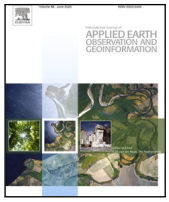
International Journal of Applied Earth Observation and Geoinformation 120 (2023) 103345

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Cost-efficient information extraction from massive remote sensing data: When weakly supervised deep learning meets remote sensing big data

Yansheng Li [a], Xinwei Li [a], Yongjun Zhang [a,*], Daifeng Peng [b], Lorenzo Bruzzone [c]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
[c] Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy

ARTICLE INFO

Keywords:
Remote sensing big data mining
Weakly supervised deep learning
Cost-efficient information extraction
Future research directions

ABSTRACT

With many platforms and sensors continuously observing the earth surface, the large amount of remote sensing data presents a big data challenge. While remote sensing data acquisition capability can fully meet the requirements of many application domains, there is still a need to further explore how to efficiently mine the useful information from remote sensing big data (RSBD). Many researchers in the remote sensing community have introduced deep learning in the process of RSBD, and deep learning-based methods have achieved better performance compared with traditional methods. However, there are still substantial obstacles to the application of deep learning in remote sensing. One of the major challenges is the generation of pixel-level labels with high quality for training samples, which is essential to deep learning models. Weakly supervised deep learning (WSDL) is a promising solution to address this problem as WSDL can utilize greedily labeled datasets that are easy to collect but not ideal to train the deep networks. In this review, we summarize the achievements of WSDL-driven cost-efficient information extraction from RSBD. We first analyze the opportunities and challenges of information extraction from RSBD. Based on the analysis of the theoretical foundations of WSDL in the computer vision (CV) domain, we conduct a survey on the WSDL-based information extraction methods under the data characteristic and task demand of RSBD in four different tasks: (i) scene classification, (ii) object detection, (iii) semantic segmentation and (iv) change detection. Finally, potential research directions are outlined to guide researchers to further exploit WSDL-based information extraction from RSBD.

1. Introduction

In recent decades, remote sensing techniques and platforms used to observe the earth surface have rapidly developed, with the exploration of airplanes, satellites, unmanned aerial vehicles, and so on Wu et al. (2021). As shown in Fig. 1, massive remote sensing data have been obtained in various spectral, spatial and temporal resolutions. Remote sensing data is fundamental for understanding the Earth (Zhang et al., 2016b) and lead to many important applications linked with biodiversity (Randin et al., 2020), humanitarian efforts (Schmieder et al., 2020), global climate change monitoring (Dirscherl et al., 2020) and so on.

It is apparent that remote sensing data already has the 4 V characteristics of big data, for instance, Velocity, Veracity, Variety and Volume (Zhang et al., 2019a). These characteristics in the remote sensing can be explained as follows: (1) *Volume*: remote sensing data are obtained from all over the world. The number of remote sensing data is huge and increases fast. For example, the Earth Science Data and Information System (ESDIS) in NASA, receives over 4TB of new data every day. Such huge volume brings big challenges to the storage, preprocessing and information extraction. (2) *Variety*: remote sensing data acquired with various imaging conditions (e.g., sensors, time, location, etc.) shows different characteristics. It is still challenging to utilize such diverse data. Generally, deep learning models trained on a specific dataset performs poorly on another dataset. (3) *Velocity*: the remote sensing data grow rapidly along with the continuous observations and increasing number of imaging sensors. In real applications, the timeliness of the applications of remote sensing demands the efficiency of processing and information extraction. (4) *Veracity*: inconsistence and incompleteness might exist in remote sensing data (Zhu et al., 2016; Zhang et al., 2019a). These characteristics make it very difficult to automatically and robustly extract valuable information from RSBD.

Despite the difficulties in information extraction, pioneers in remote sensing have achieved fruitful results. Benefited from long-accumulated image processing experience, the traditional methods with handcrafted feature descriptors are adopted to extract information from remote

* Corresponding author.
E-mail address: zhangyj@whu.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.jag.2023.103345
Received 31 July 2022; Received in revised form 24 April 2023; Accepted 5 May 2023
Available online 19 May 2023

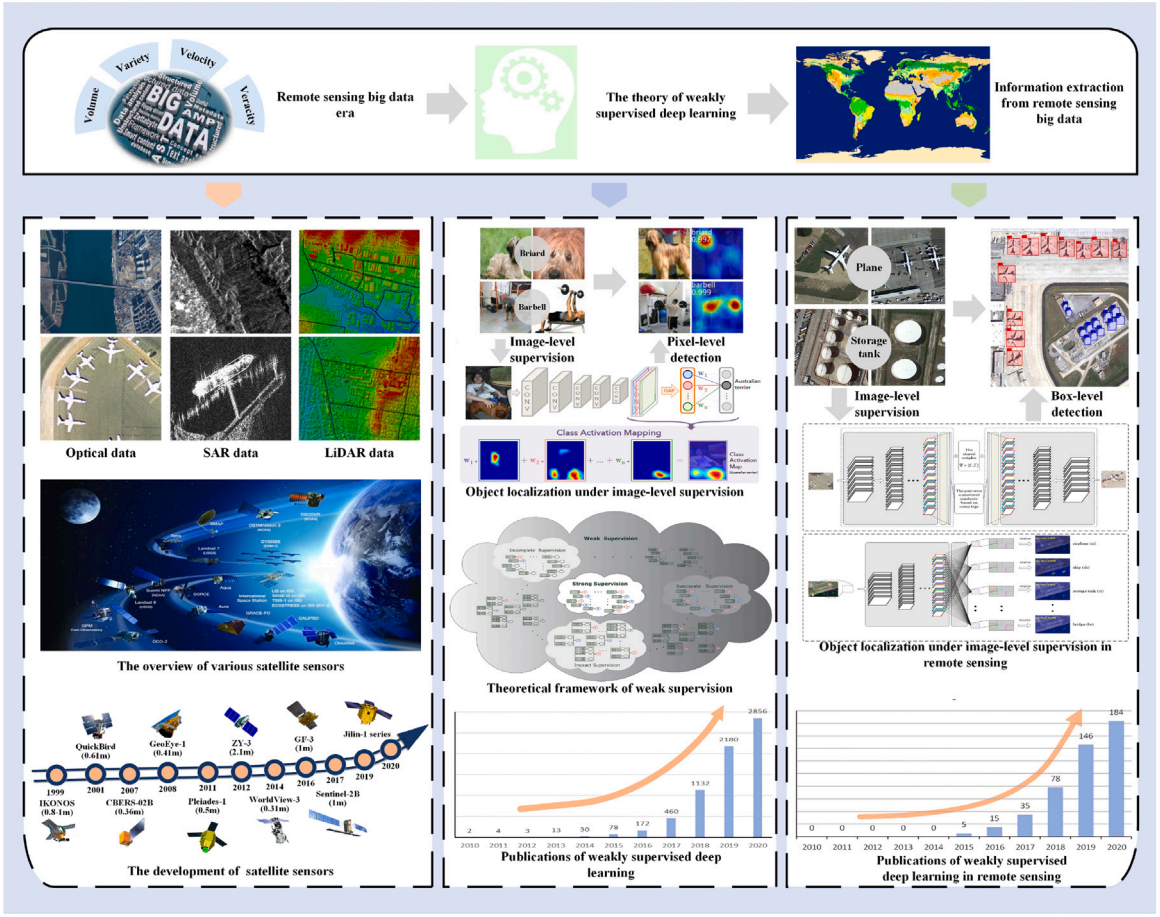


Fig. 1. Information extraction from RSBD in weakly supervised manner.

sensing images. However, these feature descriptors are only effective in some limited situations (Benediktsson et al., 2003; Jia et al., 2013), and often fail to achieve an optimal balance between the generalization and robustness especially when coping with RSBD (Zhang et al., 2016b). Thanks to the rapid development of computing power and big data with high-quality labels, deep learning has achieved great success and surpassed traditional methods and even outperformed human ability in many tasks (Zhu et al., 2017; Silver et al., 2016). Nowadays, deep learning methodologies are ubiquitous in remote sensing studies, such as change detection (Huang et al., 2019; Wang et al., 2018), object detection (Deng et al., 2018; Chen et al., 2019b), scene classification (Tong et al., 2020a; Xue et al., 2020) and semantic segmentation (Kemker et al., 2018; Diakogiannis et al., 2020).

As it is well known, high quality pixel-level labels are essential to the performance of deep learning. Taking the task for semantic segmentation as one example, a label map indicating the class of each pixel is needed for every training image. Obviously, creating such detailed labels is very time-consuming, especially for images that cover large areas and contain many objects. In the era of RSBD, the exhaustive annotation of massively increasing remote sensing data has become impossible. As a consequence, how to train deep networks in a cost-efficient manner becomes a practical issue to extract information from RSBD. In the literature, one promising solution is to leverage easily collected weak supervision information for training deep networks. Compared with full supervision, weak supervision refers to incomplete, inaccurate and inexact labels (Zhou, 2018) which is explained in detail in Section 3. Exploiting weak supervision can significantly reduce the cost of labeling. For instance, image-level ones can be used instead of pixel-level labels to train deep models for semantic segmentation tasks. Obviously, it is easier to acquire image-level labels. As summarized in Fig. 1, the number of papers about this technology is increasing each year. Many methods have been proposed to learn robust models under weak supervision. However, a performance gap between fully and weak supervision is still inevitable.

In this paper, we focus on WSDL-based cost-efficient information extraction from RSBD. First, we illustrate the challenges and opportunities in information extraction from RSBD. Second, we introduce the concept of WSDL. Third, we provide a detailed review of WSDL's achievements in four tasks of remote sensing: (i) scene classification, (ii) object detection, (iii) semantic segmentation, and (iv) change detection. Finally, we summarize some potential research directions to guide future research.

2. Information extraction from remote sensing big data: opportunities and challenges

2.1. Opportunities in information extraction from remote sensing big data

Due to the emergence of new technologies, there are many opportunities for information extraction from RSBD. Traditional methods are not able to extract all the information hidden behind big data due to the 4 V characteristics of volume, variety, velocity and veracity (Zhang et al., 2019a). Impressive performance has been achieved (LeCun et al., 2015) in natural language processing (Otter et al., 2020), natural image processing (Jiao and Zhao, 2019), and other areas. Many researches have used deep learning to process remote sensing images and have achieved considerable results (Zhang et al., 2016b). Due to the large volume of remote sensing data, deep learning technology can automatically learn the feature extraction and classification models from RSBD instead of requiring hand-crafted feature operators (Zhu et al., 2017). Moreover, the large amount of data can help to avoid overfitting.

Another opportunity in information extraction from RSBD is the variety of remote sensing data. There are different types of remote sensing data (i.e., multimodal data), such as optical imagery, multispectral imagery, and synthetic aperture radar (SAR) imagery. Multimodal data comes from different sensors and has diverse imaging principles (Zhang et al., 2021c; Ma et al., 2022; Tang et al., 2022). Therefore, fusing multimodal data can address situations where single modal data is insufficient (Li et al., 2023c). For instance, optical imagery cannot clearly capture the Earth's surface when there are clouds. However, SAR is an all-weather sensor that can collect ground information regardless of cloud coverage. The fusion of SAR and optical data complements each other's deficiencies (Li et al., 2023c).

Two crucial variables play key roles in information extraction from RSBD: storage and computation. The increasing volume of remote sensing data has raised the problem of realizing efficient storage and computation. It is impossible to load such a volume of data on the local memory (Wu et al., 2021). Fortunately, the modern distributed platforms provide the capacity of dealing with storage and processing of RSBD by distributing the heavy processing loads. Cloud computing is currently the most efficient method to process RSBD. With the developments of techniques for storing, computing and understanding RSBD, these data have been successfully applied to address many real-world problems (Chi et al., 2016), including forest monitoring at a global scale (Crowther et al., 2015), land use (Martinuzzi et al., 2007), urban planning (Deng et al., 2008; Bhatta, 2010) and so on.

2.2. Challenges in information extraction from remote sensing big data

Despite the aforementioned opportunities, there are still some challenges that hinder the information extraction from RSBD. In recent years, image retrieval from RSBD has made a lot of progress (Li et al., 2021f, 2018c,a). Compared with image retrieval, how to effectively extract information from RSBD is a more challenging problem as the information extraction needs to solve complicated classification problems and even object location. The deep learning-based methods require not only many remote sensing images but also a sufficient number of high-quality labels to supervise the training of deep learning models. This brings the following challenges to the information extraction from RSBD:

(1) Obtaining labels for the huge volume of remote sensing data requires lots of time and labor. Utilizing the massively existing land cover products or the crowdsourcing way to generate labels seems to be a promising solution. However, the quality of such data labels cannot always be guaranteed.

(2) The high velocity of RSBD means that the existing labels can be out of date very fast and need to be updated frequently. Considering the big volume of data, updating such a large volume of data is often unaffordable. Furthermore when an urgent task emerges, timely collecting sufficient images and reliable labels is quite difficult, which limits the practical applications of RSBD.

(3) Big variety of RSBD makes it more challenging to extract information from images. Images from different sensors, places and time have diverse features. When applying the deep learning models trained on one type of data to another, the performance might decrease significantly.

In the era of RSBD, the challenges mentioned above make it difficult to extract information in remote sensing. Effectively training deep networks under weak supervision is an important way to meet the challenges of information extraction from RSBD.

Except for the lack of labels, the end-to-end large-size remote sensing image interpretation also poses a challenge for information extraction. Normally, a remote sensing image has a large field-of-view, which help collecting more information and interpreting images (Li et al., 2023a,b). However, limited by the GPU memories, the most of deep-learning-based methods of information extraction fail to process one remote sensing image with large size holistically (Li et al., 2023a). One whole large-size image generally is cut into small patches and process on such small one, which might prevent the deep-learning model from complete information and leads to incorrect results of information extraction. As mentioned above, the performance of deep learning models suffers from the lack of high-quality labels. Embedding prior knowledge into the deep learning training process can be a promising solution to it, which has more interpretability compared with totally data-driven methods (Li et al., 2021c, 2022b,a). But how to sufficiently employ the prior knowledge in the deep networks still need more research.

3. Weakly supervised deep learning: basic concepts

In general, for deep learning methods, the neural networks learn weights from the training datasets guided by the validation datasets, and then predict the results on the test ones. These datasets consist of two parts, namely images and labels, and the type of labels vary greatly for different tasks. For example, in the scene classification task, the label indicates the class of the whole scene, whereas in semantic segmentation task, the label typically is a pixel-level classification map. In deep learning methods, each sample has a label and label should be error-free and intact. Only with such high-quality labels the deep learning methods can result in outstanding performance. However, as already mentioned, obtaining such desirable datasets is not easy. This motivates studying how to perform deep learning under weak supervision information for little loss of accuracy. According to the review in Zhou (2018), WSDL mainly uses three types of weak supervision information: inexact supervision, incomplete supervision and inaccurate supervision. Fig. 2 provides an illustration of these three types of weak supervision information and examples in the image processing field. The incomplete, inexact, and inaccurate supervision is specifically discussed below.

