Research Progress on Few-Shot Learning for Remote Sensing Image Interpretation

Xian Sun, *Senior Member, IEEE*, Bing Wang, *Student Member, IEEE*, Zhirui Wang, *Member, IEEE*, Hao Li, *Member, IEEE*, Hengchao Li, *Member, IEEE*, and Kun Fu, *Member, IEEE*

***Abstract*—The rapid development of deep learning brings effective solutions for remote sensing image interpretation. Training deep neural network models usually require a large number of manually labeled samples. However, there is a limitation to obtain sufficient labeled samples in remote sensing field to satisfy the data requirement. Therefore, it is of great significance to conduct the research on few-shot learning for remote sensing image interpretation. First, this article provides a bibliometric analysis of the existing works for remote sensing interpretation related to few-shot learning. Second, two categories of few-shot learning methods, i.e., the data-augmentation-based and the prior-knowledge-based, are introduced for the interpretation of remote sensing images. Then, three typical remote sensing interpretation applications are listed, including scene classification, semantic segmentation, and object detection, together with the corresponding public datasets and the evaluation criteria. Finally, the research status is summarized, and some possible research directions are provided. This article gives a reference for scholars working on few-shot learning research in the remote sensing field.**

***Index Terms*—Deep generative model, few-shot learning, meta-learning, metric learning, remote sensing, transfer learning.**

I. Introduction

IN RECENT years, deep learning methods have been rapidly developed, and the application in computer vision field is remarkable [1]–[3]. Data, computing power, and algorithm are three crucial factors for the development of deep learning. Due to the complex structure as well as a huge number of parameters to be optimized [4], [5], the performance of the deep learning model heavily relies on plenty of samples. In other words, data are the booster for the rapid development of deep learning.

Inspired by the success of deep learning in computer vision, preliminary studies have been carried out on deep learning for remote sensing image interpretation and attain significant progress [6]–[11]. Compared with natural scene images, the application of deep learning in remote sensing images faces three challenges: small geographic objects in high-resolution images, large variations in the visual appearance of objects, and complex background noise. Moreover, the total amount of available training data in remote sensing field is far less than that of natural scenes, which results from the following reasons.

1) *Lack of raw data:* Different from natural images, which can be easily obtained via web resource, remote sensing images are collected with high cost and long cycle owing to its strict requirement on imaging sensors and conditions. Besides, the number of high-value objects in remote sensing images, such as aeroplane and ship, is smaller than that of common objects in natural images. The scarcity of target sources increases the difficulty in obtaining the remote sensing data.
2) *Difficulty of dataset annotation:* Generally, the training of deep learning model depends on the manually labeled samples. However, the annotation of remote sensing samples requires high interpretation ability and even expert knowledge, such as the aircraft type recognition. In addition, considering the annotation difficulty and huge data requirement, the dataset construction process can be very time-consuming and laborious.
3) *Limitation of sensor characteristic:* The imaging results of the same object may differ greatly due to the sensor factors, such as imaging angle and resolution, especially in the synthetic aperture radar (SAR) field. However, the fixed sensor parameters lead to the lack of imaging diversity, which cannot fully reflect the complete target characteristic based on the existing samples.

Although a series of publicly available remote sensing datasets have been released, the number of images and object categories are still relative small compared with the natural scene dataset. Under this condition, the deep learning model will encounter the overfitting problem [14]. It is difficult to obtain the optimal parameters according to the rule of empirical error minimization with limited data. Therefore, it is of great importance to solve the overfitting problem and improve the

Manuscript received November 16, 2020; revised January 7, 2021; accepted January 9, 2021. Date of publication January 19, 2021; date of current version February 18, 2021. This work was supported by the National Natural Science Foundation of China under Grant 61725105 and in part by the National Major Project on High Resolution Earth Observation System under Grant GFZX0404120201. *(Corresponding author: Xian Sun.)*

Xian Sun, Bing Wang, and Kun Fu are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, also with the Key Laboratory of Network Information System Technology (NIST), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China, also with University of Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: sunxian@aircas.ac.cn; wangbing181@mails.ucas.ac.cn; fukun@mail.ie.ac.cn).

Zhirui Wang and Hao Li are with the Key Laboratory of Network Information System Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: zhirui1990@126.com; lihaoaircas@163.com).

Hengchao Li is with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: hcli@home.swjtu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3052869

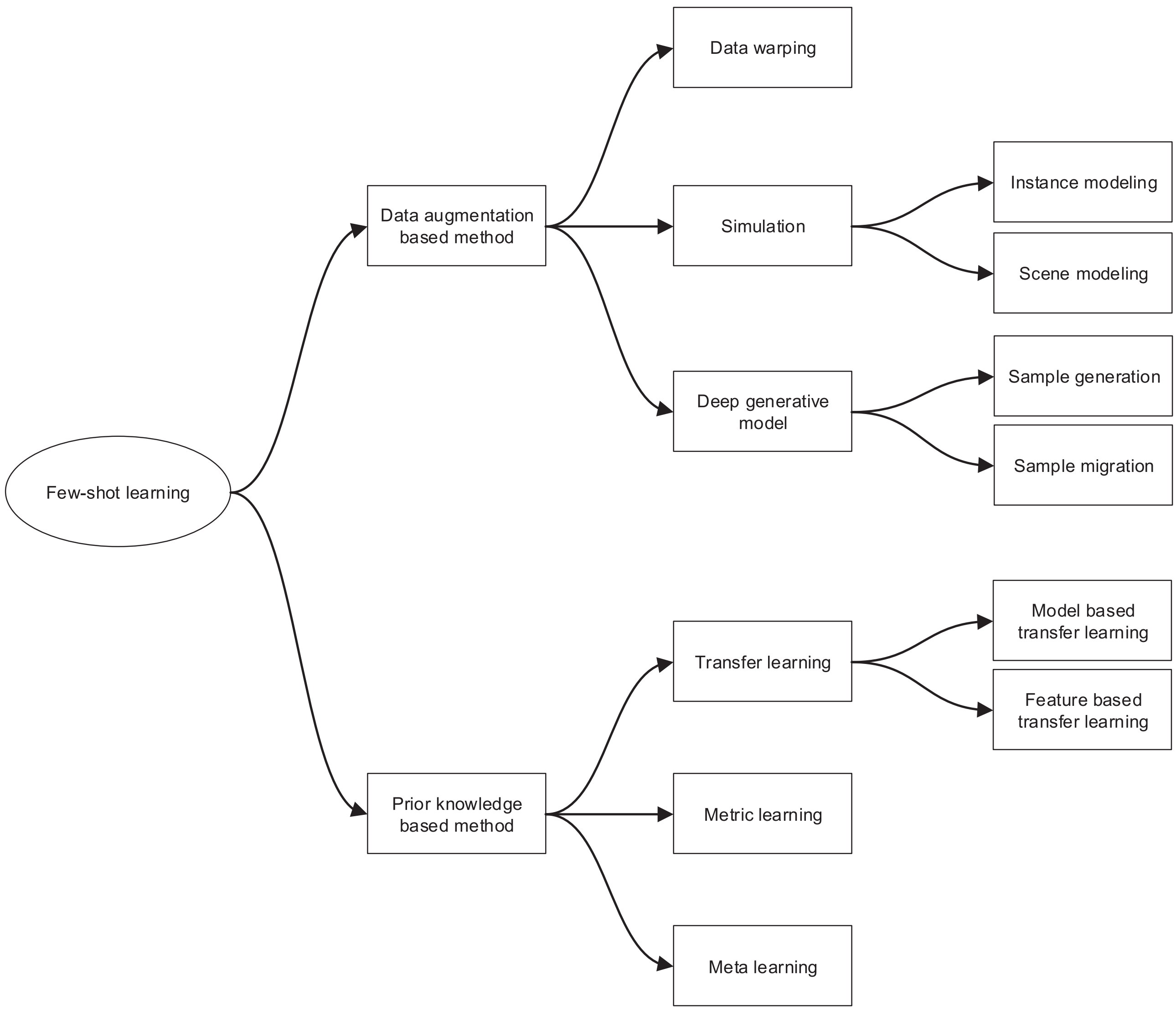


Fig. 1. Pattern of remote sensing image interpretation with limited labeled data.

generalization ability of the deep learning model on the basis of limited data in remote sensing field. As a possible solution, deriving from [15] and [16], the few-shot learning is proposed with the goal of learning from a small number of labeled samples and has gained significant improvement on the interpretation of natural images, including image classification [17], [18], object detection [19], [20], semantic segmentation [21], [22], etc.

Inspired by their successful applications in natural images, extensive efforts had been made by researchers with the aim of exploring effective few-shot learning methods for remote sensing image interpretation.

The formal definitions of few-shot learning are provided as follows. Given a learning task T, few-shot learning learns from the dataset $D = \{D_{\text{train}}, D_{\text{test}}\}$ for the current task T, which is composed of the training set $D_{\text{train}} = \{(x_i, y_i)\}_{i=1}^{K}$, where K is very small, and the test set $D_{\text{test}} = \{x_j, y_j\}_{j=1}^{M}$. The goal is to approximate the optimal embedding function $F(\cdot;\theta)$ from input x to label y by the iterative training on the training set D_{train}. Evaluation of the few-shot learning performance on the given task T is conducted by the predefined loss function $L(\hat{y}, y)$ over the test set D_{test} between the prediction $\hat{y} = F(\theta, x)$ and the ground truth label y.

It is worth mentioning that in this review, we use the term of "few-shot" as a general concept, which consists the narrow sense where only limited labeled samples for novel classes can be used (commonly adopted in computer version) and the broad sense where limited labeled samples and a certain amount of unlabeled samples for the target classes can be accessed.

As shown in Fig. 1, we divide the existing few-shot learning methods into the following two categories in light of the principles whether the amount of available labeled samples for the target classes is increased.

