript received April 19, 2019; revised August 26, 2019.

This work was supported by the Project Funded by the National Science and Technology Major Project of China under Grant TC190A3WZ-2, the Natural Science Foundation of Jiangsu Province under Grant BK20191384, the China Postdoctoral Science Foundation under Grant 2019M661896, the National Natural Science Foundation of China under Grant 61671253, the Jiangsu Specially Appointed Professor under Grant RK002STP16001, the Innovation and Entrepreneurship of Jiangsu High-level Talent under Grant CZ0010617002, the Six Top Talents Program of Jiangsu under Grant XYDXX-010, the 1311 Talent Plan of Nanjing University of Posts and Telecommunications, Nanjing University of Posts and Telecommunications Science Foundation (NUPTSF Grant No. 218085). *(Corresponding author: Guan Gui)*

W. Li, J. Wu, and Y. Jia are with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mails: liwm@njupt.edu.cn, 1326464135@qq.com, jiayan@njupt.edu.cn)

Z. Wang, Y. Wang, J. Wang and G. Gui are with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mails: {1019010409, 1018010407, 1219012920, guiguan}@njupt.edu.cn)

results show that InceptionV3 is the best mode among the three models.
- TL-DeCNN based HSRRS image scene classification is proposed in limited labeled samples case. Our proposed TL-DeCNN model is trained in a limited labeled HSRRS scene samples in a few epochs by considering fine-tune.
- DeCNN based scene classification method is also considered as benchmark method using large amount of labeled HSRRS images.

The remainder of the paper is organized as follows. An overview of CNN based HSRRS image scene classification and transfer learning based application is presented in Section II. The proposed architectures based on DeCNN model and TL-DeCNN for HSRRS image scene classification with small and large amounts of labeled data are given in Section III, respectively. In the Section IV, the HSRRS image preprocessing, the architecture based on VGG19, ResNet50 and InceptionV3 for scene classification will be described, respectively. Also the evaluation indexes of the classification model will be described. Following, the results of HSRRS scene classification with DeCNN and TL-DeCNN with a few shot samples and quantitative indicators are described in Section V. Meanwhile, the results of the large amount of labeled HSRRS image scene classification based on DeCNN are compared with that of TL-DeCNN with few shot. Finally, some concluding remarks are drawn in Section VI.

II. Related Work

HSRRS image scene classification problem can be extracted subregions into different semantic classes, and it is a fundamental task and significant for remote sensing applications, such as urban planning, object detection, and natural resource management. Many recent works have demonstrated that CNN is the most successful and widely applied deep learning method, and has been used to make HSRRS image scene classification task [18]. Especially, the DeCNN performs well in semantic features extraction with a lot of convolutional layers and a large amount of training data set. However, it is difficult to train a DeCNN model with a few samples.

HSRRS image has higher spatial resolution and fewer spectral channels compared with coarse or medium spatial resolution remote sensing data, and it is more difficult to identify subtle differences among similar land cover types. Meanwhile, the phenomenon "same object with different spectrum" and "same spectrum with different objects" of HSRRS image leads to the failure in solving lots of classification tasks with high accuracy demand. Tremendous efforts have been made to develop robust and automatic image classification methods. Machine learning approaches (eg. support vector machine, random forest, k-nearest neighbor and multilayer perceptron) have been used widely in HSRRS image classification, and lots of achievements have been gained [19]–[21].

Recently, deep learning has represented the state of the art in a variety of domains, and CNN as a typical deep learning method, has obtained excellent results in the field of computer vision [22], wireless communications [23], [24] and remote sensing image processing [18]. HSRRS image scene classification based on CNN has achieved excellent results recently. Penatti *et al*. evaluated the generalization power of CNN features from fully-connected layers and obtained a state of the art result with a public HSRRS image data sets [25]. Feature fusion strategies to integrate the multilayers features to CNNs for HSRRS image scene classification have been proposed to complete the classification tasks [18], [26]–[29]. Gong *et al*. proposed a deep structural metric based learning approach for HSRRS image scene classification [30]. Ji *et al*. proposed a model based on multilevel features and attention model for remote sensing image scene classification [31]. Bi *et al*. proposed an attention pooling based convolutional network for aerial scene classification [32]. The early works have achieved excellent results in HSRRS image scene classification with a fully training CNN model. However, training a CNN model needs a considerable amount of labeled data set, which is rather difficult for HSRRS images. Many efforts have been made to add the training samples or improve the robustness of CNN, including data enhancement, detecting adversarial perturbations [33], increasing the depth of CNN and transferring the pre-trained CNN model or knowledge into scene classification task [34].

Transfer learning is an important solution for improving the robustness of CNN based classification models. Zhang *et al*. based on the features of adjacent parallel lines searched for regions of interest and confirmed the final targets through transfer learning on the AlexNet [36]. Li *et al*. proposed a best activation model (BAM) in the end-to-end process for LULC image classification [4]. Nogueira *et al*. proposed a method by transferring parameters from a pre-trained network and retrained the new network without parameter selection [37]. Zhao *et al*. combined the pre-trained AlexNet with a multilayer perception structure to make classification [38]. Huang *et al*. constructed a semi-transfer DeCNN to make image classification [39].

III. The Proposed TL-DeCNN based Method

Deep learning based HSRRS scene classification problem is still a challenge due to the limited labeled images. In this section, a robust classification method using TL-DeCNN is proposed. The architecture for our proposed TL-DeCNN based HSRRS image scene classification method is shown in Fig. 1. We can see that the architecture can be divided into three steps, the first step is training classification model based on ImageNet2015 and transfer the knowledge to the target classification task, the second step is fine-tuning with HSRRS images, and the third is the evaluation indicators for model and results. The goal of the architecture is to transfer deep knowledge from the ImageNet2015 to the limited training HSRRS image data in urban built-up areas scene classification, and improve the accuracy of classification.