Incomplete supervision means that in the training datasets some samples are labeled but the others are unlabeled. In other words, in such supervision condition, the algorithms need to learn $f : X \longmapsto Y$ from a training dataset $D = \{(x_1, y_1), \ldots, (x_l, y_l), x_{l+1}, \ldots, x_m\}$, where l denotes the number of labeled data, m denotes the number of samples in the whole dataset, x_i denotes the sample and y_i stands for its label. This is a common situation to avoid the cost of labeling all samples. In this case, the labeled part of samples is typically error-free but it is not sufficient to train the networks. Thus, it is a desirable solution to use such unlabeled data to make up for the missing labeled data. In the CV field, semi-supervised learning is used to solve such problems as shown in Fig. 2. One of the strategies is to assign pseudo labels to unlabeled samples and train deep networks simultaneously exploiting labeled samples and unlabeled samples (Lee, 2013). The corresponding optimization function can be formulated as:

$$Loss = \frac{1}{l}\sum_{i=1}^{l} l(f(x_i;\theta), y_i) + \alpha \frac{1}{m-l}\sum_{j=l+1}^{m} l(f(x_j;\theta), y'_j) \tag{1}$$

where $l(f(x_j;\theta), y'_j)$ stands for the loss function of unlabeled samples and y'_j is the pseudo label. $lf(x_i;\theta)$ is the loss function of labeled samples. α is the parameter used to control the influence of unlabeled samples. *Inexact supervision* refers to supervision information that is not as exact as desired (Zhou, 2018). In other words, in this situation, only coarse labels are given. Formally, under such condition, $f : X \longmapsto Y$ is learned from $D = \{(X_1, y_1), \ldots, (X_i, y_i), \ldots, (X_m, y_m)\}$ where the sample $X_i = \{x_{i,1}, \ldots, x_{i,m_i}\}$ may contain several objects and y_i denotes its image-level label. For example, in the case of semantic segmentation, there are weakly supervised methods that exploit image-level labels. As shown in Fig. 2, these methods utilize much coarser labels which only give the classes of objects taking up the majority of images. The most of image-level semantic segmentation methods are based on class activation maps (CAM) (Zhou et al., 2016b) defined as:

$$M_c(X_i, y_i) = \sum_k \omega_k^c f_k(X_i, y_i) \tag{2}$$

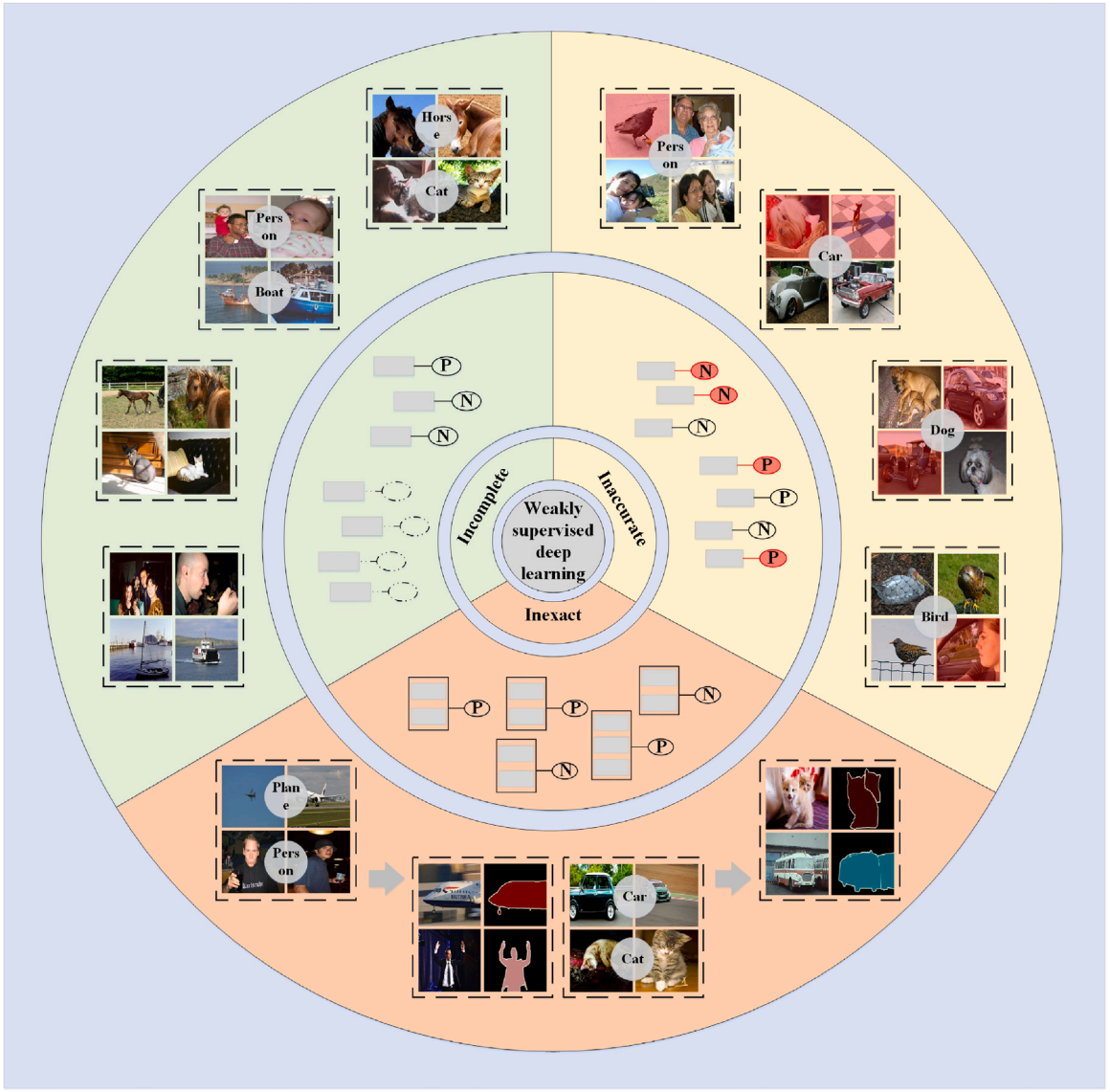


Fig. 2. The overview of weak supervision types involved in this review. 'N' and 'P' denote negative and positive samples respectively. Images with red masks are mislabeled samples.

where $M_c(X_i, y_i)$ is the class activation map for the class c. X_i is an image which is assigned the label y_i. $f_k(X_i, y_i)$ is the feature map from channel k. ω_k^c are the parameters learned from a fully connected layer. $M_c(X_i, y_i)$ can locate the major regions of objects of class c.

Inaccurate supervision means that there are errors in the labels of samples (Zhou, 2018). This means that in the dataset $D = \{(x_1, \widetilde{y_1}), \ldots, (x_i, \widetilde{y_i}), \ldots, (x_m, \widetilde{y_m})\}$, $\widetilde{y_i}$ is not necessarily the true label of the sample x_i. In the CV field, error-tolerant methods are designed to learn robust models from datasets containing mislabeled samples. To illustrate the error-tolerant approaches more intuitively in the following, the equation used for updating the reweighted loss is presented, which is one of the strategy to accomplish the error-tolerant learning (Song et al., 2020):

$$\theta_{t+1} = \theta_t - \eta \nabla \left(\frac{1}{m} \sum_{i=1}^{m} \omega(x_i, \widetilde{y_i}) l(f(x_i; \theta_t), \widetilde{y_i})\right) \tag{3}$$

where θ are the parameters of the network $f(x_i; \theta_t)$ and t indicates the epoch number. η is the learning rate. $\omega(x_i, \widetilde{y_i})l$ is the reweighted loss, in which $\omega(x_i, \widetilde{y_i})$ is the weight of the sample x_i. The samples more likely to be the noisy samples are assigned with smaller weights.

4. Weakly supervised deep learning for remote sensing big data mining

In this section, we review in detail the WSDL-based achievements for RSBD information extraction in four common tasks: semantic segmentation, scene classification, change detection and object detection, .

4.1. WSDL-based remote sensing image scene classification

the task of scene classification in remote sensing is to classify a block of remote sensing images into one or several categories (Gong et al., 2018). The task of scene classification in remote sensing has been applied widely in many fields, such as environmental monitoring and urban management (Tu et al., 2020).

There are mainly two types of methods for scene classification in remote sensing: traditional methods with handcrafted descriptors and data-driven feature-based methods (Kang et al., 2021). Traditional handcrafted feature-based methods (Yang and Newsam, 2008; Thoonen et al., 2011; Chen et al., 2016) extract low-level information of

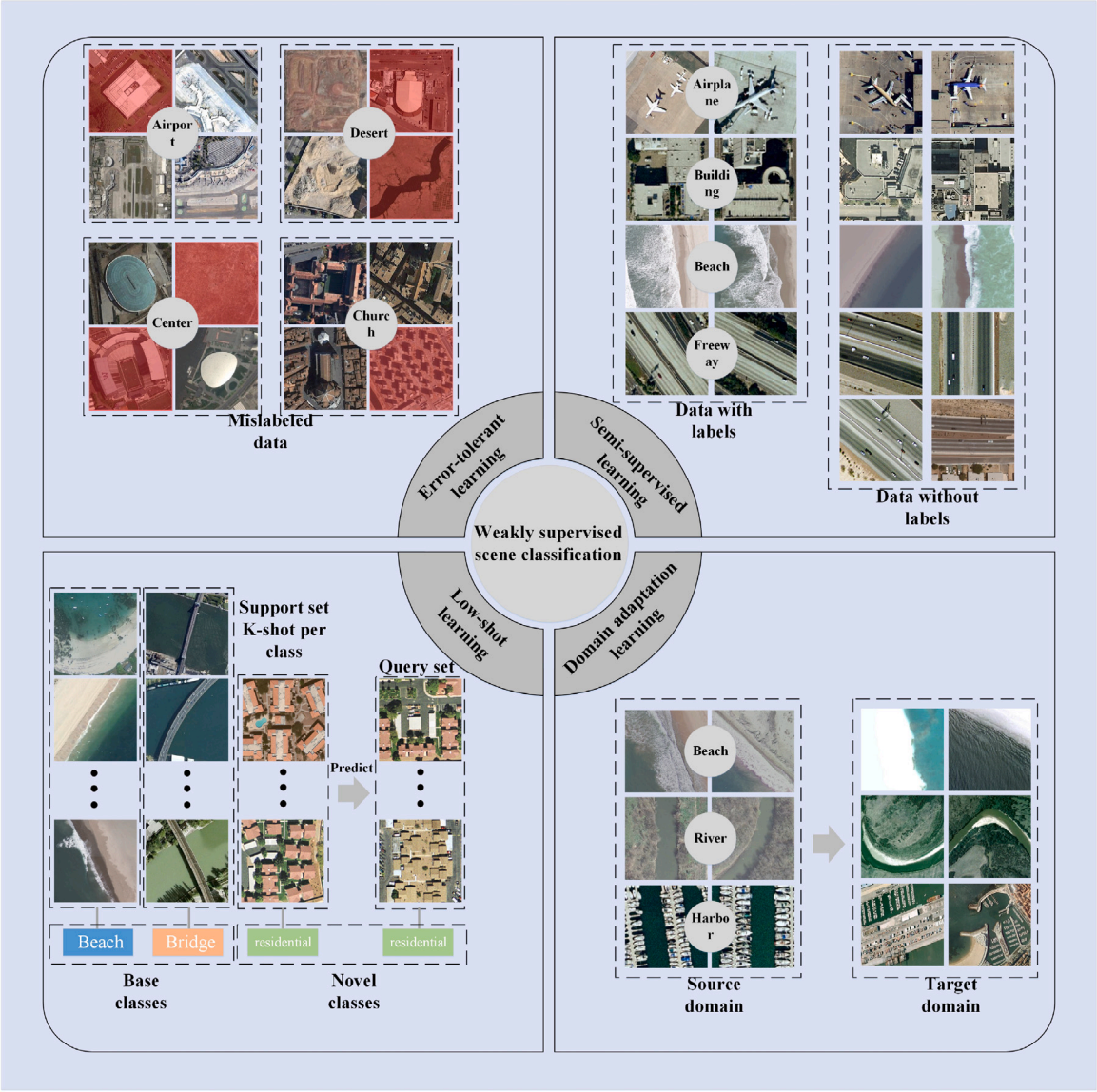


Fig. 3. Overview of weak supervision in the task of scene classification. Images with red masks are mislabeled samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

geographic objects, such as shape, spectral properties, by exploiting several descriptors. While these handcrafted descriptors are simple and interpretable, the complexity of remote sensing scenes limits the performance of these descriptors (Bratasanu et al., 2010). Deep learning is the most effective data-driven feature-based method. It automatically extracts features from data by utilizing a hierarchical structure and learning weights from huge amounts of data. The desirable performance of deep learning in natural image classification (Rawat and Wang, 2017) has attracted researchers to apply deep learning techniques to remote sensing scene classification (He et al., 2018; Hu et al., 2015; Cheng et al., 2018).

However, to achieve high accuracy it requires a huge number of labeled images. Although labels in scene classification are image-level, namely only one or few classes for image, it is still time-consuming to label all images. Therefore, there are bounds to the number of labeled images that can be obtained, preventing the fully supervised scene classification methods from the desired performance. To solve this problem, many weakly supervised methods have been proposed which could be classified into four types (see Fig. 3): (i) error-tolerant learning, (ii) semi-supervised learning, (iii) domain adaptation learning and (iv) low-shot learning. In the following, we review these four types of weakly supervised methods. Table 1 lists some datasets for weakly supervised scene classification in remote sensing that researchers can use.

4.1.1. Error-tolerant deep learning for scene classification

There are two types of methods to label data at an affordable cost in the literature (Li et al., 2021i). The first kind of approaches is based on greedy annotation, which is to cluster all remote sensing images into several classes. Images in the same cluster are assigned to the same label based on the assumption that samples within the same cluster are more likely to have the same label (Li and Ye, 2018; Xia et al., 2015). The second kind of approaches is based on crowdsourcing, which is to use crowd-sourcing information to produce labeled datasets (Li et al., 2017b; Jin et al., 2018), e.g., OpenStreetMap (Vargas-Munoz et al., 2020). All of these methods alleviate the lacking of labeled data and make it easier to annotate remote sensing images. Obviously, both types of annotating methods introduce label noises into datasets. It is worth noting that even in the samples which are annotated manually label noise might exist. Remote sensing images have high interclass similarity and high intraclass diversity (Tu et al., 2020), for which the classification of remote sensing scenes needs much expertise and experience

Table 1
Datasets for weakly supervised scene classification in remote sensing.

Datasets	Description	Supervision type	Download link
AID-GA (Li et al., 2021i)	Samples with noisy labels generated by greedy annotation algorithm (Li and Ye, 2018)	Noisy supervision for scene classification	https://captain-whu.github.io/AID
BUD-GLC (Li et al., 2021i)	5000 training samples with noisy labels generated from FROM-GLC10 product (Chen et al., 2019c)	Noisy supervision for scene classification	
Modified AID (Dai et al., 2019)	This dataset is modified from the dataset AID (Xia et al., 2017); 2% are labeled data; 98% are unlabeled data	Semi-supervision for scene classification	https://captain-whu.github.io/AID
UCM (Yang and Newsam, 2010) + RSSCN7 (Song et al., 2019)	Source domain: UC Merced dataset (Yang and Newsam, 2010); target domain: RSSCN7 (Zou et al., 2015)	Domain adaptation scene classification	UCM:weegee.vision.ucmerced.edu/datasets/landuse.html RSSCN7:https://sites.google.com/site/qinzoucn/documents
RSSDIVCS (Li et al., 2021j)	56000 samples with labels of 70 scene categories and each categories have corresponding language-level description	Scene labels and category language-level description for zero-shot learning	https://github.com/kdy2021/SR-RSKG
Modified UCM (Yuan et al., 2020)	This dataset is modified from the dataset UCM (Yang and Newsam, 2010); 16 classes as the base class for training; 5 classes as the novel class for evaluation	Few-shot supervision for scene classification	weegee.vision.ucmerced.edu/datasets/landuse.html
UCM2MAI (Hua et al., 2021)	1600 single-label samples from UCM (Yang and Newsam, 2010); 1649 multi-label samples from MAI (Hua et al., 2021)	Single-label to multi-label	https://github.com/Hua-YS/Prototype-based-Memory-Network
AID2MAI (Hua et al., 2021)	7050 single-label samples from AID (Xia et al., 2017); 3239 multi-label samples from MAI (Hua et al., 2021)	Single-label to multi-label	https://github.com/Hua-YS/Prototype-based-Memory-Network

and sometimes ground investigation is mandatory. Therefore, it is easy to introduce errors in these processes. The presence of label noises significantly reduces the performance of deep networks (Pelletier et al., 2017).

There are many error-tolerant scene classification methods in remote sensing which can be roughly classified into two categories (Li et al., 2021i), i.e., methods based on robust loss functions (Gong et al., 2018; Kang et al., 2021, 2020; Zhao et al., 2017; Hua et al., 2020) and methods based on noise correction (Tu et al., 2020; Li et al., 2021i). Methods based on robust loss functions in Gong et al. (2018), Kang et al. (2021, 2020) and Hua et al. (2020) adopt the strategy of modifying the loss function to alleviate the influence of mislabeled samples and train robust models. Zhao et al. (2017) add a noisy-label transition layer in the convolutional neural networks (CNNs), which can also be regarded as a modification of the loss function. Methods based on noise correction aim to identify the mislabeled samples and correcting them. Tu et al. (2020) model the correlation between mislabeled samples and correct samples using the covariance matrix and makes it easier to push noisy labels apart. Furthermore, Li et al. (2021i) propose a novel framework which detects the potential labeling noises and then relabels these uncertain labels utilizing a multi-view structure to iteratively train the multiple networks. In Perantoni and Bruzzone (2021), the obsolete digital maps are used as weak supervision combined with a small clean dataset. A novel training strategy is proposed to weight samples and let the most reliable labels guide the optimization.

4.1.2. Semi-supervised deep learning for scene classification

Semi-supervised deep learning aims at learning parameters from a dataset where there are few labeled samples and other unlabeled samples. Due to the more complicated imaging conditions and the need for extra geo-knowledge, it is more difficult for annotators to manually label remote sensing images. Therefore, many researchers introduced semi-supervised learning into remote sensing (Dai et al., 2019; Guo et al., 2020; Han et al., 2018; Zhang and Yang, 2020; Roy et al., 2018; Tao et al., 2020). According to how the unlabeled data are used, these methods can be divided into three types: (i) methods based on generative adversarial networks (GANs), (ii) methods based on self-labeling techniques and (iii) methods based on self-supervised representation learning.