1) *Data-augmentation-based method:* A straightforward way to solve the few-shot learning problem is to enlarge the number of training samples by data augmentation. Data augmentation generally uses transformation operations, simulation, or deep generative models to generate samples without actually collecting new data. This kind of method can improve the generalization ability of the model and suppress the risk of overfitting. At present, the key of

data argumentation-based method lies in generating new samples automatically with high fidelity, which not only expands the dataset in quantity, but also enriches the semantic information of dataset.

2) *Prior-knowledge-based method:* This kind of method mainly focuses on learning with limited labeled data, which means making full use of prior knowledge and experience to guide the learning progress of new tasks. Human beings can learn a new concept from a few pictures and even sometimes form the knowledge concepts without seeing the pictures. Inspired by the human ability, researchers hope that deep learning models can quickly learn new categories with a small number of training samples, which is the purpose of this method.

Considering the lack of a thorough survey of few-shot learning for remote sensing image interpretation, we conduct a comprehensive review of the relevant research by the detailed bibliometric analysis of the existing works, systematically review the few-shot learning approaches from the perspective of data augmentation and knowledge reuse, and discuss several promising future directions of few-shot learning for remote sensing image interpretation. This article may provide a reference for scholars working on few-shot learning research in the remote sensing field.

The remainder of this article is organized as follows. Section II provides a quantitative analysis of the few-shot learning articles in remote sensing field. Sections III and IV introduce the few-shot learning methods from the perspective of data augmentation and prior knowledge, respectively. Section V presents several typical applications of few-shot learning in remote sensing image interpretation. Section VI discusses the current issues to be addressed and proposes some promising future directions. Section VII concludes this article with final remarks.

II. Quantitative Analysis of Articles

In order to systematically analyze the trend and research hotspots of few-shot learning in remote sensing image interpretation field, the relevant articles from Web of Science (WOS) and China National Knowledge Infrastructure (CNKI) are collected and quantitatively analyzed in this section. The core collection in the WOS is selected, and the retrieval keywords are set as (title = few shot) OR (title = zero shot) OR (title = limited data) OR (title = semi-supervised) OR (title = limited labeled data) OR (title = one-shot) OR (title = metric learning) OR (title = transfer learning) OR (title = self-label) OR (title = generative adversarial networks) OR (title = auto-encoders) And (subject = remote sensing). The retrieval duration is from 2000 to 2019. After manual selection, a total of 243 valid titles can be obtained.

In the CNKI database, SCI, EI, and the core journal database are selected. The corresponding retrieval keywords are (key words = few shot learning) OR (key words = limited data) OR (key words = meta learning) OR (key words = transfer learning) OR (key words = generative adversarial networks) OR (key words = auto-encoders) AND (subject = remote sensing). The retrieval duration is from 2000 to 2019. After manual selection, a total of 40 valid titles can be obtained.

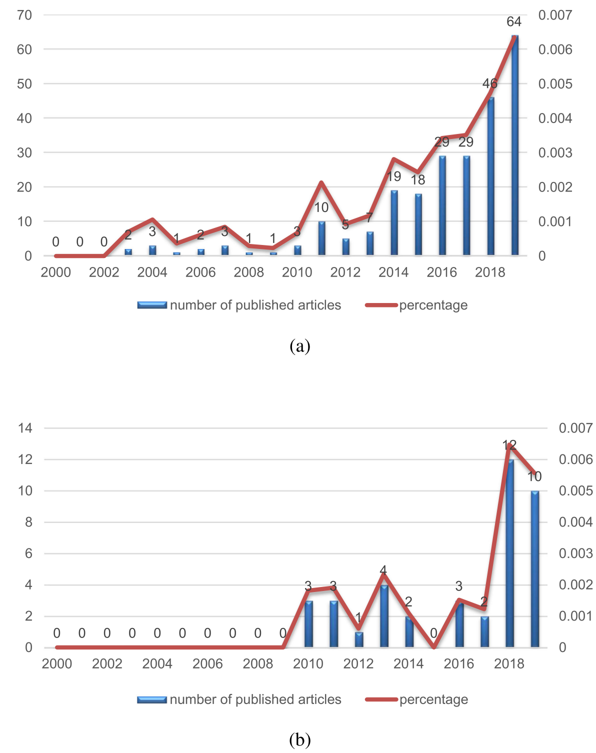


Fig. 2. Quantitative analysis of the published articles. (a) Number and proportion of published articles by year on few-shot learning in the remote sensing field (data source from WOS). (b) Number and proportion of published articles by year on few-shot learning in the remote sensing field (data source from CNKI).

A. Quantitative Analysis of Published Articles

The published articles from 2000 to 2019 are counted and analyzed. Data retrieved from WOS show that the number of few-shot learning articles has increased year by year since 2003 and reached a peak of 64 in 2019, as shown in Fig. 2(a). Furthermore, the red line in Fig. 2(a) denotes the percentage of the few-shot learning articles in the number of remote sensing topic articles each year. After comprehensive analysis, it can be found that both the annually published articles and the proportion of remote sensing few-shot learning articles have significantly increased since 2014, and there is an obvious rising trend in recent years. As for the Chinese journals, the relevant data are collected from CNKI, and the corresponding statistical results are shown in Fig. 2(b). It can be seen that Chinese articles related to the remote sensing few-shot learning began to appear around 2010. Although the number is small, the overall trend was still rising, and it reached a peak of 12 in 2018. Combined with the data obtained in WOS, it can be concluded that the few-shot learning has attracted more and more attention in both Chinese and English remote sensing articles. The number of published articles is increasing and will continue to increase in the next few years.

B. Analysis of Journal (Conference)

In this subsection, a statistical analysis of the journal and conference articles relevant to few-shot learning is presented. Concretely, by means of the paper analysis tool named Histcite,

TABLE I
NUMBERS OF PUBLISHED ARTICLES ON DIFFERENT JOURNALS AND THEIR TLCS AND TGCS IN WOS DATA

Ranking	Journal / Conference	Number of articles	TLCS	TGCS
1	IEEE International Geoscience and Remote Sensing Symposium	58	6	107
2	Remote Sensing	28	4	244
3	IEEE Transactions on Geoscience and Remote Sensing	12	7	664
4	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	10	8	179
5	ISPRS Journal of Photogrammetry and Remote Sensing	10	20	243
6	IEEE Geoscience and Remote Sensing Letters	5	2	72
7	Remote Sensing Letters	4	0	20
8	Sensors	4	0	29
9	Applied Soft Computing	3	5	89
10	IEEE Access	3	1	12

TABLE II
NUMBERS OF PUBLISHED ARTICLES IN DIFFERENT JOURNALS FROM CNKI DATA

Ranking	Journal	Number of articles
1	Acta Geodaetica et Cartographica Sinica	4
2	Bulletin of Surveying and Mapping	3
3	Computer Science	3
4	Computer Engineering and Applications	3
5	Journal of Image and Graphics	3

TABLE III
STATISTICS OF HIGH-FREQUENCY CENTRAL WORDS AND HIGHLY CENTRAL KEYWORDS

Ranking	High-frequency keywords	Frequency	High-centrality keywords	Centrality
1	classification	49	hyperspectral image classification	0.34
2	transfer learning	42	semi-supervised learning	0.33
3	semi-supervised learning	35	active learning	0.32
4	semi-supervised classification	23	algorithm	0.29
5	deep learning	21	image classification	0.24
6	image classification	20	transfer learning	0.23
7	hyperspectral image	19	framework	0.21
8	convolutional neural network	16	remote sensing	0.18
9	generative adversarial network	14	classification	0.18
10	change detection	13	change detection	0.18

the number of articles, the total location citation score (TLCS), and the total global citation score (TGCS) for each journal or conference article were obtained and summarized in Table I. Table I shows that IEEE International Geoscience and Remote Sensing Symposium (IGARSS), *Remote Sensing*, and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) have made the most contributions to the number of published articles. The number of publications on IGARSS is 58, accounting for 23.87%. The number of published articles on TGRS ranks the first. Although the number of published articles in the *ISPRS Journal of Photogrammetry and Remote Sensing* only ranks fifth, the corresponding TLCS and TGCS ranks first and third, respectively.

Endnote, an article analysis tool, was used to analyze 40 articles retrieved from CNKI. As shown in Table II, we can see that the number of Chinese articles is relatively small, and corresponding journals are relatively scattered. It can be concluded that the number of remote sensing few-shot learning in Chinese journals is relatively average and significantly less than that published in journals in the WOS database. Therefore, it can be concluded that remote sensing few-shot learning articles in Chinese journals have great growth potential both in the journals and the number of articles.

C. Analysis of Keywords

In this section, the analysis tool Citespace [23] is used to process the data obtained from WOS. The title, abstract, and keywords are selected as the source of subject words. The analysis results of high-frequency keywords and high-centrality keywords are summarized. According to Table III, the most commonly used methods include semisupervised classification learning, transfer learning, convolutional neural network (CNN), and generative adversarial network (GAN), etc.

Meanwhile, we analyze the data year by year and conduct the co-citation analysis of keywords with Citespace. The time range is set from 2000 to 2019. The node type is set as keywords, and the clustering is conducted according to the frequency of keywords. Finally, we obtain the evolution map of keywords as shown in Fig. 3. Concretely, the horizontal axis represents the earliest proposed time of the target keywords, and the vertical axis corresponds to the name of the specific keywords. And the circle size as well as the font size is proportionate to the emergence frequency of the keywords in a given slice time, which can reflect the research heat in some degree.

Based on the analysis of high-frequency keywords, high-centrality keywords, and the evolution map of keywords, the development of few-shot learning in remote sensing field can be divided into the following three stages.