A. Transfer Learning

Transfer learning is a popular training strategy to overcome the label-limited difficulty by initializing the training model with the parameters or knowledge which have been learned from other large data sets. Through fine-tuning with a small

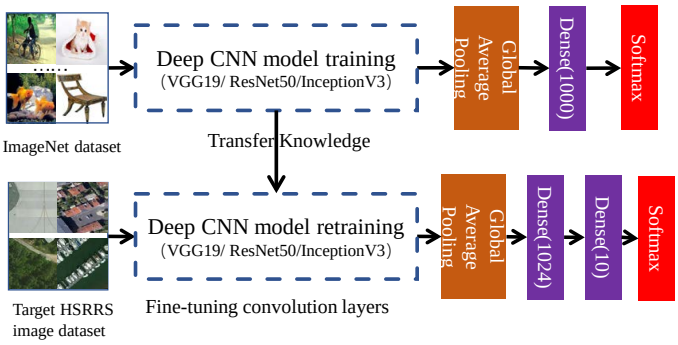


Fig. 1: The framework of our proposed TL-DeCNN based method.

amount of labeled data of the target task to obtain a better training model. The section II has show that CNN performs well in semantic information extraction and scene classification for HSRRS scene classification and object identification, and the pre-trained CNN model can be transferred to the current classification task. However, most of the researches focus on a shallow network with insufficient samples. And the DeCNN mostly focuses on object identification or classification with a large number of training samples, it needs a lot of labeled samples. When the depth of network increases, HSRRS image scene classification architecture may not feasible. In order to solve this problem, we proposed TL-DeCNN based HSRRS image scene classification methods in a few slot samples.

B. Knowledge Transfer from ImageNet2015 to HSRRS Scene Classification Task

This work is divided into three parts: model training based on ImageNet2015, feasibility of transfer learning between the ImageNet2015 and HSRRS scene classification task and the method for knowledge transfer. Firstly, the architectures of deep CNN models, VGG19, ResNet50 and InceptionV3 are applied to extract features. And then the applicable conditions of transfer learning are introduced. Finally, the extracted features are transferred into the HSRRS scene classification task.

1) DeCNN training: DeCNN contains more than one layer of CNN to extract discriminate features and accurate classification. A DeCNN usually is constructed by stacking several convolutional layers, pooling layers to form deep architecture [40]. CNN is one of the typical supervised learning methods, which need labeled data to learn and then make predictions for the unlabeled data. The input labeled data can be expressed as

$$\boldsymbol{x} = [\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(i)}, \cdots, \boldsymbol{x}^{(n)}]^T, \tag{1}$$

where $x^{(i)}$ represents i-th feature of x, and the training data is formed by pairs of feature x_i and output $f_i(x)$. Then the training function can be expressed as

$$\boldsymbol{T} = \{(x_1, f_1(x)), (x_2, f_2(x)), \cdots, (x_N, f_N(x))\} \tag{2}$$

where $\boldsymbol{x}_i = [x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(n)}]^T$. The training method of CNN is similar with (2). And the goal of CNN is to learn mapping from input features to output, which is represented by a model in application. The model can be expressed as

$$\mathscr{T} = f_{CNN}(\theta; S) \tag{3}$$

where θ is the parameters trained by CNN with samples S, and θ can be divided into two parts: θ=(θ^F, θ^{CCE}), and the former is feature extraction or learning and the latter is called classification cross-entropy (CCE) loss function, which is applied to make multi-category classification or prediction. Therefore, the equation can be written as

$$\tilde{\mathscr{T}} = f_{CNN}(\theta^F, \theta^{CCE}; S) \tag{4}$$

where $\tilde{\mathscr{T}}$ is the approximation of $\mathscr{T}$, and the formulas of CCE is given as

$$CCE(x) = -\sum_{i=1}^{C} y_i log(f_i(x)) \tag{5}$$

where C is the number of categories, y_i is the true label of i-th category, and $f_i(x)$ is the corresponding output of the model. Hence, features and classifiers will be got through CNN training with ImageNet2015. To extract deeper semantic features, VGG19, ResNet50 and InceptionV3 are applied to train the classification model, respectively. All of the training can be classified into two parts, feature extraction and classifier. As only the features or knowledge are useful for the following applications, the introduction of the approaches mainly focuses on feature extraction.

2) VGG19: One of the most popular DeCNN models is VGG19, which is developed by Simonyan *et al.* [41]. It is an influential DeCNN model, and it considers the depth of appropriate layers without increasing the total number of parameters. There are 16 convolutional layers and 3 fully connected layers in VGG19. And a series of convolutional, max pooling and rectified linear unit (ReLu) functions construct a convolutional block.

3) ResNet50: ResNet50 is one of the most common deep CNN for object detection and classification with a huge amount of samples, and it well resolves the degradation caused by the increasing number of layers in the network. It has been indicated that ResNet50 performs better in image scene classification than other CNN models in the ImageNet dataset [42]. The main idea of ResNet is to add a direct connection channel in the network, and it is called a highway network, which allows the original input information to be passed directly into the next layer. And its formula is given as

$$x_l = ReLU(f(x_{l-1}, w_l) + x_{l-1}) \tag{6}$$

where x_{l-1} and x_l is the input and output features of the l th and $(l+1)$ th layers, respectively. w_l is the weights associated with the l th layer of ResNet block. Each residual block consists of a series of layers, convolutional, batch normalization, pooling and ReLU. And it can resolve the gradient degradation and over-fitting problems very well.

4) InceptionV3: InceptionNet is proposed to increase the depth and width of the network, and finally improves the performance of the neural network. InceptionV3 is one of the most popular InceptionNet for classification [35]. It introduces the idea of factorization into small convolutions and uses branches not only in the inception module but also in the branches, which can promote high dimensional representations.

5) Transfer learning based method: Fig. 2 is the schematic diagram of transfer learning, and given a source domain $\mathcal{D}_{\mathcal{S}}$ and learning task $\mathcal{T}_{\mathcal{S}}$, a target domain $\mathcal{D}_{\mathcal{T}}$ and learning task $\mathcal{T}_{\mathcal{T}}$, transfer learning is defined to help improve the learning of the target predictive function $f_T(.)$ in $\mathcal{D}_{\mathcal{T}}$ with the knowledge in $\mathcal{D}_{\mathcal{S}}$ and $\mathcal{T}_{\mathcal{S}}$, where $\mathcal{D}_{\mathcal{S}} \neq \mathcal{D}_{\mathcal{T}}$, or $\mathcal{T}_{\mathcal{S}} \neq \mathcal{T}_{\mathcal{T}}$. What's need to be noted is that each domain is a pair $\mathcal{D}_{\mathcal{S}} = \mathcal{X}_{\mathcal{S}}$, $\mathcal{P}(\mathcal{X})_{\mathcal{S}}$ and $\mathcal{D}_{\mathcal{T}} = \mathcal{X}_{\mathcal{T}}$, $\mathcal{P}(\mathcal{X})_{\mathcal{T}}$, the condition implies for the source and target task, either the term features are different or their marginal distribution are different. Similarly, the tasks have the same requirement. Therefore, it can be inclined to that when the domains are different, either the feature spaces are different or the feature spaces between the domains are the same but the marginal probability distributions are different. And the definition implies that when there is some relationship (overt or covert) between the feature spaces of the two domains, the source and target domains are considered related, and transfer learning can be carried out between the two domains.

