IEEE.org   IEEE *Xplore*   IEEE SA   IEEE Spectrum   More Sites

Donate   Cart   Create Account   Personal Sign In

Browse ⌄   My Settings ⌄   Help ⌄   Access provided by: | Sign Out
UNIV OF STIRLING

Access provided by: Sign Out
UNIV OF STIRLING

All   ▼   🔍

ADVANCED SEARCH

PDF

Help

Semi-MCNN: A Semisupervised Multi-CNN Ensemble Learning Method for Urban Land Cover Classification Using Submeter HRRS Images

Publisher: IEEE   **Cite This**   📄 **PDF**

Runyu Fan   ; Ruyi Feng   ; Lizhe Wang   ; Jining Yan   ; Xiaohan Zhang   **All Authors** •••

35
Cites in
Papers

1584
Full
Text Views

🔓   Ⓡ   🔗   ©   📁   🔔

Alerts

Manage Content Alerts

Add to Citation Alerts

Abstract

Document Sections

I. Introduction

II. Related Work

III. Semisupervised Multi-
CNN Ensemble Learning
Method

IV. Experiment and Analysis

V. Conclusion

Authors

Figures

References

Citations

Keywords

Metrics

More Like This

📄
Downl
PDF

Abstract: Submeter high-resolution remote sensing image land cover classification could provide significant help for urban monitoring, management, and planning. Deep learning (DL)-... **View more**

▶ **Metadata**

Abstract:

Submeter high-resolution remote sensing image land cover classification could provide significant help for urban monitoring, management, and planning. Deep learning (DL)-based models have achieved remarkable performance in many land cover classification tasks through end-to-end supervised learning. However, the excellent performance of DL-based models relies heavily on a large number of well-annotated samples, which is impossible in practical land cover classification scenarios. Additionally, the training set could contain all of the different land cover types. To overcome these problems, in this article, a semisupervised multiple-CNN ensemble learning method, namely semi-MCNN, is proposed to solve the land cover classification problem. Considering the lack of labeled samples, a semisupervised learning strategy was adopted to leverage large amounts of unlabeled data. In the proposed approach, an automatic sample selection method called an ensembled teacher model dataset generation was adopted to select samples and generate a dataset from large amounts of unlabeled data automatically. To tackle the error propagation problem, an important strategy was adopted to correct the errors by pretraining on the selected unlabeled data, and fine-tuning on the labeled data. Moreover, the semisupervised idea together with the multi-CNN ensemble framework was integrated into an end-to-end architecture. This could significantly improve the generalization ability of the semisupervised model, as well as the classification accuracy. Experiments were conducted on Shenzhen's land cover data and two other public remote sensing datasets. These experiments confirmed the superior performance of the proposed semi-MCNN compared to the state-of-the-art land cover classification models.

PDF

Help

☰ Contents

🔍

ᴛT

SECTION I.
Introduction

Land cover classification is a fundamental task of intelligent interpretation of remote sensing imagery, which aims to classify each pixel into a predefined land cover category. With the rapid development of earth observation and remote sensing technology, more and more submeter high-spatial-resolution remote sensing images have been acquired. Compared with low-spatial-resolution remote sensing imagery, submeter high-spatial-resolution remote sensing imagery can capture more details of urban objects, which makes it capable of urban monitoring [1], planning [2], [3], and management [4] with a higher level of discrimination.

Currently, the widely used remote sensing imagery land cover classification methods are based on supervised learning, which train the classifier according to the training samples and then classify the testing data into presetting categories with the trained classifier. Since AlexNet [5] was proposed in 2012, many deep learning (DL)-based models have been proposed to address the task of land cover classification. These methods use different DL-based models, such as autoencoder-based models [6], multilevel DL-based model [7], semitransfer learning-based model [8], and multiscale DL-based model [9]. By utilizing the spatial and spectral information comprehensively, and the powerful learning and feature conversion capabilities of deep models, these methods have achieved very significant results in land-cover-classification-related tasks.

Through the end-to-end supervised learning manner, these DL-based models can automatically learn and transform high-level features from numerous labeled samples. These can also achieve state-of-the-art performance without any statistical information from spectral or spectral–spatial features. However, in real urban land cover classification scenarios, as the resolution of remote sensing imagery increases, more spatial details can be captured. With cities becoming more fragmented, there is significant heterogeneity in urban land cover types. The highly heterogeneous land cover types pose a considerable challenge for land cover classification tasks. Based on a data-driven idea, the DL-based methods utilize a large number of labeled samples in order to train highly nonlinear neural networks to distinguish highly heterogeneous land cover types. In real scenarios, the lack of labeled samples leads to poor classification results, especially for DL-based models.

In the real-world land cover classification scenario, it is much easier to obtain unlabeled images than manually annotated datasets. Hence, many semisupervised-learning-based methods are proposed to use the raw unlabeled images to generate a labeled dataset, which effectively makes up for the lack of well-labeled samples. These methods combine semisupervised learning strategies (such as self-training approach [10] and pseudo-labeling strategies [11]) with support vector machine (SVM) [12], graph-based method [13], generative adversarial networks (GANs) [14], transferable deep models [15], discriminative adversarial learning [16], and other models. By using a large amount of unlabeled data, the effect of land cover classification is significantly improved. However, the wrong choice of samples always occurs in the sample selection process of the semisupervised method. These errors will be propagated to the downstream land cover classification tasks, reducing the final land cover classification accuracy. Hence, most

semisupervised models reduce the impact of error propagation by manually setting certain thresholds (confidence parameters). However, since it is impossible to prevent the propagation of errors, the accuracy of land cover classification is restricted.

Inspired by the significant progress made in semisupervised learning, a semisupervised multi-CNN ensemble learning method (semi-MCNN) for land cover classification is proposed in this work. To solve the lack of labeled samples problem, a semisupervised learning strategy that started with a small labeled dataset and then spread the label to other unlabeled samples was used to leverage the unlabeled submeter high-spatial-resolution remote sensing images. In this process, the wrong selection of samples might bring about an error propagation and affect the classification accuracy. To tackle this problem, a multimodel ensemble learning method was used to fuse multiple trained CNNs to reduce the errors in the sample selection. In addition, an important strategy was adopted to correct the errors by pretraining on the selected unlabeled data and fine-tuning on the labeled data. As a result, the proposed semi-MCNN approach integrated a semisupervised manner and the multi-CNN ensemble strategy into an end-to-end architecture. This effectively compensates for the poor generalization performance of a single model and the error propagation of a semisupervised model.

To validate the effectiveness of the proposed method, a dataset built from the submeter high-resolution remote sensing (HRRS) images of Shenzhen city, the ShenzhenLC dataset, was used for land cover classification. Experimental results with the ShenzhenLC dataset and two other widely used public remote sensing datasets showed the superior performance of the proposed semi-MCNN model compared to state-of-the-art land cover classification methods. The main contributions of this article are summarized as follows.

