Generating and Sifting Pseudolabeled Samples for Improving the Performance of Remote Sensing Image Scene Classification

Xiaoliang Qian, *Member, IEEE*, Xiaohao Chen, Weichao Yue, Xianglong Liu, Jungang Guo, Zhehui Li, Yinhua Li, and Wei Wang

***Abstract*—Deep learning-based remote sensing image scene classification methods are the current mainstream, and enough labeled samples are very important for their performance. Considering the fact that manual labeling of samples requires high labor and time cost, many methods have been proposed to automatically generate pseudosamples from real samples, however, existing methods cannot directly sift the pseudosamples from the perspective of model training. To address this problem, a generating and sifting pseudolabeled samples scheme is proposed in this article. First of all, the existing SinGAN is used to generate multiple groups of pseudosamples from the real samples. Afterward, the proposed quantitative sifting measure which can evaluate both the authenticity and diversity from the perspective of model training is employed to select the best pseudosamples from the multiple generated pseudosamples. Finally, the selected pseudosamples and real samples are used to pretrain and finetune the deep scene classification network (DSCN), respectively. Moreover, the focal loss that is originally proposed for object detection is adopted to replace the traditional cross entropy loss in this article. A designed quantitative evaluation shows that the value of proposed quantitative sifting measure is proportional to the overall accuracy, which validates the effectiveness of proposed quantitative sifting measure. The comprehensive quantitative comparisons on AID and NWPU-RESISC45 datasets in terms of overall accuracy and confusion matrices demonstrate that incorporating the pseudosamples selected by proposed sifting measure and the focal loss can improve the performance of DSCN.**

***Index Terms*—Deep learning, focal loss, quantitative sifting measure, remote sensing image, scene classification.**

I. Introduction

With the continuous development of remote sensing imaging technology, a variety of resolution (spatial resolution, spectral resolution, radiation resolution, and time resolution) and higher quality aerial or satellite remote sensing images could be obtained, so the higher requirements for the understanding of remote sensing images were put forward [1], [2]. The high-resolution remote sensing image scene classification is to distinguish the land use or coverage category of remote sensing image according to the image content [3], [4], which can provide the important cues for other remote sensing image processing tasks [5]–[10]. Furthermore, it has an important role in natural disaster monitoring, environmental detection, traffic supervision, weapon guidance and urban planning [11]–[15].

Early remote sensing scene classification methods are based on handcrafted features, such as CH [16], SIFT [17], GIST [18], etc. Afterward, the coding of handcrafted features is widely used to further improve the accuracy of scene classification, the milestone work is the bag of visual words (BoVW) model [19]. The main idea of BoVW is first extracting local manual features from the image, then clustering these features to obtain a "word bag," and finally using "word bag" to encode the image to get a histogram as the feature representation of the image. A large number of scene classification methods adopt BoVW model [20]–[24], or some improved models of BoVW model, such as spatial pyramid matching (SPM) [25], sparse coding SPM (SCSPM) [26] etc.

With the development of deep learning [27]–[31], the deep learning-based methods have become the mainstream of scene classification because of the more powerful representation of deep features [32]. The scene classification methods based on deep learning can be divided into the following three categories according to supervised mode: 1) fully supervised methods; 2) semisupervised methods; and 3) weakly supervised methods.

Most of the existing deep learning-based methods can be classified into the first category. The topic model-based methods are one of the effective methods. Zhu *et al.* [33] proposed a adaptive deep sparse semantic modeling (ADSSM) framework, in which the topic model and convolutional neural network (CNN) are combined and used to extract more discriminative features for remote sensing images. Other topic model-based methods include [34], [35] etc. Many methods adopt multilevel deep features to improve the accuracy of scene classification. Yuan *et al.* [36] used the last convolution layer and the last full connection layer as the local and global features, respectively,

Manuscript received May 31, 2020; revised July 24, 2020 and August 18, 2020; accepted August 23, 2020. Date of publication August 26, 2020; date of current version September 7, 2020. This work was supported in part by the Key Science and Technology Program of Henan Province under Grant 202102210347 and Grant 202102210143, and in part by the Key Scientific Research Project of Colleges and Universities in Henan Province under Grant 19A413014. *(Corresponding authors: Yinhua Li; Wei Wang.)*

Xiaoliang Qian, Xiaohao Chen, Weichao Yue, Xianglong Liu, Yinhua Li, and Wei Wang are with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China (e-mail: qxl_sunshine@163.com; cxhclc@163.com; yue_weichao@163.com; xianglongliu@zzuli.edu.cn; zzuli412@163.com; 2014036@zzuli.edu.cn).

Jungang Guo is with the China Science Quantum Cloud Technology Group, Zhengzhou 450000, China (e-mail: guo@csqctg.com).

Zhehui Li is with the Network Technology Center, Henan Provincial Institute of Scientific & Technical Information, Zhengzhou 450000, China (e-mail: qbslzh@163.com).

Digital Object Identifier 10.1109/JSTARS.2020.3019582

then the local features are rearranged according to the similarity between them and the cluster centers generated by using the global features, the final representations of remote sensing images are obtained by fusing the global and rearranged local features. Other multilevel deep features-based methods include [37]–[41] etc. In addition, Cheng *et al.* [42] combined the deep learning with metric learning, and proposed a new loss function to train the fused deep network. Chen *et al.* [43] automatically learned the CNN architecture using the labeled datasets. Zhang *et al.* [44] combined CNN with capsnet for scene classification.

Semisupervised methods which can utilize the unlabeled samples are attractive because of the less requirement of labeled samples [45]. Han *et al.* [46] proposed a semisupervised generation framework based on deep learning features, which can automatically augment the unlabeled samples through the iterative training. This method first used the labeled samples to train the pretrained CNN, second, the deep features extracted by trained CNN are used to train the SVM, third, the trained SVM was employed to predict the tag of unlabeled samples, then the automatically labeled samples were added into the original labeled samples, above steps are iteratively implemented. To improve the annotation accuracy, multiple SVMs were jointly used to determine the tags of samples that belonged to the easily confused categories. Soto *et al.* [47] jointly used the labeled and unlabeled samples to train the generative adversarial network (GAN), the trained discriminator was used for scene classification. In addition, some works [48]–[50] constructed the feature extraction model through the unsupervised feature learning, and used the labeled samples to train the classifier.

