A Weakly Supervised Semantic Segmentation Framework for Medium-resolution Forest Classification with Noisy Labels and GF-1 WFV Images

Xueli Peng, Guojin He, Guizhou Wang, Ranyu Yin, and Jianping Wang

***Abstract*—Forests are the most widely distributed terrestrial vegetation type and play a significant role in the global carbon cycle and ecological diversity. Accurate and timely forest detection provides essential data for forest management and development. Current forest-related products differ in definition, accuracy, and spatial consistency, making them difficult to use. Therefore, it is necessary to map forest cover under a unified framework. However, detecting forests on a large scale requires high-quality and representative samples, which can be challenging. This study proposes a weakly supervised forest classification framework (WSFCF) that uses noisy labels. The WSFCF is designed to address label generation, correction, and sample location optimization. We employ a spectral-spatial network to extract forest cover accurately for medium-resolution forest classification. The experimental results show that the proposed method outperforms the compared methods, achieving an accuracy of 91.76% OA and 88.28% F1 score on 110 GF-1 WFV images. This supports the subsequent extraction of national-scale forest cover and encourages the mapping of China's forest cover using GF-1 WFV images. Moreover, the proposed method produces satisfactory outcomes for objects such as water, farmland, and built-up areas within the study area, demonstrating its effectiveness and potential for transferability.**

***Index Terms*—Weakly supervised, noisy learning, Forest classification, GF-1 WFV, LULC products.**

This work is supported by the Second Tibetan Plateau Scientific Expedition and Research Program under grant 2019QZKK030701, the Strategic Priority Research Program of the Chinese Academy of Sciences under grant XDA19090300, and the program of the National Natural Science Foundation of China under grant 61731022. *(Corresponding author: Guojin He.)*

Xueli Peng, Guojin He and Guizhou Wang are with Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and are also with University of Chinese Academy of Sciences, Beijing 100049, China (email: pengxueli211@mails.ucas.ac.cn; hegj@aircas.ac.cn; wanggz@aircas.ac.cn).

Guojin He is also with Zhongke Aerospace Information Research Institute of Kashgar, Kashgar 844000, China and Zhongke Aerospace Information Research Institute of Kashgar, Kashgar 844000, China

Ranyu Yin is with Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (yinry@aircas.ac.cn)

Jianping Wang is with Zhongke Aerospace Information Research Institute of Kashgar, Kashgar 844000, China (wangjp@aircas.ac.cn)

I. INTRODUCTION

FORESTS are an essential component of terrestrial ecosystems and are closely linked to the global carbon cycle, hydrological cycle and biodiversity. The protection and monitoring of forests is highly emphasized by the international community [1], but forest change and degradation have continued to occur at an alarming rate over the past three decades [2]. The capability for accurate and timely forest mapping is essential for the sustainable exploitation and management of forests [3]. Remote sensing technology takes an increasingly important role in forest mapping by virtue of its powerful data acquisition capability and the advantage of being able to repeat observations in a relatively short period [4].

Over the past decades, scholars have conducted substantial research on forest monitoring with remote sensing technologies, yielding vital processes [1, 5, 6]. The TreeCover for the year 2000 [1] is part of the Global Forest Change (GFC) product which is created based on Landsat satellite images. The GFC delivers the forest change maps for each year after 2000 besides TreeCover. However, obtaining the forest cover maps for each year after 2000 based on GFC is difficult. The FNF [5] produced based on ALOS PALSAR and the threshold segmentation method donates the global forest cover maps with 25m spatial resolution from 2007-2010. Following the characteristics of different vegetation zones, Zhang et al. [6] generated a global forest cover product on the Google Earth Engine platform with the RF algorithm and Landsat satellite archive. Except for the forest thematic products, forests are also reflected in all forest-related layers of land cover and land use (LULC) products, such as the forest in GlobeLand30 [7], FROM-GLC [8, 9] and GLC_FCS30 [10] and TreeCover in WorldCover [11] and Esri Land Cover [12]. Such products deliver basic data for forest research, but they are difficult to use in practice because of the large uncertainties in terms of definition, accuracy, area estimation and spatial consistency [13].

Remote sensing monitoring of forests is continually evolving; however, immense challenges remain [14] such as the need for standardization of satellite data , the need for validation data, and the need for highly automated forest monitoring procedures. These challenges have been mitigated

to some extent with the increase in computer processing power, advances in sensor technology, and improvements in image processing algorithms.