In Guo et al. (2020), Roy et al. (2018) and Zhan et al. (2017), GANs are used to boost the performance of classifier in the semi-supervised paradigm. In such configuration, the discriminator of GAN detects not only whether the input images are fake or not but also to which class each real image belongs. The generator still generates fake images from noise vectors. In Singh and Bruzzone (2021), a limited number of training samples are used to perform data augmentation and the high-quality generated data can be used as training samples in supervised classification. To exploit unlabeled samples, Han et al. (2018) utilize a self-labeling technique and label unlabeled data with the help of a small number of labeled data. After that, samples with pseudo labels are also used as supervision to achieve desirable performance in scene classification. Some self-supervised feature learning methods (Tao et al., 2020; Li et al., 2016a, 2017c) can also be categorized as semi-supervised methods. Tao et al. (2020) demonstrate that self-supervised learning is suited to remote sensing scene classification and analyze the factors that can contribute to better performance of self-supervised learning in remote sensing. In Li et al. (2016a), two layers are trained to extract features by a clustering algorithm and the support vector machine (SVM) algorithm is used to classify the learned

features. In Li et al. (2017c), both unsupervised deep networks and fully connected networks are used together to extract features from unlabeled data by achieving accurate results.

4.1.3. Domain adaptation deep learning for scene classification

As already mentioned, RSBD can greatly vary due to the differences in the sensors, in the acquisition conditions and so on Tuia et al. (2021). Typically, deep learning methods can reach great performance only when the testing and training samples belong to a similar domain. Otherwise, the performance deteriorates significantly. However, since collecting sufficient labeled samples for every new applications is not realistic, a possible solution is to utilize the existing datasets, which are not from the same domain as the target samples, and make the methods more robust across different domains (Tuia et al., 2021).

Domain adaptation deep learning aims at learning knowledge from the source domain and applying the learned knowledge to the target domain to achieve desirable performance under the condition of lack of labels in the target domain samples. In the field of CV, domain adaptation in deep learning has achieved considerable results (Wang and Deng, 2018; Hong et al., 2018). In Das and Chandran (2021), CNNs are trained by a source domain dataset and then used to initialize the classifier. The initialized classifier is fine-tuned on the target domain. In Wei et al. (2021), several classifiers are fused with adaptive weights, under the assumption that different classifiers can achieve robust domain adaptation. In Lu et al. (2020b), Lu et al. address the problem of having in the source domain dataset only some of classes in the target domain. To solve this problem, a classifier complement module is proposed to align classes from multiple sources. In order to align the source and the target domains, in Song et al. (2019), a modified CNNs architecture is proposed and a new layer is added to the network.

4.1.4. Low-shot deep learning for scene classification

There is still a category of problem that can be categorized as weakly supervised learning, namely low-shot learning. Collecting samples of rare classes is difficult (Long et al., 2017) and labeling many samples of emerging classes sometimes can be unaffordable (Guo et al., 2017). As a solution, low-shot learning aims at classifying images of novel classes which have few or even no labeled training samples by utilizing the auxiliary knowledge and base datasets which have labeled samples that do not overlap with novel classes. Both base datasets, which consist of no target classes, and the auxiliary knowledge is regarded as weakly supervised information in this review. According to the number of samples available for novel classes, the low-shot learning can be classified into two types: (i) zero-shot and (ii) few-shot.

In remote sensing, there are already some works on zero-shot learning (Li et al., 2021j; Song and Xu, 2017; Li et al., 2017d; Sumbul et al., 2017; Quan et al., 2018; Li et al., 2020b, 2021c). In Li et al. (2017d), the auxiliary knowledge is the textual information and the word2vec model is utilized to map both the seen classes and the unseen classes into the same semantic space. A generative framework is proposed in Song and Xu (2017) to construct a feature space used to understand target classes which do not exist in the training samples. As shown in Fig. 4, (Li et al., 2021j) proposes a novel locality-preservation deep cross-modal embedding networks (LPDCMENs). The quality analysis for zero-shot scene classification can be seen in Table 2. The zero-shot learning is used in Sumbul et al. (2017) to solve the problem of fine-grained object recognition. Quan et al. (2018) employ the algorithm of semi-supervised embedding to make the class structure more consistent with visual space prototypes.

Some researchers have been pursuing for few-shot learning. Alajaji et al. (2020), Cheng et al. (2021a) and Zhang et al. (2021a) adopt metric-based methods. These methods calculate the distance between the query set and the support set. In contrast to the original prototypical networks, (Cheng et al., 2021a) proposes Siamese-prototype network that shows outstanding performance and Zhang et al. (2021a) adopts a meta-learning strategy which makes the model more effective in a few-shot setting.

4.1.5. Other methods

Many scene classification datasets are single-label, namely only one category is identified for each sample, and many scene classification methods share the same assumption that every input image contains merely only one kind of objects. In practice, it is more common that there are various categories in one scene and a single label is not sufficient to describe such complex scenes. To cope with the multi-label problems, a huge multi-label dataset is essential especially for deep learning-based methods. However, it takes ten times longer to assign multiple labels to one image than to assign a single label (Hua et al., 2021). An alternative solution is to learn the multi-label tasks from the plenty of existing single-label datasets. We consider this as a type of weakly supervised learning which utilizes inexact supervision information. So far, the only paper (Hua et al., 2021) solves such problem in remote sensing. Hua et al. (2021) propose a novel method using the existing single-label datasets combined with a small amount of multi-label samples to train a deep networks, which alleviates the lack of high-quality multi-label datasets in remote sensing (Singh and Bruzzone, 2021).

4.2. WSDL-based remote sensing object detection

Object detection aims at locating all objects of interest in the input images and assigning categories to the detected objects. This task is more challenging than scene classification. Generally, for the deep learning-based object detection methods, the training data should be labeled with bounding boxes that outline all objects. Compared with the natural images, images in remote sensing usually contain smaller objects with arbitrary orientation, and the background is more complex. Therefore, labeling all the interest objects with bounding box is time-consuming and weakly supervised learning is also needed to cope with this problem. As illustrated in Fig. 5, apart from deep learning under image-level supervision for object detection (Li et al., 2018b; Zhang et al., 2016a; Wu et al., 2020; Chen et al., 2020b), there are also other types of methods that are regarded as weakly supervised object detection, including domain adaptation object detection (Pan et al., 2017; Chen et al., 2018b, 2021a), semi-supervised object detection (Weinstein et al., 2019; Xue and Tong, 2019; Gao et al., 2018) and low-shot object detection (Chen et al., 2019a). To facilitate reproducing related methods, datasets for weakly supervised object detection in remote sensing are listed in Table 3.

4.2.1. Deep learning under image-level supervision for object detection

Instead of labeling all objects contained in one remote sensing image, the image-level supervision only gives the information of whether objects of a certain category are present or not. Obviously, it greatly reduces the cost of labeling and makes it possible to leverage the large number of existing scene classification datasets to complete the object detection tasks. Meanwhile, compared with bounding box labels, image-level labels contain less information. Thus, deep learning models will be learned with less robustness. Some efforts have been made to bridge the gap between bounding box supervision and image-level supervision in remote sensing. There have been some works on object detection under image-level supervision in remote sensing (Li et al., 2018b; Zhang et al., 2016a; Chen et al., 2020b; Wu et al., 2020; Zhou et al., 2016a; Feng et al., 2020; Zhang and Ma, 2021; Qiao et al., 2020; Yao et al., 2020; Aygunes et al., 2021; Kellenberger et al., 2019). Most of these methods exploit the class activation maps (CAMs) to locate the interest objects coarsely and then refine the detected regions (Zhou et al., 2016b). Li et al. (2018b) utilize image-level labels to train classification networks which can be used to generate the CAMs indicating salient regions. This paper exploits the similarity constraints within the samples of the same class as shown in Fig. 6. Finally, object detection is conducted on segmented salient maps. Wu et al. (2020), Qiao et al. (2020) adopt a similar method to undertake object detection by using CAMs and threshold segmentation. Yao et al. (2020) introduce

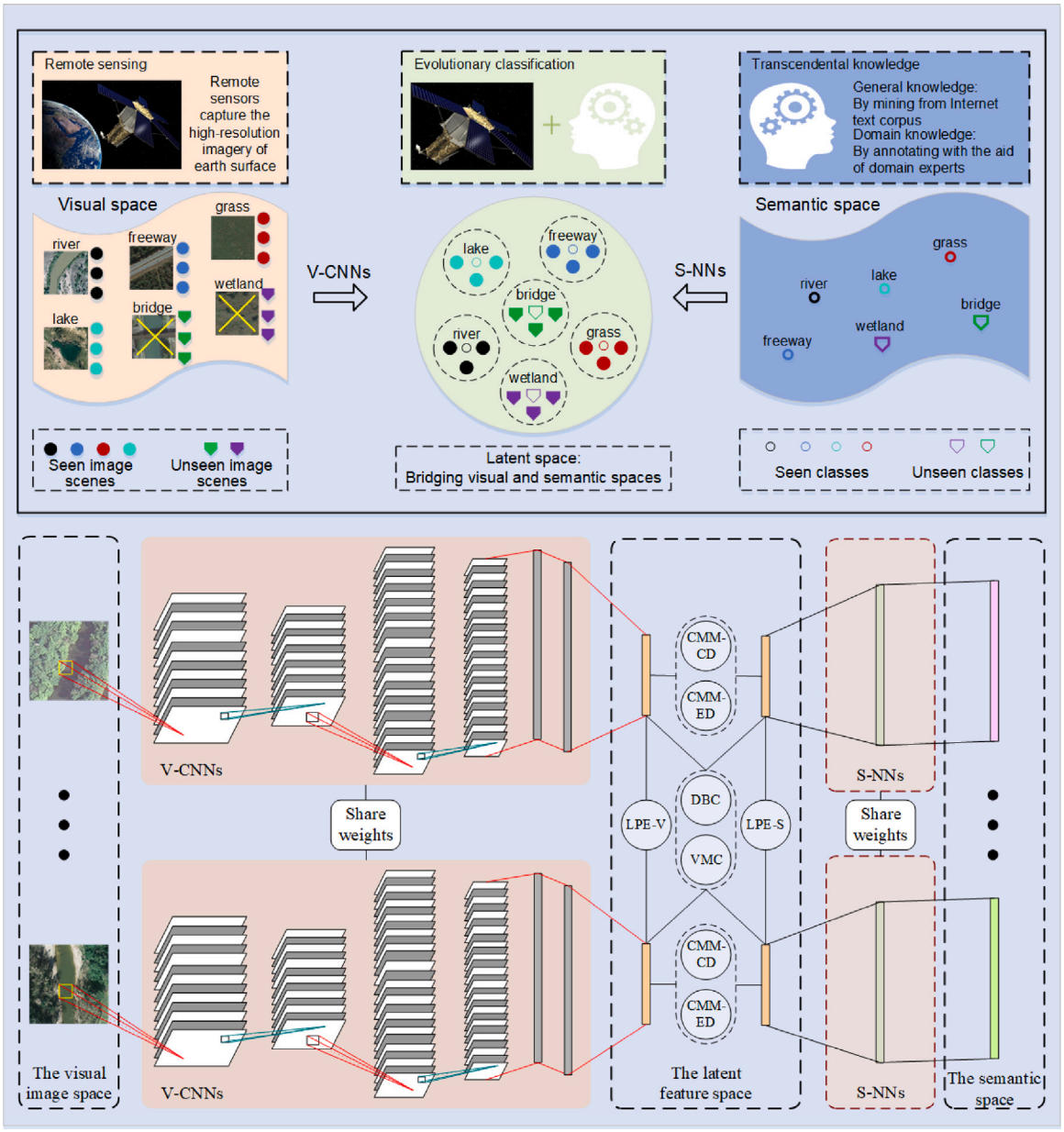


Fig. 4. The architecture of the zero-shot scene classification method. In this figure, LPE-V and LPE-S mean the locality-preservation constraint for visual samples and the semantic representations respectively. CMM-CD and CMM-ED denote the cross-modal matching constraint. DBC and VMC denote the latent feature regularization constraint.

Table 2
Quality analysis for zero-shot scene classification (Li et al., 2021j).

Knowledge type	General knowledge			Domain knowledge		
Seen/unseen ratio	40/30	50/20	60/10	40/30	50/20	60/10
SAE (V→ S) (Kodirov et al., 2017)	0.096 ±0.014	0.137 ±0.017	0.235 ±0.042	0.088 ±0.013	0.124 ±0.019	0.220 ±0.017
SAE (S→ V) (Kodirov et al., 2017)	0.052 ±0.014	0.095 ±0.016	0.167 ±0.041	0.050 ±0.013	0.080 ±0.023	0.168 ±0.044
DMaP (Li et al., 2017e)	0.104 ±0.009	0.167 ±0.022	0.260 ±0.036	0.100 ±0.008	0.156 ±0.019	0.164 ±0.019
SPLE (Tao et al., 2017)	0.098 ±0.014	0.132 ±0.019	0.201 ±0.037	0.083 ±0.020	0.132 ±0.026	0.190 ±0.038
CIZSL (Elhoseiny and Elfeki, 2019)	0.060 ±0.012	0.106 ±0.037	0.206 ±0.004	0.062 ±0.021	0.103 ±0.019	0.204 ±0.041
ZSRSSC-GP (Li et al., 2017d)	0.047 ±0.003	0.083 ±0.007	0.148 ±0.004	0.046 ±0.004	0.074 ±0.003	0.141 ±0.016
ZSRSSC-SE (Quan et al., 2018)	0.121 ±0.077	0.152 ±0.010	0.267 ±0.053	0.131 ±0.030	0.183 ±0.013	0.293 ±0.038
LPDCMENs (Li et al., 2021j)	0.158 ±0.029	0.197 ±0.036	0.342 ±0.090	0.216 ±0.038	0.249 ±0.037	0.438 ±0.073

Fig. 5. Overview of weak supervision in the task of object detection.

Table 3
Datasets for weakly supervised object detection in remote sensing.

Datasets	Description	Supervision type	Download link
WSADD (Wu et al., 2020)	600 samples with image-level labels;300 samples with bounding box labels	Image-level supervision for object detection	
ISAR dataset (Xue and Tong, 2019)	11000 samples with key points annotations	Semi-supervision for object detection	
Modified DOTA (Chen et al., 2021a)	Source domain: 834 samples with original images and labels from DOTA dataset (Ding et al., 2021);target domain: 1112 samples with brightness reduction processing from DOTA dataset	Domain adaptation for object detection	https://captain-whu.github.io/DOTA/index. html
Vehicle Dataset (Chen et al., 2019a)	12495 samples of 16 classes;12 categories as seen classes;4 categories as unseen classes	Zero-shot supervision for object detection	
Modified DIOR (Cheng et al., 2021b)	23463 samples of 20 classes from DIOR (Li et al., 2020c);5 categories as novel classes;15 categories as base classes	Few-shot supervision for object detection	www.escience.cn/people/JunweiHan/DIOR.html

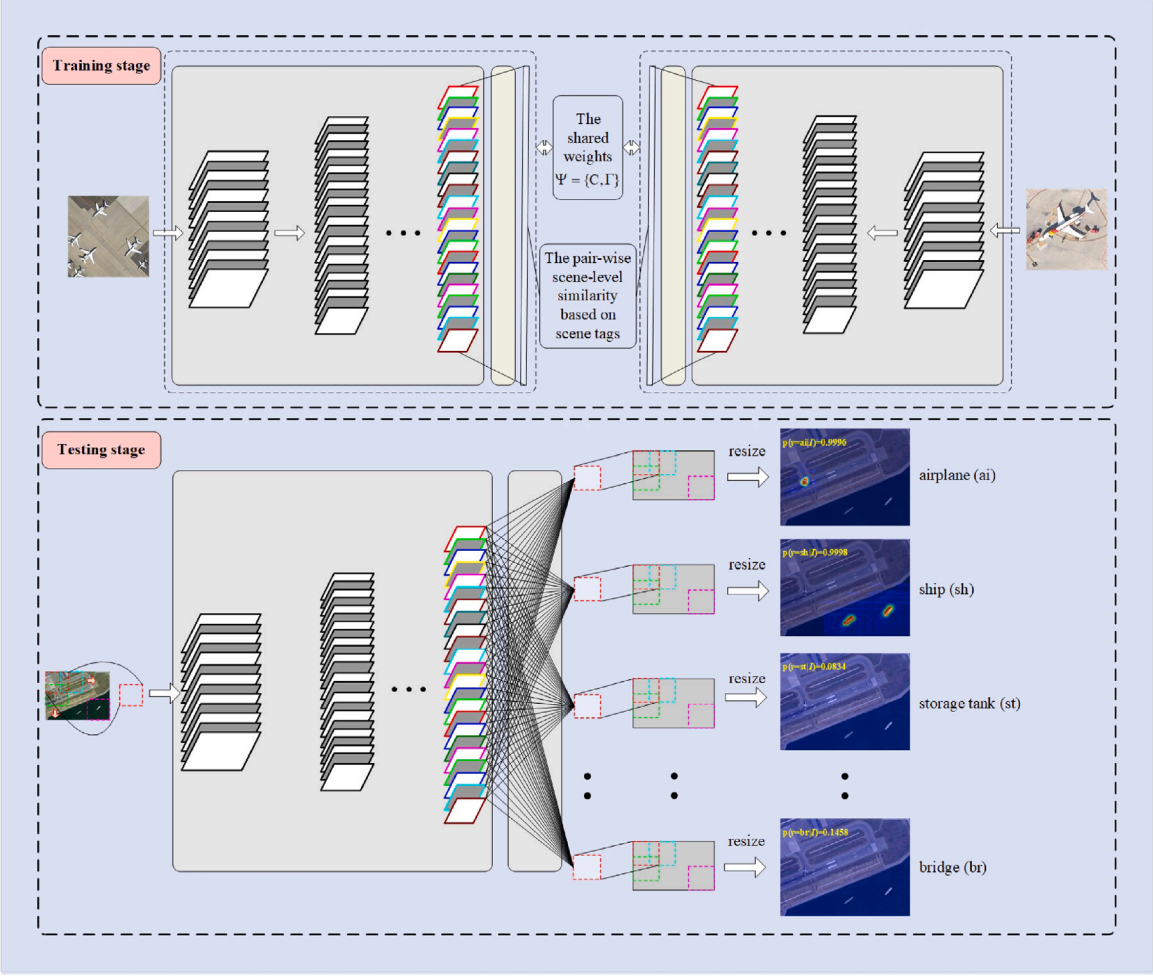


Fig. 6. Weakly supervised object detection method under image-level supervision.

