Semisupervised Center Loss for Remote Sensing Image Scene Classification

Jun Zhang, Min Zhang, Bin Pan, and Zhenwei Shi, *Member, IEEE*

***Abstract*—High-resolution remote sensing image scene classification is a scene-level classification task. Driven by a wide range of applications, accurate scene annotation has become a hot and challenging research topic. In recent years, convolutional neural networks (ConvNets) have achieved promising performance among a variety of supervised classification methods. However, due to the lack of clearly labeled remote sensing images, it may be difficult to further improve the performance of scene classification. To address this issue, we propose a novel semisupervised center loss for scene classification. The main innovation of our method is to develop a cooperative framework of supervised and unsupervised branches in an end-to-end way. Specifically, we consider the class centers as guiding factors between the supervised and unsupervised branches. The supervised branch relies on a small number of labeled samples to generate class centers, which serve as initialization centers for the unsupervised branch. Meanwhile, the unsupervised branch utilizes the easily available remote sensing images to correct the class centers for enhancing the discriminative power of supervised ConvNets. Experimental results on three public benchmarks have indicated that the proposed method is superior to supervised center loss based methods.**

***Index Terms*—Cooperative framework, convolutional neural networks (ConvNets), remote sensing scene classification, semisupervised center loss (SSCL).**

I. Introduction

With the improvement of remote sensing image quality, remote sensing images have presented great potential in a lot of significant image interpretation tasks, such as scene classification, object detection, and semantic segmentation [1]–[4]. As a basic image understanding work, scene classification has attracted increasing attention. Different from pixel/object-level image classification, the main goal of scene classification is to automatically assign high-level semantic labels (e.g., school, parking lot, and railway station) to local areas of remote sensing images for achieving scene-level classification. The major difficulty lies in obtaining the discriminative features of high-resolution remote sensing scenes.

During the past decades, existing research works for remote sensing image scene classification can be roughly divided into three levels: low-level feature-based methods, mid-level feature-based methods, and high-level feature-based methods.

Low-level feature-based methods mainly focus on developing feature descriptors, which represent the primary visual attributes of remote sensing images such as spectral characteristics, texture characteristics, and geometric structure characteristics. The widely used feature descriptors include local binary pattern [5], color histogram [6], GIST [7], scale invariant feature transform (SIFT) [8], and so on. By comparing the SIFT and Gabor [9] texture features, the literature [10] analyzed the effect of image descriptions based on local measures of saliency on labeling high-resolution remote sensing images. Xia *et al.* [11] developed novel structural feature descriptors based on the topographic map and shape for interpreting remote sensing images. Considering the combined advantages of different low-level features, the literature [12], [13] fused multiple low-level features to describe remote sensing scenes.

For enhancing the representation power of low-level features, mid-level feature-based methods were proposed to mine the scene semantic information from low-level features. Yang and Newsam [14] adopted the standard bag of visual words (BoVW) approach to summarize the SIFT descriptors, and proposed a spatial extension termed spatial co-occurrence kernel to capture the spatial features. Based on the BoVW, the researchers [15]–[20] developed the probabilistic topic model (PTM) such as probabilistic latent semantic analysis (pLSA) and latent dirichlet allocation (LDA). The PTM reduces the dimensionality of mid-level features and constructs the semantic relationship between visual words. Nevertheless, mid-level features are derived from the low-level features, which lead to some limitations of mid-level features on remote sensing scene interpretation.

In more recent years, convolutional neural networks (ConvNets) have made remarkable achievements in the field of remote sensing image scene classification [21]–[26]. Unlike low-level and mid-level features based on artificial design, ConvNets generate feature representations of images by learning a large number of training samples. In addition, due to the multilayer structure of ConvNets, the obtained deep features are high-level abstraction of remote sensing scene contents, which make convolutional networks more suitable for scene-level

Manuscript received January 2, 2020; revised February 17, 2020; accepted February 24, 2020. Date of publication March 16, 2020; date of current version April 20, 2020. This work was supported in part by the National Key R&D Program of China under Grant 2017YFC1405605, in part by the Natural Science Foundation of Tianjin under Grant 19JCZDJC40000, and in part by the National Natural Science Foundation of China under Grant 61671037. *(Corresponding author: Bin Pan.)*

Jun Zhang and Min Zhang are with the School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China, and also with the Hebei Province Key Laboratory of Big Data Calculation, Tianjin 300401, China (e-mail: zhangjun@scse.hebut.edu.cn; zhangmin.hebut@hotmail.com).

Bin Pan is with the School of Statistics and Data Science, Nankai University, Tianjin 300071, China (e-mail: panbin@nankai.edu.cn).

Zhenwei Shi is with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.2978864

classification. Usually, quite a few researchers adopt the pretrained ConvNets [27]–[30] as the basic framework of classification networks. Penatti *et al.* [31]–[34] extracted the fully connected layer features from pretrained ConvNets to obtain the global information of remote sensing scenes. In order to further utilize the local information, Li *et al.* [35] proposed a region-wise deep feature extraction algorithm based on an improved vector of locally aggregated descriptors. To overcome the weakness of using only local or global features, Yuan *et al.* [36] first rearranged local features of the last convolutional layer for VGG-19, then concatenated global features of the last fully connected layer. Furthermore, a two-stage deep feature fusion model was proposed in [37] for integrating deep features of different pretrained ConvNets.

The above various algorithms improve the classification performance of remote sensing scenes. Nevertheless, it is worth noting that these methods mainly depend on feature transformation. Essentially, the constructed ConvNets may not directly generate discriminative features. Therefore, Wen *et al.* [38] regarded intraclass compactness as the learning goal, and designed the center loss to enhance the discriminative ability of deep features. Specifically, the center loss learns a center of each class from training samples and penalizes the distances between each sample of each class and its class center. Since the center loss specifically addresses the problem of large intraclass variations, the center loss and a visual attention mechanism to force the ConvNets to generate discriminative representations were introduced [39].

Inspired by the center loss algorithm, in this article, combining the following characteristics of remote sensing images, we propose a semisupervised center loss (SSCL) algorithm for remote sensing image scene classification.

1) The aforementioned center loss-based classification algorithms require labeled samples. However, the labeled samples are scarce seriously, especially for high-resolution remote sensing images.
2) A large number of unlabeled remote sensing images can be obtained every day, but the manual annotation for them is time-consuming.
3) The performance of center loss-based remote sensing scene classification algorithms may be further improved by integrating labeled and unlabeled samples.

The overarching goal of the SSCL algorithm is to further improve the performance of the center loss algorithm in scene classification task by learning more scene information contained in labeled and unlabeled samples. In general, SSCL comprises three key strategies. First of all, an end-to-end semisupervised framework is developed, which can deal with the labeled and unlabeled samples simultaneously. Second, we propose an improved clustering algorithm that makes the scene information learned from unlabeled samples as effective as possible. Finally, we establish a cooperation mechanism between the center loss algorithm and the clustering algorithm, which guarantees that the center loss better supervises the ConvNets to generate discriminative features.

The major contributions of this article can be summarized as follows.

1) We improve center loss to a semisupervised form, and construct an end-to-end deep learning model for remote sensing scene classification.
2) We design a cooperative dual-branch architecture to integrate the labeled and unlabeled samples and optimize our SSCL-based model.