1) *2000–2007:* At this stage, the main keywords were classification, remote sensing image, semisupervised learning, etc. The concept of few-shot learning had just been formed and the semisupervised learning began to appear. However, the relevant research progress was not great, and the number of articles was also small.

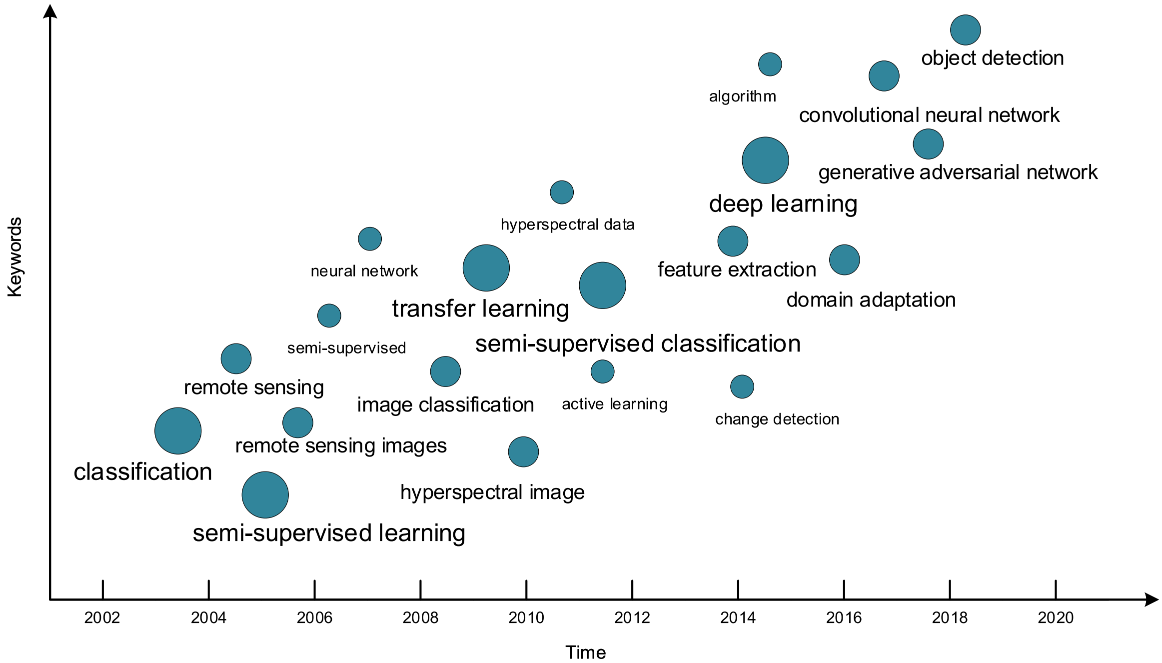


Fig. 3. Evolution map of remote sensing few-shot learning research.

TABLE IV
COMPARISON OF DIFFERENT DATA AUGMENTATION METHODS

Methods	Overview	Pros	Cons	Typical References
Data warping	Perform basic image manipulations over original data	Fast generation speed and easy implementation	Poor expansion of semantic information	[24]–[26]
Simulation	Utilize simulation and virtual imaging technology to generate synthetic image	Fast generation speed, high controlment of image information	Domain gap between synthetic data and real data	[27]–[29]
Deep generative model	Learn the probability distributions of image and generate fake samples with high fidelity	Automatic image generation without human labor	Difficulty in generating images with very high resolution for complex scenes	[30]–[32]

2) *2008–2013:* The keywords were semisupervised classification, image classification, transfer learning, hyperspectral image, etc. During this period, the semisupervised learning methods were constantly improved. Additionally, new methods such as transfer learning were introduced in the remote sensing data processing.
3) *2014–2019:* The keywords were CNN, deep learning, domain adaptation, object detection, GAN, etc. The rapid development of deep learning made great contribution to the research of few-shot learning in remote sensing field, and many new few-shot learning methods emerged successively.

III. DATA-AUGMENTATION-BASED LEARNING METHOD

Data augmentation is a well-known technique to mitigate data scarcity problem by increasing the volume and diversity of the available data instead of actually collecting new data. Based on the augmented dataset, the risk of overfitting can be obviously decreased, and the generalization ability of model can be effectively strengthened. Generally, as shown in Table IV, the existing data augmentation methods designed for remote sensing image can be divided into three categories: data-warping-based method, simulation-based method, and deep-generative-model-based method.

A. Data Augmentation Based on Data Warping

Data warping is a way of generating new samples by performing basic image manipulations based on the existing data. Commonly used transformation techniques include cropping, flipping, filtering, rotation, and noising. These transformations are easy to implement to increase the data scale. However, the new semantic information cannot be generated to increase the data diversity. The effect of this data augmentation method on improving the model performance is very limited. Hence, this kind of method cannot completely solve the sample limit problem and usually is adopted as an auxiliary technique in data preprocessing [24]–[26].

B. Data Augmentation Based on a Simulation Technique

Another data augmentation strategy is to establish the computing model and simulate the remote sensing imaging process, which can output the remote sensing images by a computer. According to the scale of modeling target, the existing simulation methods can be classified into two categories.

1) *Instance modeling:* Instance modeling indicates that the 3-D model of object is built and will be projected to a certain remote sensing background by the simulation system. This kind of method is widely adopted for synthetic image generation in target recognition and object detection tasks. Kusk *et al.* [28], [33] investigated the generation of synthetic SAR data based on Computer Simulation Technology Microwave Studio Asymptotic Solver (CST). Concretely, a set of 3-D CAD models, which contain the object radar reflectivity information, were sent to the simulation system. Then, CST estimated the complex scattered electric field components and generated the corresponding synthetic samples. Similarly, Yan *et al.* [29] proposed a simulation method for ship detection in remote sensing images. Considering the texture difference between the simulation target and the background image, Wang *et al.* [34] designed a multiscale generator network to perform domain conversion operation automatically.
2) *Scene modeling:* In a remote sensing image, the instance modeling is utilized to generate the target information and obtain its annotation. In order to acquire the pixelwise label information, the scene modeling is investigated to automatically generate the stimulated images of a certain area in an efficient way. Kemker *et al.* [35] conducted scene modeling experiments on Trona to assist in multispectral remote sensing image segmentation. With the support of Digital Imaging and Remote Sensing Image Generation modeling software, a large number of multispectral images and the corresponding label maps can be generated automatically. Besides, weather conditions, lighting conditions, as well as imaging height can be adjusted flexibly as required. To solve the problem of modeling cost, Kong *et al.* [27] developed an approach to generate synthetic overhead imagery rapidly and cheaply based on CityEngine. They released a collection of synthetic dataset Synthinel-1 for building segmentation and verified that Synthinel-1 was consistently beneficial to augment real images.

The main advantage of this augmentation method lies in its fast image generation speed and high controllability of image content information. Based on this technique, the remote sensing images can be effectively augmented with high fidelity and low cost, especially for some images that are hard to obtain in reality. However, one limitation of this augmentation technique is that there still exists the domain gap between the synthetic image and the real image. To solve the domain shift problem caused by the gap, it is required to combine transfer learning for further optimization.

C. Data Augmentation Based on the Deep Generative Model

The deep generative model can be used to learn rich probability distribution over target images and generates new samples with variations. One of the most commonly used deep generative models in remote sensing field are GANs [27]. Similar to the game theory, the GAN can learn the distribution of target data implicitly by finding the Nash equilibrium between the generator and the discriminator. According to the source of the generated samples, the existing deep-generative-model-based methods designed for data augmentation can be classified into two categories.

1) *Sample synthesis:* One typical approach of applying the deep generative model for data augmentation is sample synthesis. Specifically, the deep generative model is exposed to a certain number of real images with the aim of approximating the probability distribution characteristic of the target classes. After training, the generator is expected to be capable of generating fake images with high fidelity, which are not present in the training set, yet share the same distribution. When the GAN was first applied to the hyperspectral feature classification task, it was used as a regularization to solve the overfitting problem under the small-sample condition, and a collaborative training algorithm was proposed to fuse the adversarial samples with the real samples to improve the model performance [36]. Zheng *et al.* [32] designed vehicle synthesis GANs to generate high-quality annotated vehicles from optical remote sensing data and verified the synthesized vehicles can benefit the training of CNN-based vehicle detectors. Due to its strong deep modeling ability and high-fidelity sample generation ability, the GAN was widely used for sample synthesis in scene classification [30], [37], [38], object detection [31], semantic segmentation [39], and many other remote sensing interpretation tasks.
2) *Sample migration:* The goal of sample migration is to learn suitable embedding functions between the samples of source domain and target domain and constrain them to share the similar distributions characteristics. Thus, the existing dataset equipped with sufficient labeled samples, which may be collected for other types of interpretation tasks and possess large visual differences with the target domain, can be utilized to join the training process of current few-shot learning tasks after aligning. Different from the traditional techniques [40]–[42], which usually require manually assigning criterions on the distribution similarities and designing mapping rules across domains, GANs are capable of automatically modeling the complex embedding relations across domains through the adversarial training. Wang *et al.* [43] used the simulation results as a conditional input source of GAN to generate SAR samples by modeling embedding functions between simulation samples and authentic samples. Benjdira *et al.* [44] used CycleGAN [45] to reduce the domain gap caused by sensor variations for the task of semantic segmentation on aerial imagery. Furthermore, considering the problem of conversion inefficiency in the existing methods, Tasar *et al.* [46] designed a color mapping GAN to translate samples by learning elementwise matrix multiplication and addition, which gains much improvement in terms of accuracy and speed.