The combination of weak supervision method and deep learning is also widely used [51]–[53]. This method usually utilizes the labeled samples that are similar to the target samples to train the scene classification model. This kind of methods divide the datasets into source domain and target domain, the former is different from latter but similar, the latter can obtain the tags through various transfer learning techniques, and is further used to train the scene classification model. Related works include Song *et al*. [54], Othman *et al.* [55], Gong *et al.* [56], and Li *et al*. [57].

The supervised methods usually give the best performance among above three kinds of methods; however, a large number of labeled samples are required for training the deep model. Although the less labeled samples are required, the semisupervised methods can only use the unlabeled samples to refine the feature space constructed by the labeled samples, there is no substantial increase in discriminative information, consequently, the classification accuracy is limited. The generality of weakly supervised methods is restricted since finding the similar labeled samples for all kinds of the scene images is not easy, moreover, the accuracy is also limited because of the inherent difference between the source and target domain. In summary, having considerable number of high quality labeled samples is very important for deep learning-based scene classification methods.

Considering the fact that manual labeling of samples requires high labor and time cost, it is a reasonable choice to automatically augment the manually labeled samples, which can be divided into two steps. The first step is generating the pseudosamples from the real samples, and many effective methods have been proposed for this purpose, such as the GAN-based methods [45], [58], [59]. The second step is sifting the pseudosamples to obtain the final high-quality samples; however, some problems still existed in this step. For instance, Ma *et al.* [58] used the GAN to generate the pseudosamples and sifted the generated pseudosamples according to the true and false probability output by the discriminator, which can only ensure the authenticity of the generated samples, while the diversity cannot be guaranteed. Some popular quantitative measures, such as Fréchet perception distance (FID) [60] which taken into account both authenticity and diversity, are often used to sift the pseudosamples, however, these measures cannot directly evaluate the quality of samples from the perspective of model training. As a matter of fact, improving the training of deep model is the main motivation of augmenting the labeled samples in many occasions.

In order to address above problems, a novel sifting method which can directly evaluate the authenticity and diversity of generated pseudosamples from the perspective of model training is proposed to select high quality samples from multiple groups of generated samples in this article. The overall framework of proposed method is shown in Fig. 1. First, the SinGAN [59] is used to generate multiple groups of pseudosamples from a small number of real samples, then the final augmented samples are selected from multiple groups of pseudosamples by proposed quantitative sifting measure. Finally, the augmented and real samples are used to pretrain and finetune the DSCN, respectively. Moreover, the focal loss [61] is applied to the classification loss to further improve the accuracy of scene classification.

The SinGAN is the ICCV2019 best paper, which can generate the high quality pseudoimages while only the single image is required for training, therefore, it is adopted to generate the initial pseudosamples in this article. It is worth noting that the SinGAN is not the only choice, other methods can also play the same role, e.g., DeLiGAN [62], BigBiGAN [63], etc. As a matter of fact, any pseudosamples generated from the same real samples can be evaluated by proposed sifting measure no matter what the generating methods are adopted.

The main contributions of this article are as follows.

1) A novel quantitative sifting measure which can directly evaluate the authenticity and diversity of the generated samples from the perspective of model training is proposed to select high-quality samples from multiple groups of generated samples. It is worth noting that the proposed quantitative measure is used to evaluate any pseudo samples generated from same real samples.
2) To our best knowledge, the focal loss is first applied to the remote sensing image scene classification, which can effectively improve the scene classification accuracy.

II. Proposed Method

A. Generating Pseudosamples by SinGAN

The initial pseudosamples are generated by the SinGAN [59] which can use the single image to train the multiscale GANs and generate multilevel pseudosamples. In this section, the SinGAN is briefly introduced for the integrity of this article.

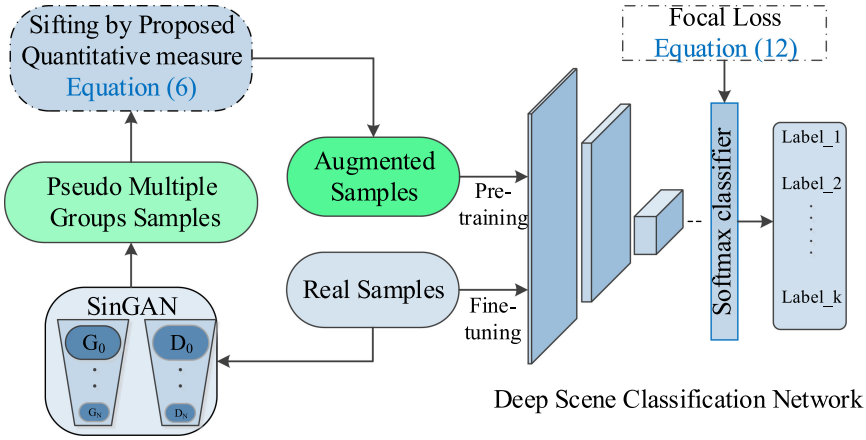

Fig. 1. Overall framework of proposed method.