1) Focusing the lack of labeled samples problem on submeter urban land cover classification, we have proposed a simple semisupervised method (semi-MCNN), which integrates multiple CNNs to enhance the sample evaluation and uses pretrain fine-tuning for error correction.

2) Taking Shenzhen city as the research area, we have constructed a large-scale land cover dataset (ShenzhenLC).

3) The proposed semi-MCNN (10% training data) achieved similar accuracy close to supervised-learning-based methods (60% training data) on the ShenzhenLC dataset. A submeter land cover classification map of Shenzhen city in 2018 has been obtained.

The rest of this article is organized as follows. Section II presents the related work. In Section III, the proposed method, a semisupervised multimodel ensemble learning model for classification, is introduced in detail. Section IV contains a description of the experimental datasets and an analysis of the experimental results. Finally, Section V concludes this article.

SECTION II.
Related Work

Land cover classification is a fundamental task in remote sensing imagery. It can be classified into two categories according to whether the samples are fully or partially labeled: supervised-learning-based methods and semisupervised-learning-based methods. The supervised-learning-based methods usually require a large number of labeled samples to train a robust classifier, while the semisupervised-learning-based methods often use small labeled samples. These methods also aim to label or learn more samples in order to tackle the small sample problem. In recent years, the sample generation method represented by the GAN has also made impressive achievements in solving the lack of labeled samples problem. Therefore, in this section, the related work will be introduced from the following three aspects.

A. Supervised-Learning-Based Methods

Currently, the widely used remote sensing land cover classification methods are supervised-learning-based methods, which train the classifier based on the prepared training samples and then utilize the classifier to label the pixels into different land cover categories. Recently, the probability statistical-based methods, traditional machine-learning-based methods, and DL-based methods have become the focus of current trends of supervised-learning-based land cover classification.

The probabilistic statistical-based methods, including the nearest neighbor classification [17], [18] and the maximum likelihood method [19], usually assume that the data should obey normal

distribution. When this condition cannot be met, classification results are unpredictable and the performance might be poor. To improve the classification results, machine-learning-based methods such as SVM [20]–[22], artificial neural networks [23]–[25], and decision trees [26]–[28] have been proposed. These methods have achieved satisfying results. However, with these methods, the manually defined features all play an important role in the final result. Selecting and combining these defined features (including spatial features and spectral features) is time-consuming and limits the generalization of these models.

To reduce the dependence on the manually defined features, DL-based models have been intensely studied. Cheng *et al.* [6] proposed a novel a single-hidden-layer autoencoder and a single-hidden-layer neural network to train coarse-to-fine shared intermediate representations. Considering that satellite images are usually multitemporal and multisourced, Kussul *et al.* [7] described a multilevel DL-based architecture that targeted land-cover- and crop-type classification from multitemporal and multisource satellite imagery. To effectively use the rich spectral information, inspired by transfer learning, Huang *et al.* [8] proposed a semitransfer deep convolutional neural network approach for multispectral remote sensing imagery. Some studies treated land cover classification as a semantic segmentation task, and the representative methods are a series of variants of FCN [29]–[32] and Unet [9], [33].

DL-based methods can learn discriminative representation of different features through an end-to-end supervised process. However, there are still some problems that need to be solved for these DL-based methods. Traditionally, the excellent performance of DL-based models relies on a large number of well-labeled samples. For a large study area such as Shenzhen city, there are significant differences in the visual representation of land cover types. Manually labeling a complete dataset containing all of the different visual representations of land cover types is expensive and time-consuming. As a result, reducing the dependence on labeled land cover samples is an important research topic.

B. GAN-Based Methods

Self-supervised-learning-based high-resolution sample generation is another way to solve the lack of labeled samples problem. Since the GAN [34] can generate a large number of high-quality samples in a self-supervised manner, many GAN-based works have been studied in depth for sample generation to promote the solution to the lack of labeled samples problem. To generate high-quality remote sensing imagery samples, Xu *et al.* [35] applied the scaled exponential linear units into the GAN. Ma *et al.* [36] designed the SiftingGAN to generate more numerous and diverse labeled samples for data augmentation. Yu *et al.* [37] introduced the attention GANs, which integrates the attention mechanism into GANs for aerial scene classification. More recently, Han *et al.* [38] have proposed the GANRSIGM, which integrates the Wasserstein distance into the GAN to create high-resolution samples for scene classification. In terms of semisupervised learning combined with generative models, Zhan *et al.*'s [14] research shows that GANs combined with semisupervised learning can also have a good hyperspectral image classification performance. Zhu *et al.* [16] proposed a semisupervised center-based discriminative adversarial learning framework for a cross-domain scene-level land cover classification of aerial images.

Most GAN-based methods use GAN for sample generation or feature learning. Compared with CNN-based methods, currently, the performance of GAN-based methods is inferior to CNN-based methods [39]. The possible reasons can be concluded as follows: First, most GAN-based classification methods cannot be trained end-to-end, because they usually require labels to train an additional classifier. Second, the training of GAN is challenging. In the case of a few samples, the performance cannot be guaranteed (for example, spatial information and spectral information may be lost). Third, the GAN-based methods generate samples from a certain distribution. In a real scene, the samples may not completely conform to the distribution or not only conform to the distribution. However, due to the powerful self-supervised feature learning ability of the GAN, the GAN-based methods still provide a promising future direction for land cover classification.

C. Semisupervised-Learning-Based Methods

To reduce the reliance on a large number of labeled samples, many semisupervised-learning-based methods for land cover classification were studied in-depth. The semisupervised-learning-based methods mainly include a semisupervised strategy and a learning model. Self-training [10] is a widely used semisupervised learning strategy. Focusing on obtaining an expanded annotation dataset, the self-training strategy annotates the most reliable predictions to unlabeled instances. In self-training, the classifier first trains on a limited number of labeled samples and then merges the most reliable instances with the initial samples into an expanded dataset. The model is finally optimized on the expanded dataset. However, existing self-training methods are based on handcrafted features, which rely heavily on manual design and cannot guarantee effectiveness. Pseudo-Label [11] technology is a simple and efficient semisupervised learning strategy for deep neural networks. During Pseudo-Label, the model is trained in a supervised process with well-

labeled data. For unlabeled data, the class that has the maximum predicted probability is picked up as if they were true labels. Self-training and Pseudo-Label show good results on semisupervised tasks, but these methods cannot guarantee reliability since they assume that the selected samples are all correct.