TABLE V
COMPARISON OF DIFFERENT PRIOR-KNOWLEDGE-BASED LEARNING METHODS

Methods	Overview	Pros	Cons	Typical References
Transfer learning	Perform basic image manipulations over original data	Easy implemenation and higher learning efficiency	Require the source domain dataset high relatedness with the target domain data	[48], [49]
Metric learning	Embed the samples to the an optimal metric space and conduct learning in an similarity-based contrast manner	Universally applicable for multiple tasks of differnet classes after training without quadratic adaptaton	Performance degradation when dealing with novel tasks which exhibits large distribution discrepancy with the training set	[50], [51]
Meta learning	Learn on the level of tasks and accumulate the task-agnostic knowledge to enhance the learning	Powerful generalization ability across classes and fast adaptation speed	Require the support of an auxiliary dataset with diverse object classes	[52], [53]

In contrast to the data-warping- and simulation-based augmentation methods mentioned above, deep generative models can automatically generate a large number of samples without human labor on the designing warping rules or computing models. Besides, generated samples are ensured to retain similar distributions to the original dataset, which can avoid the domain shift problem. In addition to the strengths mentioned above, yet there still exists several challenges for deep-generative-model-based methods that remain to be solved. Especially, considering the intraclass diversity characteristic of remote sensing images, the fidelity of the generated images remains to be improved when dealing with very high solution remote sensing imagery with complex background information. Another prerequisite should be taken into account that some more regularization techniques should be explored to alleviate the mode collapse problem [47] and improve the stability during the training of the deep generative networks.

IV. PRIOR-KNOWLEDGE-BASED LEARNING METHOD

Human's visual interpretation of remote sensing images depends on the accumulation of empirical knowledge. Similarly, as for the machine, few-shot learning on the current task can be established based on the empirical knowledge obtained from the previous tasks, which is called the prior-knowledge-based learning method. Specifically, the prior knowledge obtained from the auxiliary dataset, which indicates the prelearned knowledge of parts and relations and may exist in various forms (such as parameter initializations, pre-extracted features, etc.), is utilized to assist the current learning tasks by designing reasonable learning strategies. In this section, as shown in Table V, the prior-knowledge-based few-shot learning methods that emerged recently are divided into three categories: the transfer-learning-based-method, the metric-learning-based method, and the meta-learning-based method.

A. *Transfer-Learning-Based Method*

The main idea of transfer learning is to improve a learner in the target domain by transferring the knowledge from a related source domain [54]. The target domain means the target dataset to be interpreted, where only a small number of labeled samples are available, and the source domain corresponds to an auxiliary dataset, where labeled samples are sufficient. Based on the relatedness across the domains, transfer learning can effectively reduce the amount of data required on the target domain. Generally, the existing transfer learning methods designed for few-shot remote sensing image interpretation can be divided into two categories based on the concrete operation level of the transference.

1) Model-Based Transfer Learning: Assuming that the source tasks and the target tasks share some parameters or prior distributions of the models [55], the model based transfer learning aims to transfer the knowledge by reusing the model trained on the source task to the training of target task. Generally, there exist two kinds of strategies on model reuse including fixing the partial parameters of the pretrained model as feature extractors or using them as model initialization in the target domain. Both of them can effectively reduce the need for the number of training samples and accelerate the training process. In terms of the source to obtain the pretrained model, several typical solutions are listed below.

1) *Pretrain the model based on a similar dataset:* The ideal condition to obtain the pretrained model is from a dataset, which is highly correlated with the target data. Taking the building segmentation, for example, remote sensing images collected at different time instants, different sensors, or even different geographic locations can be utilized as the source domain dataset to pretrain the model. For some cases where no auxiliary dataset is available, another alternative paradigm is to pretrain the model from the synthetic images by additional simulation experiments [33], [35].
2) *Pretrain the model from an unlabeled dataset:* This method corresponds to the situation where only unlabeled data are available. The model can be pretrained from the unlabeled data by conducting unsupervised learning. Huang *et al.* [48] investigated the effects of this method in SAR target recognition. Specifically, they designed an assembled CNN architecture with the reconstruction pathway integrated and pretrained the network with the stacked convolutional autoencoders. Experiments demonstrated that this method can lead to improving the performance under small labeled sample conditions.

2) Feature-Based Transfer Learning: As shown in Fig. 4, the intuitive idea of this approach is to select and learn the feature representations that are generally suitable for both the source domain and the target domain. Thus, the knowledge can be transferred in the form of shared feature representation through

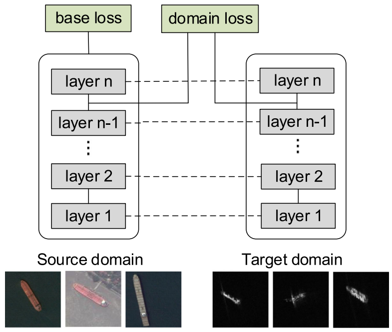


Fig. 4. Illustration of feature-based transfer learning.

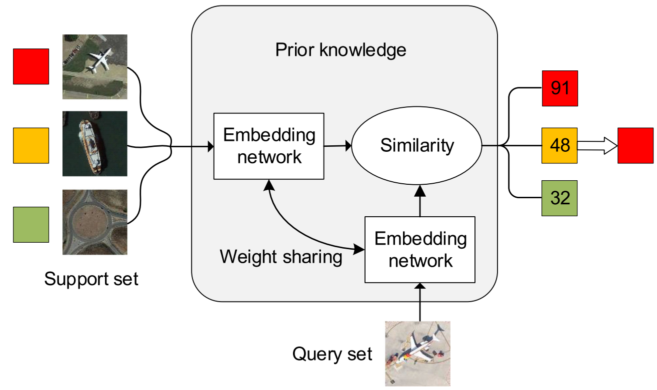


Fig. 5. Illustration of metric-learning-based few-shot learning methods.

designing specific strategies to align the feature representations across domains. By making the model focus on learning the shared feature representation in the source domain, the need for labeled training samples in the target domain can be reduced. The core issue of the feature-based transfer learning lies in the way of finding and learning the shared feature representations for the model across different domains. Bruzzone and Persello [56] proposed an approach to select the spatial invariant features for hyperspectral image classification. Specifically, a multiobjective to evaluate the feature discrimination capability as well as the feature shift across domains is explicitly optimized. The commonly adopted criteria for measuring domain distributions include Bhattacharyya distance [57], Jeffries–Matusita distance [58], maximum mean discrepancy (MMD) [59], and multikernel MMD [60]. Rostami *et al.* [49] designed a network with two deep encoders that were coupled to transfer the electrooptical (EO) domain knowledge to the SAR domain. Based on the sliced Wasserstein distance (SWD) [61], the samples were mapped to a shared embedding space and aligned class-conditionally. As a result, the model trained in the EO domain can be well generalized in the SAR domain. However, selecting and designing criteria for distribution measurement heavily depend on the laboratory experiments. Considering this disadvantage, Xu *et al.* [62] utilized the adversarial training technique to automatically select and learn the feature representation that is applicable across the domains. Concretely, a domain classifier was added to the network to distinguish the domain of input features. In addition, the model was trained to generate the feature representations that were indistinguishable enough across the domain to cheat the domain classifier. As a result, feature representations shared across the domains were learned.

Transfer learning provides a desirable paradigm to deal with the small-sample problem for remote sensing image interpretation. However, before applying transfer learning techniques, one prerequisite should be taken into account that a source domain dataset is required, which has sufficient labeled information and high relatedness with the target domain. Transferring knowledge from unrelated source data can gain little improvement and sometimes even lead to a negative impact on the target learner. To avoid this problem, the relevant research on the measurement of the transferability across the remote sensing datasets from multisources should be further studied.

B. *Metric-Learning-Based Method*

The metric-learning-based method aims to find an optimal metric space, where the distribution of similar samples is compact, while the distribution of dissimilar samples is alienated. Specifically, as illustrated in Fig. 5, a set of project functions are learned to encode the samples into lower dimensional feature embeddings. Combined with the suitable similarity measurement for feature embeddings, the samples can be easily classified in the metric space. This method can be interpreted as training the model to "learn to compare" and is capable of obtaining a good performance when dealing with a small number of samples.

Koch *et al.* [63] used the Siamese neural networks [64] as the feature extractor for few-shot learning. This two-branch architecture learned a common project function for both training samples and test samples and was proved to be suitable for the paired metric comparison. Vinyals *et al.* [65] designed the episode mechanism to mimic few-shot learning tasks by repeated subsampling classes and samples. This method can effectively improve the model's generalization ability and is widely used as the standard data organization strategy for few-shot learning. Based on the prior works mentioned above, Snell *et al.* [66] further assumed that there exists a prototype representation clustered by sample points within each category, and thus, test samples can be recognized with a fixed-nearest-neighbor classifier over the test embeddings and prototypes. Comparing this idea with the 3-D-CNN-based [67] spatial spectrum feature extraction mechanism, Tang *et al.* [50] applied the metric learning to few-shot scene classification in the hyperspectral images. The commonly used distance metric criteria for nearest-neighbor classification algorithms include Euclidean distance, cosine similarity, Manhattan distance, etc. Furthermore, Rao *et al.* [51] utilized a parameterized classifier to perform adaptive metric learning on the distance metric criterion, which effectively alleviated the defects of manually selecting metric criterion.

The main advantage of this framework lies in its simplicity and generalization ability. Concretely, a trained model can be directly applied to various kinds of learning tasks simultaneously without fine-tuning. This characteristic is based on the assumption that novel tasks share similar distribution with the previous learning tasks during training stage. Nevertheless, for new learning tasks with large discrepancy to previous learning

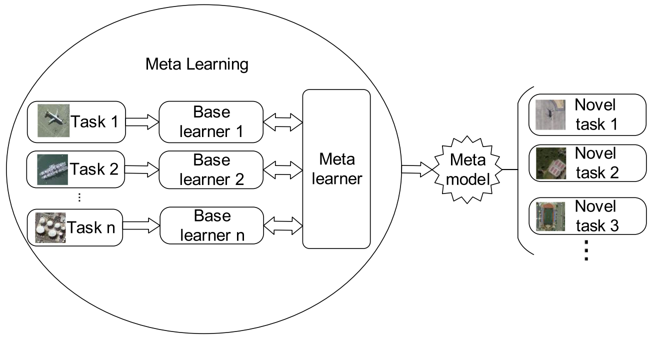


Fig. 6. Illustration of the meta-learning framework.

tasks, the generalization ability of the model can be greatly decreased. Considering the intraclass diversity and scale variations in remote sensing imagery, the techniques on enhancing the model's within-class generalization ability remain to be further studied.