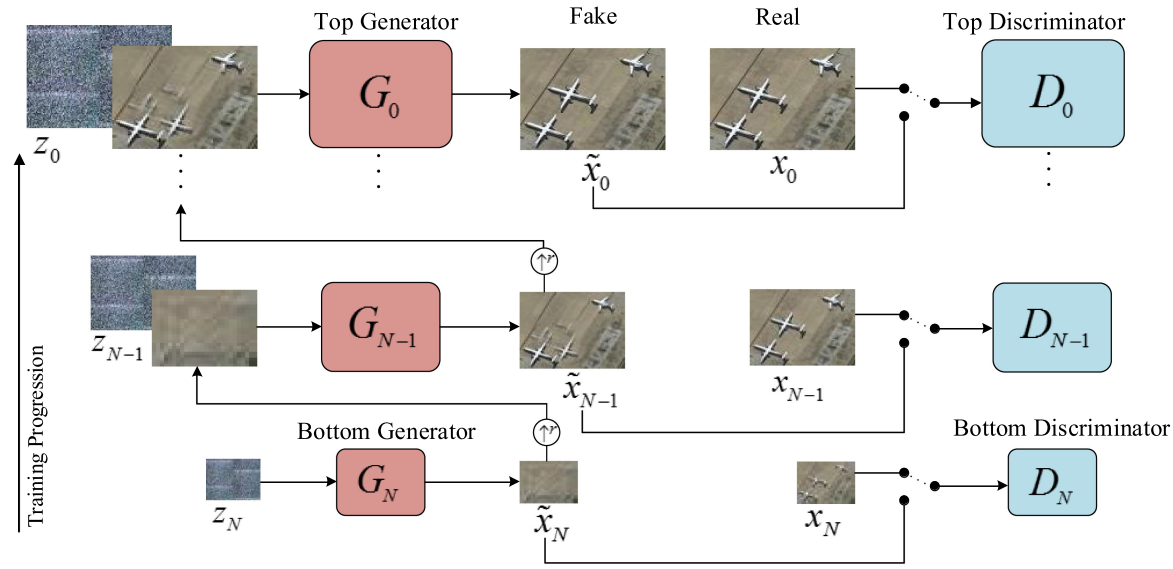

Fig. 2. Architecture of SinGAN.

1) Architecture of SinGAN

As shown in Fig 2, the SinGAN has a pyramid structure, which can be considered as the cascade of multiple GANs. The bottom and top GAN generate the coarsest and finest scale images respectively. The initial generation starts from the bottom GAN

$$\tilde{x}_N = G_N(z_N) \tag{1}$$

where N+1 is the number of GANs contained in SinGAN, G_N and G_0 denote the bottom and top generator, respectively, z_N denotes the input noise of G_N, $\tilde{x}_N$ denotes the pseudoimage generated by G_N.

The other generators except G_N have the same structure, as shown in Fig 3, they jointly use the noises and the output of last scale GANs to generate the pseudoimages

$$\begin{aligned}\tilde{x}_n &= G_n(z_n, (\tilde{x}_{n+1})\uparrow^r) \\ &= (\tilde{x}_{n+1})\uparrow^r + \psi_n(z_n + (\tilde{x}_{n+1})\uparrow^r),\ n \in [0, N-1]\end{aligned} \tag{2}$$

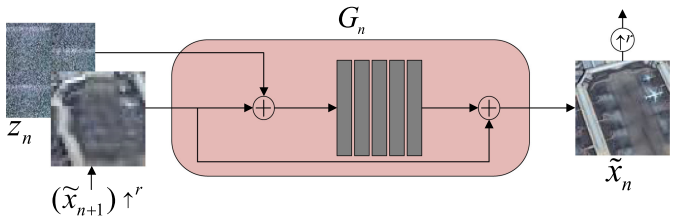

Fig. 3. Architecture of each generator contained in SinGAN except the bottom generator.

where G_n denotes the nth generator contained in SinGAN, z_n is its input noise, $\tilde{x}_n$ and $\tilde{x}_{n+1}$ denote the output of nth and (n+1)th generator, respectively, $(\tilde{x}_{n+1})\uparrow^r$ denotes the upsampling version of $\tilde{x}_{n+1}$, r is the upsampling ratio, ψ_n denotes a bank of convolution operations.

As shown in Fig 2, the discriminators $\{D_0 \cdots D_N\}$ coupled with the generators $\{G_0 \cdots G_N\}$ are used to the distinguish x_n

from $\tilde{x}_n$ at the patch level [64], where x_n is the *n*th downsampling version of real image *x*.

2) Training of SinGAN

The SinGAN is trained by single image, which is an important characteristic compared with other GANs. All of the GANs contained in SinGAN are trained one by one, from bottom GAN to top GAN. The parameters of each GAN will not change again once it has been trained.

The WGAN-GP [65] is used to define the loss functions of D_n

$$LD_n = D(x_n) - D(\tilde{x}_n) + \mu \underset{\hat{x}\sim\chi}{E} \left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1\right)^2 \quad (3)$$

where $\chi$is the joint sampling space of x_n and $\tilde{x}_n$, the last term is the gradient penalty team, and $\mu$is its weight coefficient.

The loss function of G_n is defined as follows:

$$LG_n = D(\tilde{x}_n) + LR_n \quad (4)$$

where LR_n denotes the reconstruction loss which is used to ensure that the real image *x* can be generated by a set of fixed noises $\{z_N^{rec}, z_{N-1}^{rec} \cdots z_0^{rec}\} = \{z^*, 0 \cdots 0\}$, the formulation of LR_n is as follows:

$$\begin{cases} LG_n = \left\|G_n(0, (\tilde{x}_{n+1}^{rec})\uparrow^r - x_n)\right\|^2, n < N \\ LG_n = \|G_n(z^*) - x_n\|^2, n = N \end{cases} \quad (5)$$

where $\tilde{x}_{n+1}^{rec}$ denotes the pseudoimages generated by the (*n*+1)th generator using above fixed noises.

3) Generating Pseudosamples

The generation of pseudosamples can started from any one of trained generators in the testing stage. When generating from the G_n, the input noise of G_N G_{n+1} is $\{z_N^{rec}, z_{N-1}^{rec}, \ldots, z_{n+1}^{rec}\}$, while the input noise of G_n G_0 is randomly determined.