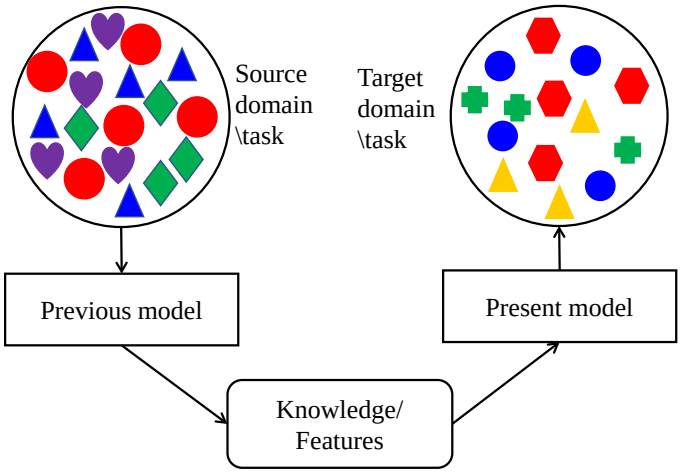


Fig. 2: The schematic diagram of transfer learning.

There are three topics in transfer learning, the first one is what to transfer, the second is how to transfer and the third is when to transfer. What to transfer means which part of knowledge can be transferred across domains or tasks. How to transfer means developing algorithms to transfer the knowledge and when to transfer asks in which situations, transfer learning should be done. In this paper, we aim to achieve good performance in the target HSRRS scene classification task by transferring knowledge from the source ImageNet2015 task, and as there are labeled data both in source and task domains, it belongs to the inductive transfer learning setting [43]. Meanwhile, the preliminary trained model based on DeCNN with ImageNet2015 is also geared to deep transfer learning. Compared with the non-deep approach, deep transfer learning automatically extracts more expressive features and meets the requirement of end-to-end in practical applications [43].

C. Fine-tuning for HSRRS Image Scene Classification Task

Fine-tuning is the process to initialize the HSRRS scene classification task network with the trained knowledge, which is transferred from the ImageNet2015. And the model is trained with the labeled HSRRS images further, the adjustment of parameters is the same with that in scratch training. It requires the layer of the initial network is the same with that of the source network, including the same layer name, types, setting parameters and so on. The fine-tuning is a vital process for HSRRS scene classification, not only make the network converge as quickly as possible but also make generic features contribute to a specific task. Compared with the learning rate in model training with ImageNet2015 (0.005), the fine-tuning learning rate is smaller (0.001), this setting can improve the accuracy of the HSRRS scene classification.

D. Accuracy Verification

The evaluation metrics include confusion matrix, OA, KC and precision. The confusion matrix is the most commonly used indicator for evaluating the performances. The OA is an indicator for evaluating the proportion correctly classified. The KC calculation using the confusion matrix is applied to check consistency and evaluate classification precision. It considers not only the overall accuracy but also the imbalance of the number of samples in each category. The precision is an indicator measuring the accuracy of each class, and it means the number classified into a certain class, which actually belongs to the true class.

IV. Experiments

In this section, to check the performance of the proposed TL-DeCNN, experiments have been conducted on three aspects. The first one is few shot HSRRS image scene classification based on VGG19, ResNet50, and InceptionV3, respectively. The second one is limited labeled HSRRS image scene classification based on TL-DeCNN, which means transferring the knowledge trained by VGG19, ResNet50, and InceptionV3 based on ImageNet2015, to the target limited labeled HSRRS image data set to make classification, respectively. And the third one is a large amount of labeled HSRRS images for scene classification based on VGG19, ResNet50, and InceptionV3, respectively.

A. Data Description

The HSRRS images collected in urban built-up areas are extracted from the UC merced land use dataset [45] and the remote sensing image classification benchmark (RSI-CB) dataset [46]. There are 10 categories objects needed to be classified in our experiments, and the sample size of training and testing for few, TL-DeCNN-few and large amount labeled samples are shown in Tab. I, respectively. All of the testing sample sizes are the same, and it is 100 samples for each

category. The few and TL-DeCNN-few amount of labeled samples for training is randomly selected in the large number of labeled samples. The training samples for TL-DeCNN-few not only contains the few HSRRS image samples but also includes the knowledge transferred from the ImageNet2015. Therefore, it combines the prior knowledge with the target to make an identification. It is notice that effective data augmentation has been made for the large number of labeled samples to enlarge the number of training samples and increase their diversity [18].

TABLE I: The sample size in our experiments.

C [1] \ SS [2]	FS [3]		TLDCNN-FS [4]		LS [5]	
	Train	Test	Train	Test	Train	Test
Airport	10	100	K[6]+10	100	578	100
Avenue	10	100	K+10	100	444	100
Bridge	10	100	K+10	100	369	100
Building	10	100	K+10	100	914	100
Roadside tree	10	100	K+10	100	321	100
Road	10	100	K+10	100	367	100
Marina	10	100	K+10	100	266	100
Parking lot	10	100	K+10	100	367	100
Residents	10	100	K+10	100	710	100
Storeroom	10	100	K+10	100	1207	100

[1] Category
[2] Sample size
[3] Few shot learning
[4] TLDCNN-few shot learning
[5] Large amount labeled sample
[6] Knowledge

B. HSRRS Image Scene Classification in a Few Shot

In this experiment, VGG19, ResNet50 and InceptionV3 are applied for HSRRS image scene classification in few shot case, respectively.