Table 4
Quality analysis for object detection under image-level supervision (Yao et al., 2020).

Method	Airplane	Ship	Storage tank	Baseball diamond	Tennis court	Basketball court	Ground track field	Harbor	Bridge	Vehicle	mAP
Fast-RCNN (Girshick, 2015)	0.9091	0.9060	0.8929	0.4732	1.0000	0.8585	0.8486	0.8822	0.8029	0.6984	0.8272
RICNN (Cheng et al., 2016)	0.8871	0.7834	0.8633	0.8909	0.4233	0.5685	0.8772	0.6747	0.6231	0.7201	0.7312
RCNN (Girshick et al., 2013)	0.8537	0.8888	0.6278	0.1973	0.9066	0.5823	0.6795	0.7987	0.5422	0.4992	0.6576
Transfer CNN (Li et al., 2017a)	0.6603	0.5713	0.8501	0.8093	0.3511	0.4552	0.7937	0.6257	0.4317	0.4127	0.5961
COPD (Cheng et al., 2014)	0.6225	0.6937	0.6452	0.8213	0.3413	0.3525	0.8421	0.5631	0.1643	0.4428	0.5488
WSDDN (Bilen and Vedaldi, 2016)	0.3008	0.4172	0.3498	0.8890	0.1286	0.2385	0.9943	0.1394	0.0192	0.0360	0.3512
OICR (Tang et al., 2017)	0.1366	0.6735	0.5716	0.5516	0.1364	0.3966	0.9280	0.0023	0.0184	0.0373	0.3452
PCL (Tang et al., 2018)	0.2600	0.6376	0.0250	0.8980	0.6445	0.7607	0.7794	0	0.013	0.1567	0.3941
MELM (Wan et al., 2018)	0.8086	0.6930	0.1048	0.9017	0.1284	0.2014	0.9917	0.1710	0.1417	0.0868	0.4229
DCL (Yao et al., 2020)	0.7270	0.7425	0.3705	0.8264	0.3688	0.4227	0.8395	0.3957	0.1682	0.3500	0.5211

the curriculum learning into the weakly supervised object detection and achieve accurate performance. Kellenberger et al. (2019) combine the weak and full supervision, which means that not only a small part of fully labeled samples but also the rest of samples with image-level labels contribute to training deep models.

To intuitively visualize performance of the existing achievements, a comparison of different object detection methods under image-level supervision is listed in Table 4. Experiments show that the results can be competitive with fully supervised methods.

4.2.2. Semi-supervised deep learning for object detection

In remote sensing, some works apply the semi-supervised learning to object detection in order to alleviate the lack of high quality labeled datasets. In the semi-supervised learning, we only need to label with high-quality small amount of samples. However also in this case, such small labeled data are not sufficient to train deep networks. While obtaining labeled data is difficult, it is much easier to get a large number of unlabeled data filling in missing information during training. One intuitive method is to extend the labels of a few labeled data to other unlabeled data. One type of semi-supervised methods in object detection for remote sensing is the co-training strategy. Gao et al. (2018) adopt such strategy to implement a semi-supervised object detection task. They propose a co-training method to train two classifiers and predict the labels of the unlabeled samples. The predictions with high confidence are added to the labeled set and used in the later training of object detection networks. Another solution is to utilize a robust loss function to achieve better performance under weak supervision. In Xue and Tong (2019), both samples with key point annotations and unlabeled samples are used to train the object detection networks. A pairwise-ranking loss is exploited to process the weak key point

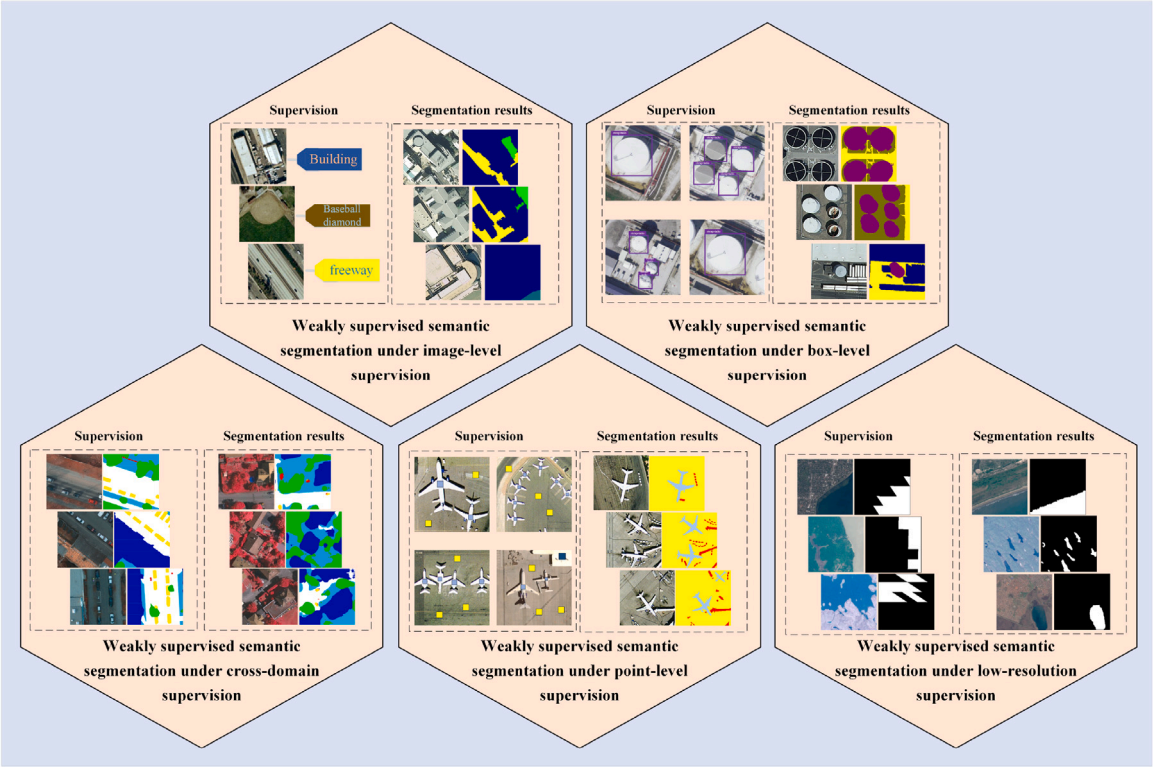


Fig. 7. Overview of weak supervision in the task of semantic segmentation.

annotations and a triplet-ranking loos is exploited to process unlabeled data according to this work. Unlike the previous methods that mine information from both unlabeled and labeled data, Weinstein et al. (2019) leverage light detection and ranging (LIDAR) data to recover the initial labels of plenty of unlabeled data in the task of tree-crown detection task and obtain the desirable performance.

4.2.3. Domain adaptation deep learning for object detection

Considering that the acquisition of vast training samples with bounding box labels is somewhat difficult, labeling just a few samples is an acceptable option. However, as aforementioned, such few labeled samples are not sufficient to accomplish this task, especially for deep learning methods where the models contain many parameters. Domain adaptation can address the insufficiency of labeled samples. By transferring the knowledge learned in the source domain to the target domain. Compared with source domain, images in the target domain have related but different features. Therefore, using samples from the source domain directly as auxiliary supervision for the target domain can lead to a reduction of performance. Domain adaptation object detection methods bridge the gap between two different domains. The methods in Pan et al. (2017) and Chen et al. (2018b) adopt the same strategy. Firstly a deep network is learned from the source domain which has a sufficient number of samples. Then a small number of target domain samples is used to fine tune the trained model. Chen et al. (2021a) not only consider the diversity between different datasets but also the difference between the training and test datasets. In this regard, a novel domain adaptation faster R-CNN algorithm is proposed to transfer between training and test domains.

4.2.4. Low-shot deep learning for object detection

For CV, there has been a lot of research about few-shot object detection (Lake et al., 2013; Dixit et al., 2017; Karlinsky et al., 2019; Wang et al., 2019b; Kang et al., 2019; Chen et al., 2018a; Wang et al., 2019a; Yan et al., 2019a; Hsieh et al., 2019). But when applied to remote sensing images, aforementioned methods could not achieve desirable performance. There has been some excellent works on few-shot object detection in remote sensing (Cheng et al., 2021b; Xiao et al., 2020, 2021; Li et al., 2021b). In Cheng et al. (2021b), a prototype learning network is used to learn the prototypes of each class which contribute to the generation of candidate boxes and boost the detection performance. In Xiao et al. (2020), a feature attention highlighting module is proposed to achieve accurate results in a simple way under the few-shot condition. Li et al. (2021b) introduce the meta-learning to reweight the feature maps in novel classes. In Xiao et al. (2021), a self-adaptive attention network is proposed to leverage the knowledge learned from base classes and apply them to the novel classes. In spite of many zero-shot object detection methods in CV (Li et al., 2019; Rahman et al., 2019; Bansal et al., 2018; Yan et al., 2020), research on the zero-shot object detection in the remote sensing is very scarce except for this work (Chen et al., 2019a). In Chen et al. (2019a), a coarse-to-fine framework is proposed which firstly locates the target in fine-grained features and then recognizes the target in a coarse-grained manner.

4.3. WSDL-based remote sensing semantic segmentation

Semantic segmentation aims at classifying every pixel in an image and plays an essential role in understanding images. Semantic segmentation using deep learning techniques usually need pixel-level annotations and outputs pixel-level predictions. However, pixel-level annotations are not easy to obtain. In the field of CV, to overcome the obstacle of acquisition of accurate labels, many methods of semantic segmentation are proposed leveraging weak supervision information, including image-level labels (Wei et al., 2016b,a; Zeng et al., 2019; Shimoda and Yanai, 2016; Wei et al., 2018), bounding boxes (Dai et al., 2015; Papandreou et al., 2015; Khoreva et al., 2017; Ibrahim et al., 2020), scribbles (Lin et al., 2016), points (Bearman et al., 2016), which are illustrated in Fig. 7. The datasets available in remote sensing are listed in Table 5. These types of labels either already exist in large numbers or have easy access. As Fu et al. (2018) indicated, the labeling

Table 5
Datasets for weakly supervised semantic segmentation in remote sensing.

Datasets	Description	Supervision type	Download link
Water dataset (Fu et al., 2018)	9409 samples with pixel-level and image-level labels	Image-level supervision for semantic segmentation	
Cloud dataset (Fu et al., 2018)	8705 samples with pixel-level and image-level labels	Image-level supervision for semantic segmentation	
WDCD Dataset (Li et al., 2020a)	206384 training samples with image-level labels; 30 large test samples with pixel-level labels	Image-level supervision for semantic segmentation	https://github.com/weichenrs/WDCD
GID-flcs (Tong et al., 2020b)	30000 training samples with image-level labels; 10 large test samples with pixel-level labels	Image-level supervision for semantic segmentation	https://x-ytong.github.io/project/GID.html
Points dataset (Zhang et al., 2021b)	16844 points with true classes for training; 17600 points with true classes for test	Point-level supervision for semantic segmentation	
Road center point Dataset (Lian and Huang, 2021)	240000 road center points for training	Point-level supervision for semantic segmentation	
Modified crowdai mapping challenge (Mohanty et al., 2020)	280741 training samples with horizontal bounding boxes;60317 validation samples with pixel-level labels	Box-level supervision for semantic segmentation	https://www.crowdai.org/challenges/mapping-challenge
Potsdam dataset + Vaihingen dataset (Li et al., 2021g)	Source domain: Potsdam dataset; Target domain: Vaihingen dataset	Domain adaptation for semantic segmentation	https://github.com/te-shi/MUCSS
SEN12MS (Schmitt et al., 2019)	180662 samples with low resolution labels	Mixed weak supervision for semantic segmentation	https://mediatum.ub.tum.de/1474000
TimeSen2Crop (Weikmann et al., 2021)	More than 1 million pixel-based samples of Sentinel 2 time series associated to 16 crop types.	Mixed weak supervision for semantic segmentation	https://rslab.disi.unitn.it/timesen2crop/

of image-level labels saves at least one hundred times more time than that of pixel-level labels for natural images.

Since pixel-level annotation of remote sensing images requires additional expertise (e.g., identifying thin and thick clouds) or even the on-ground investigation, the labeling procedure is even more time-consuming (Reichstein et al., 2019). Therefore, remote sensing image processing needs weak supervision methods to solve the problem of pixel-level label shortage. However, directly applying CV methods to remote sensing imagery cannot achieve desirable performance (Chan et al., 2021).

4.3.1. Deep learning under image-level supervision for semantic segmentation

Some of the methods using image-level supervision are based on the research in Zhou et al. (2016b). This work proposes a method to extract class activation maps (CAMs) from trained classification networks which show the most discriminative regions for a given class. The method establishes a connection between image-level labels and pixel-level labels and makes it possible to extract spatial information from a classification network which is exactly what image-level labels lack.

The CAM-based methods can be summarized into two phases. First, class activation maps are generated to locate the salient objects. In this phase, the activation regions may not be able to cover the whole objects of interest and may generate inaccurate boundaries. Next, the classification of every pixel is generated based on the CAMs.

In the first phase of CAM-based methods, class activation maps are typically generated by the last convolutional layer which has a large receptive field and reflects high-level semantic features. Therefore, due to its large receptive field, class activation maps probably lose some local information which is essential to reconstruct the accurate boundaries. Meanwhile, the scale of objects in the remote sensing images varies greatly (Fu et al., 2018). The methods extracting activation maps just from one layer are not able to handle remote sensing images properly. To address this, Zhang et al. (2019b) propose to generate multiscale class-specific activation maps from different convolutional layers of a trained classification network. And after superpixel segmentation and low-rank matrix recovery, multiscale saliency maps are fused to produce pixel-level classification. The method adopted in Fu et al. (2018) can also be classified as a two phase method. The first step is to gain saliency maps which combine outputs of different convolutional layers, which makes this method more suitable for remote sensing images. Next, the maps are segmented into superpixels to generate new labels. During the testing stage, the trained network classifies every pixel. Li et al. (2021h) demonstrate the effectiveness of training a segmentation network after generating class activation maps. In order to recover accurate boundaries Chen et al. (2020a) add a superpixel pooling layer into the network from which the class activation maps are extracted. Since the superpixel combines pixels based on low-level shape information, it can make the CAMs more accurate. As shown in Fig. 8, Li et al. (2020a) exploit a new global convolutional pooling to train classification models generating CAMs. In the phase of recovering pixel-level classification from CAMs, instead of training a new segmentation network, this work proposes a novel local pooling pruning strategy. All of local poolings are discarded, which makes CAMs accurate enough to serve as pixel-level classification. Fig. 9 gives a qualitative comparison of weakly supervised semantic segmentation methods under image-level supervision.