Since the semisupervised-learning-based methods have good performance in the case of limited samples, in recent years, many semisupervised-learning-based methods have been proposed for remote sensing image classification tasks. Camps-Valls *et al.* [13] designed a semisupervised graph-based method to handle the special characteristics of hyperspectral images for hyperspectral image classification. In terms of semisupervised learning combined with discriminative models, Liu *et al.* [12] proposed a novel semisupervised SVM model that utilized the self-training approach in order to address the problem of remote sensing land cover classification. Han *et al.* [40] presented the SSGA-E based on the cotraining strategy and DL. Inspired by transfer learning, recently, Tong *et al.* [15] have proposed a semisupervised-learning-based pseudo-labeling and sample selection scheme to train transferable deep models for land use classification with HRRS images. Yalniz *et al.* [41] presented a web-scale semisupervised teacher/student paradigm pipeline with large convolutional networks to leverage billions of images on the Internet to enhance image and video classification. Due to the dependence on manual preset parameters and thresholds, these methods still have room for improvement in real land cover classification scenarios.

These semisupervised-learning-based methods successfully solved the problem of reducing the dependence on a large number of labeled samples. However, they just only consider using a single feature space to evaluate samples. A single feature space has insufficient ability to distinguish samples. Thus, it is easy to select the wrong samples, which may ultimately affect the classification accuracy. Especially when dealing with highly heterogeneous land cover types, the overfitting problems and the poor generalization performance are prone to occur. In this article, we use multiple CNN ensembles to obtain different feature spaces through different neural networks. In this way, the error propagation problem in semisupervised sample selection can be largely alleviated. In addition, though these semisupervised methods can reduce error propagation by optimizing the sample selection method, there were still different levels of error propagation problems. To solve these problems, an automatic sample selection method called ensembled teacher model dataset generation (EMDG) is proposed. The proposed EMDG can reduce the propagation of errors in the semisupervised sample selection by automatically selecting samples with higher accuracy, without any manual setting of thresholds and parameters. Moreover, a strategy was used to further correct the propagated errors by pretraining on the selected unlabeled data and fine-tuning on the labeled data. In Section III, more details will be introduced for the proposed algorithm.

SECTION III.
Semisupervised Multi-CNN Ensemble Learning Method

To tackle the lack of labeled samples problem, as well as to enhance the generalization of the classification models, a semisupervised multimodel ensemble learning model (namely, semi-MCNN) is proposed in this article. This method aimed to identify the land cover types for submeter HRRS imagery. Formally, an HRRS image $I = \{x_{ij}\}$ was given, where i and j are the width and length of image I, respectively. The output $Y = \{y_{ij}\}$ with y_{ij} was the land cover label of x_{ij}.

The flowchart of the proposed semi-MCNN is shown in Fig. 1. It consists of a semisupervised learning strategy and ensembles multiple DL-based CNN models. Specifically, the proposed method contains two processes: training and testing. During the training process, the model utilizes the small samples and trains them with the semisupervised strategy. Then, the initial classification model can be obtained. For the testing process, the test dataset is input into the initial model and output with the probability value through the ensembled multiple DL-based model. The final label for each pixel is decided with the maximum probability value in the probability distribution. The details are introduced in the following two subsections.
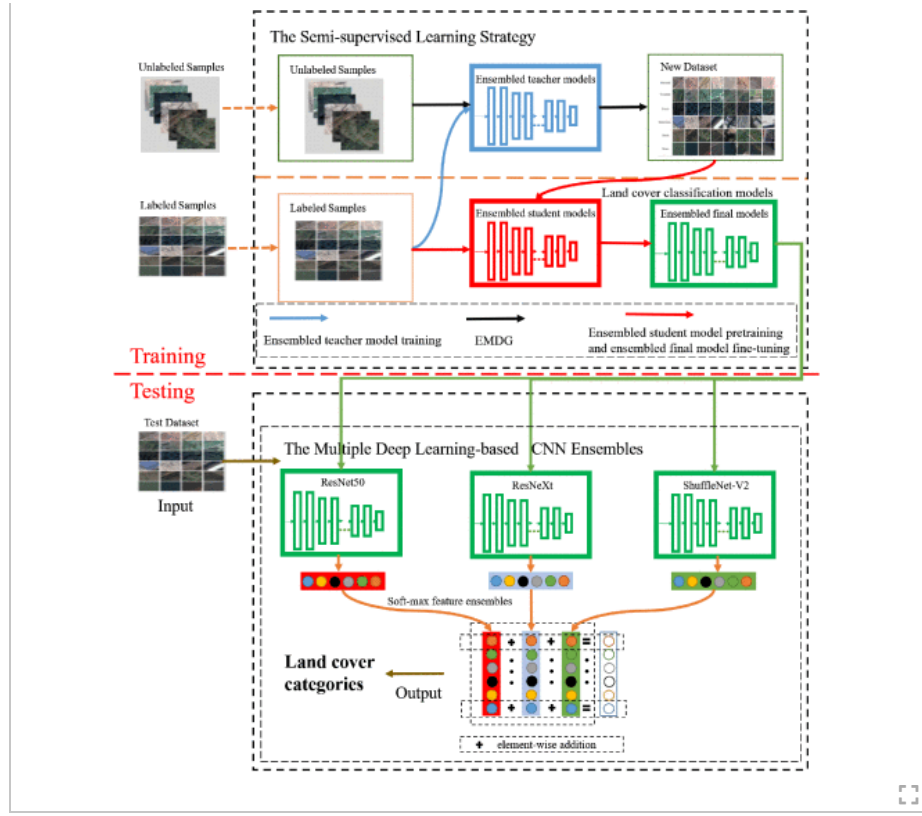
Fig. 1.
Flow of the proposed semi-MCNN.