1) VGG19: There are 16 convolutional layers mainly using 3×3 convolutional kernels and 3 fully connected layers. The combination of convolutional, BN and ReLu layers constructs a convolutional block. The max pooling layer is applied in every two or three convolutional blocks. And the convolutional blocks are followed by the dense layers, which are set as 4096, 4096 and 10 in our experiment. Finally, the softmax is applied to make a classification. The accuracy and loss in the training and testing stages are shown in Fig. 3(a). It is easier to see that the accuracy in training is nearly to 100% and that in testing is lower than 40%. Meanwhile, the loss is close to 0 and fluctuating around 8 in training and testing stages, respectively, which means the VGG19 model is over-fitting in HSRRS image scene classification with limited labeled samples.

2) ResNet50: As illustrated in Fig. 1, the limited labeled HSRRS images are input into the ResNet50 model. And the accuracy and loss in training and testing phases are shown in Fig. 3(b). It can be seen that the training accuracy is nearly to 100%, and the testing accuracy is about 75% which is below 80% after training and testing process is stabilized. Meanwhile, the training loss is nearly to 0, and the test loss is larger than 2 when the model is stable. Compared with the accuracy and loss of VGG19, ResNet50 obtains a better performance, which reduces the over fitting phenomenon to some extent. However, the ResNet50 proposed for HSRRS scene classification with few shot still demonstrates a certain over-fitting problem.

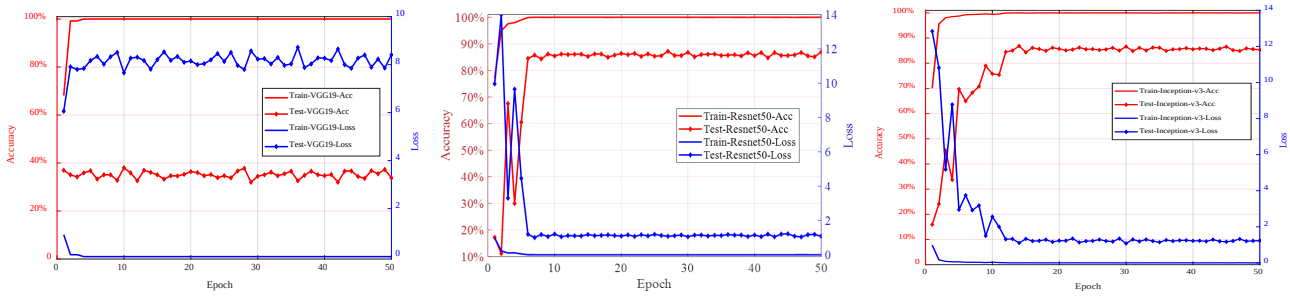
3) InceptionV3: To solve the over-fitting problem further, InceptionV3 is applied to the limited labeled HSRRS scene classification task. As described in section III, the idea of InceptionV3 is the fractorization, which promotes high dimensional representations. The accuracy and loss during training and testing stages are shown in Fig. 3(c). It shows that the accuracy is 100% and 83.0% in training and testing after stabilization, respectively. And the loss is 0 and about 1.8 in training and testing phases, respectively. Compared with the accuracies and losses of VGG19 and ResNet50, the InceptionV3 is better in solving the over-fitting problem. But the testing result is still much worse than that of training, and there is still over-fitting for InceptionV3 model with few shot.

C. TL-DeCNN based HSRRS Image Scene Classification Method

The TL-DeCNN is proposed to solve the over-fitting problem with limited training HSRRS images. Similar with that of few shot experiments, TL-DeCNN experiment is carried out based on limited labeled HSRRS image and knowledge transferred from ImageNet2015. Three typical deep CNN models VGG19, ResNet50 and InceptionV3 are considered in this experiment.

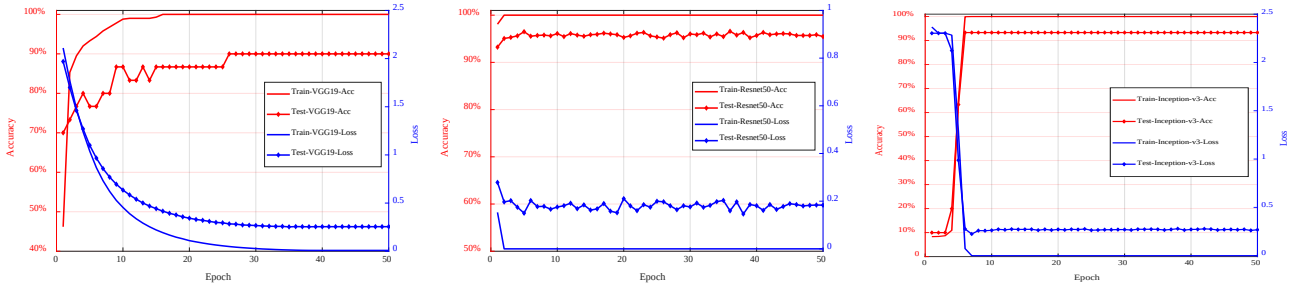
1) VGG19: The architecture of HSRRS scene classification based on transfer learning and VGG19 (TLVGG19) model can be seen from Fig. 1. The knowledge trained by VGG19 with ImageNet2015 is transferred to the limited labeled HSRRS scene classification task. The accuracy and loss during the task training and testing are shown in Fig. 4(a). When the process is stabilized, the training accuracy is 100%, and the testing accuracy is 90.0%. Meanwhile, the training loss is 0, and the testing loss is nearly to 0.25. Compared with that without transferred knowledge, the HSRRS scene classification task based on TLVGG19 performs better in accuracy and loss. The testing accuracy increases from about 40% to 90.0%, and the testing loss decreases from about 8 to 0.25. It demonstrates that the proposed approach can greatly reduce the effect of over fitting problems with limited labeled HSRRS images.

2) ResNet50: In few shot HSRRS image scene classification task, the architecture of transfer learning based on ResNet50 (TLResNet50) is also shown in Fig. 1. Similar to TLVGG19, the architecture transfers the knowledge trained with ImageNet2015 to the target HSRRS scene classification task. And the added one fully connected layer is able to map features to result better. The accuracy and loss during task training and testing are shown in Fig. 4(b), the training accuracy is 100% and the testing accuracy is about 93.3% when the processes are stable. The loss is 0 and 0.65 in the training and testing phase after the process stabilized, respectively. Compared with that without transfer knowledge, the testing accuracy increases about 18%, and the loss decreases about 74%. This result indicates that the TLResNet50 solve the effect of over-fitting problem well with limited labeled HSRRS image.