Given the a group of real labeled samples RS, the *N*+1 groups of pseudo samples $\{PS_N, \ldots, PS_n, \ldots, PS_0\}$ are generated in this article, where PS_n denotes the pseudosamples generated by using the G_n as the initial generator.

B. Sifting Pseudosamples

This section introduce a novel quantitative sifting measure to select the best pseudosamples from $\{PS_N, \ldots, PS_n, \ldots, PS_0\}$, which is the most important contribution of this article.

A novel quantitative measure $M \in [0, 1]$ is proposed to evaluate the quality of each group of pseudosamples

$$M_n = \alpha NFID_n + \beta TR_n, n \in [0, N] \quad (6)$$

where M_n denotes the quantified score of PS_n, $NFID_n \in [0, 1]$ and $TR_n \in [0, 1]$ denote the normalized FID score and training evaluation score of PS_n, respectively, α and β separately denote the weight of $NFID_n$ and TR_n with the constraint of $\alpha + \beta = 1$. The default value of α and β is 0.5, which indicates that $NFID_n$ and TR_n give the equal contributions to M_n.

The $NFID_n$ is the normalized version of FID_n

$$NFID_n = \frac{\underset{i\in[0,N]}{Min}(FID_i)}{FID_n} \quad (7)$$

where FID_n denotes the FID [60] score of PS_n, $Min(\cdot)$ denotes the minimum operation since the quality of pseudosamples is inversely proportional to FID value.

The FID is a good quantitative measure to evaluate the quality of the generated image, which can directly evaluate the authenticity and diversity of the generated image. The formulation of FID_n is as follows:

$$FID_n = \|F_{RS} - F_{PS_n}\|_2^2 + Tr\left(CF_{RS} + CF_{PS_n} - 2(CF_{RS}CF_{PS_n})^{1/2}\right) \quad (8)$$

where F_{RS} and F_{PS_n} denote the mean value of the feature vectors of RS and PS_n, respectively, CF_{RS} and CF_{PS_n} denote covariance matrix of the feature vectors of RS and PS_n, respectively, $Tr(\cdot)$ denotes the trace of a matrix. The feature vectors are extracted by the Inception V3 model which has been trained on ImageNet.

The FID is a popular quantitative measure for evaluating the quality of generated samples; however, it cannot directly evaluate the generated samples from the perspective of improving the training quality which is the core motivation of generating pseudo samples. Consequently, the TR_n is proposed to be jointly used with FID for sifting pseudosamples. Inspired by [66], the TR_n includes two parts

$$TR_n = \lambda NSIM_n + \eta NDIV_n$$
$$NSIM_n = \frac{SIM_n}{\underset{i\in[0,N]}{\text{Max}}(SIM_n)},\ NDIV_n = \frac{DIV_n}{\underset{i\in[0,N]}{\text{Max}}(DIV_n)} \quad (9)$$

where SIM_n denotes the similarity between RS and PS_n, DIV_n denotes the relative diversity of PS_n with reference to *RS*, $NSIM_n \in [0, 1]$ and $NDIV_n \in [0, 1]$ are the normalization version of SIM_n and DIV_n, respectively, λ and η separately denote the weight of $NSIM_n$ and $NDIV_n$ with the constraint of $\lambda + \eta = 1$. The default value of λ and η is 0.5, which indicates that $NSIM_n$ and $NDIV_n$ give the equal contributions to TR_n.

The SIM_n and DIV_n can be calculated through following:

$$\begin{aligned} SIM_n &= OA\left(DNN(RS),\ PS_n\right) \\ DIV_n &= OA\left(DNN(PS_n),\ RS\right) \end{aligned} \quad (10)$$

where $DNN(RS)$ and $DNN(PS_n)$ denote the deep neural networks (DNNs) trained by RS and PS_n, respectively, $OA(DNN(RS),\ PS_n)$ denotes the overall accuracy (OA) of $DNN(RS)$ tested on PS_n, $OA(DNN(PS_n),\ RS)$ denotes OA of $DNN(PS_n)$ tested on RS.

The designing idea of SIM_n and DIV_n is that the real samples and pseudosamples are used as the training samples and testing samples for each other. On the one hand, if the pseudosamples are similar to the real samples, the DNN trained on real samples can obtain high score when tests on pseudosamples, i.e., SIM_n gets the high value. On the other hand, if the diversity of pseudosamples is not enough, the pseudosamples cannot fully cover the distribution of real samples, the DNN trained on pseudosamples cannot obtain the high score when tests on real samples, i.e., DIV_n gets the low value.

It is worth noting that the value range of FID_n, SIM_n, and DIV_n is different, therefore, the normalization operation given in (7) and (9) is imposed on them to ensure their value range is restricted to [0, 1]. Furthermore, the $\alpha + \beta = 1$ and $\lambda + \eta = 1$ can ensure that the value range of M_n and TR_n is also restricted to [0,1].

The best pseudosamples PS_j can be obtained through following:

$$j = \arg\max_n \ M_n. \tag{11}$$

C. Focal Loss-Based Training

As shown in Fig. 1, the selected pseudosamples PS_j is used to pretrain the DSCN and the real samples RS is further used to finetune the DSCN. To further improve the accuracy of scene classification, the focal loss [61] is applied to the loss function of DSCN.

The focal loss is original applied to the classification branch of one-stage object detection model for improving the accuracy of object detection. The motivation of focal loss is reducing the weight of easy samples, which is equivalent to increasing the relative weight of hard samples. The formulation of focal loss is as follows:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \tag{12}$$

where p_t denotes the output probability of DSCN that samples belong to the true class, γ is a tunable parameter to control the extent of "focusing." The α_t is used to address the problem that the numbers of positive samples and negative samples are unbalanced for object detection, as a matter of fact, the similar problem that the number of samples of different scenes is usually unbalanced, is also existed in scene classification, consequently, the α_t is remained to address the problem of class imbalance.