The rest of this article is organized as follows. In Section II, we will first outline the proposed SSCL method. Subsequently, the overall framework, the details of SSCL algorithm, and the optimization algorithm are elaborated in Section II-B, II-C, and II-D, respectively. Section III reports the experimental results. Finally, the conclusion is drawn in Section IV.

II. Proposed Method

A. Overview of the Proposed Method

The developed SSCL algorithm is illustrated in Fig. 1. The datasets utilized in our method are composed of labeled and unlabeled high-resolution remote sensing images. Among them, different datasets (WHU-RS19 [11], UC-Merced dataset (UCM) [14], AID [40]) contain different categories, and the typical scene categories mainly include airport, residential, industrial, railway station, and storage tanks. Based on the overarching goal, our approach focuses on the problem of within-class diversity and between-class similarity for remote sensing scene classification, which can be abstracted as

$$J = D_w(f(I_i;\theta), f(I_j;\theta)) - D_b(f(I_i;\theta), f(\bar{I}_i;\theta)) \quad (1)$$

where I_i and I_j represent different scene images of the same class. I_i and $\bar{I}_i$ represent scene images from different classes. θ is the parameter to be learned. $f(\cdot)$ represents the features extracted from the ConvNets. D_w and D_b denote within-class and between-class differences. The task of our algorithm is to minimize the within-class difference D_w, and maximize the between-class distance D_b as much as possible.

The whole algorithm consists of three components: the overall framework, the SSCL, and optimization.

B. Overall Framework

Motivated by the superior performance of residual learning, we construct the SSCL based on the ResNet. Specifically, our backbone comprises 16 stacked residual blocks, and shortcut connections are inserted into all residual blocks. The architecture of each residual block is presented at the bottom right of Fig. 1.

In SSCL, the training data consists of the labeled sample set $S_{la} = \{x_i, y_i\}_{i=1}^{n}$ and unlabeled sample set $S_{un} = \{\widetilde{x}_j\}_{j=1}^{m}$, where x_i, y_i denotes the *i*th sample and its label, respectively. $\widetilde{x}_j$ denotes an unlabeled sample. *n* is the number of labeled samples, and *m* is the number of unlabeled samples. As shown in Fig. 1, SSCL contains two branches, where the yellow line is the supervised branch and the red line is the unsupervised branch. To obtain the discriminative features from supervised ConvNets, the labeled samples $\{x_i, y_i\}_{i=1}^{t_{la}}$ and unlabeled samples $\{\widetilde{x}_j\}_{j=1}^{t_{un}}$ are simultaneously fed into the backbone. The batch sizes of labeled samples and unlabeled samples are t_{la} and t_{un}, respectively. Then, the average pooling layer *N*-dimensional features

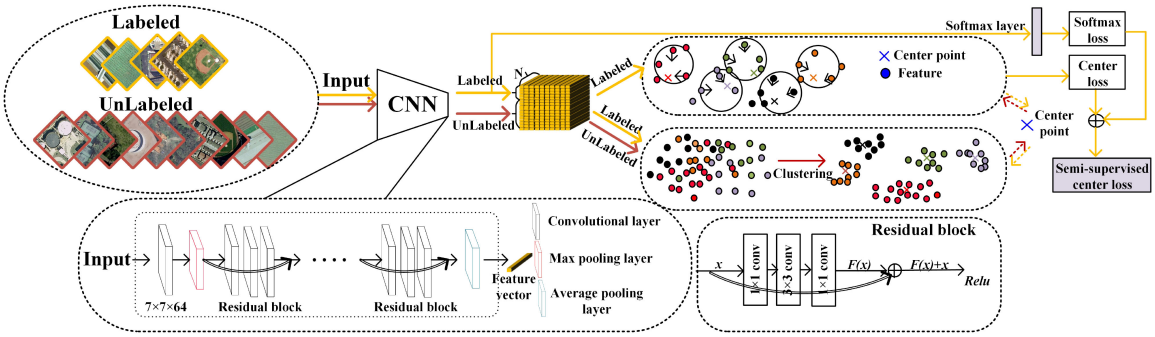


Fig. 1. Overall architecture of the proposed framework with SSCL. The yellow line represents the supervised branch, and the red line is the unsupervised branch. Details of the residual block are presented in the bottom right corner.

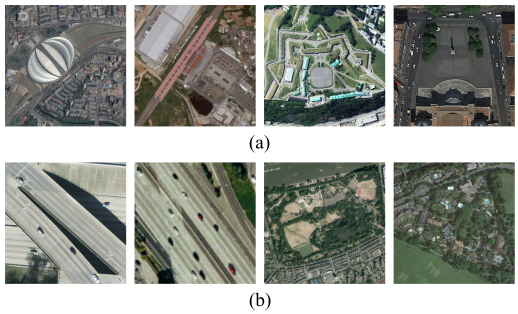


Fig. 2. Two major characteristics of high-resolution remote sensing images: (a) large intraclass variations, and (b) small interclass dissimilarity. From the top-left corner to the bottom-right corner, the corresponding categories of remote sensing images are railway station, square, overpass, freeway, park, and resort.

$\{f_i\}_{i=1}^{t_{la}}$ and $\{\widetilde{f}_j\}_{j=1}^{t_{un}}$ are extracted. In the training phase, we first generate original class centers $\{C_k\}_{k=1}^{C}$ based on the obtained features $\{f_i\}_{i=1}^{t_{la}}$ and center loss algorithm. The original class centers are regarded as the initialization centers for unsupervised clustering (Fig. 1, yellow line, Step 7 of Algorithm 1). Subsequently, clustering algorithm iterates p times based on $\{f_i\}_{i=1}^{t_{la}}$ and $\{\widetilde{f}_j\}_{j=1}^{t_{un}}$ to generate correctional class centers $\{\widetilde{C}_k\}_{k=1}^{C}$, which are fed back to supervised branch for updating the original class centers $\{C_k\}_{k=1}^{C}$ (Fig. 1, red line). Finally, we apply the updated class centers to calculate the center loss, and SSCL is optimized by standard stochastic gradient descent (SGD) algorithm under the joint supervision of softmax loss and center loss. It is worth noting that the SSCL framework is an end-to-end semisupervised framework.

C. Semisupervised Center Loss

Generally, scene classification task mainly focuses on high-resolution remote sensing images with scene information. Nevertheless, due to the high spatial resolution, remote sensing images present large intraclass variations and small interclass dissimilarity (see Fig. 2). In order to learn discriminative features from all available remote sensing images (including labeled and unlabeled) for reducing intraclass distance and increasing interclass distance, we improve the center loss. The center loss function is defined by

$$L_C = \frac{1}{2}\sum_{i=1}^{t_{la}} \|f_i - C_{y_i}\|_2^2 \tag{2}$$

where $C_{y_i} \in R^{N\times 1}$ denotes the y_ith class center of average pooling layer features. $f_i \in R^{N\times 1}$ represents the ith labeled sample features. Obviously, the center loss function [see (2)] effectively penalizes the intraclass distance. The update equation of C_{y_i} is computed as

$$C_k = C_k - \alpha \cdot \Delta C_k \tag{3}$$

$$\Delta C_k = \frac{\sum_{i=1}^{t_{la}} \delta(y_i = k)\cdot(C_k - f_i)}{1 + \sum_{i=1}^{t_{la}} \delta(y_i = k)}. \tag{4}$$

In (3), C_k represents the kth class center to be updated. α ($\alpha \le 1$) is a positive parameter to control the update rate of centers. ΔC_k denotes the gradient of C_k. More correctly, δ represents an indicator function. $\delta(condition) = 1$ if the *condition* $(y_i = k)$ is satisfied, and $\delta(condition) = 0$ if not. Therefore, the numerator of ΔC_k refers to the difference between the samples of the kth class and C_k. $\sum_{i=1}^{t_{la}} \delta(y_i = k)$ represents the number of samples belonging to the kth class. Due to the possibility of $\sum_{i=1}^{t_{la}} \delta(y_i = k)$ being 0, the center loss improved $\sum_{i=1}^{t_{la}} \delta(y_i = k)$ to $1 + \sum_{i=1}^{t_{la}} \delta(y_i = k)$.