4.3.2. Deep learning under point-level supervision for semantic segmentation

Point-level labels, where points are labeled with true classes, indicate not only high-level category information but also information about the appropriate localization of objects, which contains stronger supervision than image-level labels. Based on this consideration, Zhang et al. (2021b) propose a weakly towards strongly (WTS) supervised learning framework exploiting labeled points as supervision, namely an image having many labeled points. Given many points with labels,

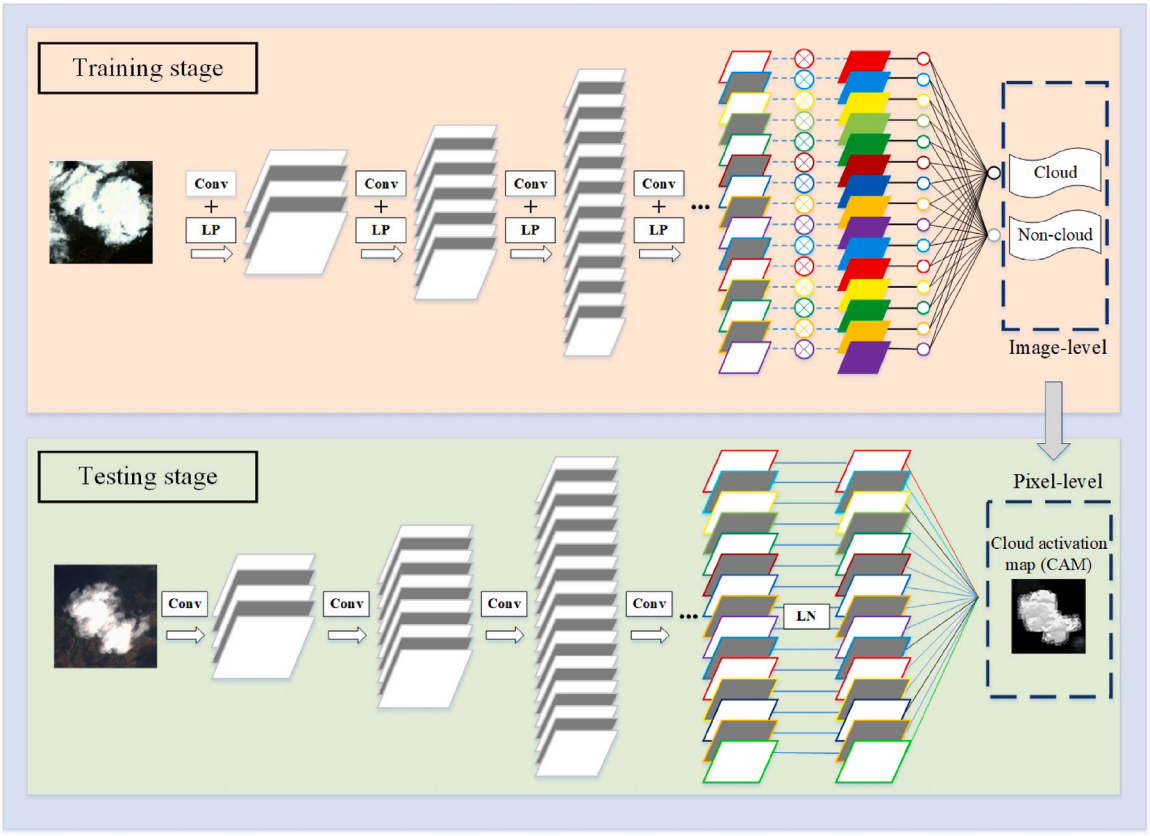


Fig. 8. Weakly supervised semantic segmentation method under image-level supervision.

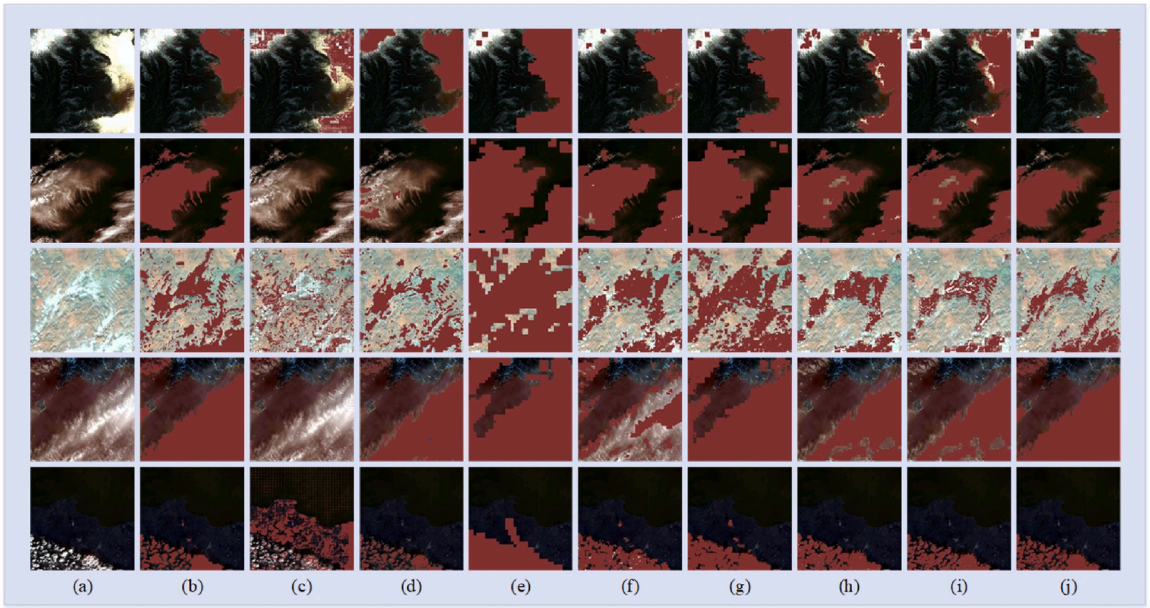


Fig. 9. Qualitative results for image-level supervised semantic segmentation. (a) Test images, (b) corresponding GT, (c) GCM, (d) PRS, (e) CAA, (f) CAM with GAP, (g) CAM with TSL, (h) CAM with GCP, (i) CAM with GCP + LPP, (j) CAM with GCP + LPP*. The red regions denote the detected cloud regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SVM is trained to produce initial seed points. After that, to expand the labeled points, a segmentation model is trained on these seed points. The conditional random field (CRF) and seeded region growing (SRG) algorithm are utilized to generate new seed points from the output of trained segmentation network. The seed points are updated by repeating this process until convergence. Finally, the trained segmentation model predicts every pixel. Wang et al. (2020) utilize simpler point-level labels. Instead of labeling many points in an image, every image

only has one point with true class label, which is more similar to image-level labels. This work trains a segmentation model directly on labeled points, masking out other unlabeled points. And in the testing stage, the trained segmentation model is used to classify every pixel.

For some applications in remote sensing, scribble labels can be more accessible and suitable than other types of weak supervision, such as road surface extraction. Scribble labels can be regarded as point labels with many points for each class. Wei and Ji (2021b) propose a scribble-based weakly supervised road surface extraction method. The proposed method firstly generates masks by a road label propagation algorithm and then, by minimizing the joint loss, a deep model is trained under the supervision of the generated masks and boundary.

4.3.3. Deep learning under box-level supervision for semantic segmentation

There is another type of weak supervision for semantic segmentation, namely horizontal bounding boxes for objects of interest. Substituting bounding boxes for pixel-level labels not only saves labeling time, but also establishes a bridge between object detection and semantic segmentation. In other words, the outputs of methods of object detection, which are usually bounding boxes, can supervise the semantic segmentation task to generate pixel-level predictions. On the basis of this, there are some works (Dai et al., 2015; Khoreva et al., 2017; Ibrahim et al., 2020) about semantic segmentation under box-level supervision in the CV field. But in the remote sensing field, little research has been done in this area except for (Rafique and Jacobs, 2019). Rafique and Jacobs (2019) present a method aiming at bridging this gap. Typically, in the field of CV, there is the major problem of overlapping objects for weakly supervised semantic segmentation exploiting bounding boxes. However, in remote sensing, land cover segmentation has exclusively one label per region. Based on this fact, this method generates pixel-level dense masks by applying a Gaussian distribution to bounding boxes, which can supervise the training of semantic segmentation networks.

4.3.4. Domain adaptation deep learning for semantic segmentation

Another potential solution to an inadequate number of high-quality labels is to utilize domain adaptation methods. In fact, there are several remote sensing datasets with pixel-level labels. However, due to different acquisition conditions, the models trained on existing datasets generally perform badly on target tasks (Li et al., 2021g). The domain adaptation methods utilize existing datasets with pixel-level labels to train semantic segmentation models and then transfer them to the target tasks that we need to solve. There are some works about domain adaptation learning in the CV (Saito et al., 2018; Tsai et al., 2018). And some pioneers have introduced this strategy into remote sensing. Li et al. (2021g) propose a novel loss function exploiting multiple weakly constraints to learn the cross-domain remote sensing semantic segmentation model, including pseudo-label constraint, rotation consistency constraint and transfer invariant constraint. The trained model performs well in the target domain. Researchers in (Benjdira et al., 2019; Ji et al., 2020; Yan et al., 2018, 2019b) utilize the GANs to align the target and source domains and achieve accurate performance on samples from the target domain. In Yan et al. (2021), the authors point out that some unsupervised domain adaptation methods using adversarial learning neglect the pixels without pseudo-labels. Thus a cross mean teacher (CMT) method is proposed to exploit not only pixels with pseudo-labels but also without pseudo-labels.

4.3.5. Deep learning under mixed weak supervision for semantic segmentation

As Grekousis et al. (2015) summarized, while there are many products of land cover maps in remote sensing, the resolution and accuracy of these land cover maps vary greatly. In other words, these are potential supervision for semantic segmentation in remote sensing (Schmitt et al., 2020) which can also be called weak supervision. Valuable land cover products are available both at continental and global scale but extraction reliable information from these potentially inaccurate and obsolete maps is challenging (Bruzzone, 2019). Under this situation, low resolution labels are used as supervision information to train a robust model and then the robust model produces high resolution predictions which are consistent with input images. It can be regarded as a mixture of multiple weak supervision information, including inexact supervision, which means that low resolution labels are not sufficient to achieve desirable performance, and inaccurate supervision, which means even in the low-resolution labels there will be errors. Errors can also be introduced into the labels by the automatic process of generation of training dataset (Paris et al., 2021).

4.4. WSDL-based remote sensing change detection

Change detection in remote sensing aims at identifying changed regions between two or more images over the same area but at different times and in some cases figuring out the specific change types. Deep learning based supervised change detection methods generally require high quality ground truth, which need annotators not only to determine which category each pixel belongs to but to compare different temporal images and analyze the changes. Such a process is too time consuming (Peng et al., 2020). In many applications, we need to identify the change regions as quickly and accurately as possible (e.g., in natural disaster evaluation Lu et al., 2020a). Therefore, weakly supervised change detection is very helpful to solve this problem in remote sensing.

In the literature, there exists three types of weakly supervised change detection methods based on deep learning in remote sensing: (i) semi-supervised deep learning for change detection, (ii) transferred deep learning for change detection and (iii) deep learning under mixed weak supervision for change detection. They are illustrated in Fig. 10. To facilitate conducting the experiments, the datasets for weakly supervised change detection in remote sensing are summarized in Table 6.

4.4.1. Semi-supervised deep learning for change detection

There has been several methods to detect change regions by combining labeled and unlabeled data which can be divided into three types: graph-based methods, methods based on adversarial framework and methods based on self-supervised representation learning. In Gong et al. (2019), a self-supervised representation learning method is presented, where unlabeled data are used to learn an encoder and then labeled data and trained encoder are used to train a classifier to predict change regions. In Liu et al. (2019), a novel unsupervised method is proposed to perform the change detection. The unchanged data can be rearranged in the correct order which makes the identification of the changed pixels easier. In Peng et al. (2020) and Saha et al. (2020b), an adversarial framework is proposed, which exploits not only labeled and unlabeled data but also fake data generated by GAN. Then the adversarial training strategy is used to enhance the performance of change detection. After obtaining initial prediction maps from labeled and unlabeled data, two discriminators are utilized to make the feature distribution of prediction maps consistent between labeled and unlabeled data. A graph model with GANs is presented in Yang et al. (2019a) to solve the semi-supervised change detection problems. First, multi-temporal remote sensing images are transformed into a graph which contains labeled and unlabeled nodes. Then, the unlabeled nodes are assigned labels by a graph learning algorithm which learns the knowledge from labeled and unlabeled nodes. The change regions can be inferred from the graph. In Malkin et al. (2021), multitemporal images are encoded as graphs which are processed through GCN and the information from labeled samples is propagated to the unlabeled ones. Table 7 gives a comparison between different weakly supervised change detection methods. Based on the experiments of Peng et al. (2020), we added two latest semi-supervised change detection (Chen et al., 2021b; Wang et al., 2022).

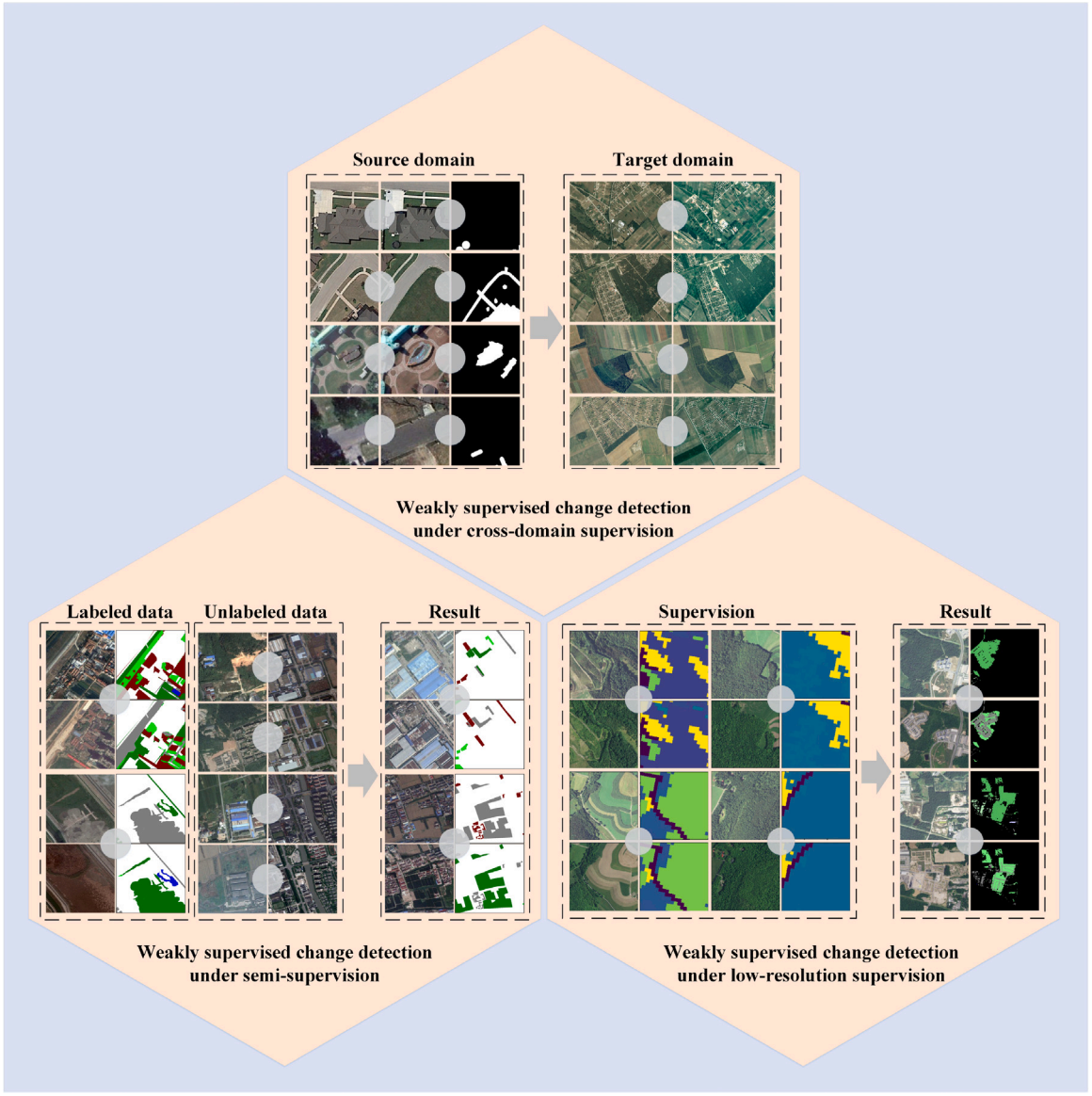


Fig. 10. Overview of weak supervision in the task of change detection.

Table 6
Datasets for weakly supervised change detection in remote sensing.

Datasets	Description	Supervision type	Download link
Modified WHU Building Dataset (Peng et al., 2020)	1922 pairs of 0.075m-resolution samples from 2012 and 2016;the ratio of labeled samples is set differently	Semi-supervision for change detection	https://study.rsgis.whu.edu.cn/pages/download/building_ dataset.html
Mixed transfer dataset (Connors and Vatsavai, 2017)	58514 samples with labels from source domain;103440 samples without labels from target domain	Transferred supervision for change detection	
DFC-MSD dataset (Saha et al., 2020a)	2250 samples with high-resolution images and low-resolution labels; images: two temporal 1m -resolution 4-band images from 2013 and 2017, five temporal 30m-resolution 9-band images from 2013, 2014, 2015, 2016 and 2017; labels: two temporal 30m-resolution landcover labels from 2013 and 2016	Mixed weak supervision for change detection	https://dfc2021.blob.core.windows.net/competition-data/dfc2021_index.txt

4.4.2. *Transferred deep learning for change detection*

Another branch of works in remote sensing change detection exploit transferred deep learning to solve the scarcity of labeled data. In this case, knowledge learned from the source domain samples is transferred to the target domain, which helps fill the missing information. In the field of change detection in remote sensing, the transferred deep learning mainly adopts model transfer methods, namely training a change detection model using source domain data in a fully supervised manner and then exploiting this trained model and target domain unlabeled data to generate change maps. In Connors and Vatsavai (2017) and

Table 7
Quality analysis for semi-supervised change detection.