C. Meta-Learning-Based Method

Based on the meta-learning, the model can automatically accumulate general knowledge during the process of continuous training. The learning ability of the model will be more proficient with more experience. With the assistance of the obtained general knowledge, the model is capable of fulfilling fast learning with a small amount of labeled data. The basic learning unit of meta-learning is composed of several few-shot learning tasks in a minibatch. In each task, the meta-train data, the meta-val data, and the meta-test data are randomly split without overlapping. As shown in Fig. 6, the meta-learning framework consists of several base learners and a meta-learner. The former base learners are expected to achieve quick knowledge learning for a single task. Then, the latter meta-learner is used to accumulate the general knowledge gradually, which can be shared across different tasks. This learning schema divides the process of learning into the task-specific adaptation and the cross-task generalization.

Ravi and Larochelle [68] analyzed the weakness of the traditional gradient update mechanism under small-sample conditions and proposed a network named Meta-LSTM for few-shot image classification. Meta-LSTM selected LSTM [69] as the meta-learner to provide the initial parameters shared among tasks. The base learner was implemented by a deep CNN classifier. The task-specific parameter update provided by the meta-learner is regarded as input. The base learner can realize quick convergence over multiple novel tasks. Different from the prior meta-learning methods that usually designed special networks to learn update functions, Finn *et al.* [52] proposed a model-agnostic meta-learning (MAML) algorithm. As a widely used meta-learning framework, MAML retains the compatibility of the traditional gradient-based learning mechanism and is applicable for any model architecture with the SGD optimization. The model parameters are meta-learned by MAML through the optimization so as to obtain the maximal accuracy within several iterations in the new tasks. Based on MAML, several improvements have been made recently. Li *et al.* [70] further proposed to train the learnable learning rates for each parameter in the base learner. Both the update direction and the step size of the model's parameters can be simultaneously optimized in a single meta-learning process, which significantly improved the automation and capacity of MAML. Lee *et al.* [71] investigated the linear classifiers as the base learner for MAML. Based on the linear classifier rule, the optimization objective in the inner loop is convex and can be effectively solved according to the theories of dual formulation and Karush–Kuhn–Tucker conditions. Alajaji and Alhichri [53] introduced MAML to the task of remote sensing scene classification and verified the effectiveness of MAML through a series of few-shot classification experiments.

Compared with the transfer-learning-based method, the meta-learning-based method can train a meta model applicable for multiple tasks with the characteristic of fast adaptation. This method avoids manually selecting the source domain data with high correlation in transfer learning. Furthermore, meta-learning can be integrated with multiple models for classification, regression, and reinforcement learning. Due to these advantages, meta-learning becomes a desirable paradigm for few-shot learning of remote sensing image interpretation. However, one limitation of this method is that meta-learning requires an auxiliary dataset that is rich in category diversity to ensure the generalization ability of the meta-model. Nevertheless, the existing publicly available remote sensing datasets only contain a small number of geospatial object categories. The high requirement of dataset seriously limits the application of meta-learning algorithms.

D. Other Methods

In addition to the three few-shot learning approaches mentioned above, there exist other complementary solutions for data scarcity problems. The lightweight design and structural optimization of models [72], [73] can be used to reduce the risk of overfitting by decreasing the amount of model parameters. Moreover, the self-taught learning [74], [75] and self-training [76] techniques can be utilized to conduct knowledge mining from unlabeled data and enhance the supervised learning process on the limited training set.

V. Typical Applications of Few-Shot Learning in Remote Sensing Image Interpretation

In this section, this article summarizes the applications of few-shot learning for remote sensing image interpretation, including scene classification, semantic segmentation, and object detection. The research status of application is introduced below from the perspective of experimental datasets, application cases, and evaluation metrics.

A. Experimental Datasets

This article collects and summarizes the existing published few-shot learning datasets for scene classification, semantic segmentation, and object detection. The final statistical results of experimental datasets are shown in Table IV. For detailed information about these datasets, the corresponding references listed in the tables can be referred. In terms of the interpretation task, the existing work mainly focuses on the study of scene

TABLE VI
SUMMARY OF PUBLIC DATASETS FOR EVALUATING FEW-SHOT LEARNING ALGORITHMS

Datasets	Type of data	Information of datasets					Applications	References
		Categories	Images	Resolution(m)	Bands	Image size		
UC Merced Land Use	Optical image	21	2100	0.3	RGB	256×256	Scene classification	[77]
WHU-RS19	Optical image	19	1013	up to 0.5	RGB	600×600 600×600	Scene classification	[78]
AID	Optical image	30	10000	0.5∼8	RGB	600×600	Scene classification	[79]
NWPU-RESISC45	Optical image	45	31500	0.2∼30	RGB	256×256	Scene classification	[80]
MSTAR	SAR image	10	5950	0.3	X-band	128×128	Military object recognition and classification	[81]
OpenSARShip	SAR image	3	11346	-	IW mode	-	Ship recognition	[82]
BUAA-SID1.0	Simulation image	20	9200	-	8-bit grayscale	320×240	Space object recognition	[83]
RIT-18	Hyperspectral image	18	3	0.047	6	9000×6000	Semantic segmentation	[35]
Indian Pines	Hyperspectral image	16	1	20	224	145×145	Semantic segmentation	[84]
Salinas	Hyperspectral image	16	1	3.7	224	512×217	Semantic segmentation	[84]
Pavia University	Hyperspectral image	9	1	1.3	103	610×610	Semantic segmentation	[84]
Pavia Centre	Hyperspectral image	9	1	1.3	102	1096×1096	Semantic segmentation	[84]
Kennedy Space Center	Hyperspectral image	13	1	18	224	512×614	Semantic segmentation	[85]
Chikusei	Hyperspectral image	19	1	2.5	128	2517×2335	Semantic segmentation	[86]
Botswana	Hyperspectral image	14	1	30	242	1476×256	Semantic segmentation	[85]

classification and semantic segmentation. For the object detection task, there is lack of publicly available datasets suitable for few-shot learning. In terms of the remote sensing data to be interpreted, the existing few-shot learning research appears unbalanced. The scene classification task mainly focuses on the optical images, and the semantic segmentation task focuses on the hyperspectral images.

B. Typical Applications of Remote Sensing Image Interpretation

1) Scene Classification: Scene classification, which aims to automatically assign remote sensing images with the predefined semantic labels, is of great importance for the comprehension of huge and complex remote sensing images and has been widely used in many fields such as urban planning, environment monitoring, and land resource management. Considering the cost of collecting and labeling large amounts of data, exploring practical scene classification algorithms under small-sample conditions has become an important research topic.

For the scene classification task in optical images, Marmanis *et al.* [87] designed a two-stage transfer learning framework to investigate the potential of using the large pretrained neural network for earth observation classification. The experimental results proved that this method can lead to a significant performance improvement as well as alleviating the overfitting problem. Since there exists a severe limitation on collecting sufficient annotation datasets in the field of remote sensing, Han *et al.* [88] presented a semisupervised generative framework, which combined the deep learning features, a self-label technique, and a discriminative evaluation method to finish the task of scene classification of high-resolution remote sensing images. As a result, the valuable knowledge can be effectively learned from unlabeled samples to improve the classification ability. Similarly, Yao *et al.* [89] proposed a local manifold constrained self-paced deep learning method. The model was trained in an incremental manner by gradually selecting and feeding the easy samples from generated pseudo samples. Meanwhile, to guarantee the credibility of pseudo samples, a local manifold constraint was introduced to ensure the consistency between the pseudo labels and the true labels. This technique can significantly improve the model accuracy and reduce the cost on manually labeling. In recent years, metric learning and meta-learning were investigated to address the small-sample problems and became an important focus in the field of remote sensing image interpretation. Cheng *et al.* [90] introduced a metric learning regularization term on the CNN features to make the model more discriminative and verified that the model in this method can obtain more information from the training data of smaller size. Yang *et al.* [91] presented a discriminative deep nearest-neighbor neural network for fine-grained few-shot space target recognition. They further introduced a center loss as an intraclass compactness principle to increase the feature robustness to the intraclass variation of space target. With limited space target samples, their algorithm was more efficient than the traditional few-shot learning methods and produced the state-of-the-art results. Zhai *et al.* [92] designed a lifelong few-shot learning model for remote sensing scene recognition. Based on the meta-learning framework, the model was able to recognize new classes using only a few labeled images. Besides, the knowledge learned from one dataset can be easily and rapidly applied to another dataset in spite of big disparities among images from various sources, which is of great significance for lifelong learning.

For the classification task in SAR images, the few-shot learning in this field is still dominated by the transfer-learning-based method. Zhang *et al.* [93] investigated the effect of the model-based transfer learning technique by fine-tuning the model pretrained on MSTAR for the task of ship recognition. Huang *et al.* [48] presented an assembled CNN to perform the knowledge transfer from the sufficient unlabeled SAR scene images to limited labeled SAR target data. Specifically, a large amount of unlabeled SAR data were used to train the reconstruction pathway with the stacked convolutional autoencoders. In addition to using the pretrained convolutional layers as part of the model initializations, they introduced a reconstruction loss as a regularization term to preserve the information of input images. Shang *et al.* [94] presented a memory CNN to record the spatial features of training samples and used the spatial similarity mechanism to predict unknown sample labels. Besides, a transfer parameter technique was utilized to solve the issue of nonconvergence during training. Rostami *et al.* [49] explored to conduct the knowledge transfer from optical images to SAR images. Based on the SWD criterion [95], the model was constrained to learn a shared invariant cross-domain embedding space. Experiments on ship classification validated that the trained classifier generalized well in both the optical domain and the SAR domain. Based on the metric learning paradigm, Tang *et al.* [96] proposed a Siamese network architecture for few-shot SAR target recognition and achieved a significant improvement in the inference time as well as the classification accuracy.