A. Semisupervised Learning Strategy

In the proposed algorithm, to label the unlabeled samples more efficiently and precisely, a semisupervised learning strategy was adopted, as shown in Fig. 2. The semisupervised strategy was based on the idea of the teacher–student model. First, an ensembled teacher model was used to train the training samples initially to obtain a preliminary sample discrimination. Then, the fine-tuned ensembled teacher model was applied to select samples and generate a new dataset. Finally, the newly generated dataset was used to train a student model to get the final model. The generated dataset used by the student model was selected by the teacher model from unlabeled samples and would cause an error propagation problem. Therefore, a fine-tuning operation was adopted by the student model with a labeled dataset in order to optimize the final results. The semisupervised strategy is described as follows.
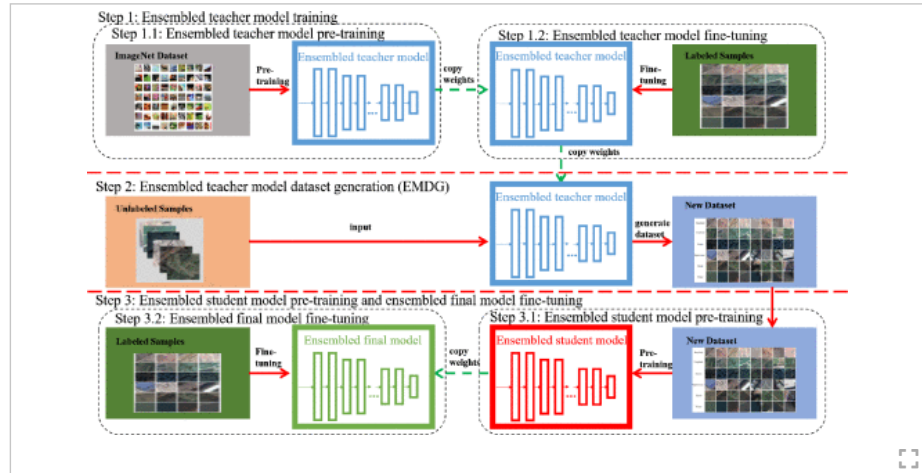

Fig. 2.
Semisupervised learning strategy.

Algorithm 1: Ensembled Teacher Model Dataset Generation.

Input: The lowest single category accuracy rate in the ensembled teacher model: $acc_{teacher}$, Unlabeled data: $U = \{u_1, u_2, \ldots, u_I\}$, Ensembled teacher model: T

Output: $NewDataSet\ D = \{d_1, d_2, \ldots, d_K\}$

1:   **function** EMDG($acc_{teacher}, U, T$)

2:   $D \leftarrow \varnothing$

3:   **for** $i = 1$ to I **do**

4:   Input u_i into the model T and calculate the corresponding category and its accuracy, and denote it as acc

5:   **if** $acc \geq acc_{teacher}$ **then**

6:   Add u_i to D according to the corresponding category

7:   **end if**

8:   **end for**

9:   Calculate the number of samples in each subset in D. The number of samples in the smallest subset are called the minimum number N

10:   Sort the samples in each subset in D in decreasing order according to acc.

11:   Keep the samples within the top N of each subset in D, and discard the remaining samples.

12:
13:   **return** D

14:   **end function**

1) Ensembled Teacher Model Training
In the semisupervised learning strategy, the first step is to train the teacher model. A good teacher model will significantly reduce the error transmission problem in the semisupervised process and directly affect the final performance of the proposed method. Therefore, two strategies are used to guarantee a better teacher model. The first one is the use of transfer learning. The parameters trained by the ImageNet dataset are adopted in order to initialize the teacher model. Next, a fine-tuning operation is carried out for the teacher model with the training samples. The second strategy is the multimodel ensemble work. Because a single model can easily fall into a local optimum, and the generalization ability is not enough, in the semi-MCNN, a variety of different CNN models are used for the ensemble methods. The features extracted and transformed by different CNN models are ensembled to obtain different features. Through the above two strategies, a good ensembled teacher model is obtained for the semisupervised-learning-based classification.

2) Ensembled Teacher Model Dataset Generation
After the teacher model is obtained, labeling the large unlabeled dataset and generating an augmented training sample dataset is another task. Currently, the sample generation methods mostly rely on thresholds, which are preset manually. These include the top-K [41] and the SSGA-E [40] method, and the performance might be greatly affected by the manually preset thresholds. To improve these circumstances, a semisupervised dataset generation method, namely, the EMDG, is presented in this article.

The process of EMDG is shown in Algorithm 1. Here, the lowest single category accuracy rate in the ensembled teacher model ($acc_{teacher}$), the unlabeled dataset, $U = \{u_1, u_2, \ldots, u_I\}$, and the ensembled teacher model T are the input. The purpose of the EMDG is to obtain a generated dataset: $NewDataSet\ D = \{d_1, d_2, \ldots, d_K\}$. First, the unlabeled sample is input into the ensembled teacher model T to calculate the category and the corresponding accuracy rate, which is denoted as acc. To reduce the error of the sample selection with teacher model T, samples with a higher confidence value are selected (lines 5–8 in Algorithm 1). In the newly generated dataset, the number of samples in each category might be inconsistent, which will cause the category imbalance problem and affect the final land cover classification accuracy. To avoid this problem, the number of samples in each category, which are denoted as N, are calculated to ensure the minimum number for each category. Then, the first N samples in descending order of acc for each class are selected to obtain the final dataset D (lines 9–12 in Algorithm 1).

3) Ensembled Student Model Pretraining and Ensembled Final Model Fine-Tuning

Finally, the ensembled student model is trained with this newly generated training dataset. Because the newly generated dataset D is automatically selected from the unlabeled samples, it will inevitably contain some wrong samples, which will result in error propagation problems. To solve this problem, the weights of the ensembled student model are inherited and used to initialize the final model. Then, the final model is fine-tuned with the original labeled dataset. By pretraining on the selected unlabeled data and fine-tuning on the labeled data, the error propagation problems can be better solved. Through progressive training from coarse (the ensembled teacher model) to fine (the ensembled final model), the performance can be greatly improved.

B. Multiple DL-Based CNN Ensembles

To enhance the generalization of the semisupervised-learning-based classification methods, multiple CNN models are to be ensembled. Many works have shown that different structures of CNN can learn variant features, and some are good at extracting spatial features. Others are more skilled at extracting spectral features. Different components of CNN have different feature perception capabilities, such as kernel size, activation function, depth, and the number of hidden units. Therefore, the features generated by different CNNs can be combined to improve the land cover classification performance. Among them, residual networks [42] have efficient feature extraction capabilities, and the ResNet50 network is widely used for various image classification tasks due to its strong feature extraction and generalization capabilities. The ResNet50 network uses skip connections, which enables the network to be trained under a very deep architecture (hundreds of layers) and significantly improves the accuracy of the model. The ResNeXt [43], as a supplement and an upgraded version of residual networks, has multiscale analysis capabilities. The RexNeXt network proposes the aggregated residual transformations combined with the deep residual network to more effectively utilize the parameters. Compared with the residual network, RexNeXt has better classification ability with fewer parameters. ShuffleNet-V2 [44] is a lightweight network with efficient feature extraction capability and few parameters, which is very suitable for model fusion. Therefore, the ResNet50, ResNeXt, and ShuffleNet-V2 models are suitable to apply to remote sensing imagery for feature extraction.

The proposed semisupervised method, shown in Fig. 2, is first used to train the above three models independently and save the optimal model for each network. The above models are used to extract the softmax feature, which was the output vector of the softmax layer (denoted as $\text{output} = \{a_1, a_2, \ldots, a_k\}$, and k is the number of categories), where

$$a_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}, \quad i \in [1, k] \tag{1}$$

View Source ⊘

z is the input vector of the softmax layer.