Although the center loss function can specifically address the problem of large intraclass distance, it still relies on labeled samples. In other words, the classic center loss is a supervised algorithm. In the case of fewer labeled samples, the center loss algorithm may not perform well.

In order to make the center loss algorithm better applicable to high-resolution remote sensing image scene classification, in this article, we develop a novel semisupervised framework. The general design idea of the framework is to establish a cooperation mechanism between the supervised branch and the unsupervised branch, so that the two branches can complement each other's strengths. Specifically, in the supervised branch, we adopt the classical center loss algorithm. However, L_C is no longer dependent on C_k, but is based on the correctional $\widetilde{C}_k$ of unsupervised branch. In the unsupervised branch, we propose an improved clustering algorithm. Particularly, the cluster centers are no longer randomly initialized, but are based on the class centers of supervised branch learning by (3). Since the center loss algorithm of supervised branch utilizes the correctional class centers, it can better promote the ConvNets to learn discriminative features. Similarly, the strategy can also help the clustering algorithm of unsupervised branches to avoid the impact of random initialization centers on the clustering results. Through the cooperation of the two branches, we can obtain

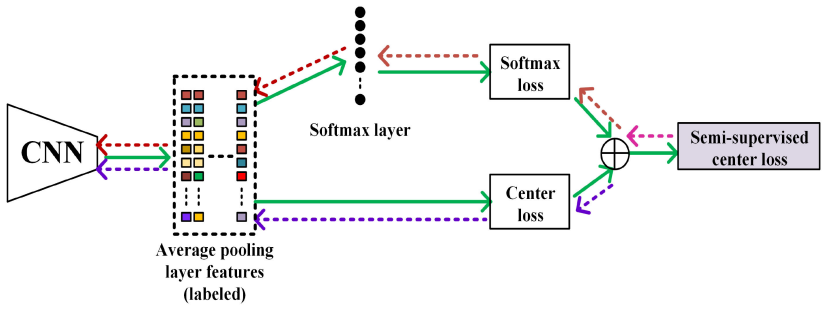


Fig. 3. Schematic diagram of forward and back propagation for SSCL framework. Among them, the solid line represents forward propagation, and the dotted line represents back propagation. Only labeled samples participate in the optimization process.

Algorithm 1: Semisupervised Center Loss.

Input:
labeled feature set $\{f_i, y_i\}_{i=1}^{t_{la}}$, unlabeled feature set $\{\widetilde{f}_j\}_{j=1}^{t_{un}}$, learning rate of centers α.
Output:
center loss value L_C.
Initialization:
Initialize original class centers $\{C_k\}_{k=1}^{C}$ with zero matrix $\widetilde{O} \in R^{N\times C}$.
Initialize $\{f\widetilde{f}_l\}_{l=1}^{t_{la}+t_{un}}$ by $\{f_i\}_{i=1}^{t_{la}} \bigcup \{\widetilde{f}_j\}_{j=1}^{t_{un}}$.
Compute original class centers $\{C_k\}_{k=1}^{C}$:
For $k = 1$ to C
$C_k = C_k - \alpha \cdot \frac{\sum_{i=1}^{t_{la}} \delta(y_i=k)\cdot(C_k - f_i)}{1+\sum_{i=1}^{t_{la}} \delta(y_i=k)}$;
Correct class centers:
Repeat
{
for k =1 to C
$D_k = \max_{y_i=C_k} \|f_i - C_{y_i}\|_2$;
for $j = 1$ to t_{un}
$d_j, label(\widetilde{f}_j) = \min_k \|\widetilde{f}_j - C_k\|_2$;
if $d_j > D_{label(\widetilde{f}_j)}$ then
$label(\widetilde{f}_j) = -1$;
for $i = 1$ to t_{la}
$label(f_i) = y_i$;
for $k = 1$ to C
$\widetilde{C}_k = \text{mean}_{label(f\widetilde{f}_l)=C_k} f\widetilde{f}_l$;
for $k = 1$ to C
$C_k \leftarrow \widetilde{C}_k$;
}
Compute center loss L_C**:**
$L_C = \frac{1}{2}\sum_{i=1}^{t_{la}} \| f_i - \widetilde{C}_{y_i} \|_2^2$.
return L_C.

more accurate class centers, then compute an effective intraclass loss L_C. The overall algorithm is summarized in Algorithm 1.

D. Optimization

To maximize the interclass distance, softmax loss and center loss with correctional centers jointly supervise the proposed scene classification framework. The optimization schematic of SSCL is shown in Fig. 3. The joint loss function is expressed as follows:

$$L_{SC} = L_S + \beta \cdot L_C \tag{5}$$

$$L_S = -\sum_{i=1}^{t_{la}} \log \frac{e^{W_{y_i}^T f_i + b_{y_i}}}{\sum_{k=1}^{C} e^{W_k^T f_i + b_k}} \tag{6}$$

$$L_C = \frac{1}{2}\sum_{i=1}^{t_{la}} \|f_i - \widetilde{C}_{y_i}\|_2^2 \tag{7}$$

where L_{SC} is the total loss function. L_S and L_C represent the softmax loss and the center loss function, respectively. $W_k^T \in R^{1\times N}$ denotes the kth row of the weight $W^T \in R^{C\times N}$ in the softmax layer and $b_k \in R^{1\times 1}$ represents the kth term of the bias $b \in R^{C\times 1}$. β is weight coefficient for balancing the two cost functions. $\widetilde{C}_{y_i} \in R^{N\times 1}$ is the y_ith correctional class center. Obviously, we can conclude from Fig. 3 and (5) that the convolutional layer parameters are based on the joint supervision of softmax loss and center loss, but the softmax layer parameters are only based on the supervision of softmax loss.

In the training phase, we select the SGD optimization algorithm, and the optimization process is summarized in Algorithm 2. Through end-to-end training, all parameters of the proposed SSCL framework will be determined.

In the testing phase, we feed the test set images into the trained SSCL model batch by batch. The SSCL model will assign a predicted label to each image. Comparing the true labels of the test set with the predicted labels given by SSCL, we can obtain the classification accuracy of the test set.

III. Experiments

In order to evaluate the performance of the proposed algorithm for high-resolution remote sensing image scene classification, SSCL was compared with six different scene classification methods. For all methods, we conducted experiments on public UCM, WHU-RS19 dataset, and AID dataset.

A. Datasets

1) UCM Dataset: The original images of UCM dataset are manually extracted from the United States Geological Survey National Map Urban Area Imagery collection for various urban areas around the country, such as Birmingham, Boston, Buffalo,

Fig. 4. Some sample images of the UCM dataset. From the top-left corner to the bottom-right corner, the corresponding categories of images are agricultural, airplane, baseball diamond, beach, building, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile homepark, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis court.