Method	Labeled Ratio											
	5%			10%			20%			50%		
	F1	OA	Kappa	F1	OA	Kappa	F1	OA	Kappa	F1	OA	Kappa
FC-EF-Res (Gao et al., 2019)	0.7715	0.9116	0.7177	0.8145	0.9328	0.7737	0.8451	0.9467	0.8129	0.8561	0.9513	0.8268
FCN-PP (Hou et al., 2017)	0.7694	0.9212	0.7219	0.8110	0.9351	0.7719	0.8475	0.9512	0.8186	0.8539	0.9504	0.8240
Unet++_att (Saha et al., 2020c)	0.7749	0.9297	0.7308	0.8265	0.9448	0.7938	0.8499	0.9502	0.8200	0.8673	0.9577	0.8458
AdvNet (Daudt et al., 2019)	0.7599	0.9238	0.7149	0.8159	0.9404	0.7839	0.8400	0.9483	0.8093	0.8760	0.9584	0.8510
CycleGAN (Lei et al., 2019)	0.7255	0.9191	0.6794	0.7807	0.9315	0.7405	0.8017	0.9338	0.7620	0.8330	0.9450	0.8000
s4GAN (Zhou et al., 2018)	0.8174	0.9425	0.7836	0.8493	0.9501	0.8194	0.8557	0.9523	0.8272	0.8772	0.9593	0.8528
SemiCDNet (Peng et al., 2020)	0.8290	0.9434	0.7960	0.8528	0.9517	0.8240	0.8657	0.9559	0.8403	0.8774	0.9595	0.8538
CPS (Chen et al., 2021b)	0.7459	0.9207	0.7033	0.8081	0.9330	0.7604	0.8221	0.9407	0.7859	0.8420	0.9483	0.8179
UPL (Wang et al., 2022)	0.8100	0.9395	0.7743	0.8437	0.9475	0.8060	0.8514	0.9519	0.8259	0.8636	0.9574	0.8360

Hung et al. (2018), a similar strategy is adopted in which pre-trained models are fine-tuned by target domain data. In Mondal et al. (2019), the deep network is pretrained under a natural image dataset and then, by utilizing representations generated from the pre-trained CNN, the binarized change maps are obtained by using the salient change region which is generated by using low-rank decomposition and a simple threshold segmentation method. In Mittal et al. (2019), a novel unsupervised change detection method is proposed which can effectively model contextual information and handle the large number of bands by grouping spectral bands into spectral-dedicated band groups.

4.4.3. Deep learning under mixed weak supervision for change detection

The available open source high-resolution change detection datasets are very scarce. However, in remote sensing, there are some existing landcover products, e.g., 30 m national land cover database (NLCD) in the United States and 500 m moderate-resolution imaging spectroradiometer (MODIS) land cover (Saha et al., 2020a). These products do not mark out the change regions directly. Change labels can be produced by comparing landcover maps at different times. But the generated labels in this way are noisy at low-resolution and the result contains many errors. The 2021 IEEE GRSS Data Fusion Contest (Saha et al., 2020a) proposes to train high-resolution change detectors with the input of high-resolution images and the supervision of low-resolution landcover labels. This kind of weak supervision contains inaccurate and inexact supervision and therefore is called mixed weak supervision. And after the contest (Saha et al., 2020a), we believe more and more researches will be made on this field.

5. Potential research directions

Along with the advancement of multiple fundamental research directions such as CV, machine learning and knowledge engineering, there are still some potential research directions that are promising in the field of WSDL-driven RSBD mining. Several potential research directions are discussed in the following.

5.1. Domain knowledge-guided weakly supervised deep learning

As the classical representative of data-driven methods, deep learning has achieved great progress compared with previous traditional methods in many fields. Inevitably, data-driven technique relies heavily on the given data and is susceptible to low-quality data particularly when the supervision is weak and incomplete. Leveraging the domain knowledge in the remote sensing is a potential solution to make up for the weak supervision. For example, as (Karniadakis et al., 2021) summarizes, additional information from enforcing the physical laws and constraints can be integrated with data to improve both the accuracy and the reliability. The prior domain knowledge may compensate for the lack of required information due to the weak supervision in the WSDL. As another research hotspot in the deep learning, knowledge graphs (Sarker et al., 2017; Andrés S. Arvor et al., 2017; Amiri and Farah, 2018) work by representing the domain concepts and relationships as a collection of triples and have strong knowledge representation capabilities and semantic reasoning capabilities. However, how to construct remote sensing knowledge graphs and apply knowledge graphs to WSDL still need much more exploration.

5.2. Multi-modal data classification by weakly supervised deep learning

Compared with full supervision, weak supervision naturally contains less information which can cause undesirable performance. Multimodal data can provide richer information to some extent, which means that multispectral, hyperspectral, SAR, multitemporal and multangular images acquired over the same scene can be used together in the weakly supervised methods (Gómez-Chova et al., 2015). However, images in different modalities present different characteristics. For example, typically, hyperspectral samples have higher spectral resolution but lower spatial resolution compared with optical samples. Even the imaging mode of sensors can be totally different such as in radar and optical images. These factors significantly hinder the fusion of multi-modal samples and make it difficult to reach the expected performance. Furthermore, social media data can also be utilized to fill the information which the earth observation instruments miss due to the limitations in the spatial and temporal resolutions, especially in the applications that need real-time response (Li et al., 2021e). There has been a lot of works about fusing multi-modal data in Gibril et al. (2018), Rasti et al. (2017), Matasci et al. (2015), Ghamisi and Höfle (2016), Rosser et al. (2017) and Huang et al. (2018), but rare methods consider fusing multi-modal data using WSDL.

5.3. Generalized representation learning by self-supervised constraints

The aforementioned weakly supervised methods, including semi-supervised learning, domain adaptation learning and so on, can reduce the cost of labeling to some extent. But these methods are difficult to be applied in general tasks (Li et al., 2021d). Even for the domain adaptation methods, the performance may be poor when the gap between target and source domains is too large. To solve these problems, self-supervised learning can be a fairly good choice. In such methods, first, the parameters of deep networks are learned from plenty of unlabeled data which can cover global areas, multi-resolution, multi-season and multi-spectral images (Li et al., 2021d). Then, the initialized networks are transferred to specific tasks with just a limited number of labeled data (Jing and Tian, 2020). In remote sensing, the existing self-supervised learning methods (Chen and Bruzzone, 2021) are very scarce and mainly focus on scene classification (Zhao et al., 2020; Stojnic and Risojevic, 2021). How to employ self-supervised learning in different remote sensing tasks needs more study.

5.4. Automatic vectorization mapping by error-tolerant deep learning

As vector maps play an important role in survey and remote sensing, automated vectorization mapping has attracted much attention from the researchers (Sirmacek and Unsalan, 2008; Zhang, 1999). However,

these traditional methods cannot achieve desirable results. With the help of deep learning (LeCun et al., 2015), vectorization mapping can reach high performance. Typically, there are two steps before obtaining vector maps, i.e. segmentation and vectorization (Wei and Ji, 2021a). But compared with vector maps, the segmentation maps are less accurate at the edges of objects in the regions of interest. That means, there are many errors in the input data for vectorization. Obviously, the noisy data can significantly affect the performance of vectorization. The error-tolerant deep learning can be an alternative solution. There have been some works about automated vectorization mapping techniques (Wei and Ji, 2021a; Chen et al., 2021c), but how to combine the vectorization and error-tolerant learning need further exploration.

5.5. Global land-cover mapping by weakly supervised deep learning

High-quality global land-cover maps are very important for humans to understand the state of the Earth surface (Robinson et al., 2021). For example, the global land-cover maps are used as the input for many global circulation models for global climate simulations (Mora et al., 2014). In the RSBD era, the increasing remote sensing data and the emerging deep learning methods make it possible to generate the global maps. However, the state-of-the-art classification methods mostly adopt the fully supervised techniques, which means that large number of labels for each remote sensing image is needed. Collecting sufficient labeled data is unaffordable, especially for large-scale mapping (Robinson et al., 2021). That is exactly what WSDL can solve. In the contest (Robinson et al., 2021), low-resolution global maps are used as the weak supervision and high-resolution land-cover maps are generated. And Li et al. (2021a) utilize the representation learning to update the global land-cover maps from the old ones more generally. Hence, WSDL can be expected to be a promising solution to generate the global landcover product. In Paris et al. (2021), an interactive strategy combining the active learning and self-paced learning techniques is proposed to greatly reduce the need for manually labeling.

5.6. Small object recognition by weakly supervised deep learning

For deep learning methods, sufficient and high-quality data are essential to achieve the great performance. However, in many applications (e.g. detection of hot spots, military applications, etc.), high-quality samples are scarce and we must train networks under a small training set (Yang et al., 2019b), which can lead to poor results and overfitting. In this situation, the preferable solution is transfer learning which can learn parameters from plenty of source domain samples and transfer them to the target domain. Moreover, in many applications, the objects of interest are typically small and difficult to identify as they are affected by noise. Even a small amount of noise can greatly influence small object recognition due to the size of objects. Although there has been some works about small target detection (Li and Zhang, 2018; Li et al., 2016b; Dong et al., 2019), how to learn robust networks under various noise conditions (e.g. speckle) for small objects detection deserves further researches.

6. Conclusion

With the arrival of RSBD era, the data acquisition technology develops fast and the volume, variety and velocity of remote sensing data increase rapidly, while the ability of processing such data still remain relatively backward and the acquisition of labels supervision is very laborious. In this review, the challenges and opportunities of information extraction from RSBD have been illustrated and discussed. After that, the concrete achievements of WSDL have been systematically reviewed for four major tasks in remote sensing, including scene classification, change detection, object detection and semantic segmentation. Different approaches have been presented and the advantages and disadvantages analyzed. In order to facilitate researchers to identify open challenges in WSDL, some potential research directions about WSDL in remote sensing have been introduced and discussed.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 41971284, the Fundamental Research Funds for the Central Universities, China under Grant 2042022kf1201, and Zhizhuo Research Fund on Spatial–Temporal Artificial Intelligence under Grant ZZJJ202210.

CRediT authorship contribution statement

Yansheng Li: Writing - review and editing, Funding acquisition. **Xinwei Li:** Writing – original draft, Visualization. **Yongjun Zhang:** Writing - review and editing. **Daifeng Peng:** Writing - review and editing. **Lorenzo Bruzzone:** Writing - review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Alajaji, D., Alhichri, H.S., Ammour, N., Alajlan, N., 2020. Few-shot learning for remote sensing scene classification. In: 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium. M2GARSS, IEEE, pp. 81–84.

Amiri, K., Farah, M., 2018. Graph of concepts for semantic annotation of remotely sensed images based on direct neighbors in rag. Can. J. Remote Sens. 44, 551–574.

Andrés S. Arvor, D., Mougenot, I., Libourel, T., Durieux, L., 2017. Ontology-based classification of remote sensing images using spectral rules. Comput. Geosci. 102, 158–166.

Aygunes, B., Cinbis, R.G., Aksoy, S., 2021. Weakly supervised instance attention for multisource fine-grained object recognition with an application to tree species classification. ISPRS J. Photogramm. Remote Sens. 176, 262–274.

Bansal, A., Sikka, K., Sharma, G., Chellappa, R., Divakaran, A., 2018. Zero-Shot Object Detection. In: Lecture Notes in Computer Science, vol. 11205, pp. 397–414.

Bearman, A., Russakovsky, O., Ferrari, V., Fei-Fei, L., 2016. What's the point: Semantic segmentation with point supervision. In: European Conference on Computer Vision. Springer, pp. 549–565.

Benediktsson, J.A., Pesaresi, M., Amason, K., 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. IEEE Trans. Geosci. Remote Sens. 41, 1940–1949.

Benjdira, B., Bazi, Y., Koubaa, A., Ouni, K., 2019. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. Remote Sens. 11 (1369).

Bhatta, B., 2010. Analysis of Urban Growth and Sprawl from Remote Sensing Data. Springer Science & Business Media.

Bilen, H., Vedaldi, A., 2016. Weakly supervised deep detection networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2846–2854.

Bratasanu, D., Nedelcu, I., Datcu, M., 2010. Bridging the semantic gap for satellite image annotation and automatic mapping applications. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 4, 193–204.

Bruzzone, L., 2019. Multisource labeled data: An opportunity for training deep learning network. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 4799–4802.

Chan, L., Hosseini, M.S., Plataniotis, K.N., 2021. A comprehensive analysis of weakly-supervised semantic segmentation in different image domains. Int. J. Comput. Vis. 129, 361–384.

Chen, Y., Bruzzone, L., 2021. Self-supervised change detection in multi-view remote sensing images. ArXiv preprint arXiv:05969.

Chen, J., He, F., Zhang, Y., Sun, G., Deng, M., 2020a. Spmf-net: Weakly supervised building segmentation by combining superpixel pooling and multi-scale feature fusion. Remote Sens. 12 (1049).

Chen, H., Luo, Y., Cao, L., Zhang, B., Guo, G., Wang, C., Li, J., Ji, R., 2019a. Generalized zero-shot vehicle detection in remote sensing imagery via coarse-to-fine framework. IJCAI 687–693.

Chen, S., Shao, D., Shu, X., Zhang, C., Wang, J., 2020b. Fcc-net: A full-coverage collaborative network for weakly supervised remote sensing object detection. Electronics 9 (1356).