(a) (b) (c)

Fig. 3: The accuracy and loss during training and testing for (a) VGG19, (b) Resnet50, and (c) InceptionV3.



(a) (b) (c)

Fig. 4: The accuracy and loss during training and testing for (a) TLVGG19, (b) TLResnet50, and (c) TLInceptionV3.

3) InceptionV3: The architecture of transfer learning combined with InceptionV3 (TLInceptionV3) for limited labeled HSRRS image is also shown in Fig. 1. The accuracies and losses in training and testing processes are shown in Fig. 4(c). After the process is stabilized, the testing accuracy and loss is about 93.3% and 0.26 , respectively. Compared with the InceptionV3 without transferred knowledge, the testing accuracy increases by 10.3%, and the testing loss decreases from 1.8 to 0.26, which indicates that the approach we proposed is effective in solving the over fitting problem with limited labeled HSRRS images.

D. HSRRS Image Scene Classification in a Large Number of Labeled Samples

From the above experiments IV-B and IV-C, it has been found that the TL-DeCNN architectures, including TLVGG19, TLResNet50 and TLInceptionV3 are efficient and effective in solving the over-fitting problem. However, whether the accuracy and loss of TL-DeCNN can compare with that of a large number of labeled samples based on DeCNN. This experiment is carried out with augmented HSRRS images using VGG19, ResNet50 and InceptionV3, respectively.

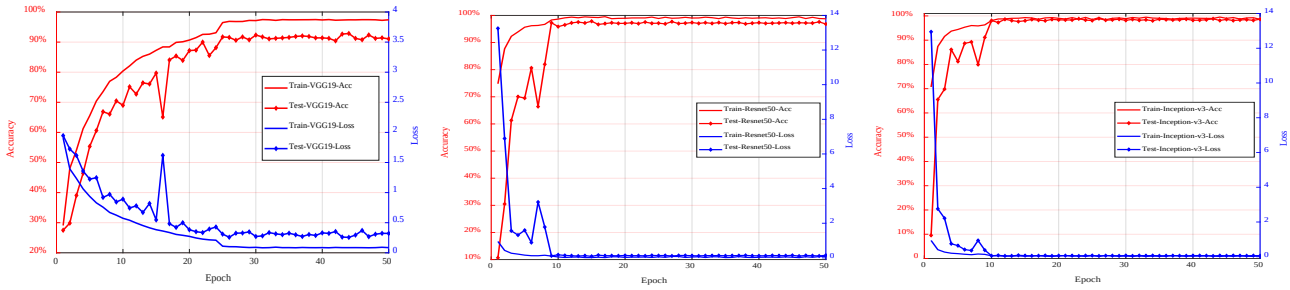
1) VGG19: As described in IV-A, there are more than 1064 samples (the size of the fewest samples is 266, and geometric transformations have been applied for data augmentation) for training in each category in the large amount of labeled data experiment. The accuracies and losses in training and testing are shown in Fig. 5(a), and it can be seen that the testing accuracy is about 90% and the testing loss is about 0.38, which is similar with that of TLVGG19. Therefore, it indicates that compared with the VGG19 based HSRRS scene classification trained with a large number of labeled samples, the TLVGG19 with few shot could obtain similar results, and reduces the effect of over fitting problem.

2) ResNet50: The ResNet50 is suitable for scene classification with a large number of labeled samples. The accuracies and losses in training and testing are shown in Fig. 5(b). After about 10 epochs, the testing accuracy and loss are stable, and the testing accuracy is close to 98% and the testing loss is nearly to 0. Compared with the testing accuracy and loss in TLResNet50 with few shot, ResNet50 architecture with a large number of labeled samples is better for HSRRS scene classification task. It demonstrates that the transfer learning contributes to the classification task, and the testing effect is inferior to the approach based on ResNet50 with large amount labeled samples.

3) InceptionV3: The InceptionV3 is a typical DeCNN for deep features extraction. It is good at extracting deep features from a large number of labeled samples. The accuracies and losses of InceptionV3 with a large amount of labeled HSRRS images in training and testing are shown in Fig. 5(c). It can be seen that after about 15 epoches, the testing accuracy and loss are stable, and the former is stable around 99%, the latter is stable around 0.1, which is better than that in TLInceptionV3.

V. Results and Discussions

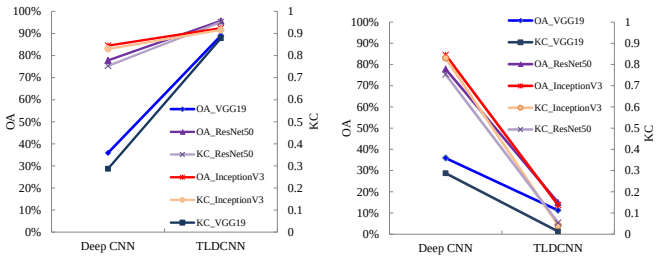
First of all, we present the confusion matrix of each DeCNN classifier. Fig. 6 shows the confusion matric of HSRRS image classification with VGG19, ResNet50 and InceptionV3 based on limited labeled samples. The OA of classification is 35.9%,



(a) (b) (c)

Fig. 5: The accuracy and loss during training and testing for (a) VGG19, (b) Resnet50, and (c) InceptionV3.

77.8%, and 87.0% for VGG19, ResNet50 and InceptionV3 architecture, respectively. The KC is 0.287, 0.753 and 0.856 for VGG19, ResNet50 and InceptionV3 architecture, respectively. Fig. 7 shows the confusion matric of HSRRS image classification with TLVGG19, TLResNet50 and TLInception-V3 based on limited labeled samples. The OA is 89.0%, 95.7% and 92.4% and the KC is 0.878, 0.952 and 0.916 for TLVGG19, TLResNet50 and TLInceptionV3 architecture, respectively. Fig. 8(a) shows the OA and KC of DeCNN and TL-DeCNN with fine-tuning. From the figure, we can see that the transferred knowledge improves the OA and KC for TL-DeCNN classification models. Transfer learning improves the OA of VGG19 (increases by 53.1%) most obviously, and has the least effect on the OA of InceptionV3 (increases by 5.4%). Meanwhile, for few shot learning, InceptionV3 obtains the best OA and KC, and after adding the transferred knowledge the TLResNet50 gets the best performance in OA and KC. And Fig. 8(b) is the corresponding OA and KC without fine-tuning, the best OA and KC is 14.9% and 0.054, respectively, for the three TL-DeCNN models, it may indicate that fine-tuning is a key step for ensuring forward transfer learning.