Algorithm 2: Optimization.

Input:

labeled sample set $\{x_i, y_i\}_{i=1}^{t_{la}}$, unlabeled sample set $\{\widetilde{x}_j\}_{j=1}^{t_{un}}$, parameters ϕ_S in softmax layer, parameters ϕ_C in convolutional layers, class centers $\{C_k\}_{k=1}^{C}$, learning rate *lr*, weight coefficient β and the number of iteration γ.

Output:

updated parameters ϕ_S, ϕ_C.

while not converge **do**
  $\gamma = \gamma + 1$.
  Compute the total loss value by $L_{SC}^{\gamma} = {L_S}^{\gamma} + \beta \cdot {L_C}^{\gamma}$.
  Compute the backpropagation error $\frac{\partial L_{SC}^{\gamma}}{\partial x_i^{\gamma}}$ of labeled sample x_i.
  Update class centers $\{C_k\}_{k=1}^{C}$:
  For $k = 1$ to C
  $C_k^{\gamma+1} \leftarrow \widetilde{C}_k^{\gamma}$;
  Update the parameters ϕ_S by $\phi_S^{\gamma+1} = \phi_S^{\gamma} - lr \cdot \frac{\partial L_S^{\gamma}}{\partial \phi_S^{\gamma}}$.
  Update the parameters ϕ_C by $\phi_C^{\gamma+1} = \phi_C^{\gamma} - lr \cdot \sum_{i=1}^{t_{la}} \left(\frac{\partial L_S^{\gamma}}{\partial x_i^{\gamma}} \cdot \frac{\partial x_i^{\gamma}}{\partial \phi_C^{\gamma}} + \beta \cdot \frac{\partial L_C^{\gamma}}{\partial x_i^{\gamma}} \cdot \frac{\partial x_i^{\gamma}}{\partial \phi_C^{\gamma}} \right)$.
end while
return ϕ_S, ϕ_C.

Columbus, and Dallas. The dataset contains 21 classes of land-use images with a pixel resolution of one foot. Each class consists of 100 images with a size of 256 × 256 pixels. Some examples of the UCM dataset are shown in Fig. 4. The UCM dataset holds a significant semantic overlap between several urban scenes such as building, dense residential, medium residential, sparse residential, and mobile home park, which makes the dataset extremely challenging for classification tasks.

2) WHU-RS19 Dataset: The WHU-RS19 dataset includes totally 1005 images divided into 19 scene classes, and per scene class has about 50 images. The size of each image is 600 × 600 pixels. All images are collected from Google Earth with spatial resolution up to 0.5 m and spectral bands of red, green, and blue. In comparison, the WHU-RS19 dataset is small in scale, and the total number of images is less than half of the UCM dataset. The problem of insufficient labeled data makes it difficult for supervised classification methods to obtain high classification accuracy. Some examples of the WHU-RS19 dataset are shown in Fig. 5.

3) AID Dataset: The AID dataset is also downloaded from Google Earth, which contains 10 000 scene images labeled into 30 aerial scene types (e.g., airport, school, and railway station). The image number of each class varies from 220 to 420, and the size of each image is 600 × 600 pixels. To increase intraclass variations, the scene images per class of the AID dataset are collected from different countries and regions (e.g., *China*, *the United States*, *England*) around the world at different time and seasons under different imaging conditions. To reduce the interclass dissimilarity, the AID dataset increases the scene classes to 30, and different scene classes share similar objects and spatial distributions. The higher intraclass variations and smaller interclass dissimilarity make the AID dataset closer to the remote sensing scene images in practical applications. Some examples of the AID dataset are presented in Fig. 6.

B. Experimental Setup

To verify the effectiveness of the proposed SSCL algorithm under the condition of less labeled samples, the UCM dataset is split into 20% for validation, 20% for test, and 60% for training. Among them, all training samples are further divided into 10% labeled training data and 50% unlabeled training data. Considering the imbalance in the number of samples for per category in the WHU-RS19 dataset, we randomly select 50

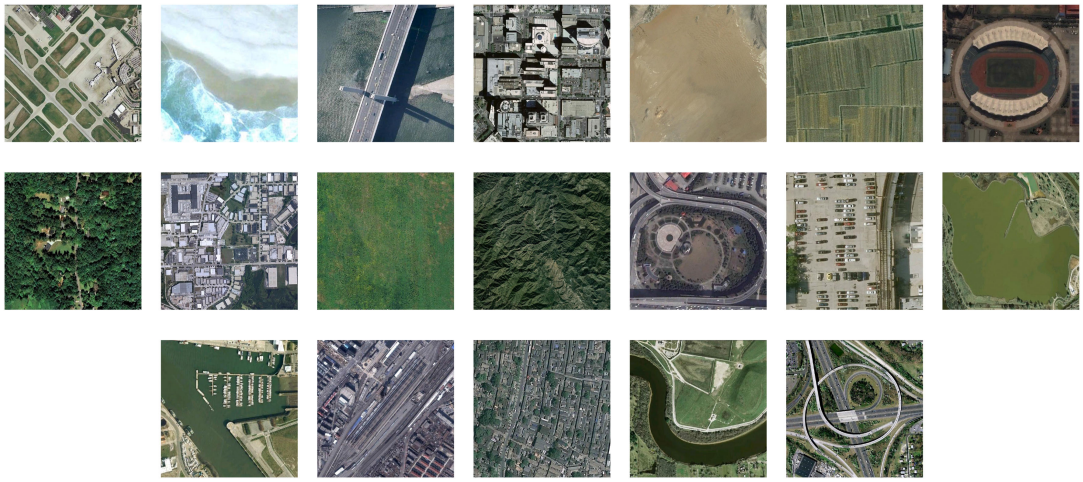


Fig. 5. Some sample images of the WHU-RS19 dataset. From the top-left corner to the bottom right corner, the corresponding categories of images are airport, beach, bridge, commercial, desert, farmland, football field, forest, industrial, meadow, mountain, park, parking, pond, port, railway station, residential, river, and viaduct.

Fig. 6. Some sample images of the AID dataset. From the top-left corner to the bottom-right corner, the corresponding categories of images are airport, bare land, baseball field, beach, bridge, center, church, commercial, dense residential, desert, farmland, forest, industrial, meadow, medium residential, mountain, park, parking lot, playground, pond, port, railway station, resort, river, school, sparse residential, square, stadium, storage tanks, and viaduct.

images from each category. The new dataset contains a total of 950 images. For the WHU-RS19 dataset, we set the same ratio as the UCM dataset. Similarly, for the AID dataset, we set the ratio of the labeled training data to 10%. The ratio of validation and test is 20%, respectively, and the rest of the dataset is used as the unlabeled training set. In the actual training, all images are resized to 256×256 pixels with bicubic interpolation.