III. Experiments and Analysis

A. Experiments Setup

1) Dataset: Two large scale datasets AID [67] and NWPU-RESISC45 [68] are employed to evaluate the effectiveness of proposed method.

The AID dataset includes 10 000 images and 30 categories, the size of each image is 600 × 600 pixels, the number of images per category ranges from 220 to 420. The spatial resolution is about 8 to 0.5 m per pixel. The images are captured from different countries, at different seasons and under different imaging conditions, which bring some challenges for scene classification methods.

The NWPU-RESISC45 dataset includes 31 500 images and 45 categories, the size of each image is 256 × 256 pixels, the number of images per category is 700. The images are captured from more than 100 countries, the spatial resolution of which is about 30 to 0.2 m per pixel. To our best knowledge, the NWPU-RESISC45 dataset is the largest public scene classification dataset, which can comprehensively evaluate the remote sensing image scene classification methods.

2) Quantitative measures: The OA and confusion matrix are the popular measures in the field of scene classification, and are also used for the quantitative evaluation in this article.

The *OA* is defined as the ratio of the number of correctly classified testing images to the total number of testing images, which is used to evaluate the overall performance of scene classification methods.

The confusion matrix is used to quantitatively evaluate the degree of confusion between each category. The rows and columns of confusion matrix represent the real and predicted scenes respectively. Moreover, the kappa coefficient (*KC*) which can evaluate the overall degree of confusion is adopted in this article, the value of *KC* is inversely proportional to the degree of confusion.

3) Implementation details: 20% and 50% images in AID dataset are used for training, and the remaining 80% and 50% images are used for testing [67]. For the NWPU-RESISC45 dataset, 10% and 20% images are used for training, and the remaining 90% and 80% images are used for testing [68].

In Section II-A, all of the input real images are resized to 224 × 224, the SinGAN adopts the default parameter setting [59], i.e., $N = 8$, 9 groups of pseudosamples are generated, denoted as $\{PS_8, \ldots, PS_0\}$, the ratio of the number of real images RS to generated images PS_n, $n \in [0, 8]$ is 1:10.

In Section II-C, the ResNet50 [69] which has been pretrained on the ImageNet is selected as the backbone of DSCN, the only difference of them is that the 1000 categories classification is replaced by 30 or 45 categories, the α_t and γ in (12) adopt the default value [61], i.e., $\alpha_t = 0.25$, $\gamma = 2$. The learning rate of the last layer of DSCN is 0.01, the rest is 0.001. The adaptive moment estimation and asynchronous stochastic gradient descent are selected as optimizers for the training of SinGAN and DSCN, respectively. The batch size is set to 32 in the training process of DSCN.

All the experiments are conducted on PyTorch framework, and running on a workstation with two E5-2650V4 CPUs (2.2 GHz, a total of 12 × 2 cores), 512 GB memory, and 8 NVIDIA TITAN RTX GPUs (a total of 24 GB × 8 video memory).

B. Evaluation of Proposed Quantitative Sifting Measure

A novel quantitative measure for sifting pseudosamples is proposed in this article, therefore, an experiment conducted on NWPU-RESISC45 dataset is used to validate its effectiveness in terms of the overall accuracy. Specifically, the 10% images in NWPU-RESISC45 dataset are considered as the real images, nine groups of pseudosamples $\{PS_8, \ldots, PS_0\}$ can be obtained as shown in Table I, the values of quantitative measure *M* of $\{PS_8, \ldots, PS_0\}$ is shown in the second row. The real images and each group of pseudo samples are jointly used to train the DSCN in turn, and the OA and KC of 9 trained DSCNs tested on remaining 90% images in NWPU-RESISC45 are shown on in third row and the fourth row of Table I, respectively. Apparently, as shown in Table I, the higher of *M*, the higher the corresponding

TABLE I
VALIDATING THE EFFECTIVENESS OF PROPOSED QUANTITATIVE SIFTING MEASURE ON NWPU-RESISC45 DATASET DATASET IN TERMS OF OVERALL ACCURACY

Pseudo samples	PS_8	PS_7	PS_6	PS_5	PS_4	PS_3	PS_2	PS_1	PS_0
M	**0.9950±0.15**	0.9257±0.13	0.9166±0.13	0.9149±0.12	0.9133±0.11	0.9112±0.12	0.9093±0.16	0.9073±0.17	0.9057±0.18
OA	**88.35±0.10**	86.24±0.11	86.22±0.16	86.19±0.19	86.12±0.23	86.05±0.15	85.91±0.17	85.87±0.21	85.82±0.19
KC	**0.8808±0.10**	0.8593±0.12	0.8590±0.15	0.8586±0.20	0.8580±0.21	0.8573±0.18	0.8559±0.17	0.8556±0.23	0.8549±0.21

Bold entities denote best results.

TABLE II
QUANTITATIVE COMPARISONS ON AID DATASET IN TERMS OF OVERALL ACCURACY

Methods	Training ratio	
	20%	50%
RS (Baseline)	87.16±0.28	90.78±0.19
PS	87.68±0.17	91.28±0.17
RS+PS	88.29±0.28	91.74±0.16
RS+FL	87.32±0.25	90.90±0.18
PS+FL	88.39±0.12	91.38±0.13
RS+FS+FL (Ours)	**88.52±0.11**	**92.06±0.13**

Bold entities denote best results.

TABLE III
QUANTITATIVE COMPARISONS ON NWPU-RESISC45 DATASET IN TERMS OF OVERALL ACCURACY

Methods	Training ratio	
	10%	20%
RS (Baseline)	86.08±0.28	88.33±0.15
PS	88.26±0.17	89.89±0.13
RS+PS	88.35±0.10	89.98±0.08
RS+FL	86.83±0.25	89.63±0.11
PS+FL	88.47±0.12	90.87±0.12
RS+FS+FL (Ours)	**88.71±0.11**	**91.21±0.05**

Bold entities denote best results.