In general, the research of few-shot scene classification in remote sensing shows an evolutionary trend from specialization for a single task to generalization for multiple tasks. The prerequisite of few-shot learning methods also becomes more realistic with the adaptive transitions to the practical applications. Designing algorithms with higher learning efficiency and stronger generalization ability for small-sample datasets is an important direction for future research in this field.

2) Semantic Segmentation: In contrast to the scene classification task that assigns each image with a single semantic label, semantic segmentation aims to extract multiple classes of geospatial objects within a single image and finish the dense prediction task in pixel level. Under small-sample conditions, the main challenge of this task is to extract the discriminative features for multiple classes of targets effectively and ensure the strong generalization of the learned features.

For the semantic segmentation task of multi/hyperspectral image data, two main strategies have been adopted in the existing few-shot learning works. The first typical solution is to utilize the transfer learning technique to conduct knowledge transfer to augment the supervised experience. Kemker *et al.* [35] investigated using the simulation techniques to generate synthetic imagery as the source domain to provide the pretrained models for multispectral remote sensing images. The experimental results verified that this method can decrease the probability of model overfitting. Considering the issue of performance degradation caused by the domain gap, Zhou and Prasad [97] employed a feature-based alignment mechanism to realize the learned feature transfer from the source domain to the target domain. Furthermore, Luo and Ma [98] and Peng *et al.* [99] proposed to combine the MMD strategy and the manifold regularization to align per-class features in the mapped subspace as well as preserve the local manifold structure of data. Experimental results demonstrated that this combination can lead to a performance boost for unsupervised domain adaptation in hyperspectral images.

One necessary prerequisite for transfer learning techniques is that there exists a source domain dataset that has strong relevance to the target domain dataset. Many algorithms even require that the source domain has the same label space as the target domain since the alignment process is conducted class-by-class (e.g., ship samples of the source domain to those of the target domain), which will bring much inconvenience in practical applications. Considering this disadvantage, the metric-learning-based methods emerged. For the target categories that do not exist in the source domain, they can still realize fast learning under small-sample conditions. Zhang *et al.* [100] proposed to learn a generic metric space that was generalizable across domain and then generate the predicted labels by performing the nearest-neighbor algorithms in the metric space. Rao *et al.* [51] further suggested that using the convolution-based parameterized classifier as the metric criterion can lead to better performance than the nonparametric nearest-neighbor classification methods.

In terms of remote sensing data type that have been studied, the existing work related to few-shot segmentation mainly focuses on the interpretation of hyperspectral images, where abundant spectral information is provided. For other types of remote sensing data where only limited spectral information can be available (e.g., SAR images and optical images), the corresponding research of few-shot segmentation appears more challenging and remains uninvestigated. Although some methods have been proposed to deal with the small-sample problem in the remote sensing image semantic segmentation, the accuracy and generalization ability of the model remain to be improved.

3) Object Detection: Object detection is another challenging task for remote sensing image interpretation with the aim of localizing ground objects of interest and assigning the predefined category labels to the detected regions. In the research of few-shot object detection where only a few annotated samples are available, the main difficulty lies in avoiding the overfitting problems under small-sample conditions as well as learning discriminative features that are robust to the multiscale problems in remote sensing images.

For object detection tasks in optical images, Chen *et al.* [101] proposed to jointly use data augmentation and transfer learning techniques for aircraft detection. They emphasized the compatibility of rotation and affine transformation augmentation techniques for remote sensing object detection. Besides, VGG16 [102] pretrained on ImageNet was used for model initialization, which can effectively reduce the risk of overfitting problem caused by small samples. However, it still took some time for data augmentation and model training, which can lead to poor learning efficiency. Considering this problem, Zhang *et al.* [103] proposed a training-free one-shot geospatial object detection framework. Concretely, VGG16 pretrained on NWPU-RESISC45 was selected as the feature extractor and fixed to provide the remote sensing domain knowledge. Based on metric learning techniques, the similarity scores between query vectors and target features were computed by a series of convolutional operations to locate and classify the regions of interest in the target image. Experiments on sewage treatment

TABLE VII
EVALUATION METRICS FOR FEW-SHOT LEARNING ALGORITHMS

Indicators	Overview	Common scenes		
		Scene classification	Semantic segmentation	Object detection
Confusion Matrix	The records in the data set are summarized in a matrix form according to the real category and the classification judgment made by the classification model.	✓	✓	✓
Precision	Ratio of true positive (TP) and the total number of predicted positives.	✓	✓	✓
Recall	Ratio of true positive (TP) and the total number of ground truth positives.	✓	✓	✓
F1 score	A weighted average of precision and recall.	✓	✓	✓
Overall Accuracy	Reflect the probability that a sample will be correctly classified. Calculated by the sum of the true positives plus true negatives divided by the total number of individuals tested.	✓	✓	
Average Accuracy	The average calculation result of per-class accuracy.	✓	✓	
Kappa Coefficient	A statistical analysis to measure the agreement between existing classification results and random classification results.	✓	✓	
PR curve	A statistical graph with precision values on the y-axis and recall values on the x-axis.	✓	✓	✓
AP (Average Precision)	The corresponding area under the PR curve.			✓
mAP (mean Average Precision)	Statistically average the AP values of all classes to suppress evaluation deviation caused by class imbalance.			✓

plant and airport detection tasks verified the effectiveness of this method. The main advantage of this algorithm was that it unified the training process of small-sample detection tasks for various types of targets on the pretraining of the feature extractor and highly simplified the procedure of model deployment. In terms of detection tasks in SAR images, Wang *et al.* [104] presented a target detection framework based on the single-shot multibox detector [105] architecture. Two kinds of data augmentation methods suitable for SAR target detection was proposed to deal with the problem of insufficient labeled images. Besides, the subaperture decomposition technique was utilized to transform one-channel SAR images into three-channel images so that the pretrained network model on natural images can be reused in SAR target detection tasks. The experimental results on SAR vehicle detection tasks verified that this method can achieve fewer false alarms and higher accuracy.

At the present stage, the task of few-shot object detection for remote sensing images is still in its infancy, and there is a lack of public available remote sensing datasets for supporting the relevant research. The existing methods usually have some strict prerequisites on auxiliary datasets, which can be hardly satisfied in reality. These problems are worthy of future research.

C. Evaluation Metrics

In traditional learning tasks with sufficient training samples, the dataset is randomly split into the training set and the test set, and the specific evaluation metrics are obtained on the basis of the fixed test set, which is a widely used evaluation pipeline. However, this kind of evaluation method is not suitable for few-shot learning algorithms since data distribution appears extremely imbalanced between the training set and the test set in small-sample conditions. For example, in one split result of the original dataset, easy samples may concentratedly emerge in the training set, while in another split result, hard samples are allocated to the training set. This variability may decrease the persuasiveness of the evaluation results for measuring the models' few-shot learning ability. To make the evaluation results more reliable, one reasonable evaluation strategy is to repeatedly conduct the evaluation algorithms over multiple dataset split results and take the average of evaluation indexes as the final result. To help researchers conduct further studies and perform fair comparisons, this article also summarizes evaluation metrics that have been utilized in the existing few-shot learning works in Table VII.

VI. FUTURE WORKS

Considering the development of earth observation technique and research progress of few-shot learning, some unaddressed tasks as well as the promising future directions are suggested in this section.

1) *Build remote sensing datasets with more diversity in category and unify the experimental evaluation protocols:* Although various large-scale datasets have been constructed for the development of interpretation algorithms, the category information is limited, which can hardly satisfy the requirement of few-shot learning. On the one hand, the diversity of object category in the training set is of great importance since the model's learning ability under small-sample conditions is gained by repeated training over various kinds of learning tasks. On the other hand, based on the dataset with rich category information, the evaluation experiments can be conducted over more testing categories, which can prove the effectiveness of few-shot learning algorithms. Besides, in the existing works of few-shot learning for remote sensing image interpretation, the experiment settings as well as evaluation protocols appear disorganized, which block the fair comparison between different algorithms. Therefore, the construction of public experiment datasets and the unification of the standard evaluation criterions are essential for the future development of few-shot learning research.