Finally, the softmax features are ensembled to calculate the final output by normalizing the elementwise addition of these softmax features, which are defined as follows:

$$\text{output}_{\text{final}} = \frac{\sum_{i=1}^{N} \text{output}_i}{N} \tag{2}$$

View Source ⊘

where N is the number of models. The category corresponding to the largest value in output is the predicted land cover type.

SECTION IV.
Experiment and Analysis

To validate the effectiveness of the proposed framework, in this section, ResNet50 [42], ResNeXt [43], and ShuffleNet-V2 [44] are adopted and compared to the proposed method. More information about the study area, as well as the classification system, will be introduced in Section IV-A. Details about the experimental settings are presented in Section IV-B. The analysis on the experimental results is shown in Sections IV-C– IV-E, respectively.

A. Shenzhen Data and Classification System

1) Classification System

In this article, the Chinese Land Use Classification Criteria (GB/T21010-2017) is referred as our land cover classification system, and the Shenzhen data are classified into six categories: bare land, cropland, impervious, shrub, water, and forest, shown as Fig. 3.

Fig. 3.
Samples of land cover categories in Shenzhen city.

2) Shenzhen Data

In this study, Shenzhen is used as an example of urban land cover classification because it is a typical international city that attracts industry, trade, tourism, and finance. The city is China's first special economic zone, which covers an area of about 2020.5 km^2, and is located in the southern part of Guangdong Province, China. Shenzhen city borders Hong Kong in the south, Huizhou in the north and northeast, Dongguan in the north and northwest, the Pearl River in the west, and Dapeng Bay in the east. Over the past three decades, Shenzhen city has been one of the fastest growing cities in the world, and its land use and land cover have undergone tremendous changes. The houses, natural growth, and artificially planted vegetation have different visual appearances. Urban planning, construction, and development have made the land cover fragmented and heterogeneous. Affected by factors such as light, temperature, and humidity, land covers in different regions also show significant differences. These will increase the heterogeneity of land cover in Shenzhen city and affect its classification accuracy. Therefore, to better estimate the land cover types of Shenzhen city, in this study, Shenzhen is treated as a research area for land cover classification research.

To better analyze the land cover situation in Shenzhen city, we used submeter HRRS images of Shenzhen city obtained from Google earth and manually constructed a submeter (0.59 m) land cover classification dataset of Shenzhen (ShenzhenLC). The dataset contains six classes of land cover types, and each class contains 1000 images, as shown in Table I. In DL-based land cover classification experiments, the images are usually reshaped to the size of 28, 56, 128, 224, or 256 pixels. In this study, considering that the image patches should contain enough information (not too small), and also should not bring too many parameters to the model (not too large), all of the image patches are reshaped to 56×56.

TABLE I Statistics for the Land Cover Classification Dataset of Shenzhen City (ShenzhenLC)

Land cover categories	bareland	cropland	forest	impervious	shrub	water
Number of samples per category	1000	1000	1000	1000	1000	1000

B. Experimental Settings

To assess the performance of the proposed framework, ResNet50 and ResNeXt, together with ShuffleNet-V2, were adopted to analyze the land cover situation of Shenzhen city using the ShenzhenLC dataset. In addition, to prove the effectiveness and generalization of the proposed

Semi-MCNN, two experiments are also designed based on the public benchmark remote sensing datasets: the NWPU-RESISC45 dataset [45] and the aerial image dataset (AID) [46].

The semisupervised strategy in the proposed semi-MCNN model includes three training steps: ensembled teacher model training, EMDG, and ensembled student model pretraining and ensembled final model fine-tuning. In the ensembled teacher model training step, only 10% of the labeled samples are used to train models, without any unlabeled samples. In the EMDG step, we pick up unlabeled samples using the proposed EMDG algorithm. The number of samples picked for the ensembled student model pretraining is 337 per class ($337 \div 1000 = 33.7\%$) for the ShenzhenLC dataset, 61 per class ($61 \times 30/10000 = 18.3\%$) for the AID dataset, and 171 per class ($171 \div 700 = 24.2\%$) for the NWPU-RESISC45 dataset. In the ensembled final model fine-tuning step (the final training process), only 10% of the labeled samples (without any unlabeled samples) are used.

The parameters of the model were finally determined after many attempts. During the training, the Adam optimization method [47] was used. The learning rate was set to 0.0001 for the pretraining period, and 0.00001 for the fine-tuning period. Although the decay was set to 0.000001, the batch size was set to 32 and a total of 100 epochs were trained. In order to facilitate a comparison to other models, the following strategy was used to split the dataset: 10% as the training set, 50% as the unlabeled set, and 40% as the testing set. All implementations were based on Pytorch (Version 1.5.0) and an NVIDIA RTX 2080Ti GPU. In order to evaluate the proposed method, we assessed the experimental results using the overall accuracy (OA) and the confusion matrix, which are widely used evaluation criteria.

C. Experiment 1: Land Cover Classification Using the ShenzhenLC Dataset

Table II shows the OA of different models for the ShenzhenLC dataset. It can be seen that the proposed method had the highest OA, with nearly 3% improvement on the OA when using just 10% of the training set compared to other models. It is worth mentioning that the OA of the proposed method using only 10% of the training set outperformed the other models. This occurred even when 50% of the training set was used. It was also close to the other methods that used 60% of the training set. The experimental results using the ShenzhenLC dataset indicate that the proposed approach can effectively generate samples using a semisupervised learning strategy, and that the multiple DL-based CNN ensembles can simultaneously improve the accuracy of land cover classification.

TABLE II Results for the Proposed Models on the ShenzhenLC Dataset

	Model Name	OA (%)		
		10% training set	50% training set	60% training set
Supervised Methods	ResNet50	90.50	92.26	92.54
	ResNeXt	90.79	92.73	**93.17**
	ShuffleNet-V2	89.75	92.83	92.88
		OA (%)		
		10% training set and 50% unlabeled set		
Semi-MCNN	Semi-MCNN (ensembled teacher model)	92.04		
	Semi-MCNN (ensembled student model)	92.92		
	Semi-MCNN (ensembled final model)	**93.04**		

To test the effectiveness of the semisupervised strategy in the proposed model, models with different training stages were used. The final results with the OA values are shown in Table III. It can be seen that the proposed semi-MCNN (final) model achieved the highest OA value. The proposed semi-MCNN (final) model achieved an improvement of nearly 2% on OA compared to the semi-MCNN (teacher) model, and a nearly 1% improvement compared to the semi-MCNN (student) model. Still, the results of the additional experiment indicate that the proposed approach can effectively reduce the error propagation problem and enhance the classification accuracy. The classification map of Shenzhen city in 2018 is shown in Fig. 4. The original classification maps exceed 40 GB. In order to show it in this article, we resample it to 3508×2480 pixels.