TABLE IV
QUANTITATIVE COMPARISONS ON AID DATASET IN TERMS OF KC

Methods	Training ratio	
	20%	50%
RS (Baseline)	0.8687±0.25	0.9057±0.17
PS	0.8739±0.16	0.9108±0.12
RS+PS	0.8802±0.26	0.9156±0.13
RS+FL	0.8703±0.22	0.9069±0.16
PS+FL	0.8813±0.14	0.9119±0.13
RS+FS+FL (Ours)	**0.8826±0.11**	**0.9187±0.12**

Bold entities denote best results.

TABLE V
QUANTITATIVE COMPARISONS ON NWPU-RESISC45 DATASET IN TERMS OF KC

Methods	Training ratio	
	10%	20%
RS (Baseline)	0.8573±0.21	0.8808±0.20
PS	0.8803±0.18	0.8966±0.15
RS+PS	0.8808±0.13	0.8975±0.13
RS+FL	0.8653±0.25	0.8941±0.10
PS+FL	0.8820±0.14	0.9066±0.10
RS+FS+FL (Ours)	**0.8845±0.10**	**0.9102±0.09**

Bold entities denote best results.

OA and *KC*, which can directly validate the effectiveness of proposed quantitative sifting measure *M*. The PS_8 of which the *M* value is highest is selected as the final augmented samples for training the DSCN, i.e., the bottom GAN in SinGAN is adopted as the initial GAN for generating the pseudosamples.

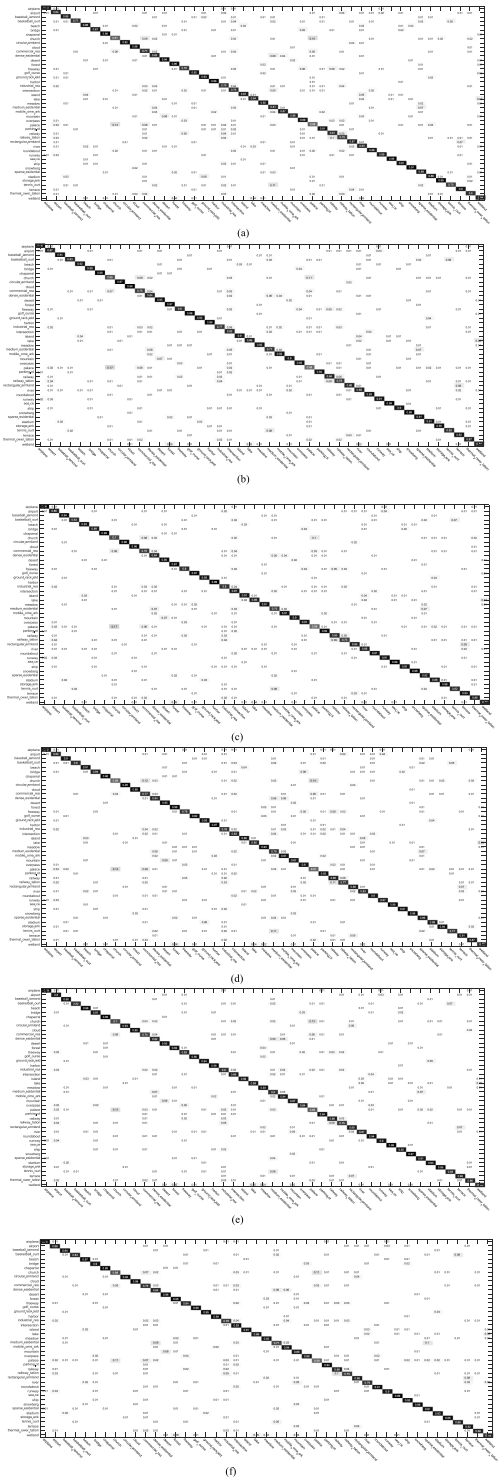

Fig. 4. Comparisons of confusion matrices on NWPU-RESISC45 dataset with the training ratio of 10%. (a) RS (Baseline). (b) PS. (c) RS+PS. (d) RS+FL. (e) PS+FL. (f) RS+FS+FL (Ours).

C. Evaluation of Pseudosamples and Focal Loss

To validate that incorporating the pseudosamples and focal loss can improve the performance of DSCN, the proposed method is quantitatively compared with five methods which have different configuration with reference to the baseline method. As shown in Tables II and III, RS denotes the DSCN model trained by the real samples only, which is also considered as the baseline method, the PS uses the pseudosamples for training instead of real samples, the RS+PS jointly uses the real and pseudosamples to train the DSCN. The RS+FL, PS+FL, and RS+FS+FL are obtained by replacing the traditional cross entropy loss of RS, PS, RS+PS by focal loss, respectively, where RS+FS+FL denotes proposed method.

The quantitative comparison results on AID dataset are shown in Table II, the overall performance of PS is superior to RS, which demonstrates that the generated pseudo samples have good quality and can improve the performance of DSCN, the comparison between RS+PS and PS indicates that the combination of PS and RS can further improve the performance. The comparisons between RS+FL, PS+FL, RS+FS+FL and RS, PS, RS+PS demonstrate that replacing the traditional cross entropy by focal loss can also improve the performance of remote sensing image scene classification besides the object detection. The similar conclusion can be derived from Table III.

Moreover, the quantitative comparisons in terms of confusion matrix are also provided to further evaluate the effectiveness of proposed method. As shown in Fig. 4, only the confusion matrices obtained from NWPU-RESISC45 dataset with the training ratio of 10% are presented because of the space limitation. Therefore, the KC which can evaluate the overall degree of confusion in a form of single value is also used for quantitative evaluation, as shown in Tables IV and V. The conclusions derived from Fig. 4, Tables IV and V are similar to the one derived from Tables II and III.