All experiments were performed with the tensorflow platform on the Ubuntu 16.04 operation system with eight Intel Xeon Silver at 2.10-GHz CPU. In addition, we train our model using a GPU of NVIDIA GeForce RTX 2080Ti for acceleration.[1]

1) Evaluation Parameters: In our experiments, the overall accuracy (OA), standard deviation (Std), producer's

[1]The codes are available at https://github.com/HEBUT-ZM/SSCL

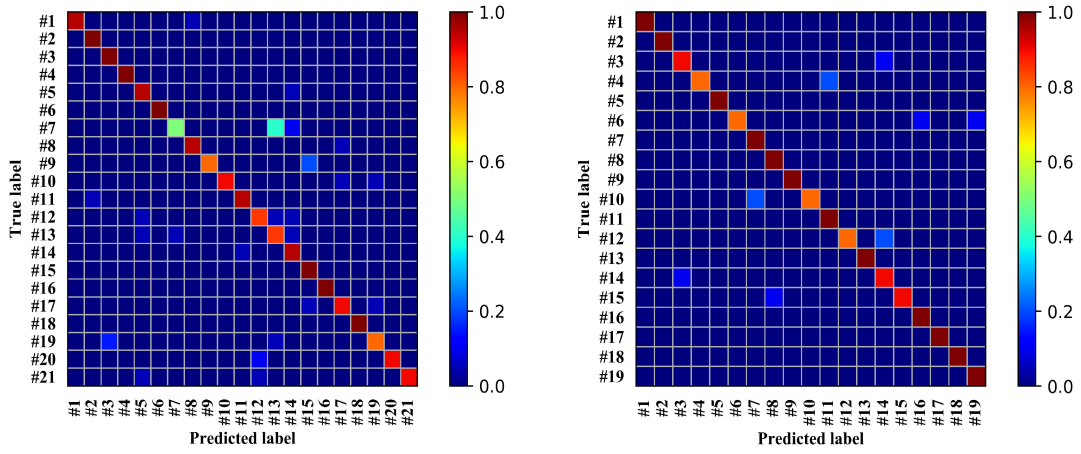


Fig. 7. Confusion matrices of the proposed SSCL algorithm. (a) UCM dataset with OA = 91.19%. (b) WHU-RS19 dataset with OA = 94.21%.

TABLE I
HYPERPARAMETERS OF THE PROPOSED SSCL FRAMEWORK

Hyperparameters	Value
lr	0.001
λ	0.00004
β	0.001
α	0.01
D_{MA}	0.9997

TABLE II
OVERALL CLASSIFICATION ACCURACIES OF DIFFERENT SCENE CLASSIFICATION METHODS

Methods	OA (%)	
	UCM (10%)	WHU-RS19 (10%)
VLAD [41]	58.23 ± 1.67	43.81 ± 2.87
IFK [41]	61.37 ± 1.25	45.12 ± 4.10
DCA [42]	80.97 ± 1.28	83.81 ± 2.06
GoogLeNet [28]	83.17 ± 0.99	78.98 ± 3.20
ResNet [29]	86.90 ± 0.47	91.58 ± 0.63
ResNet-Center Loss	87.62 ± 0.19	92.63 ± 0.52
SSCL	**91.19 ± 0.24**	**94.21 ± 0.53**

The best results are in bold.

accuracy (PA), and average accuracy (AA) are used as evaluation parameters. OA is defined as the number of correctly classified samples divided by the total number of samples in the test set. PA is the number of correctly classified samples for each class divided by the total number of samples for the corresponding class. AA is the average of the PA for all classes. To test the robustness of the SSCL framework, we repeat experiments ten times for each dataset, and Std is reported to indicate the experimental results.

2) Parameters Setting: In the proposed SSCL framework, five hyperparameters require to validate, which include the learning rate lr, the weight coefficient β for balancing the two cost functions, the learning rate of centers α, the weight decay rate λ, and moving average decay rate D_{MA}. The λ and D_{MA} are aimed at the overfitting problem. All hyperparameters are validated on the validation set. The hyperparameters mentioned above are determined as Table I.

3) Experimental Design: In this article, we conduct comparative experiments on the UCM, WHU-RS19, and AID datasets. Specifically, we compare our method with other scene classification methods, such as vector of locally aggregated descriptors (VLAD) [40] and discriminant correlation analysis (DCA) [41]. In particular, the classic center loss algorithm (ResNet-Center Loss, R-CL) is taken for comparison, which also adopts the ResNet as the backbone network. To illustrate the effectiveness of the proposed algorithm with fewer labeled samples, the classification effects of SSCL based on different data proportions are analyzed. In addition, we visualize the features generated by ResNet, ResNet-Center Loss, and SSCL separately to indicate that the SSCL algorithm can improve the discrimination of features for the ConvNets.

C. Experimental Results

1) Experimental Results of the UCM and WHU-RS19 Dataset: Table II has presented the overall classification accuracies of different scene classification methods on the UCM and WHU-RS19 datasets. The proportion of labeled data for each dataset is 10%. According to the experimental results of two datasets, the ConvNets-based classification algorithm DCA [41], GoogLeNet [28], ResNet [29], ResNet-Center Loss, and SSCL far outperform the VLAD and IFK approaches based on artificial designed features. The results have shown that it is difficult to capture the characteristics of complex remote sensing scenes using only artificial designed features. For the UCM and WHU-RS19 datasets, the proposed SSCL framework obtains the highest classification accuracies with only 10% labeled data. In addition, compared with the baseline, ResNet, our method improves the classification accuracies of the UCM and WHU-RS19 datasets by 4.29% and 2.63%, respectively. To better evaluate our algorithm, Fig. 7 shows confusion matrices of the two datasets. The confusion matrix is the classification result of one of the experiments.

TABLE III
OVERALL CLASSIFICATION ACCURACIES OF DIFFERENT SCENE CLASSIFICATION METHODS ON THE AID DATASET (%)

Category	VLAD	IFK	DCA	GoogLeNet	ResNet	ResNet-Center Loss	SSCL
Airport	41.08	45.43	69.35	71.60	86.11	93.06	**95.83**
Bare land	60.65	65.84	82.47	81.18	**91.93**	90.32	90.32
Baseball field	35.86	42.32	75.35	78.64	93.18	88.64	**97.92**
Beach	55.50	64.83	91.28	88.28	96.25	96.25	**98.75**
Bridge	33.83	50.77	77.25	80.12	93.05	97.22	**98.61**
Center	23.38	34.74	50.60	52.86	**75.00**	69.23	73.07
Church	48.80	59.58	65.51	67.64	**87.50**	87.50	85.42
Commercial	64.29	63.46	63.52	58.00	81.43	90.00	**91.43**
Dense residential	67.29	69.59	83.09	84.15	85.37	82.93	**89.02**
Desert	50.48	63.19	87.07	86.44	91.67	93.33	**95.00**
Farmland	36.37	52.58	88.41	89.58	95.95	94.59	**98.65**
Forest	79.91	79.69	87.69	87.29	98.00	96.00	**98.00**
Industrial	49.17	49.03	69.23	69.23	**94.87**	88.46	82.05
Meadow	65.04	70.87	87.38	89.29	**98.25**	98.21	98.21
Medium residential	60.31	55.17	71.11	74.67	81.03	87.38	**89.65**
Mountain	83.66	80.62	96.31	94.54	97.05	**99.26**	98.53
Park	59.11	60.41	64.73	66.41	71.43	80.00	**81.43**
Parking lot	91.48	90.60	93.99	92.79	98.07	99.35	**99.35**
Playground	38.26	44.23	83.21	83.24	87.84	**94.59**	93.24
Pond	49.39	54.10	76.40	74.02	94.05	**98.80**	95.24
Port	70.38	77.87	80.06	79.50	94.74	97.37	**98.68**
Railway station	48.29	51.92	65.43	61.32	73.08	61.54	**82.69**
Resort	26.93	28.20	46.55	51.88	63.79	70.69	**74.14**
River	45.61	52.60	82.44	81.63	86.58	89.02	**95.12**
School	31.30	33.48	46.44	47.56	**70.00**	46.67	51.67
Sparse residential	72.04	71.44	84.52	89.26	**95.00**	95.00	93.33
Square	37.37	42.36	57.58	51.85	66.67	74.24	**88.33**
Stadium	52.57	57.16	75.90	77.09	79.31	**81.03**	75.86
Storage tanks	51.79	51.02	82.75	82.50	93.06	**97.22**	93.05
Viaduct	76.72	81.88	93.81	94.05	96.43	**99.40**	97.61
AA	53.56	58.17	75.98	76.22	86.28	88.69	**90.01**
OA	54.22	58.83	76.86	76.96	87.85	88.85	**90.35**
Std	0.66	0.64	0.42	0.66	0.25	0.36	0.55

The best results are in bold.