TABLE III Results for Different Models of the Proposed Semi-MCNN on the ShenzhenLC Dataset

Model Name	ResNet50	OA (%)
Semi-MCNN (ensembled teacher model)	ResNet50	90.50
	ResNeXt	90.79
	ShuffleNet-V2	89.75
Semi-MCNN (ensembled student model)	ResNet50	91.75
	ResNeXt	91.83
	ShuffleNet-V2	91.67
Semi-MCNN (ensembled final model)	ResNet50	**92.29**
	ResNeXt	**92.67**
	ShuffleNet-V2	**92.50**
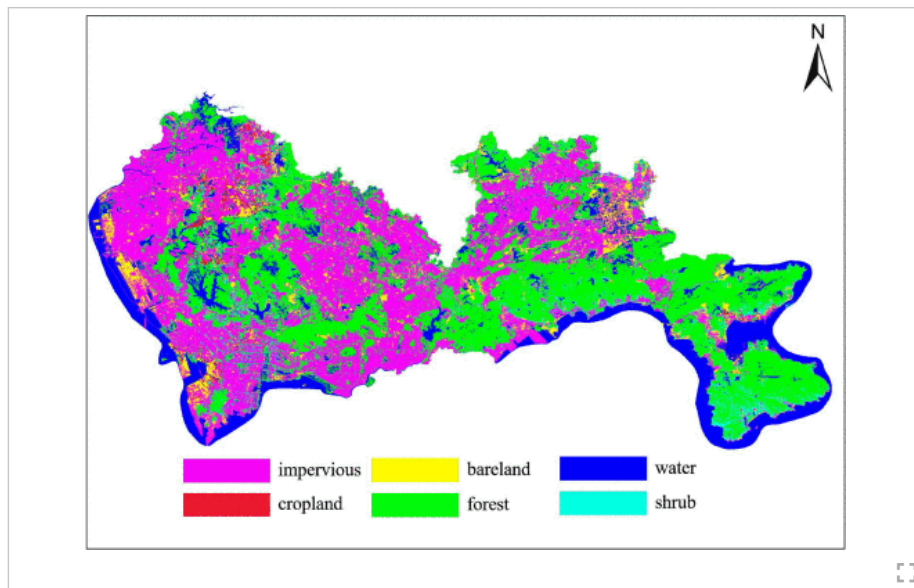
Fig. 4.
Land cover classification map of Shenzhen city.

Table IV shows the time cost of different models as well as the proposed semi-MCNN on the ShenzhenLC dataset. As can be seen from Table IV, for training time per epoch, the time cost of the proposed semi-MCNN model (6.40 s per epoch) on a single GPU is approximately the sum of the other three models. To reduce the overall time cost, the three different CNN networks were independently trained on different machines or GPUs. When the proposed model is trained on three GPUs separately, the time consumption (2.40 s per epochs) is close to the training time of these single models (ResNet50, ResNeXt, and ShuffleNet-V2). For the ShenzhenLC test data, the time cost of the proposed model on the entire test dataset (8.73 s) is about twice of the single models. Considering that for the neural networks, the time cost of training is much greater than the time cost of testing, so we believe that the twice of testing time will not have a great impact on the real application outcome of the proposed semi-MCNN.

TABLE IV Time Cost of Different Models and the Proposed Semi-MCNN on the ShenzhenLC Dataset

Model	Training Time per Epochs (s)	Time cost of the ShenzhenLC test data (s)
ResNet50	2.20	4.90
ResNeXt	2.40	5.12
ShuffleNet-V2	2.35	4.96
Semi-MCNN (One GPU)	6.40	8.73
Semi-MCNN (Three GPU)	2.40	-

The confusion matrices obtained by ResNet50 [42], ResNeXt [43], ShuffleNet-V2 [44], and the proposed method are shown in Fig. 5(a)–(f). It can be seen that for all models, the forest and shrub

categories were heavily confused, perhaps because of the difficulty in identifying different visual features, such as texture and structural information. For water, impervious surfaces, and bare land, all models had a high classification score. It demonstrated that visual features can help classify these categories well. It can be seen that the proposed method had the highest accuracy rate in all of the categories except for the shrub category, which obtained the second-highest OA accuracy.
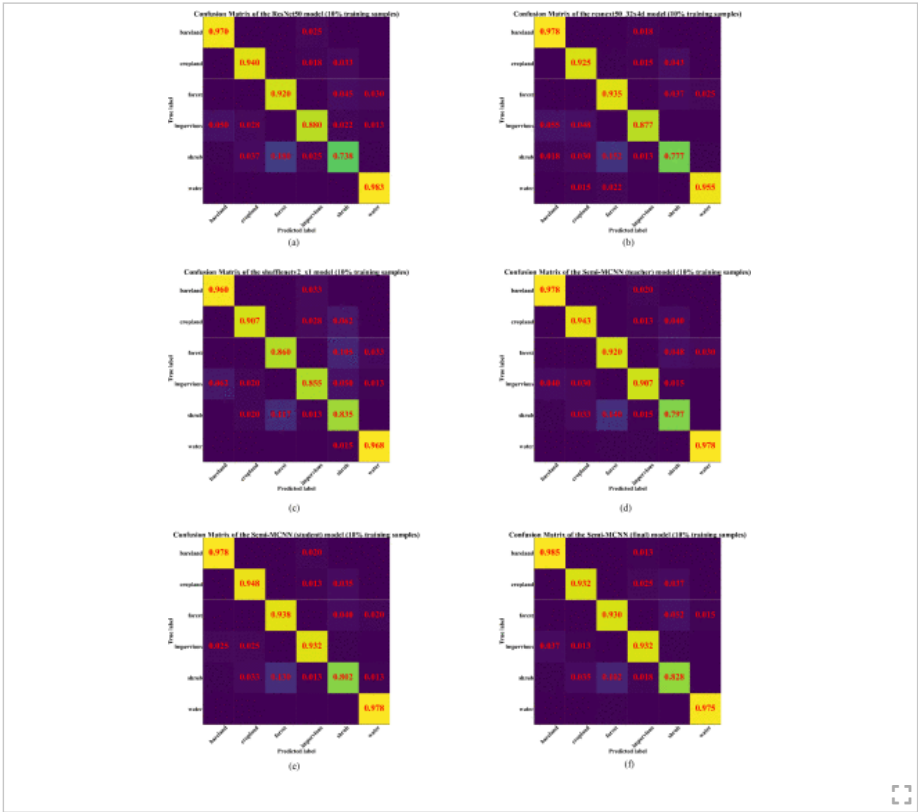

Fig. 5.
Confusion matrix of the mentioned methods for the ShenzhenLC dataset. (a) Confusion matrix of ResNet50. (b) Confusion matrix of ResNeXt. (c) Confusion matrix of ShuffleNet-V2. (d) Confusion matrix of semi-MCNN (teacher model). (e) Confusion matrix of Semi-MCNN (student model). (f) Confusion matrix of semi-MCNN (final model).