(a) (b)

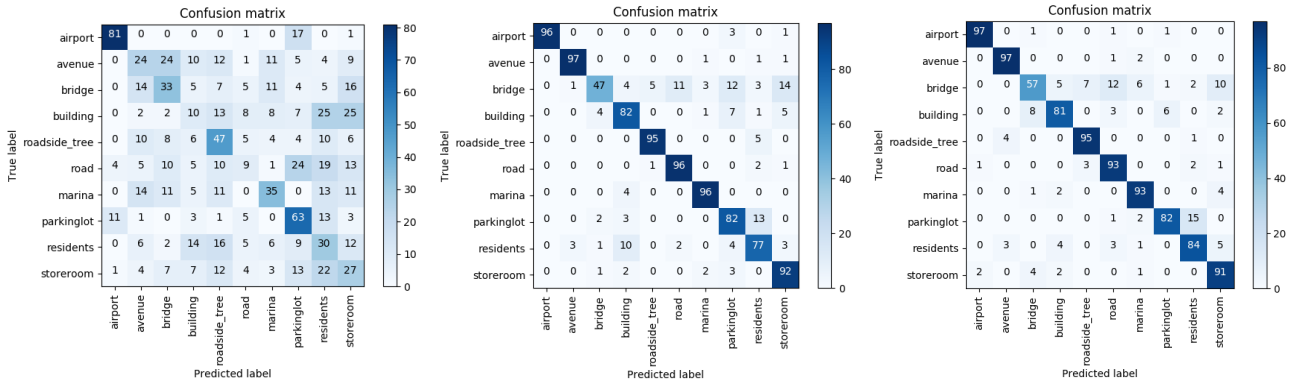
Fig. 8: The OA and KC of DeCNN and TL-DeCNN with (a) fine-tuning, and (b) without fine-tuning.

Then, the precision of each category with VGG19, ResNet50, InceptionV3, TLVGG19, TLResNet50 and TLInceptionV3 is labeled in the Tab. II. VGG19 obtains the lowest precision 9.0% for "road" identification, at the same time ResNet50 and InceptionV3 gets 99.0% and 91.0% precision for the same class. The same phenomenon appears in "roadside tree" and "marina" classes, and it may indicate that ResNet50 and InceptionV3 perform better for these objects identification. When the transferred knowledge is considered, the precisions for each category obtained by TLVGG19 are greatly improved, and the category with the greatest growth is "avenue", from 24.0% to 99.0%. Compared with VGG19, ResNet50 obtains better precision for all categories. The lowest precision is "bridge" 46.0%, and after the knowledge transferred into the model, the precision increases to 96.0%. Similar to the situation of VGG19, when the transferred knowledge is considered, the precisions of all categories are improved. The lowest precision is 41% of the InceptionV3 model for "bridge" identification. After the knowledge transferred into the architecture, the precision increases to 83%. Most of the precisions are improved, but the precision of the "airport" category decreases from 98% to 89%. It may be caused by the transferred knowledge which is extracted from huge airport information in ImageNet2015. The transferred knowledge contains intricate airport information, which is not similar or the same with our task "airport" in features. In short, the transferred knowledge improves the precisions of most of the categories for DeCNN scene classification tasks.

Finally, to evaluate the performance gap between TL-DeCNN based on limited labeled samples and DeCNN based on a large amount of labeled HSRRS images, the VGG19, ResNet50, and InceptionV3 are applied to make HSRRS scene classification with a large amount of labeled samples, respectively. The OA and KC is 96.1%, 97.1%, 99.4%, 0.956, 0.968, and 0.993 for VGG19, ResNet50 and InceptionV3, respectively. It's obvious that the OA and KC are both larger than that obtained by TL-DeCNN, among which the InceptionV3 obtains the best result for a large number of labeled samples, and for few shot samples TLResNet50 is the best architecture.

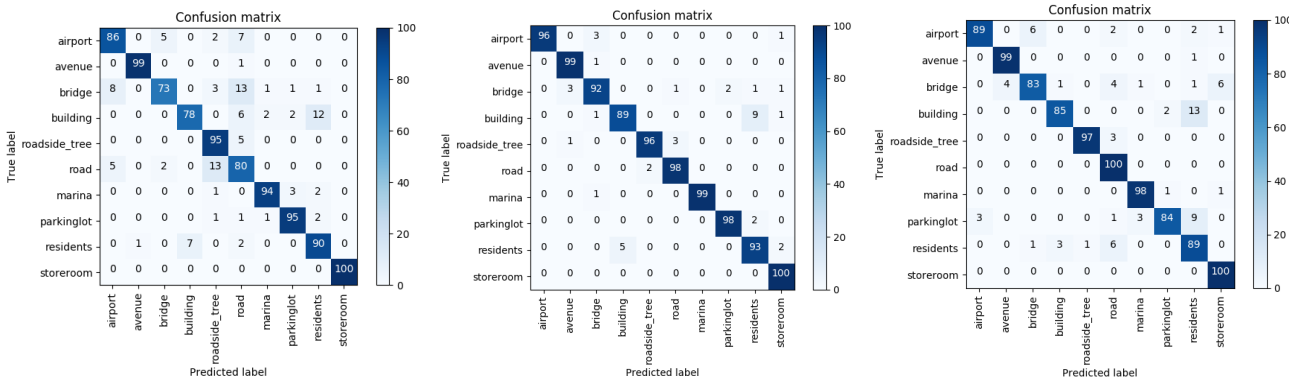
VI. Conclusion

In this paper, three TL-DeCNN models, i.e., TLVGG19, TLResNet50 and TLInceptionV3 are proposed for HSRRS scene classification in urban built-up areas. The main idea of our work is to solve the over fitting and gradient disappearance problems with limited labeled HSRRS images. Three experiments have been carried out, first is the DeCNN based HSRRS scene classification with few shot, the second is the TL-DeCNN based scene classification with the same few shot, and the third one is DeCNN based HSRRS scene classification with a large number of labeled samples. The results show that for few shot HSRRS scene classification, all of the three architectures TLVGG19, TLResNet50 and



(a) (b) (c)

Fig. 6: The confusion matric of limited HSRRS image samples based on (a) VGG19, (b) ResNet50, and (c) InceptionV3.