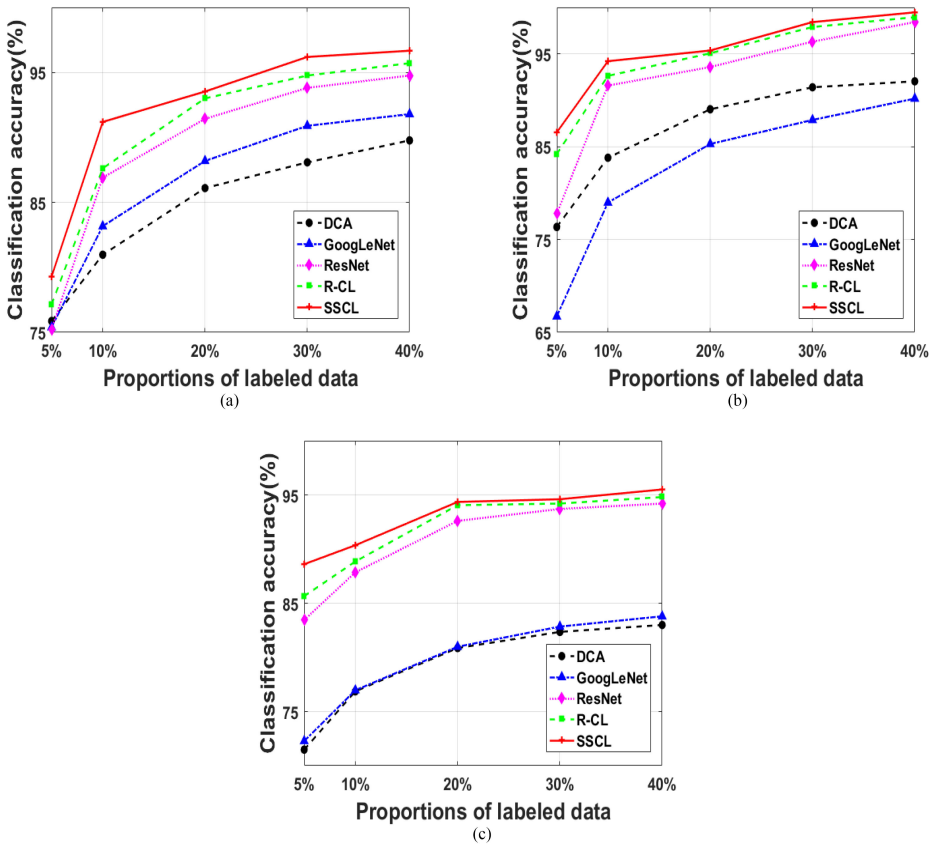


Fig. 8. Overall classification accuracies of different scene classification methods with different proportions of labeled data (%). (a) UCM dataset. (b) WHU-RS19 dataset. (c) AID dataset.

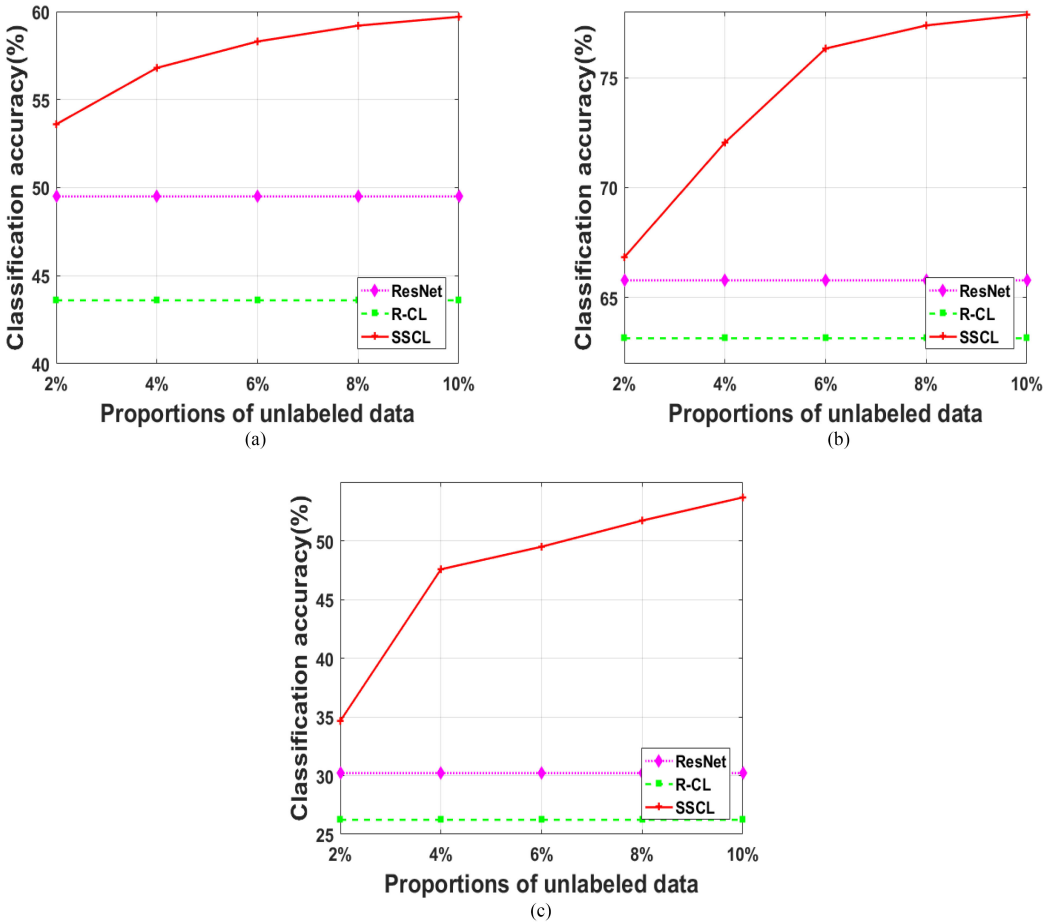


Fig. 9. Overall classification accuracies of different scene classification methods with different proportions of unlabeled data (%). (a) UCM dataset. (b) WHU-RS19 dataset. (c) AID dataset.

2) Experimental Results of the AID Dataset: In Table III, we report the classification results of the AID dataset. Compared to the UCM, WHU-RS19 dataset, the AID dataset contains more categories (30 categories), so the PA is listed to compare the classification performances of different algorithms. In terms of PA, the SSCL algorithm improves the classification accuracies of 17 categories. Moreover, the classification results of 20 categories exceed the ResNet, and the results of 18 categories are better than the ResNet-Center Loss. Similarly, SSCL also obtains the best OA (90.35%) with 10% labeled data. Our method greatly improves the classification accuracies of the square and railway station with higher intraclass variations, as well as resort and park with smaller interclass dissimilarity. The experimental results of the AID dataset further show that the SSCL algorithm can utilize the scene information contained in the unlabeled data to correct the class centers, thereby effectively penalizing the intraclass distance and increasing the interclass distance.