D. Experiment 2: NWPU-RESISC45 Dataset

The NWPU-RESISC45 dataset is a large-scale dataset that was used for the remote sensing imagery scene classification task, which contains 31 500 remote sensing images with 45 categories. The images were obtained from satellite as well as aerial photography. The spatial resolution varies from about 0.2–30 m. Due to the diversity of different perspectives, occlusions, poses, spatial resolutions, this dataset contains certain similarities between the different categories, as well as certain differences within the same categories. Recently, the NWPU-RESISC45 dataset has been widely used to evaluate the performance of different scene classification models. Therefore, in this study, it was also used to assess the proposed model with other state-of-the-art models.

Table V shows the quantitative results of the different models in our experiments. It can be seen that the proposed semi-MCNN outperformed the other semisupervised methods, with an improvement in OA ranging from 5% to 12%. The proposed semi-MCNN had the highest OA value compared to all of the supervised-learning-based models, with an improvement of nearly 1–3.5% in the case of the 10% training set. It can be seen that the semi-MCNN, in the case of the 10% training set, had a higher OA value than some of the supervised-learning-based models in the case of the 20% training set. It was also close to HW-CNN [51] and Hygra [52] in the case of the 20% training set, which are the current outstanding models for remote sensing image classification. The results also indicated that the proposed framework had a very competitive performance with small samples.

TABLE V Results for the Proposed Models on the NWPU-RESISC4 Dataset

Model Name	OA (%)	
	10% training set	20% training set
Supervised Methods		
D-CNN [48]	89.22 ± 0.50	91.89 ± 0.22
SF-CNN with GoogleNet [49]	87.43 ± 0.13	90.51 ± 0.13
SF-CNN with VGGNet [49]	89.89 ± 0.16	92.55 ± 0.14
Inception-v3-CapsNet [50]	89.03 ± 0.21	92.6 ± 0.11
HW-CNN [51]	-	94.38 ± 0.16
ResNet [52]	89.24 ± 0.75	91.96 ± 0.71
Hygra: an ensemble of CNNs [52]	**92.44 ± 0.34**	**94.51 ± 0.21**

	OA (%)
	10% training set and 50% unlabeled set
Semi-supervised Methods	
Self-training (VGG-S) [40]	81.46 ± 0.68
Self-training (ResNet) [40]	85.82 ± 1.30
Co-training [40]	87.25 ± 0.95
SSGA-E [40]	88.60 ± 0.95
Semi-MCNN	
Semi-MCNN (ensembled teacher model)	92.28
Semi-MCNN (ensembled student model)	91.13
Semi-MCNN (ensembled final model)	**93.48**

The confusion matrix of the proposed method for the NWPU-RESISC45 dataset is present in Fig. 6. It can be seen that some pairs of classes, e.g., church and palace, mountain and desert, rectangular farmland and terrace, and lake and wetland, were slightly confused. The reason for this is because there is a certain visual similarity between these categories. Some categories, such as church, freeway, commercial area, medium residential, palace, railway station, and thermal power station, where the OA was lower than 90%, were difficult to correctly classify due to their high interclass diversity.
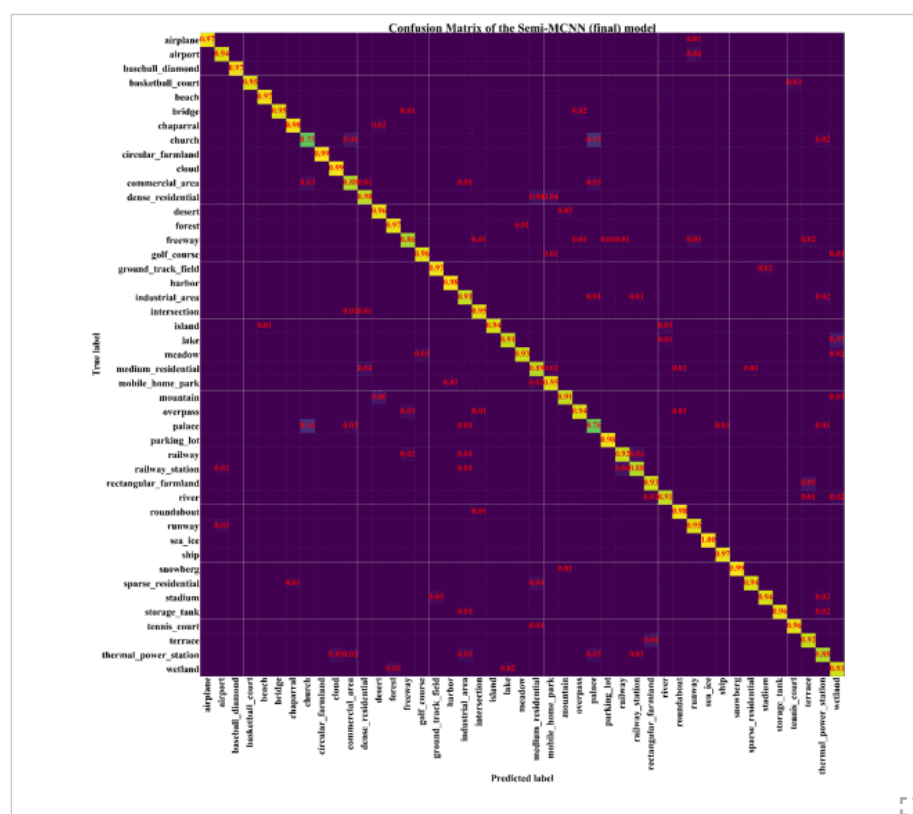

Fig. 6.
Confusion matrix of the proposed semi-MCNN on the NWPU-RESISC45 dataset.

E. Experiment 3: Aerial Image Dataset

The AID is different from the NWPU-RESISC45 dataset, as it is a large-scale AID collected from Google Earth images, which consists of 30 categories of aerial scenes with 10 000 images. The images in the AID are multisourced, as these come from different sensors. They are acquired at different locations, times, and seasons, and therefore, the different imaging conditions lead to a certain diversity within each category. This can bring certain challenges to different scene classification algorithms. In this section, the AID dataset was also used to evaluate the proposed method.