(a) (b) (c)

Fig. 7: The confusion matric of limited HSRRS image samples based on (a) TLVGG19, (b) TLResNet50, and (c) TLInceptionV3.

TABLE II: The precision of each category with 6 different architectures.

C [1] \ P [2]	VGG19	ResNet50	Inception V3	TLVGG19	TLResNet50	TLInceptionV3
Airport	81%	97%	98%	86%	99%	89%
Avenue	24%	79%	99%	99%	99%	99%
Bridge	33%	46%	41%	73%	96%	83%
Building	10%	77%	80%	78%	84%	85%
Roadside tree	47%	94%	95%	95%	98%	97%
Road	9%	99%	91%	80%	100%	100%
Marina	35%	93%	92%	94%	100%	98%
Parking lot	63%	66%	77%	95%	91%	84%
Residents	30%	58%	75%	90%	90%	89%
Storeroom	27%	95%	97%	100%	100%	100%

[1] Category
[2] Precision

TLInceptionV3 greatly improve the performance compared with that without transferred knowledge. And the ResNet50 is more suitable for transfer learning applications compared with VGG19 and InceptionV3, and InceptionV3 could reduce the over fitting and gradient disappearance problems to a certain degree and it performs better with few shot. Meanwhile, DeCNN based HSRRS scene classification with a large amount of labeled HSRRS images show that their performance are better compared with Tl-DeCNN with few shot. It indicates that there is still space for improvement of classification performance for transfer learning and DeCNN with few shot.

REFERENCES

[1] Y. Zhong, X. Han, and L. Zhang, "Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 138, pp. 281–294, 2018.

[2] H. Liu, X. Y. Huang, *et al.*, "Hybrid polarimetric GPR calibration and elongated object orientation estimation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 07, pp. 2080-2087, 2019.

[3] H. Liu, H. Y. Xia, *et al*., "Reverse time migration of acoustic waves for imaging based defects detection for concrete and CFST structures," *Mechanical System and Signal Processing*, vol. 117, pp. 210-220, 2019.

[4] B. Li, W. Su, H. Wu, R. Li, W. Zhang, W. Qin, S. Zhang, and J. Wei, "Further exploring convolutional neural networks potential for land-use scene classification," *IEEE Geoscience and Remote Sensing Letters*, doi: 10.1109/LGRS.2019.2952660, [Online].

[5] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2811–2821, 2018.

[6] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image cassification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11 pp. 6712–6722, 2018.

[7] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and Fisher discriminative convolutional neural networks for object detection," *IEEE Transactions on Image Processing*, vol. 28, no, 1, pp. 265–278, 2019.

[8] G. Cheng, Z. Li, X. Yao, and L. Guo, and Z. Wei, "Remote sensing image scene classification using bag of convolutional features," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1735–1739, 2017.

[9] H. Huang, J. Yang, Y. Song, H. Huang, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system", *IEEE Transactions on Vehicluar Technology*, vol. 67, no. 9, pp. 8549–8560, 2018.

[10] H. Huang, *et al*., "Fast beamforming design via deep learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1065–1069, 2020.

[11] G. Gui, *et al*., "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, 2018.

[12] F. Tang, *et al*., "An intelligent traffic load prediction-based adaptive channel assignment algorithm in SDN-IoT: A deep learning approach," *IEEE Internet of Things Jounral*, vol. 5, no. 6, pp. 5141–5154, 2018.

[13] B. Mao, *et al*., "A novel non-supervised deep learning based network traffic control method for software defined wireless networks," *IEEE Wireless Communications Magazine*, vol. 25, no. 4, pp. 74–81, 2018.

[14] Y. Wang, M. Liu, J. Yang and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4074–4077, 2019.

[15] N. Kato, *et al*., "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Communications Magazine*, vol. 24, no. 3, pp. 146–153, 2017.

[16] Z. Shao, J. Cai, P. Fu, L. Hu, T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sensing of Environment*, vol. 235, Dec. 2019, doi: 10.1016/j.rse.2019.111425.

[17] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, vol. 27, pp. 3320–3328, 2014.

[18] W. Li, H. Liu, Y. Wang, Z. Li, Y. Jia, and G. Gui, "Deep learning-based classification methods for remote sensing images in urban built-up areas," *IEEE Access*, vol. 7, no. 1, pp. 36274–36284, 2019.

[19] C. Zhang, T. Wang, P. M. Atkinson, X. Pan, and H. Li, "A novel multi-parameter support vector machine for image classification," *International Journal of Remote Sensing*, vol. 36, pp. 1890–1906, 2015.

[20] M. E. Mavroforakis, and S. Theodoridis, "A geometric approach to Support Vector Machine (SVM) classification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 671–682, May 2006.

[21] M. Belgiu, and L. Dragut, "Random forest in remote sensing: a review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[22] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 133–144, 2018.

[23] J. Sun, W. Shi, Z. Han, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10348–10356, Nov. 2019.

[24] M. Liu, T. Song, G. Gui, J. Hu, and H. Sari, "Deep cognitive perspective: resource allocation for NOMA based heterogeneous IoT with imperfect SIC," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2885–2894, Apr. 2019.

[25] O. A. Penatti, K. Nogueira, J. A. Dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, MA, USA, 12 June 2015, pp. 44–51.

[26] E. Li, J. Xia, P. Du, C. Lin, and A. Samat, "Integrating multilayer features of convolutional neural networks for remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5653–5665, 2017.

[27] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for VHR remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4775–4784, 2017.

[28] C. Ma, X. Mu, and D. Sha, "Multi-layers feature fusion of convolutional neural network for scene classification of remote sensing," *IEEE Access*, vol. 7, no. 1, pp. 121685–121694, 2019.

[29] H. Sun, S. Li, X. Zheng, and X. Lu, "Remote sensing scene classification by gated bidirectional network," *IEEE Transactions on Geoscience and Remote Sensing*, in press, doi: 10.1109/TGRS.2019.2931801.