D. Analysis and Discussions

1) Effect of the Labeled Sample Ratio on Classification Accuracy: Fig. 8(a), (b) and (c) shows the classification results of different methods on the UCM, WHU-RS19, and AID datasets with the proportion of labeled data as 5%, 10%, 20%, 30%, and 40%. To ensure a fair comparison, we only compare with the methods that are close to our classification results. On the three datasets, the classification accuracies of GoogLeNet and DCA are lower than that of other algorithms. In particular, classification accuracies of the AID dataset are lower than 85%, in the case of 40% labeled data. On the one hand, the AID dataset is more difficult to classify scenes because of its multisource and multiresolution characteristics. On the other hand, the performances of backbone networks for GoogLeNet and DCA are weaker than the ResNet. For the WHU-RS19 dataset, the classification accuracies of all proportions for GoogLeNet are lower than the DCA algorithm. This is because the WHU-RS19 dataset contains only 950 images, which cannot meet the deeper GoogLeNet. It further shows the dependence of ConvNets on training data. Observing the SSCL results of three datasets in Fig. 8, it is obvious that our method is more effective with less annotated data. With the increase of labeled data, the classification accuracies of all methods are gradually improved, and the advantages of the proposed SSCL over other methods gradually decrease. Therefore, we can draw a conclusion that when labeled data are sufficient for classification tasks, the introduction of unlabeled data for classification models will no longer be the optimal choice.

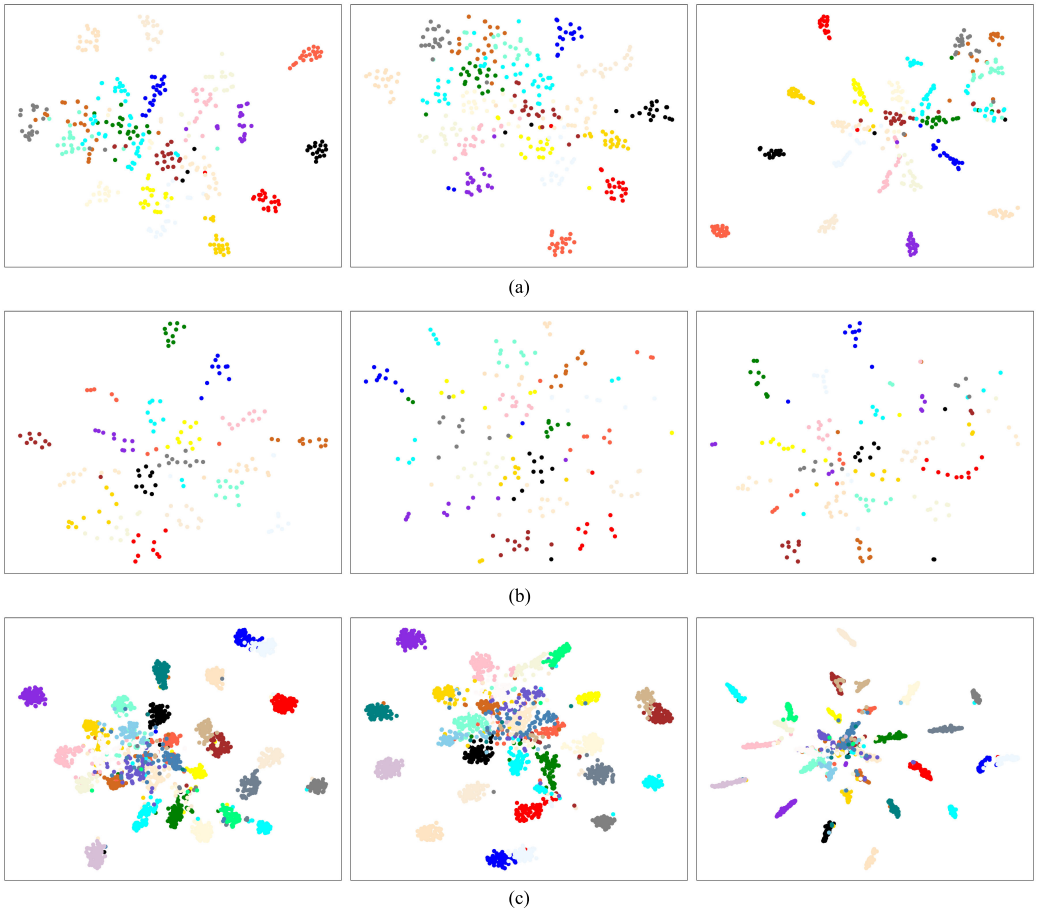


Fig. 10. 2-D scatter plots of the average pooling layer features on three datasets via t-sne. The points with different colors denote features of different categories. (a) UCM. (b) WHU-RS19. (c) AID.

2) Effect of the Unlabeled Sample Ratio on Classification Accuracy: To further investigate the impact of the number of unlabeled samples on the proposed algorithm, we only select one labeled sample from each category to construct the labeled training set. For the UCM dataset, the number of labeled training samples is 21. For the WHU-RS19 dataset, the number of labeled training samples is 19. Similarly, the number of labeled training samples in the AID dataset is 30. Besides, we randomly choose 2%, 4%, 6%, 8%, and 10% images from per category as the unlabeled samples. In Fig. 9, we report the experimental results of three datasets with different proportions of unlabeled data. In order to highlight the advantages of our method over the baseline and supervised center loss based method, we only present the ResNet, ResNet-Center Loss, and SSCL algorithm in Fig. 9. Among them, the ResNet and ResNet-Center Loss only rely on the labeled data. As the number of unlabeled data increases, their classification accuracies remain unchanged. Since the labeled training data of each dataset contain only one sample per category, the ResNet-Center Loss algorithm determines class centers based on one sample per class. Inaccurate class centers make classification accuracies of the ResNet-Center Loss algorithm on three datasets lower than the baseline. In general, the classification accuracies of SSCL on all datasets are progressively improved with the increase of unlabeled data. The results indicate that in the case of less labeled data, unlabeled data are of importance for the classification model. Unfortunately, we discover that the growth rate of SSCL algorithm gradually decreases. It also shows that the unlabeled data can only improve the classification accuracy of ConvNets to a certain extent, and may not ensure its infinite increase.

3) Visualization Results of Different Methods: To show the experimental results of SSCL algorithm on the UCM, WHU-RS19, and AID datasets more clearly, we extract the average pooling layer features of three test sets from ResNet, ResNet-Center Loss, and SSCL. The t-sne algorithm is used to reduce the extracted high-dimensional features to 2-D features. Fig. 10 displays the scatter plots of the 2-D features. In Fig. 10(a), (b), and (c), the left column is the visualization result of ResNet. The middle column is the visualization result of ResNet-Center Loss, and the right column is the visualization result of SSCL.

Comparing the feature distributions of three algorithms, we can discover that the deep features of the same category generated by the SSCL framework are more compact and the different categories are more separated. In the three test sets, the visualization effects of UCM and AID test sets are obvious. Since the WHU-RS19 test set contains a small number of samples, the clustering results of different algorithms are similar. Through

feature visualization analysis, we conclude that the discriminative power of deep features based on the SSCL algorithm can be significantly enhanced.