Chen, J., Sun, J., Li, Y., Hou, C., 2021a. Object detection in remote sensing images based on deep transfer learning. Multimedia Tools Appl. 1–17.
Chen, J., Wan, L., Zhu, J., Xu, G., Deng, M., 2019b. Multi-scale spatial and channel-wise attention for improving object detection in remote sensing imagery. IEEE Geosci. Remote Sens. Lett. 17, 681–685.
Chen, H., Wang, Y., Wang, G., Qiao, Y., 2018a. Lstd: A low-shot transfer detector for object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence.
Chen, B., Xu, B., Zhu, Z., Yuan, C., Suen, H.P., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., 2019c. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Sci. Bull. 64, 370–373.
Chen, X., Yuan, Y., Zeng, G., Wang, J., 2021b. Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622.
Chen, Z., Zhang, T., Ouyang, C., 2018b. End-to-end airplane detection using transfer learning in remote sensing images. Remote Sens. 10 (139).
Chen, C., Zhang, B., Su, H., Li, W., Wang, L., 2016. Land-use scene classification using multi-scale completed local binary patterns. Signal Image Video Process. 10, 745–752.
Chen, D., Zhong, Y., Zheng, Z., Ma, A., Lu, X., 2021c. Urban road mapping based on an end-to-end road vectorization mapping network framework. ISPRS J. Photogramm. Remote Sens. 178, 345–365.
Cheng, G., Cai, L., Lang, C., Yao, X., Chen, J., Guo, L., Han, J., 2021a. Spnet: Siamese-prototype network for few-shot remote sensing image scene classification. IEEE Trans. Geosci. Remote Sens.
Cheng, G., Han, J., Zhou, P., Guo, L., 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS J. Photogramm. Remote Sens. 98, 119–132.
Cheng, G., Yan, B., Shi, P., Li, K., Yao, X., Guo, L., Han, J., 2021b. Prototype-cnn for few-shot object detection in remote sensing images. IEEE Trans. Geosci. Remote Sens.
Cheng, G., Yang, C., Yao, X., Guo, L., Han, J., 2018. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns. IEEE Trans. Geosci. Remote Sens. 56, 2811–2821.
Cheng, G., Zhou, P., Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 54, 7405–7415.
Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: Challenges and opportunities. Proc. IEEE 104, 2207–2219.
Connors, C., Vatsavai, R.R., 2017. Semi-supervised deep generative models for change detection in very high resolution imagery. In: 2017 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, IEEE, pp. 1063–1066.
Crowther, T.W., Glick, H.B., Covey, K.R., Bettigole, C., Maynard, D.S., Thomas, S.M., Smith, J.R., Hintler, G., Duguid, M.C., Amatulli, G., 2015. Mapping tree density at a global scale. Nature 525, 201–205.
Dai, J., He, K., Sun, J., 2015. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1635–1643.
Dai, X., Wu, X., Wang, B., Zhang, L., 2019. Semisupervised scene classification for remote sensing images: A method based on convolutional neural networks and ensemble learning. IEEE Geosci. Remote Sens. Lett. 16, 869–873.
Das, A., Chandran, S., 2021. Transfer learning with res2net for remote sensing scene classification. In: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, pp. 796–801.
Daudt, R.C., Saux, B.Le., Boulch, A., Gousseau, Y., 2019. Multitask learning for large-scale semantic change detection. Comput. Vis. Image Underst. 187, 102783.
Deng, X., Huang, J., Rozelle, S., Uchida, E., 2008. Growth, population and industrialization, and urban land expansion of china. J. Urban Econ. 63, 96–115.
Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., Zou, H., 2018. Multi-scale object detection in remote sensing imagery with convolutional neural networks. ISPRS J. Photogramm. Remote Sens. 145, 3–22.
Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J. Photogramm. Remote Sens. 162, 94–114.
Ding, J., Xue, N., Xia, G.S., Bai, X., Yang, W., Yang, M.Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M., 2021. Object detection in aerial images: A large-scale benchmark and challenges. ArXiv preprint arXiv:2102.12219.
Dirscherl, M., Dietz, A.J., Dech, S., Kuenzer, C., 2020. Remote sensing of ice motion in antarctica - a review. Remote Sens. Environ. 237.
Dixit, M., Kwitt, R., Niethammer, M., Vasconcelos, N., 2017. Aga: Attribute-guided augmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7455–7463.
Dong, R.C., Xu, D.Z., Zhao, J., Jiao, L.C., An, J.G., 2019. Sig-nms-based faster r-cnn combining transfer learning for small target detection in vhr optical remote sensing imagery. IEEE Trans. Geosci. Remote Sensing 57, 8534–8545.
Elhoseiny, M., Elfeki, M., 2019. Creativity inspired zero-shot learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5784–5793.
Feng, X., Han, J., Yao, X., Cheng, G., 2020. Progressive contextual instance refinement for weakly supervised object detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. 58, 8002–8012.
Fu, K., Lu, W., Diao, W., Yan, M., Sun, H., Zhang, Y., Sun, X., 2018. Wsf-net: Weakly supervised feature-fusion network for binary segmentation in remote sensing image. Remote Sens. 10 (1970).
Gao, Y., Gao, F., Dong, J., Wang, S., 2019. Transferred deep learning for sea ice change detection from synthetic-aperture radar images. IEEE Geosci. Remote Sens. Lett. 16, 1655–1659.
Gao, F., Yang, Y., Wang, J., Sun, J., Yang, E., Zhou, H., 2018. A deep convolutional generative adversarial networks (dcgans)-based semi-supervised method for object recognition in synthetic aperture radar (sar) images. Remote Sens. 10, 846.
Ghamisi, P., Höfle, X.X., 2016. Hyperspectral and lidar data fusion using extinction profiles and deep convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 3011–3024.
Gibril, M.B.A., Idrees, M.O., Shafri, H.Z.M., Yao, K., 2018. Integrative image segmentation optimization and machine learning approach for high quality land-use and land-cover mapping using multisource remote sensing data. J. Appl. Remote Sens. 12, 016036.
Girshick, R., 2015. Fast r-cnn. Comput. Sci.
Girshick, R., Donahue, J., Darrell, T., Malik, J., 2013. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 580–587.
Gómez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G., 2015. Multimodal classification of remote sensing images: A review and future directions. Proc. IEEE 103, 1560–1584.
Gong, X., Xie, Z., Liu, Y., Shi, X., Zheng, Z., 2018. Deep salient feature based anti-noise transfer network for scene classification of remote sensing imagery. Remote Sens. 10, 410.
Gong, M., Yang, Y., Zhan, T., Niu, X., Li, S., 2019. A generative discriminatory classified network for change detection in multispectral imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 321–333.
Grekousis, G., Mountrakis, G., Kavouras, M., 2015. An overview of 21 global and 43 regional land-cover mapping products. Int. J. Remote Sens. 36, 5309–5335.
Guo, Y., Ding, G., Han, J., Gao, Y., 2017. Synthesizing samples fro zero-shot learning. In: IJCAI.
Guo, D., Xia, Y., Luo, X., 2020. Gan-based semisupervised scene classification of remote sensing image. IEEE Geosci. Remote Sens. Lett.
Han, W., Feng, R., Wang, L., Cheng, Y., 2018. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. ISPRS J. Photogramm. Remote Sens. 145, 23–43.
He, N., Fang, L., Li, S., Plaza, A., Plaza, J., 2018. Remote sensing scene classification using multilayer stacked covariance pooling. IEEE Trans. Geosci. Remote Sens. 56, 6899–6910.
Hong, W.X., Wang, Z.Z., Yang, M., Yuan, J.S., 2018. Conditional generative adversarial network for structured domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1335–1344.
Hou, B., Wang, Y., Liu, Q., 2017. Change detection based on deep features and low rank. IEEE Geosci. Remote Sens. Lett. 14, 2418–2422.
Hsieh, T.I., Lo, Y.C., Chen, H.T., Liu, T.L., 2019. One-shot object detection with co-attention and co-excitation. ArXiv preprint arXiv:1911.12529.
Hu, F., Xia, G.S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens. 7, 14680–14707.
Hua, Y., Lobry, S., Mou, L., Tuia, D., Zhu, X.X., 2020. Learning multi-label aerial image classification under label noise: A regularization approach using word embeddings. In: IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 525–528.
Hua, Y., Mou, L., Lin, J., Heidler, K., Zhu, X.X., 2021. Aerial scene understanding in the wild: Multi-scene recognition via prototype-based memory networks. ISPRS J. Photogramm. Remote Sens. 177, 89–102.
Huang, X., Wang, C., Li, Z., 2018. A near real-time flood-mapping approach by integrating social media and post-event satellite imagery. Ann. GIS 24, 113–123.
Huang, F., Yu, Y., Feng, T., 2019. Hyperspectral remote sensing image change detection based on tensor and deep learning. J. Vis. Commun. Image Represent. 58, 233–244.
Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H., 2018. Adversarial learning for semi-supervised semantic segmentation. ArXiv preprint arXiv:1802.07934.
Ibrahim, M.S., Vahdat, A., Ranjbar, M., Macready, W.G., 2020. Semi-supervised semantic image segmentation with self-correcting networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12715–12725.
Ji, S., Wang, D., Luo, M., 2020. Generative adversarial network-based full-space domain adaptation for land cover classification from multiple-source remote sensing images. IEEE Trans. Geosci. Remote Sens. 59, 3816–3828.
Jia, X., Kuo, B.C., Crawford, M.M., 2013. Feature mining for hyperspectral image classification. Proc. IEEE 101, 676–697.
Jiao, L., Zhao, J., 2019. A survey on the new generation of deep learning in image processing. IEEE Access 7, 172231–172263.
Jin, P., Xia, G.S., Hu, F., Lu, Q., Zhang, L., 2018. Aid++: An updated version of aid on scene classification. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 4721–4724.
Jing, L., Tian, Y., 2020. Self-supervised visual feature learning with deep neural networks: A survey. IEEE Trans. Pattern Anal. Mach. Intell.

Kang, J., Fernandez-Beltran, R., Duan, P., Kang, X., Plaza, A.J., 2020. Robust normalized softmax loss for deep metric learning-based characterization of remote sensing images with label noise. IEEE Trans. Geosci. Remote Sens.

Kang, J., Fernandez-Beltran, R., Kang, X., Ni, J., Plaza, A., 2021. Noise-tolerant deep neighborhood embedding for remotely sensed images with label noise. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 2551–2562.

Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., Darrell, T., 2019. Few-shot object detection via feature reweighting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8420–8429.

Karlinsky, L., Shtok, J., Harary, S., Schwartz, E., Aides, A., Feris, R., Giryes, R., Bronstein, A.M., 2019. Repmet: Representative-based metric learning for classification and few-shot object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5197–5206.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nat. Rev. Phys. 3, 422–440.

Kellenberger, B., Marcos, D., Tuia, D., 2019. When a few clicks make all the difference: improving weakly-supervised wildlife detection in uav images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0.

Kemker, R., Salvaggio, C., Kanan, C., 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS J. Photogramm. Remote Sens. 145, 60–77.

Khoreva, A., Benenson, R., Hosang, J., Hein, M., Schiele, B., 2017. Simple does it: Weakly supervised instance and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 876–885.

Kodirov, E., Xiang, T., Gong, S., 2017. Semantic autoencoder for zero-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3174–3183.

Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B., 2013. One-shot learning by inverting a compositional causal process. Adv. Neural Inf. Process. Syst.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lee, D.H., 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on Challenges in Representation Learning. ICML, p. 896.

Lei, T., Zhang, Y., Lv, Z., Li, S., Liu, S., Nandi, A.K., 2019. Landslide inventory mapping from bitemporal images using deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 16, 982–986.

Li, W., Chen, K., Chen, H., Shi, Z., 2021a. Geographical knowledge-driven representation learning for remote sensing images. ArXiv preprint arXiv:2107.05276.

Li, Y., Chen, W., Huang, X., Gao, Z., Li, S., He, T., Zhang, Y., 2023a. Mfvnet: a deep adaptive fusion network with multiple field-of-views for remote sensing image semantic segmentation. Sci. China Inf. Sci. 66, 140305.

Li, Y., Chen, W., Zhang, Y., Tao, C., Xiao, R., Tan, Y., 2020a. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. Remote Sens. Environ. 250, 112045.

Li, K., Cheng, G., Bu, S., You, X., 2017a. Rotation-insensitive and context-augmented object detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. 56, 2337–2348.

Li, Y., Dang, B., Li, W., Zhang, Y., 2023b. Glh-water: A large-scale dataset for global surface water detection in large-size very-high-resolution satellite imagery. ArXiv preprint arXiv:2303.09310.

Li, X., Deng, J., Fang, Y., 2021b. Few-shot object detection on remote sensing images. IEEE Trans. Geosci. Remote Sens.

Li, H., Dou, X., Tao, C., Hou, Z., Chen, J., Peng, J., Deng, M., Zhao, L., 2017b. Rsi-cb: A large scale remote sensing image classification benchmark via crowdsource data. ArXiv preprint arXiv:1705.10450.

Li, Y., Huang, X., Liu, H., 2017c. Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images. Photogramm. Eng. Remote Sens. 83, 567–579.

Li, Y., Kong, D., Zhang, Y., Ji, Z., Xiao, R., 2020b. Zero-shot remote sensing image scene classification based on robust cross-domain mapping and gradual refinement of semantic space. Acta Geod. Cartogr. Sin. 49, 1564.

Li, Y., Kong, D., Zhang, Y., Tan, Y., Chen, L., 2021c. Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification. ISPRS J. Photogramm. Remote Sens. 179, 145–158.

Li, H., Li, Y., Zhang, G., Liu, R., Huang, H., Zhu, Q., Tao, C., 2021d. Remote sensing images semantic segmentation with general remote sensing vision model via a self-supervised contrastive learning method. ArXiv preprint arXiv:2106.10605.

Li, J., Liu, Z., Lei, X., Wang, L., 2021e. Distributed fusion of heterogeneous remote sensing and social media data: A review and new developments. Proc. IEEE.

Li, A., Lu, Z., Wang, L., Xiang, T., Wen, J.R., 2017d. Zero-shot scene classification for high spatial resolution remote sensing images. IEEE Trans. Geosci. Remote Sens. 55, 4157–4167.

Li, Y., Ma, J., Zhang, Y., 2021f. Image retrieval from remote sensing big data: A survey. Inf. Fusion 67, 94–115.

Li, Y., Ouyang, S., Zhang, Y., 2022a. Combining deep learning and ontology reasoning for remote sensing image semantic segmentation. Knowl.-Based Syst. 243, 108469.

Li, Y., Shi, T., Zhang, Y., Chen, W., Wang, Z., Li, H., 2021g. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. ISPRS J. Photogramm. Remote Sens. 175, 20–33.

Li, Y., Tao, C., Tan, Y., Shang, K., Tian, J., 2016a. Unsupervised multilayer feature learning for satellite image scene classification. IEEE Geosci. Remote Sens. Lett. 13, 157–161.

Li, K., Wan, G., Cheng, G., Meng, L., Han, J., 2020c. Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS J. Photogramm. Remote Sens. 159, 296–307.

Li, Y., Wang, D., Hu, H., Lin, Y., Zhuang, Y., 2017e. Zero-shot recognition using dual visual-semantic mapping paths. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3279–3287.

Li, Y., Wei, F., Zhang, Y., Chen, W., Ma, J., 2023c. Hs2p: Hierarchical spectral and structure-preserving fusion network for multimodal remote sensing image cloud and shadow removal. Inf. Fusion 94, 215–228.

Li, Z.H., Yao, L.N., Zhang, X.Q., Wang, X.Z., Kanhere, S., Zhang, H.X., 2019. Zero-shot object detection with textual descriptions. In: Thirty-Third Aaai Conference on Artificial Intelligence / Thirty-First Innovative Applications of Artificial Intelligence Conference / Ninth Aaai Symposium on Educational Advances in Artificial Intelligence. AAAI.

Li, Y., Ye, D., 2018. Greedy annotation of remote sensing image scenes based on automatic aggregation via hierarchical similarity diffusion. IEEE Access 6, 57376–57388.

Li, Y., Zhang, Y., 2018. Robust infrared small target detection using local steering kernel reconstruction. Pattern Recognit. 77, 113–125.

Li, Y., Zhang, Y.J., Huang, X., Ma, J.Y., 2018a. Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval. IEEE Trans. Geosci. Remote Sens. 56, 6521–6536.

Li, Y., Zhang, Y., Huang, X., Yuille, A.L., 2018b. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. ISPRS J. Photogramm. Remote Sens. 146, 182–196.

Li, Y., Zhang, Y.J., Huang, X., Zhu, H., Ma, J.Y., 2018c. Large-scale remote sensing image retrieval by deep hashing neural networks. IEEE Trans. Geosci. Remote Sens. 56, 950–965.

Li, Z., Zhang, X., Xiao, P., Zheng, Z., 2021h. On the effectiveness of weakly supervised semantic segmentation for building extraction from high-resolution remote sensing imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 3266–3281.

Li, Y., Zhang, Y., Yu, J.G., Tan, Y., Tian, J., Ma, J., 2016b. A novel spatio-temporal saliency approach for robust dim moving target detection from airborne infrared image sequences. Inform. Sci. 369, 548–563.

Li, Y., Zhang, Y., Zhu, Z., 2021i. Error-tolerant deep learning for remote sensing image scene classification. IEEE Trans. Cybern. 51, 1756–1768.

Li, Y., Zhou, Y., Zhang, Y., Zhong, L., Wang, J., Chen, J., 2022b. Dkdfn: Domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. ISPRS J. Photogramm. Remote Sens. 186, 170–189.

Li, Y., Zhu, Z., Yu, J.G., Zhang, Y., 2021j. Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification. IEEE Trans. Geosci. Remote Sens.

Lian, R., Huang, L., 2021. Weakly supervised road segmentation in high-resolution remote sensing images using point annotations. IEEE Trans. Geosci. Remote Sens.

Lin, D., Dai, J., Jia, J., He, K., Sun, J., 2016. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3159–3167.

Liu, J., Chen, K., Xu, G., Li, H., Yan, M., Diao, W., Sun, X., 2019. Semi-supervised change detection based on graphs with generative adversarial networks. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 74–77.

Long, Y., Liu, L., Shao, L., Shen, F., Ding, G., Han, J., 2017. From zero-shot learning to conventional supervised classification: Unseen visual data synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1627–1636.

Lu, N., Chen, C., Shi, W., Zhang, J., Ma, J., 2020a. Weakly supervised change detection based on edge mapping and sdae network in high-resolution remote sensing images. Remote Sens. 12 (3907).

Lu, X.Q., Gong, T.F., Zheng, X.T., 2020b. Multisource compensation network for remote sensing cross-domain scene classification. IEEE Trans. Geosci. Remote Sens. 58, 2504–2515.

Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., Ma, Y., 2022. Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer. IEEE/CAA J. Autom. Sin. 9, 1200–1217.

Malkin, N., Robinson, C., Jojic, N., 2021. High-resolution land cover change from low-resolution labels: Simple baselines for the 2021 ieee grss data fusion contest. ArXiv preprint arXiv:2101.01154.

Martinuzzi, S., Gould, W.A., González, O.M.R., 2007. Land development, land use, and urban sprawl in puerto rico integrating remote sensing and population census data. Landsc. Urban Plan. 79, 288–297.

Matasci, G., Volpi, M., Kanevski, M., Bruzzone, L., Tuia, D., 2015. Semisupervised transfer component analysis for domain adaptation in remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 53, 3550–3564.

Mittal, S., Tatarchenko, M., Brox, T., 2019. Semi-supervised semantic segmentation with high-and low-level consistency. IEEE Trans. Pattern Anal. Mach. Intell.

Mohanty, S.P., Czakon, J., Kaczmarek, K.A., Pyskir, A., Tarasiewicz, P., Kunwar, S., Rohrbach, J., Luo, D., Prasad, M., Fleer, S., 2020. Deep learning for understanding satellite imagery: An experimental survey. Front. Artif. Intell. 3.

Mondal, A.K., Agarwal, A., Dolz, J., Desrosiers, C., 2019. Revisiting cyclegan for semi-supervised segmentation. ArXiv preprint arXiv:1908.11569.

Mora, B., Tsendbazar, N.E., Herold, M., Arino, O., 2014. Global Land Cover Mapping: Current Status and Future Trends. Springer, pp. 11–30.

Otter, D.W., Medina, J.R., Kalita, J.K., 2020. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Netw. Learn. Syst. 32, 604–624.

Pan, B., Tai, J., Zheng, Q., Zhao, S., 2017. Cascade convolutional neural network based on transfer-learning for aircraft detection on high-resolution remote sensing images. J. Sensors 2017.