IV. Conclusion

The existing methods for sifting the pseudosamples cannot directly evaluate the quality of generated pseudosamples from the perspective of model training. A novel quantitative sift measure is proposed to address this problem in this article, which can directly evaluate both the authenticity and diversity from the perspective of model training. Moreover, the focal loss that is originally proposed for object detection is first applied to the remote sensing image scene classification in this article. The quantitative evaluation of proposed sifting measure in terms of OA shows that the value of sifting measure is proportional to the overall accuracy, which validates the effectiveness of proposed sifting measure. The quantitative comparisons on AID and NWPU-RESISC45 dataset in terms of OA and confusion matrices demonstrate that incorporating the augmented pseudosamples and focal loss can improve the performance of DSCN. It is worth noting that the proposed quantitative sifting measure can applied to the evaluation of any pseudosamples which are generated from same real samples.

In addition to the SinGAN, the other generation models, such as DeliGAN [62] etc., will also be adopted to generate more diverse samples in the following works. Moreover, some lightweight CNN models, such as MobileNet V3 [70] etc., will be used as the backbone of DSCN instead of ResNet50 to improve the efficiency of scene classification.

References

[1] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.

[2] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, Jun. 2016.

[3] X. Qian, J. Li, G. Cheng, X. Yao, S. Zhao, Y. Chen, and L. Jiang, "Evaluation of the impact of feature extraction strategy on the high-resolution remote sensing image scene classification," *J. Remote Sens.*, vol. 22, no. 5, pp. 758–776, Sep. 2018.

[4] Z. Lv, G. Li, Z. Jin, J. A. Benediktsson, and G. M. Foody, "Iterative training sample expansion to increase and balance the accuracy of land classification from VHR imagery," in *Proc. IEEE Trans. Geosci. Remote Sens.*, Jun. 2020, pp. 1–12.

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

[6] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[7] X. Qian, S. Lin, G. Cheng, X. Yao, H. Ren, and W. Wang, "Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion," *Remote Sens.*, vol. 12, pp. 143–163, Jan. 2020.

[8] Z. Y. Lv, T. F. Liu, P. Zhang, J. A. Benediktsson, T. Lei, and X. Zhang, "Novel adaptive histogram trend similarity approach for land cover change detection by using bitemporal very-high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9554–9574, Aug. 2019.

[9] Z. Lv, T. Liu, and J. A. Benediktsson, "Object-oriented key point vector distance for binary land cover change detection using VHR remote sensing images,"in *Proc. IEEE Trans. Geosc. Remote Sens.*, Mar. 2020, pp. 1–10.

[10] X. Lv, D. Ming, Y. Chen, and M. Wang, "Very high resolution remote sensing image classification with SEEDS-CNN and scale effect analysis for superpixel CNN classification," *Int. J. Remote Sens.*, vol. 40, no. 2, pp. 506–531, Sep. 2018.

[11] S. Xu, X. Mu, P. Zhao, and J. Ma, "Scene classification of remote sensing image based on multiscale feature and deep neural network," *Acta Geodaetica et Cartographica Sinica*, vol. 45, no. 7, pp. 834–840, Jul. 2016.

[12] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.

[13] G. Cheng, J. Han, L. Guo, Z. Liu, S. Bu, and J. Ren, "Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4238–4249, Aug. 2015.

[14] W. Zhou, D. Ming, X. Lv, K. Zhou, H. Bao, and Z. Hong, "SO–CNN based urban functional zone fine division with VHR remote sensing image," *Remote Sens. Environ.*, vol. 236, pp. 111458–111478, Jan. 2020.

[15] T. Lu, D. Ming, X. Lin, Z. Hong, X. Bai, and J. Fang, "Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network," *Remote Sens.*, vol. 10, no. 9, pp. 1496–1515, Sep. 2018.

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

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

[18] 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, Jan. 2001.

[19] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2005, pp. 524–531.

[20] L. Zhao, P. Tang, and L. Huo, "A 2-D wavelet decomposition-based bag-of-visual-words model for land-use scene classification," *J. Remote Sens.*, vol. 35, no. 6, pp. 2296–2310, Mar. 2014.

[21] H. Sridharan and A. Cheriyadat, "Bag of lines (BoL) for improved aerial scene representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 676–680, Mar. 2015.

[22] J. Hu, T. Jiang, X. Tong, G.-S. Xia, and L. Zhang, "A benchmark for scene classification of high spatial resolution remote sensing imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 5003–5006.

[23] L. Chen, W. Yang, K. Xu, and T. Xu, "Evaluation of local features for scene classification using VHR satellite images," in *Proc. Joint Urban Remote Sens. Event*, 2011, pp. 385–388.

[24] 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, Nov. 2015.

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

[26] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2009, pp. 1794–1801.

[27] H. Junwei, Z. Dingwen, H. Xintao, G. Lei, R. Jinchang, and W. Feng, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.

[28] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.

[29] J. Han, X. Yao, G. Cheng, X. Feng, and D. Xu, "P-CNN: Part-based convolutional neural networks for fine-grained visual categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[30] J. Han *et al.*, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Jul. 2013.

[31] X. Qian, J. Li, J. Cao, Y. Wu, and W. Wang, "Micro-cracks detection of solar cells surface via combing short-term and long-term deep features," *Neural Netw.*, vol. 127, pp. 132–140, Jul. 2020.

[32] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," in *Proc. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, Jun. 2020, pp. 3735–3756.

[33] 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. 26, no. 10, pp. 1–16, Oct. 2018.

[34] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.

[35] Q. Zhu, Y. Zhong, L. Zhang, and D. Li, "Scene classification based on the fully sparse semantic topic model," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5525–5538, Oct. 2017.

[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] N. He, L. Fang, S. Li, A. Plaza, and J. Plaza, "Remote sensing scene classification using multilayer stacked covariance pooling," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 6899–6910, Dec. 2018.

[38] J. Xie, N. He, L. Fang, and A. Plaza, "Scale-free convolutional neural network for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6916–6928, Sep. 2019.