IV. Conclusion

This article proposes a semisupervised algorithm for remote sensing image scene classification called SSCL. In general, SSCL integrates a cooperative dual-branch structure into ConvNets to learn discriminative information from unlabeled data. To perform effective cooperation between supervised and unsupervised branches, the proposed dual-branch structure adopts class centers as guiding factors, and an improved clustering algorithm is developed for the unsupervised branch. Based on the problem of large intraclass distance and small interclass distance in high-resolution remote sensing scenes, we optimize the ConvNets under the joint supervision of SSCL and softmax loss. In summary, our algorithm has the following two contributions. 1) We construct an end-to-end SSCL framework for remote sensing scene classification. 2) We propose an improved clustering algorithm and design a dual-branch structure based on it, which can fuse discriminative information of all available remote sensing images to optimize our SSCL-based model.

To validate the effectiveness of the SSCL algorithm, we performed experiments on the UCM, WHU-RS19, and AID datasets. Experimental results have demonstrated the proposed method improves the classification performance of high-resolution remote sensing scenes, especially superior to supervised center loss based methods.

However, according to the experimental results of visualization analysis, our method still requires to be further optimized to increase the interclass distance. Therefore, in our future work, we will further consider how to improve the joint loss function to minimize confusion of different categories.

Acknowledgment

The authors would like to thank S. Newsam from the University of California at Merced, and G. Xia from Wuhan University, for providing the UCM, WHU-RS19, and AID datasets in this article, respectively.

References

[1] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.

[2] C. Wang, X. Bai, S. Wang, J. Zhou, and P. Ren, "Multiscale visual attention networks for object detection in VHR remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 310–314, Feb. 2019.

[3] B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang, and X. Zhu, "Coinnet: Copy initialization network for multispectral imagery semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 816–820, May 2019.

[4] B. Pan, X. Xu, Z. Shi, N. Zhang, H. Luo, and X. Lan, "DSSNET: A simple dilated semantic segmentation network for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, 2020, doi: 10.1109/LGRS.2019.2960528.

[5] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[6] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.

[7] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[9] S. Newsam, L. Wang, S. Bhagavathy, and B. S. Manjunath, "Using texture to analyze and manage large collections of remote sensed image and video data," *Appl. Opt.*, vol. 43, no. 2, pp. 210–217, 2004.

[10] Y. Yang and S. Newsam, "Comparing sift descriptors and gabor texture features for classification of remote sensed imagery," in *Proc. 15th IEEE Int. Conf. Image Process.*, 2008, pp. 1852–1855.

[11] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître, "Structural high-resolution satellite image indexing," in *Proc. ISPRS TC VII Symp. Years ISPRS*, 2010, vol. 38, pp. 298–303.

[12] B. Luo, S. Jiang, and L. Zhang, "Indexing of remote sensing images with different resolutions by multiple features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 1899–1912, Aug. 2013.

[13] A. Avramović and V. Risojević, "Block-based semantic classification of high-resolution multispectral aerial images," *Signal, Image Video Process.*, vol. 10, no. 1, pp. 75–84, 2016.

[14] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.

[15] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification via PLSA," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 517–530.

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[17] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent Dirichlet allocation," *IEEE Geosci.Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, Jan. 2010.

[18] Q. Zhu, Y. Zhong, L. Zhang, and D. Li, "Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6180–6195, Oct. 2018.

[19] Y. Zhong, M. Cui, Q. Zhu, and L. Zhang, "Scene classification based on multifeature probabilistic latent semantic analysis for high spatial resolution remote sensing images," *J. Appl. Remote Sens.*, vol. 9, no. 1, 2015, Art. no. 095064.

[20] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

[21] K. Nogueira, O. A. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.

[22] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[23] Y. Liu, Y. Zhong, F. Fei, Q. Zhu, and Q. Qin, "Scene classification based on a deep random-scale stretched convolutional neural network," *Remote Sens.*, vol. 10, no. 3, 2018, Art. no. 444.

[24] W. Zhang, P. Tang, and L. Zhao, "Remote sensing image scene classification using CNN-capsnet," *Remote Sens.*, vol. 11, no. 5, 2019, Art. no. 494.

[25] S. Song, H. Yu, Z. Miao, Q. Zhang, Y. Lin, and S. Wang, "Domain adaptation for convolutional neural networks-based remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1324–1328, Aug. 2019.

[26] W. Teng, N. Wang, H. Shi, Y. Liu, and J. Wang, "Classifier-constrained deep adversarial domain adaptation for cross-domain semisupervised classification in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, 2019, doi: 10.1109/LGRS.2019.2931305.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[28] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 1–9.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2016, pp. 770–778.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 4700–4708.

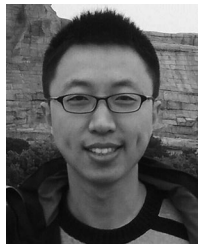
[31] O. A. Penatti, K. Nogueira, and J. A. Dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshops*, 2015, pp. 44–51.
[32] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.
[33] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," 2015. [Online]. Available: https://arxiv.org/abs/1508.00092
[34] B. Zhao, B. Huang, and Y. Zhong, "Transfer learning with fully pretrained deep convolution networks for land-use classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1436–1440, Sep. 2017.
[35] P. Li, P. Ren, X. Zhang, Q. Wang, X. Zhu, and L. Wang, "Region-wise deep feature representation for remote sensing images," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 871.
[36] Y. Yuan, J. Fang, X. Lu, and Y. Feng, "Remote sensing image scene classification using rearranged local features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1779–1792, Mar. 2019.
[37] Y. Liu, Y. Liu, and L. Ding, "Scene classification based on two-stage deep feature fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 183–186, Feb. 2018.
[38] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 499–515.
[39] J. Li, D. Lin, Y. Wang, G. Xu, and C. Ding, "Deep discriminative representation learning with attention map for scene classification," 2019. [Online]. Available: https://arxiv.org/abs/1902.07967
[40] G.-S. Xia *et al.*, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
[41] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for VHR remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4775–4784, Aug. 2017.

Jun Zhang received the B.S. and Ph.D. degrees from the Hebei University of Technology (HEBUT), Tianjin, China, in 1999 and 2011, respectively.

He is currently an Associate Professor with the School of Artificial Intelligence, HEBUT. His research interests include machine learning and intelligent computing.

Min Zhang received the B.S. degree in computer science and technology from Langfang Normal University, Hebei, China, in 2017. She is currently working toward the master's degree with the School of Artificial Intelligence, Hebei University of Technology, Tianjin, China.

Her research interests include machine learning and intelligent computing.



Bin Pan received the B.S. and Ph.D. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2013 and 2019, respectively.

Since 2019, he has been an Associate Professor with the School of Statistics and Data Science, Nankai University, Tianjin, China. His research interests include machine learning, remote sensing image processing, and multiobjective optimization.

Zhenwei Shi (Member, IEEE) received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005.

From 2005 to 2007, he was a Postdoctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China. From 2013 to 2014, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing, China. He has authored or coauthored more than 100 scientific papers in related journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE Conference on Computer Vision and Pattern Recognition. His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi received the Best Reviewer Award for his service to IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2017. He has been an Associate Editor for the *Infrared Physics and Technology* since 2016.