Table VI shows the quantitative results of different algorithms with the AID. Unlike the previous experiments on the NWPU-RESISC45 where the OA only reached 94.51%, on the AID, the OA of the best methods is higher than 95% and 97% with the labeled training set ratio of 20% and 50%. Compared with the NWPU-RESISC45, the AID contains fewer categories and smaller diversities and variations, which enables DL-based models to achieve better classification results. It can be observed that the semi-MCNN outperformed the other semisupervised methods with an improvement rate of 7–14%. Different from the results for other semisupervised comparison methods, the supervised-learning-based methods generally achieved an OA of higher than 90%, due to the use of more samples. Among them, the Hygra [52] achieved the best OA in the case of 20% and 50% ratio of training sets. The semi-MCNN in the case of the 10% training set achieved the highest OA value. It had an improvement rate of nearly 0.2–5% in the case of the 20% training set for supervised methods. In addition, in the case of the 10% training set, the proposed final model reached a higher OA than some of the supervised-learning-based models in the case of the 50% training set. When the proposed method used the 10% training set, the results were only 1–1.7% lower than the results of the current best classification methods ((D-CNN [48], SF-CNN with VGGNet [49], HW-CNN [51], and Hygra [52]) with the 50% training set.

TABLE VI Result of Proposed Models on the AID

	Model Name	OA (%)		
		10% training set	20% training set	50% training set
Supervised Methods	D-CNN [48]	-	90.82 ± 0.16	96.89 ± 0.10
	SF-CNN with GoogleNet [49]	-	91.83 ± 0.11	95.53 ± 0.09
	SF-CNN with VGGNet [49]	-	93.60 ± 0.12	96.66 ± 0.11
	VGG-16-CapsNet [50]	-	91.63 ± 0.19	94.64 ± 0.17
	HW-CNN [51]	-	-	96.98 ± 0.33
	Hygra: an ensemble of CNNs [52]	-	$\mathbf{95.50 \pm 0.27}$	$\mathbf{97.40 \pm 0.10}$
		OA (%)		
		10% training set and 50% unlabeled set		
Semi-supervised Methods	Self-training (VGG-S) [40]	81.46 ± 0.68		
	Self-training (ResNet) [40]	85.82 ± 1.30		
	Co-training [40]	87.25 ± 0.95		
	SSGA-E [40]	88.60 ± 0.95		
Semi-MCNN	Semi-MCNN (ensembled teacher model)	0.9410		
	Semi-MCNN (ensembled student model)	0.9490		
	Semi-MCNN (ensembled final model)	**0.9573**		

The confusion matrix of the semi-MCNN on the AID dataset is provided in Fig. 7. It can be observed that the resort and park were easily misclassified, and that the categories of school, resort, commercial, and square were slightly confused, as these have a high intraclass similarity.
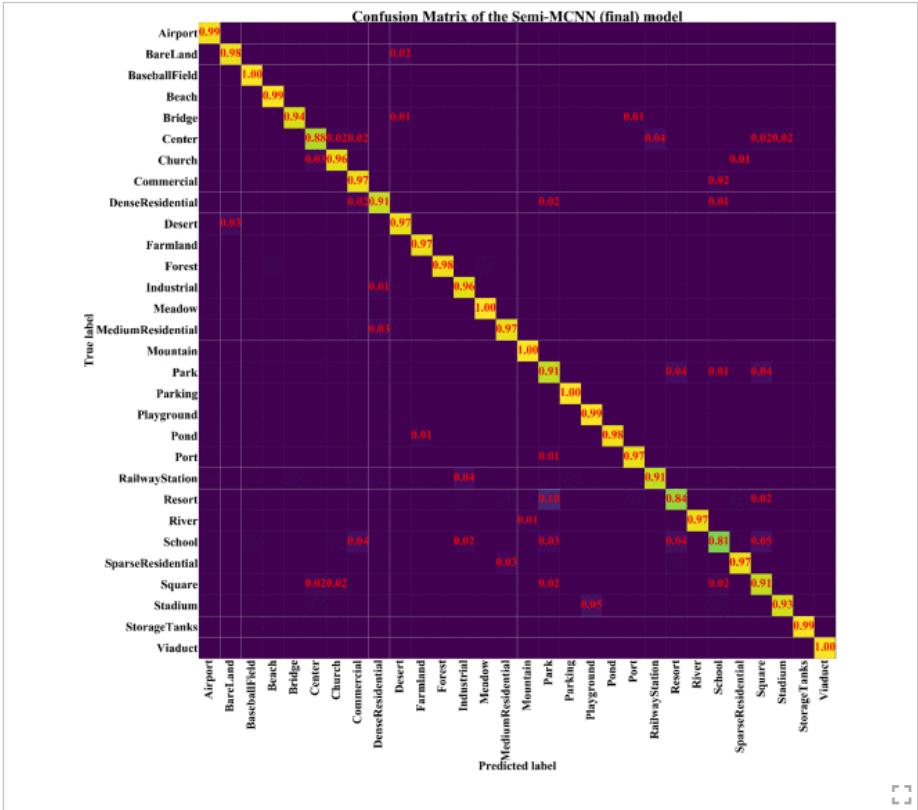
Fig. 7.
Confusion matrix of the proposed semi-MCNN on the AID.

SECTION V.
Conclusion

In this article, the semi-MCNN was proposed for urban land cover classification, which integrated a semisupervised strategy and a multi-CNN ensemble strategy into an end-to-end architecture. A semisupervision strategy was adopted to leverage lots of unlabeled images to labeled samples, and the EMDG was proposed to automatically select samples and generate a dataset from unlabeled data without any manual setting thresholds. In addition, pretraining on the selected unlabeled data and fine-tuning on the labeled data operation were taken in the proposed framework and used to solve the error propagation problem. To evaluate the performance of the proposed semi-MCNN method, several state-of-the-art land cover classification models were compared to the ShenzhenLC dataset, together with the public HRRS scene classification benchmarks, i.e., the NWPU-RESISC45 dataset and the AID. All of the experimental results demonstrated a consistent conclusion that the proposed semi-MCNN had better performance results compared to other state-of-the-art models in both quality and quantity. In addition, it can achieve comparative accuracy with small samples compared to the supervised model with big training sample sets and also proves the superiority of the semisupervised strategy and the multi-CNN fusion strategy of the semi-CNN in handling land cover classification tasks.

Authors	⌄
Figures	⌄
References	⌄
Citations	⌄
Keywords	⌄
Metrics	⌄

PDF

Help

More Like This

Coupled Adversarial Training for Remote Sensing Image Super-Resolution
IEEE Transactions on Geoscience and Remote Sensing
Published: 2020

Simple and Efficient: A Semisupervised Learning Framework for Remote Sensing Image Semantic Segmentation
IEEE Transactions on Geoscience and Remote Sensing
Published: 2022

Show More

PDF

Help

PDF

Help