[39] X. Lu, H. Sun, and X. Zheng, "A feature aggregation convolutional neural network for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7894–7906, Oct. 2019.

[40] Y. Liu, Y. Zhong, and Q. Qin, "Scene classification based on multiscale convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7109–7121, Dec. 2018.

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

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

[43] J. Chen *et al.*, "Convolution neural network architecture learning for remote sensing scene classification," 2020, *arXiv:2001.09614v1*.

[44] W. Zhang, P. Tang, and L. Zhao, "Remote sensing image scene classification using CNN-CapsNet," *Remote Sens.*, vol. 11, no. 5, pp. 494–515, Feb. 2019.

[45] X. Qian *et al.*, "Hardness recognition of robotic forearm based on semi-supervised generative adversarial networks," *Front Neurorobot.*, vol. 13, no. 73, pp. 1–10, Sep. 2019.

[46] W. Han, R. Feng, L. Wang, and Y. Cheng, "A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, no. Part A, pp. 23–43, Nov. 2018.

[47] P. J. Soto, J. D. Bermudez, P. N. Happ, and R. Q. Feitosa, "A comparative analysis of unsupervised and semi-supervised representation learning for remote sensing image categorization," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W7, pp. 167–173, Sep. 2019.

[48] Y. Yu, X. Li, and F. Liu, "Attention GANs: Unsupervised deep feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 519–531, Jan. 2020.

[49] D. Lin, K. Fu, Y. Wang, G. Xu, and X. Sun, "MARTA GANs: Unsupervised representation learning for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2092–2096, Nov. 2017.

[50] X. Lu, X. Zheng, and Y. Yuan, "Remote sensing scene classification by unsupervised representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, Sep. 2017.

[51] X. Yao, X. Feng, J. Han, G. Cheng, and L. Guo, "Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning," in *Proc. IEEE Trans. Geosci. Remote Sens.*, May 2020, pp. 1–11.

[52] G. Cheng, J. Yang, D. Gao, L. Guo, and J. Han, "High-quality proposals for weakly supervised object detection," in *Proc. IEEE Trans Image Process.*, Apr. 2020, pp. 5794–5804.

[53] X. Feng, J. Han, X. Yao, and G. Cheng, "Progressive contextual instance refinement for weakly supervised object detection in remote sensing images," in *Proc. IEEE Trans. Geosci. Remote Sens.*, Apr. 2020, pp. 1–11.

[54] 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.

[55] E. Othman, Y. Bazi, F. Melgani, H. Alhichri, N. Alajlan, and M. Zuair, "Domain adaptation network for cross-scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4441–4456, Aug. 2017.

[56] Z. Gong, P. Zhong, Y. Yu, and W. Hu, "Diversity-promoting deep structural metric learning for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 371–390, Nov. 2018.

[57] A. Li, Z. Lu, L. Wang, T. Xiang, and J.-R. Wen, "Zero-shot scene classification for high spatial resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4157–4167, Jul. 2017.

[58] D. Ma, P. Tang, and L. Zhao, "SiftingGAN: Generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1046–1050, Jul. 2019.

[59] T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. IEEE Int. Conf. Comput. Vision*, 2019, pp. 4570–4580.

[60] D. Dowson, and B. Landau, "The Frećhet distance between multivariate normal distributions," *J. Multivariate Anal.*, vol. 12, no. 3, pp. 450–455, Sep. 1982.

[61] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 2980–2988.

[62] S. Gurumurthy, R. K. Sarvadevabhatla, and R. V. Babu, "DeLiGAN: Generative adversarial networks for diverse and limited data," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4941–4949.

[63] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–35.

[64] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 1125–1134.

[65] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[66] K. Shmelkov, C. Schmid, and K. Alahari, "How good is my GAN?," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 218–234.
[67] 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.
[68] 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.
[69] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.
[70] A. Howard *et al.*, "Searching for MobileNetV3," in *Proc. IEEE Int. Conf. Comput. Vision*, 2019, pp. 1314–1324.

Xiaoliang Qian (Member, IEEE) received the Ph.D. degree from the School of Automation, Northwestern Polytechnical University, Xi'an, China, in 2013.

He is currently an Associate Professor with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include remote sensing image understanding, computer vision, pattern recognition, and machine learning.

Xiaohao Chen received the B.S. degree from the College of Electrical and Information Engineering, Henan University of Engineering, Zhengzhou, China, in 2018. He is currently pursuing the M.S. degree with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China.

His research interests include remote sensing image scene classification and deep learning.

Weichao Yue received the Ph.D. degree from the School of Automation, Central South University, Changsha, China, in 2019.

He is currently a Lecturer with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include artificial intelligence, process control, knowledge representation, and reasoning.

Xianglong Liu received the Ph.D. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China, in 2019.

He is currently a Lecturer with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include signal processing, electromagnetic tomography, electromagnetic sensing and imaging, and nondestructive testing.

Jungang Guo received the B.S. degree from Zhengzhou University, Zhengzhou, China, in 2013.

He is currently the Chairman of China Science Quantum Cloud Technology Group Corporation, Ltd., Zhengzhou, China. His research interests include multidimensional sensor, edge computing, and data fusion.

Zhehui Li received the B.S. degree from the Institute of economics and management, Henan Agricultural University, Zhengzhou, China, in 2000.

She is currently a Senior Engineer with the Network Technology Center, Henan Provincial Institute of Scientific & Technical Information, Xinxiang, China. Her research interests include computer software applications and data statistical analysis.

Yinhua Li received the M.S. degree from the School of Electrical Engineering, Southeast University, Nanjing, China, in 1994.

He is currently a Professor with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include artificial intelligence, computer vision, and intelligent instrument.

Wei Wang received the Ph.D. degree from Concordia University, Montreal, QC, Canada, in 2002.

He is currently a Professor with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include artificial intelligence, pattern recognition, computer vision, remote sensing image understanding, and tactile sensor.