[30] Z. Gong, P. Zhong, Y. Yu, and W. Hu, "Diversity promoting deep structural metric learning for remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 371-390, 2018.

[31] J. Ji, T. Zhang, L. Jiang, W. Zhong, and H. Xiong, "Combining multilevel features for remote sensing image scene classification with attention model," *IEEE Geoscience and Remote Sensing Letters*, in press, doi: 10.1109/LGRS.2019.2949253.

[32] Q. Bi, K. Qin, H. Zhang, J. Xie, Z. Li, and K. Xu, "APDC-Net: attention pooling-based convolutional network for aerial scene classification," *IEEE Geoscience and Remote Sensing Letters*, in press, doi: 10.1109/LGRS.2019.2949930.

[33] W. Li, Z. Li, J. Sun, Y. Wang, H. Liu, J. Yang, and G. Gui, "Spear and shield: attack and detection for CNN-based high spatial resolution remote sensing images identification," *IEEE Access*, vol. 7, pp. 94583-94592, 2019.

[34] F. Hu, G. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, no. 11, pp. 14680-14707, 2015.

[35] C. Wang, D. Chen, L. Hao, X. Liu, Y. Zeng, J. Chen, and G. Zhang, "Pulmonary image classification based on inception-v3 transfer learning model," *IEEE Access*, vol. 7, no. 1, pp. 146533–146541, 2019.

[36] P. Zhang, X. Niu, Y. Dou, and F. Xia, "Airport detection on optical satellite images using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1183-1187, 2017.

[37] K. Nogueira, Otavio A. B. Penatti, and J. A. Dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539–556, Jan. 2017.

[38] B. Zhao, B. Huang, and Y. Zhong, "Transfer learning with fully pretrained deep convolution networks for land-use classification," *IEEE Geoscience and Remote Sensing Letter*, vol. 14, no. 9, pp. 1436–1440, 2017.

[39] B. Huang, B. Zhao, and Y. Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sensing Environment*, vol. 214, pp. 73–86, Sep. 2018.

[40] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," Cambridge, MA, USA: MIT Press, 2016.

[41] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representations*, San Diego, CA, USA, 7–9 May, 2015.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, Jun. 2016, pp. 770–778.

[43] S. J. Pan, and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[44] A. Andreas, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in *Proceedings of the 19th International Conference on Neural Information Processing Systems*, MIT Press, 2006.

[45] Y. Yang, and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proccedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.

[46] H. Li, C. Tao, Z. Wu, J. Chen, J. Gong, and M. Deng, "RSI-CB:Alarge scale remote sensing image classification benchmark via crowdsource data," [Online]. Available: https://arxiv.org/abs/1705.10450

Wenmei Li (M'18) received the M.S. degree, PhD degree from Nanjing University and Chinese Academy of Forestry in 2010 and 2013, respectively. She is associate professor with School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications. And she is working for her postdoctoral studies (2018-) in Nanjing University of Posts and Telecommunications. Her research interests include deep learning, optimization, image reconstruct, and their application in land remote sensing.

Ziteng Wang (S'19) received the B.E degree in electronic information engineering from the North China Institute of Aerospace Engineering, China, in 2018. He is currently pursuing the masters degree with the Nanjing University of posts and Telecommunications, Nanjing, China. His research interests include deep learning, optimization, and their applications in remote sensing image processing.

Yu Wang (S'18) received his B.S. degree in Communication Engineering from Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China in 2018. He has published more 15 papers in peer-reviewed IEEE journal/conferences and received 2 best paper awards (i.e., CSPS 2018 and ICEICT 2019). He is currently working toward the Ph.D. degree in NJUPT. His research interests include deep learning, optimization, and its application in wireless communications.

Jiaqi Wu received the B.S. degree in geographic information science from the Nanjing University of Posts and Telecommunications(NJUPT), Nanjing, China, in 2019, where he is currently pursuing the M.A. degree. His research interests include spatial-temporal fusion, time series, deep learning and their applications in land remote sensing.

Juan Wang (S'19) is currently pursuing the master's degree in communication and information engineering with the Nanjing University of Posts and Telecommunications, Nanjing, China. She has published more than 5 peer-reviewed IEEE journal/conference papers in deep learning based physical layer wireless techniques. Her research interest includes machine learning for wireless communications.

Yan Jia received the double M.S. degree in telecommunications engineering and computer application technology from Politecnico di Torino, Turin, Italy, and Henan Polytechnic University, in 2013. Her Ph.D. degree was awarded in electronics engineering from Politecnico di Torino in 2017. Now she is working in Nanjing University of Posts and Telecommunications. Her research interests include microwave remote sensing, soil moisture retrieval, Global Navigation Satellite System Reflectometry (GNSS-R) applications to land remote sensing and antenna design.

Guan Gui (M'11–SM'17) received the Dr. Eng degree in Information and Communication Engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2012. From 2009 to 2014, he joined the wireless signal processing and network laboratory (Prof. Adachi laboratory), Department of Communications Engineering, Graduate School of Engineering, Tohoku University as for research assistant as well as postdoctoral research fellow, respectively. From 2014 to 2015, he was an Assistant Professor in Department of Electronics and Information System, Akita Prefectural University. Since 2015, he has been a professor with Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China.

He is currently engaged in research of deep learning, compressive sensing and advanced wireless techniques. Dr. Gui has published more than 200 international peer-reviewed journal/conference papers and received night best paper awards, e.g., ICNC 2018, ICC 2017, ICC 2014 and VTC 2014-Spring. He received Member and Global Activities Contributions Award (2018), and Top Editor Award of IEEE Transactions on Vehicular Technology (2020). He was also selected as for Jiangsu Specially-Appointed Professor (2016), Jiangsu High-level Innovation and Entrepreneurial Talent (2016), Jiangsu Six Top Talent (2018), Nanjing Youth Award (2018). Dr. Gui was an Editor of Security and Communication Networks (2012 2016). He has been the Editor of IEEE Transactions on Vehicular Technology, since 2017, the Editor of IEEE Access, since 2018, the Editor of Physical Communication, since 2019, the Editor of KSII Transactions on Internet and Information Systems since 2017, the Editor of Journal of Communications, since 2019, and the Editor-in-Chief of EAI Transactions on Artificial Intelligence, since 2018. He is IEEE Senior Member.