Papandreou, G., Chen, L.C., Murphy, K.P., Yuille, A.L., 2015. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1742–1750.

Paris, C., Orlandi, L., Bruzzone, L., 2021. An interactive strategy for the training set definition based on active self-paced learning implemented on a cloud-computing platform. IEEE Geosci. Remote Sens. Lett.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Sicre, C.Marais., Dedieu, G., 2017. Effect of training class label noise on classification performances for land cover mapping with satellite image time series. Remote Sens. 9, 173.

Peng, D., Bruzzone, L., Zhang, Y., Guan, H., Ding, H., Huang, X., 2020. Semicdnet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. IEEE Trans. Geosci. Remote Sens.

Perantoni, G., Bruzzone, L., 2021. A novel technique for robust training of deep networks with multisource weak labeled remote sensing data. IEEE Trans. Geosci. Remote Sens.

Qiao, R., Ghodsi, A., Wu, H., Chang, Y., Wang, C., 2020. Simple weakly supervised deep learning pipeline for detecting individual red-attacked trees in vhr remote sensing images. Remote Sens. Lett. 11, 650–658.

Quan, J., Wu, C., Wang, H., Wang, Z., 2018. Structural alignment based zero-shot classification for remote sensing scenes. In: 2018 IEEE International Conference on Electronics and Communication Engineering. ICECE, IEEE, pp. 17–21.

Rafique, M.U., Jacobs, N., 2019. Weakly supervised building segmentation from aerial images. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 3955–3958.

Rahman, S., Khan, S., Porikli, F., 2019. Zero-Shot Object Detection: Learning to Simultaneously Recognize and Localize Novel Concepts. In: Lecture Notes in Computer Science, vol. 11361, pp. 547–563.

Randin, C.F., Ashcroft, M.B., Bolliger, J., Cavender-Bares, J., Coops, N.C., Dullinger, S., Dirnbock, T., Eckert, S., Ellis, E., Fernandez, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., Price, B., Rocchini, D., Schaepman, M., Schmid, B., Verburg, P., Wilson, A., Woodcock, P., Yoccoz, N., Payne, D., 2020. Monitoring biodiversity in the anthropocene using remote sensing in species distribution models. Remote Sens. Environ. 239.

Rasti, B., Ghamisi, P., Gloaguen, R., 2017. Hyperspectral and lidar fusion using extinction profiles and total variation component analysis. IEEE Trans. Geosci. Remote Sens. 55, 3997–4007.

Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput. 29, 2352–2449.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., 2019. Deep learning and process understanding for data-driven earth system science. Nature 566, 195–204.

Robinson, C., Malkin, K., Jojic, N., Chen, H., Qin, R., Xiao, C., Schmitt, M., Ghamisi, P., Hänsch, N., 2021. Global land-cover mapping with weak supervision: Outcome of the 2020 ieee grss data fusion contest. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 3185–3199.

Rosser, J.F., Leibovici, D., Jackson, M., 2017. Rapid flood inundation mapping using social media, remote sensing and topographic data. Nat. Hazards 87, 103–120.

Roy, S., Sangineto, E., Sebe, N., Demir, B., 2018. Semantic-fusion gans for semi-supervised satellite image classification. In: 2018 25th IEEE International Conference on Image Processing. ICIP, IEEE, pp. 684–688.

Saha, S., Bovolo, F., Bruzzone, L., 2020a. Change detection in image time-series using unsupervised lstm. IEEE Geosci. Remote Sens. Lett.

Saha, S., Mou, L., Zhu, X.X., Bovolo, F., Bruzzone, L., 2020b. Semisupervised change detection using graph convolutional network. IEEE Geosci. Remote Sens. Lett. 18, 607–611.

Saha, S., Solano-Correa, Y.T., Bovolo, F., Bruzzone, L., 2020c. Unsupervised deep transfer learning-based change detection for hr multispectral images. IEEE Geosci. Remote Sens. Lett. 18, 856–860.

Saito, K., Watanabe, K., Ushiku, Y., Harada, T., 2018. Maximum classifier discrepancy for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3723–3732.

Sarker, M.K., Xie, N., Doran, D., Raymer, M., Hitzler, P., 2017. Explaining trained neural networks with semantic web technologies: First steps. ArXiv preprint arXiv:1710.04324.

Schmieder, M., Holl, F., Fotteler, M.L., Ortl, M., Buchner, E., Swoboda, W., 2020. Remote sensing and on-site characterization of wetlands as potential habitats for malaria vectors - A pilot study in southern Germany. In: IEEE Global Humanitarian Technology Conference Proceedings. IEEE.

Schmitt, M., Hughes, L.H., Qiu, C., Zhu, X.X., 2019. Sen12ms–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. ArXiv preprint arXiv:1906.07789.

Schmitt, M., Prexl, J., Ebel, P., Liebel, L., Zhu, X.X., 2020. Weakly supervised semantic segmentation of satellite images for land cover mapping–challenges and opportunities. ArXiv preprint arXiv:2002.08254.

Shimoda, W., Yanai, K., 2016. Distinct class-specific saliency maps for weakly supervised semantic segmentation. In: European Conference on Computer Vision. Springer, pp. 218–234.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., 2016. Mastering the game of go with deep neural networks and tree search. Nature 529, 484–489.

Singh, A., Bruzzone, L., 2021. Sigan: Spectral index generative adversarial network for data augmentation in multispectral remote sensing images. IEEE Geosci. Remote Sens. Lett.

Sirmacek, B., Unsalan, C., 2008. Building detection from aerial images using invariant color features and shadow information. In: 2008 23rd International Symposium on Computer and Information Sciences. IEEE, pp. 1–5.

Song, H., Kim, M., Park, D., Shin, Y., Lee, J.G., 2020. Learning from noisy labels with deep neural networks: A survey. ArXiv preprint arXiv:2007.08199.

Song, Q., Xu, F., 2017. Zero-shot learning of sar target feature space with deep generative neural networks. IEEE Geosci. Remote Sens. Lett. 14, 2245–2249.

Song, S., Yu, H., Miao, Z., Zhang, Q., Lin, Y., Wang, S., 2019. Domain adaptation for convolutional neural networks-based remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. 16, 1324–1328.

Stojnic, V., Risojevic, V., 2021. Self-supervised learning of remote sensing scene representations using contrastive multiview coding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1182–1191.

Sumbul, G., Cinbis, R.G., Aksoy, S., 2017. Fine-grained object recognition and zero-shot learning in remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 56, 770–779.

Tang, L., Deng, Y., Ma, Y., Huang, J., Ma, J., 2022. Superfusion: A versatile image registration and fusion network with semantic awareness. IEEE/CAA J. Autom. Sin. 9, 2121–2137.

Tang, P., Wang, X., Bai, X., Liu, W., 2017. Multiple instance detection network with online instance classifier refinement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2843–2851.

Tang, P., Wang, X., Bai, S., Shen, W., Bai, X., Liu, W., Yuille, A., 2018. Pcl: Proposal cluster learning for weakly supervised object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42, 176–191.

Tao, C., Qi, J., Lu, W., Wang, H., Li, H., 2020. Remote sensing image scene classification with self-supervised paradigm under limited labeled samples. IEEE Geosci. Remote Sens. Lett.

Tao, S.Y., Yeh, Y.R., Wang, Y.C.F., 2017. Semantics-preserving locality embedding for zero-shot learning. In: BMVC.

Thoonen, G., Mahmood, Z., Peeters, S., Scheunders, P., 2011. Multisource classification of color and hyperspectral images using color attribute profiles and composite decision fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5, 510–521.

Tong, W., Chen, W., Han, W., Li, X., Wang, L., 2020a. Channel-attention-based densenet network for remote sensing image scene classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4121–4132.

Tong, X.Y., Xia, G.S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L., 2020b. Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sens. Environ. 237, 111322.

Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M., 2018. Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7472–7481.

Tu, B., Kuang, W., He, W., Zhang, G., Peng, Y., 2020. Robust learning of mislabeled training samples for remote sensing image scene classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 5623–5639.

Tuia, D., Persello, C., Bruzzone, L., 2021. Recent advances in domain adaptation for the classification of remote sensing data. ArXiv preprint arXiv:2104.07778.

Vargas-Munoz, J.E., Srivastava, S., Tuia, D., Falcao, A.X., 2020. Openstreetmap: Challenges and opportunities in machine learning and remote sensing. IEEE Geosci. Remote Sens. Mag. 9, 184–199.

Wan, F., Wei, P., Jiao, J., Han, Z., Ye, Q., 2018. Min-entropy latent model for weakly supervised object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1297–1306.

Wang, S., Chen, W., Xie, S.M., Azzari, G., Lobell, D.B., 2020. Weakly supervised deep learning for segmentation of remote sensing imagery. Remote Sens. 12 (207).

Wang, M., Deng, W., 2018. Deep visual domain adaptation: A survey. Neurocomputing 312, 135–153.

Wang, Y.X., Ramanan, D., Hebert, M., 2019a. Meta-learning to detect rare objects. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9925–9934.

Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X., 2022. Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4248–4257.

Wang, Q., Zhang, X., Chen, G., Dai, F., Gong, Y., Zhu, K., 2018. Change detection based on faster r-cnn for high-resolution remote sensing images. Remote Sens. Lett. 9, 923–932.

Wang, T., Zhang, X., Yuan, L., Feng, J., 2019b. Few-shot adaptive faster r-cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7173–7182.

Wei, S., Ji, S., 2021a. Graph convolutional networks for the automated production of building vector maps from aerial images. IEEE Trans. Geosci. Remote Sens.

Wei, Y., Ji, S., 2021b. Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images. IEEE Trans. Geosci. Remote Sens.

Wei, Y., Liang, X., Chen, Y., Jie, Z., Xiao, Y., Zhao, Y., Yan, S., 2016a. Learning to segment with image-level annotations. Pattern Recognit. 59, 234–244.

Wei, Y., Liang, X., Chen, Y., Shen, X., Cheng, M.M., Feng, J., Zhao, Y., Yan, S., 2016b. Stc: A simple to complex framework for weakly-supervised semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2314–2320.

Wei, H., Ma, L., Liu, Y., Du, Q., 2021. Combining multiple classifiers for domain adaptation of remote sensing image classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 1832–1847.

Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., Huang, T.S., 2018. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7268–7277.

Weikmann, G., Paris, C., Bruzzone, L., 2021. Timesen2crop: A million labeled samples dataset of sentinel 2 image time series for crop-type classification. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 14, 4699–4708.

Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. Remote Sens. 11, 1309.

Wu, Z., Sun, J., Zhang, Y., Wei, Z., Chanussot, J., 2021. Recent developments in parallel and distributed computing for remotely sensed big data processing. Proc. IEEE.

Wu, Z.Z., Weise, T., Wang, Y., Wang, Y., 2020. Convolutional neural network based weakly supervised learning for aircraft detection from remote sensing image. IEEE Access 8, 158097–158106.

Xia, G.S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X., 2017. Aid: A benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. 55, 3965–3981.

Xia, G.S., Wang, Z., Xiong, C., Zhang, L., 2015. Accurate annotation of remote sensing images via active spectral clustering with little expert knowledge. Remote Sens. 7, 15014–15045.

Xiao, Z., Qi, J., Xue, W., Zhong, P., 2021. Few-shot object detection with self-adaptive attention network for remote sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 4854–4865.

Xiao, Z., Zhong, P., Quan, Y., Yin, X., Xue, W., 2020. Few-shot object detection with feature attention highlight module in remote sensing images. In: 2020 International Conference on Image, Video Processing and Artificial Intelligence. International Society for Optics and Photonics, p. 115840Z.

Xue, W., Dai, X., Liu, L., 2020. Remote sensing scene classification based on multi-structure deep features fusion. IEEE Access 8, 28746–28755.

Xue, B., Tong, N., 2019. Diod: Fast and efficient weakly semi-supervised deep complex isar object detection. IEEE Trans. Cybern. 49, 3991–4003.

Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., Lin, L., 2019a. Meta r-cnn: Towards general solver for instance-level low-shot learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9577–9586.

Yan, L., Fan, B., Liu, H., Huo, C., Xiang, S., Pan, C., 2019b. Triplet adversarial domain adaptation for pixel-level classification of vhr remote sensing images. IEEE Trans. Geosci. Remote Sens. 58, 3558–3573.

Yan, L., Fan, B., Xiang, S., Pan, C., 2018. Adversarial domain adaptation with a domain similarity discriminator for semantic segmentation of urban areas. In: 2018 25th IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1583–1587.

Yan, L., Fan, B., Xiang, S., Pan, C., 2021. Cmt: Cross mean teacher unsupervised domain adaptation for vhr image semantic segmentation. IEEE Geosci. Remote Sens. Lett.

Yan, C.X., Zheng, Q.H., Chang, X.J., Luo, M.N., Yeh, C.H., Hauptman, A.G., 2020. Semantics-preserving graph propagation for zero-shot object detection. IEEE Trans. Image Process. 29, 8163–8176.

Yang, M., Jiao, L., Liu, F., Hou, B., Yang, S., 2019a. Transferred deep learning-based change detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. 57, 6960–6973.

Yang, Y., Newsam, S., 2008. Comparing sift descriptors and gabor texture features for classification of remote sensed imagery. In: 2008 15th IEEE International Conference on Image Processing. IEEE, pp. 1852–1855.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 270–279.

Yang, Z., Yu, W., Liang, P., Guo, H., Xia, L., Zhang, F., Ma, Y., Ma, J., 2019b. Deep transfer learning for military object recognition under small training set condition. Neural Comput. Appl. 31, 6469–6478.

Yao, X., Feng, X., Han, J., Cheng, G., Guo, L., 2020. Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning. IEEE Trans. Geosci. Remote Sens. 59, 675–685.

Yuan, Z., Huang, W., Li, L., Luo, X., 2020. Few-shot scene classification with multi-attention deepemd network in remote sensing. IEEE Access 9, 19891–19901.

Zeng, Y., Zhuge, Y., Lu, H., Zhang, L., 2019. Joint learning of saliency detection and weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7223–7233.

Zhan, Y., Hu, D., Wang, Y., Yu, X., 2017. Semisupervised hyperspectral image classification based on generative adversarial networks. IEEE Geosci. Remote Sens. Lett. 15, 212–216.

Zhang, Y., 1999. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. ISPRS J. Photogramm. Remote Sens. 54, 50–60.

Zhang, P., Bai, Y., Wang, D., Bai, B., Li, Y., 2021a. Few-shot classification of aerial scene images via meta-learning. Remote Sens. 13, 108.

Zhang, B., Chen, Z., Peng, D., Benediktsson, J.A., Liu, B., Zou, L., Li, J., Plaza, A., 2019a. Remotely sensed big data: Evolution in model development for information extraction [point of view]. Proc. IEEE 107, 2294–2301.

Zhang, F., Du, B., Zhang, L., Xu, M., 2016a. Weakly supervised learning based on coupled convolutional neural networks for aircraft detection. IEEE Trans. Geosci. Remote Sens. 54, 5553–5563.

Zhang, L., Ma, J., 2021. Salient object detection based on progressively supervised learning for remote sensing images. IEEE Trans. Geosci. Remote Sens.

Zhang, L., Ma, J., Lv, X., Chen, D., 2019b. Hierarchical weakly supervised learning for residential area semantic segmentation in remote sensing images. IEEE Geosci. Remote Sens. Lett. 17, 117–121.

Zhang, W., Tang, P., Corpetti, T., Zhao, L., 2021b. Wts: A weakly towards strongly supervised learning framework for remote sensing land cover classification using segmentation models. Remote Sens. 13, 394.

Zhang, H., Xu, H., Tian, X., Jiang, J., Ma, J., 2021c. Image fusion meets deep learning: A survey and perspective. Inf. Fusion 76, 323–336.

Zhang, K., Yang, H., 2020. Semi-supervised multi-spectral land cover classification with multi-attention and adaptive kernel. In: 2020 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1881–1885.

Zhang, L., Zhang, L., Du, B., 2016b. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4, 22–40.

Zhao, J., Guo, W., Liu, B., Cui, S., Zhang, Z., Yu, W., 2017. Convolutional neural network-based sar image classification with noisy labels. J. Radars 6, 514–523.

Zhao, Z., Luo, Z., Li, J., Chen, C., Piao, Y., 2020. When self-supervised learning meets scene classification: Remote sensing scene classification based on a multitask learning framework. Remote Sens. 12, 3276.

Zhou, Z.H., 2018. A brief introduction to weakly supervised learning. Natl. Sci. Rev. 5, 44–53.

Zhou, P., Cheng, G., Liu, Z., Bu, S., Hu, X., 2016a. Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. Multidimens. Syst. Signal Process. 27, 925–944.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016b. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2921–2929.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. Unet++: A Nested U-Net Architecture for Medical Image Segmentation. Springer, pp. 3–11.

Zhu, J., Shi, Q., Chen, F., Shi, X., Do, Z., Qin, Q., 2016. Research status and development trends of remote sensing big data. J. Image Graph. 21, 1425–1439.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36.

Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. 12, 2321–2325.