2) *Improve the generalization ability of algorithms:* The existing data-augmentation-based approaches and the transfer-learning-based approaches are designed to apply to the specific learning tasks with fixed category. This setting is somewhat unrealistic and inefficient due to the diversity and variability of learning tasks in reality. Relatively, the metric learning and the meta-learning provide a more desirable paradigm with strong generalization ability as they can be applied over multiple categories. In the future, the model generalization can be strengthened from the following two aspects. In the situation of domain shift caused by the difference of imaging time, locations, and platforms, the performance of model learning needs to be robust enough under small-sample conditions. Additionally, a feasible few-shot learning paradigm can not only target at the specific object category with fixed types, but also generalize for multiple objects learning task with the fast adaptation ability.
3) *Enhance the robustness of algorithms to withstand the label noise:* Considering the impact of imaging quality and human annotation error, there usually exist some inaccurate labeled samples in the practical scenarios, especially for the densely prediction tasks such as object detection and semantic segmentation. However, most of the existing few-shot learning techniques lack the ability to deal with the noisy labels since they are originally developed based on the publicly benchmark dataset equipped with ideal imaging conditions and accurate labels. Due to the severe dependence on the precise supervisory information, the few-shot learning algorithms can be easily disturbed by the irrelevant noisy features and lead to poor learning effects. Consequently, it is necessary to improve the robustness of algorithms to learn from the coarsely labeled samples.
4) *Few-shot learning and zero-shot learning:* The existing algorithms usually define the small samples as hundreds of labeled samples. This experiment setting can be unrealistic sometimes. When only dozens of images can be gained, the number of training samples required by the algorithms is still large compared with the real applications. Moreover, the existing algorithms still suffer from the burden of data labeling, especially for some interpretation tasks with dense prediction requirements, such as semantic segmentation and object detection. Considering these issues, few-shot learning and zero-shot learning suitable for remote sensing images are promising directions as they are corresponding to the urgent need for decreasing the training samples for image interpretation.
5) *Expand the research field of remote sensing image interpretation under small-sample conditions:* Currently, the research of few-shot learning remains in the initial stage, and its distribution over target data type as well as application type appears unbalanced. Considering the applications of few-shot learning, the existing works mainly focus on the scene classification task, while other typical interpretation tasks, such as semantic segmentation and object detection, can be further explored.

VII. Conclusion

In this article, we gave a comprehensive overview of the recent research progress in few-shot learning for remote sensing image interpretation. Through a bibliometric analysis, we first presented a systematic review of the existing works related to this field. Subsequently, we categorized the existing methods into two groups, including data-augmentation-based methods and prior-knowledge-based methods, and described them, respectively. Besides, we also introduced the application of few-shot learning for remote sensing image interpretation by summarizing and listing the experimental datasets, application cases, and evaluation metrics. Finally, we discussed the challenges of current studies and gave some promising research directions in future. Generally, the few-shot learning for remote sensing image interpretation is in its infancy and is of great need for further improvements. This survey can be beneficial for the researchers to better understand this research field.

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 3431–3440.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1106–1114.

[3] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 580–587.

[4] G. V. Pérez, C. Q. Camargo, and A. A. Louis, "Deep learning generalizes because the parameter-function map is biased towards simple functions," in *Proc. 7th Int. Conf. Learn. Represent.*, New Orleans, LA, USA, 2019, pp. 1–35.

[5] R. Caruana, S. Lawrence, and C. L. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Denver, CO, USA, 2000, pp. 402–408.

[6] R. Tao, X. Zhao, W. Li, H. Li, and Q. Du, "Hyperspectral anomaly detection by fractional fourier entropy," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.

[7] X. Zhao *et al.*, "Joint classification of hyperspectral and lidar data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.

[8] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogrammetry Remote Sens.*, vol. 159, pp. 296–307, 2020.

[9] G. Cheng, X. Xie, J. Han, L. Guo, and G. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020.

[10] F. Xu, C. Hu, J. Li, A. Plaza, and M. Datcu, "Special focus on deep learning in remote sensing image processing," *Sci. China Inf. Sci.*, vol. 63, 2020, Art. no. 140300.

[11] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, "FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–19, 2020.

[12] Q. Song and F. Xu, "Zero-shot learning of SAR target feature space with deep generative neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2245–2249, Dec. 2017.

[13] Q. Song, H. Chen, F. Xu, and T. J. Cui, "EM simulation-aided zero-shot learning for SAR automatic target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1092–1096, Jun. 2020.

[14] I. J. Goodfellow, Y. Bengio, and A. C. Courville, *Deep Learning, (ser. Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: MIT Press, 2016.

[15] F. Li, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006.

[16] M. Fink, "Object classification from a single example utilizing class relevance metrics," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2004, pp. 449–456.

[17] Q. Sun, Y. Liu, T. Chua, and B. Schiele, "Meta-transfer learning for few-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, 2019, pp. 403–412.

[18] R. Hou, H. Chang, B. Ma, S. Shan, and X. Chen, "Cross attention network for few-shot classification," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2019, pp. 4005–4016.

[19] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell, "Few-shot object detection via feature reweighting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Seoul, South Korea, 2019, pp. 8419–8428.

[20] Q. Fan, W. Zhuo, C. Tang, and Y. Tai, "Few-shot object detection with attention-RPN and multi-relation detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 2020, pp. 4012–4021.

[21] A. Shaban, S. Bansal, Z. Liu, I. Essa, and B. Boots, "One-shot learning for semantic segmentation," in *Proc. Brit. Mach. Vis. Conf.*, London, U.K., 2017, pp. 167.1–167.13.

[22] N. Dong and E. P. Xing, "Few-shot semantic segmentation with prototype learning," in *Proc. Brit. Mach. Vis. Conf.*, Newcastle, U.K., 2018, p. 79.

[23] C. Chen, "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 57, no. 3, pp. 359–377, 2006.

[24] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep CNN," *IEEE Geosci. Remote. Sens. Lett.*, vol. 16, no. 4, pp. 593–597, Apr. 2019.

[25] J. Nalepa, M. Myller, and M. Kawulok, "Training- and test-time data augmentation for hyperspectral image segmentation," *IEEE Geosci. Remote. Sens. Lett.*, vol. 17, no. 2, pp. 292–296, Feb. 2020.

[26] X. Yu, X. Wu, C. Luo, and P. Ren, "Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework," *GISci. Remote Sens.*, vol. 54, no. 5, pp. 741–758, 2017.

[27] F. Kong, B. Huang, K. Bradbury, and J. M. Malof, "The Synthinel-1 dataset: A collection of high resolution synthetic overhead imagery for building segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Snowmass Village, CO, USA, 2020, pp. 1803–1812.

[28] A. Kusk, A. Abulaitijiang, and J. Dall, "Synthetic SAR image generation using sensor, terrain and target models," in *Proc. 11th Eur. Conf. Synth. Aperture Radar,* 2016, pp. 1–5.

[29] Y. Yan, Z. Tan, and N. Su, "A data augmentation strategy based on simulated samples for ship detection in RGB remote sensing images," *ISPRS Int. J. Geo Inf.*, vol. 8, no. 6, 2019, Art. no. 276.

[30] J. Guo, B. Lei, C. Ding, and Y. Zhang, "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1111–1115, Jul. 2017.

[31] J. Howe, K. Pula, and A. A. Reite, "Conditional generative adversarial networks for data augmentation and adaptation in remotely sensed imagery," *Proc. SPIE*, vol. 11139, 2019, Art. no. 111390G.

[32] K. Zheng, M. Wei, G. Sun, B. Anas, and Y. Li, "Using vehicle synthesis generative adversarial networks to improve vehicle detection in remote sensing images," *ISPRS Int. J. Geo Inf.*, vol. 8, no. 9, 2019, Art. no. 390.

[33] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geosci. Remote. Sens. Lett.*, vol. 14, no. 9, pp. 1484–1488, Sep. 2017.

[34] K. Wang, G. Zhang, and H. Leung, "SAR target recognition based on cross-domain and cross-task transfer learning," *IEEE Access*, vol. 7, pp. 153 391–153399, 2019.

[35] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 60–77, 2018.

[36] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote. Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.

[37] J. Wang, J. Li, B. Sun, and Z. Zuo, "SAR image synthesis based on conditional generative adversarial networks," *J. Eng.*, vol. 2019, no. 21, pp. 8093–8097, 2019.

[38] Y. Zhang, H. Sun, J. Zuo, H. Wang, G. Xu, and X. Sun, "Aircraft type recognition in remote sensing images based on feature learning with conditional generative adversarial networks," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1123.

[39] H. Gao *et al.*, "A hyperspectral image classification method based on multi-discriminator generative adversarial networks," *Sensors*, vol. 19, no. 15, 2019, Art. no. 3269.

[40] S. Inamdar, F. Bovolo, L. Bruzzone, and S. Chaudhuri, "Multidimensional probability density function matching for preprocessing of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 46, no. 4, pp. 1243–1252, Apr. 2008.

[41] D. Tuia, J. Muñoz-Marí, L. Gómez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote. Sens.*, vol. 51, no. 1, pp. 329–341, Jan. 2013.

[42] A. A. Nielsen and M. J. Canty, "Kernel principal component and maximum autocorrelation factor analyses for change detection," in *Proc. SPIE*, vol. 7477, 2009, Art. no. 74770T.

[43] G. Wang, G. Dong, H. Li, L. Han, X. Tao, and P. Ren, "Remote sensing image synthesis via graphical generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, 2019, pp. 10027–10030.

[44] B. Benjdira, A. Ammar, A. Koubaa, and K. Ouni, "Data-efficient domain adaptation for semantic segmentation of aerial imagery using generative adversarial networks," *Appl. Sci.*, vol. 10, no. 3, 2020, Art. no. 1092.

[45] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, 2017, pp. 2242–2251.

[46] O. Tasar, S. L. Happy, Y. Tarabalka, and P. Alliez, "Colormapgan: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7178–7193, Oct. 2020.

[47] D. Bau *et al.*, "Seeing what a GAN cannot generate," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 4502–4511.

[48] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote. Sens.*, vol. 9, no. 9, 2017, Art. no. 907.

[49] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, "Deep transfer learning for few-shot SAR image classification," *Remote. Sens.*, vol. 11, no. 11, 2019, Art. no. 1374.

[50] H. Tang, Y. Li, X. Han, Q. Huang, and W. Xie, "A spatial-spectral prototypical network for hyperspectral remote sensing image," *IEEE Geosci. Remote. Sens. Lett.*, vol. 17, no. 1, pp. 167–171, Jan. 2020.

[51] M. Rao, P. Tang, and Z. Zhang, "Spatial-spectral relation network for hyperspectral image classification with limited training samples," *IEEE J. Sel. Top. Appl. Earth Observ. Remote. Sens.*, vol. 12, no. 12, pp. 5086–5100, Dec. 2019.

[52] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 1126–1135.

[53] D. Alajaji and H. Alhichri, "Few shot scene classification in remote sensing using meta-agnostic machine," in *Proc. 6th Conf. Data Sci. Mach. Learn. Appl.*, 2020, pp. 77–80.

[54] K. R. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, 2016, Art. no. 9.

[55] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[56] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Trans. Geosci. Remote. Sens.*, vol. 47, no. 9, pp. 3180–3191, Sep. 2009.

[57] J. A. Richards and J. Richards, *Remote Sensing Digital Image Analysis*, vol. 3. Berlin, Germany: Springer, 1999.

[58] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Amsterdam, The Netherlands: Elsevier, 2013.

[59] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474v1*.

[60] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, 2015, pp. 97–105.

[61] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 214–223.

[62] S. Xu, X. Mu, X. Zhang, and D. Chai, "Unsupervised remote sensing domain adaptation method with adversarial network and auxiliary task," *Acta Geodaetica Et Cartographica Sinica*, vol. 46, no. 12, pp. 1969–1977, 2017.

[63] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, 2015, vol. 2, pp. 1–8.

[64] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a "Siamese" time delay neural network," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 1994, pp. 737–744.

[65] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Barcelona, Spain, 2016, pp. 3630–3638.

[66] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 4077–4087.

[67] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[68] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. 5th Int. Conf. Learn. Represent.*, Toulon, France, 2017, pp. 1–11.

[69] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[70] Z. Li, F. Zhou, F. Chen, and H. Li, "Meta-SGD: Learning to learn quickly for few shot learning," 2017, *arXiv:1707.09835v2*.

[71] K. Lee, S. Maji, A. Ravichandran, and S. Soatto, "Meta-learning with differentiable convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA,2019, pp. 10657–10665.

[72] B. Pan, Z. Shi, and X. Xu, "R-VCANet: A new deep-learning-based hyperspectral image classification method," *IEEE J. Sel. Top. Appl. Earth Observ. Remote. Sens.*, vol. 10, no. 5, pp. 1975–1986, May 2017.

[73] P. Bin, S. Zhenwei, and X. Xia, "MugNet: Deep learning for hyperspectral image classification using limited samples," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 108–119, 2018.

[74] R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote. Sens.*, vol. 55, no. 5, pp. 2693–2705, May 2017.

[75] R. Kemker, R. Luu, and C. Kanan, "Low-shot learning for the semantic segmentation of remote sensing imagery," *IEEE Trans. Geosci. Remote. Sens.*, vol. 56, no. 10, pp. 6214–6223, Oct. 2018.

[76] Y. Wu, G. Mu, C. Qin, Q. Miao, W. Ma, and X. Zhang, "Semi-supervised hyperspectral image classification via spatial-regulated self-training," *Remote. Sens.*, vol. 12, no. 1, 2020, Art. no. 159.

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

[78] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître, "Structural high-resolution satellite image indexing," in *Proc. Adv. Remote Sens. Sci.*, Vienna, Austria, 2010, pp. 298–303.

[79] G. 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.

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

[81] S. N. Laboratory, "Air force moving and stationary target recognition database," [EB/OL]. Accessed: Nov. 1, 2020. [Online]. Available: http://www.sdms.afrl.af.mil/datasets/mstar/

[82] L. Huang *et al.*, "Opensarship: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote. Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018.

[83] H. Zhang, Z. Liu, Z. Jiang, M. An, and D. Zhao, "BUAA-SID1. 0 space object image dataset," *Spacecr. Recovery Remote Sens.*, vol. 34, no. 4, pp. 69–75, 2010.

[84] M. Graña, M. Veganzons, and B. Ayerdi, "Hyperspectral remote sensing scenes," [EB/OL]. Accessed: Feb. 3, 2021. [Online]. Available: http://www.ehu.eus/ccwintco/index.php/

[85] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote. Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[86] S. Mahesh, D. Jayas, J. Paliwal, and N. White, "Hyperspectral imaging to classify and monitor quality of agricultural materials," *J. Stored Products Res.*, vol. 61, pp. 17–26, 2015.

[87] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using imagenet pretrained networks," *IEEE Geosci. Remote. Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016.

[88] 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, pp. 23–43, 2018.

[89] X. Yao, L. Yang, G. Cheng, J. Han, and L. Guo, "Scene classification of high resolution remote sensing images via self-paced deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 521–524.

[90] 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.

[91] X. Yang, X. Nan, and B. Song, "D2N4: A discriminative deep nearest neighbor neural network for few-shot space target recognition," *IEEE Trans. Geosci. Remote. Sens.*, vol. 58, no. 5, pp. 3667–3676, May 2020.

[92] M. Zhai, H. Liu, and F. Sun, "Lifelong learning for scene recognition in remote sensing images," *IEEE Geosci. Remote. Sens. Lett.*, vol. 16, no. 9, pp. 1472–1476, Sep. 2019.

[93] D. Zhang, J. Liu, W. Heng, K. Ren, and J. Song, "Transfer learning with convolutional neural networks for SAR ship recognition," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, 2018, vol. 322, Art. no. 0 72001.

[94] R. Shang, J. Wang, L. Jiao, R. Stolkin, B. Hou, and Y. Li, "SAR targets classification based on deep memory convolution neural networks and transfer parameters," *IEEE J. Sel. Top. Appl. Earth Observ. Remote. Sens.*, vol. 11, no. 8, pp. 2834–2846, Aug. 2018.

[95] A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein, "Scale space and variational methods in computer vision," in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis.*, 2012, pp. 373–385.

[96] J. Tang, F. Zhang, Y. Zhou, Q. Yin, and W. Hu, "A fast inference networks for SAR target few-shot learning based on improved siamese networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, 2019, pp. 1212–1215.

[97] X. Zhou and S. Prasad, "Deep feature alignment neural networks for domain adaptation of hyperspectral data," *IEEE Trans. Geosci. Remote. Sens.*, vol. 56, no. 10, pp. 5863–5872, Oct. 2018.

[98] C. Luo and L. Ma, "Manifold regularized distribution adaptation for classification of remote sensing images," *IEEE Access*, vol. 6, pp. 4697–4708, 2018.

[99] J. Peng, W. Sun, L. Ma, and Q. Du, "Discriminative transfer joint matching for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote. Sens. Lett.*, vol. 16, no. 6, pp. 972–976, Jun. 2019.

[100] C. Y. Zhang, J. Yue, and Q. Qin, "Deep quadruplet network for hyperspectral image classification with a small number of samples," *Remote. Sens.*, vol. 12, no. 4, 2020, Art. no. 647.

[101] Z. Chen, T. Zhang, and C. Ouyang, "End-to-end airplane detection using transfer learning in remote sensing images," *Remote. Sens.*, vol. 10, no. 1, 2018, Art. no. 139.

[102] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–14.

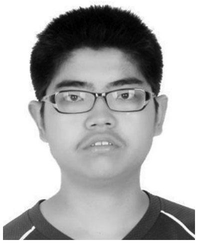
[103] T. Zhang *et al.*, "A training-free, one-shot detection framework for geospatial objects in remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, 2019, pp. 1414–1417.

[104] Z. Wang, L. Du, J. Mao, B. Liu, and D. Yang, "SAR target detection based on SSD with data augmentation and transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 150–154, Jan. 2019.

[105] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

Xian Sun (Senior Member, IEEE) received the B.Sc. degree in electronic and information engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2004, and the M.Sc. and Ph.D. degrees in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2009.

He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include computer vision, geospatial data mining, and remote sensing image understanding.



Bing Wang (Student Member, IEEE) received the B.Sc. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 2018. He is currently working toward the Ph.D. degree with the University of Chinese Academy of Sciences, Beijing, China, and the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing.

His research interests include computer vision, geospatial data mining, and remote sensing image understanding.

Zhirui Wang (Member, IEEE) received the B.Sc. degree in electronic information engineering from the Harbin Institute of Technology, Harbin, China, in 2013, and the Ph.D. degree in information and communication engineering from Tsinghua University, Beijing, China, in 2018.

He is an Assistant Researcher with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include synthetic aperture radar (SAR) terrain classification, and SAR target detection and recognition.

Hao Li (Member, IEEE) received the B.E. degree from Jilin University, Changchun, China, in 2014, and the M.Sc. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2017, all in information and communication engineering.

He is an Assistant Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include computer vision and remote sensing image processing.

Hengchao Li (Member, IEEE) received the B.Sc. and M.Sc. degrees from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008, all in information and communication engineering.

From 2013 to 2014, he was a Visiting Scholar with Prof. William J. Emery with the University of Colorado, Boulder, CO, USA. He is currently a Professor with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu. His research interests include the statistical analysis of synthetic aperture radar images, remote sensing image processing, and signal processing in communications.

Dr. Li received several scholarships or awards, especially including the Special Grade of the Financial Support from the China Postdoctoral Science Foundation, in 2009 and the New Century Excellent Talents in University from the Ministry of Education of China, in 2011. He is an Associate Editor for the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. He has been a Reviewer for several international journals and conferences, such as the IEEE Transactions on Geoscience and Remote Sensing, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, *IEEE Geoscience and Remote Sensing Magazine*, IEEE Transactions on Image Processing, IEEE Sensors Journal, *IET Radar*, *Sonar and Navigation*, *IET Signal Processing*, *IET Image Processing*, *Electronics Letters*, *Pattern Recognition*, *Signal Processing*, *International Journal of Remote Sensing*, *Remote Sensing Letters*, *Remote Sensing*, *Sensors*, *Journal of Applied Remote Sensing*, and *Canadian Journal of Remote Sensing*.

Kun Fu (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 1995, 1999, and 2002, respectively, all in information and communication engineering.

He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision, remote sensing image understanding, and geospatial data mining and visualization.