Advances in classification methods have improved the accuracy of classification results[8]. The traditional remote sensing-based forest mapping mostly adopt shallow machine learning methods [15], including decision tree (DT) [16], random forest (RF) [17, 18], multilayer perceptual machines, support vector machines, and maximum likelihood classification. In particular, DT and RF [19, 20] are widely used in forest classification for their robustness to noise and their advantages in handling multidimensional data [21, 22]. Classical classification algorithms rely on manually selecting features (e.g., spectral features, spectral index features, texture features, etc.) as inputs to the classifier. The representativeness of the input features significantly affects the accuracy of classification.

In recent years, many deep learning methods have been used for forest monitoring, with excellent performance in applications such as independent tree crown extraction [23], forest mapping [24] and change detection[25, 26], monitoring leaf phenology[27] and forest tree species[28, 29]. However, most existing deep learning-based forest studies focus on the high spatial resolution unmanned aerial vehicle (UAV) or airborne platform image data [30], with few studies using medium resolution satellite images. Moreover, such algorithms rely heavily on massive finely labeled samples which are challenging to acquire. The recent development of weakly supervised technology [31-33] reduces the demand for sample labeling, demonstrating the robustness of the models to inaccurate, incomplete and inexact labels [34]. This suggests that the deep learning methods are promising for forest research [30, 35].

Although the classification method or classifier has a direct impact on the classification results, research has found that the selection of training data has a greater influence on the classification results than the classification method[36]. It is a matter of deeper reflection on how to select representative and informative training data[37], including training data distribution, size and balance and outliers in the training data. However, it is challenging to collect comprehensive training data on the land surface, which is one of the most time-consuming components of the entire task [14]. The training data are usually derived through (stratified/proportional) random sampling [37] and visual interpretation [38], but this is time-consuming and labor-intensive. Moreover, the quality of the sample annotation is affected by the experience of the annotators.

In response to the difficulty of sample annotation, some studies take advantage of publicly available LULC products as the annotation of the training data [39, 40], which greatly reduces the cost of manual annotation and visual interpretation. The LULC products allow pixel-level annotation of the land cover. Although such an approach might lead to errors in the training data, a certain percentage of errors barely affects the classification results [9, 14]. There are some methods that are usually used to reduce the labeling error to improve the accuracy of the samples [17, 32], such as voting [41] and D-S evidence theory [42]. A few recent studies have successfully realized the large-scale LULC classification with existing LULC products as labels[39, 40, 43].

Inspired by these studies, this work explores the potential of weakly supervised deep learning algorithms for medium-resolution forest classification. Focusing on issues faced in large-scale forest cover classification tasks including the dependence on a large number of high-quality labeled samples, and the need for automated procedures for large-scale forest classification, this paper proposes a weakly supervised framework for large-scale forest classification with existing LULC products as labels. The key components of the framework consist of four elements such as the deep classification model, the pseudo label generation module, the sample selection and optimization module and the pseudo label correction module. Specifically, forest/non-forest labels with noise are first obtained via these freely available forest-related products. Then, a weakly supervised semantic segmentation algorithm is employed to distinguish forest and non-forest areas. The contributions of this work include the following three aspects.

(1) A weakly supervised forest classification framework (WSFCF) for automatic forest classification is proposed in this study, which provides a reliable solution for large-scale forest mapping. The framework can correct the noisy samples automatically and achieve accurate forest classification even if the samples have problems such as inaccurate and incomplete label.

(2) A simple and efficient method was designed to obtain relatively high-precision land cover samples with less human intervention to address the inconsistency and difficult-to-use problems of existing LULC products. It not only solves the problem of difficult sample labeling but also provides an effective reference for the use of existing products.

(3) A strategy is designed to ensure the representativeness and informativeness of the samples which proceeds to the reliability of the samples in terms of sample location selection and optimization and dynamic correction of the pseudo labels.

(4) A semantic segmentation network (the Spectral-Spatial network, SSNet) is purposely employed in the proposed framework based on the analysis of forest characteristics and medium-resolution images. The SSNet composes of a spectral module and a spatial module for fully leveraging the spectral and textural information of the forest/non-forest.

This study focuses on the land cover rather than land use. The forest definition partially follows the Food and Agriculture Organization of the United Nations (FAO), that is, land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10 percent, or trees able to reach these thresholds in situ. The rest of this paper is organized as follows. The related works are summarized in Section II. The study areas and materials are described in Section III. The Section IV presents the details of the proposed forest classification framework. In Section V, we compare the

performance of SSNet with commonly used deep networks and analyze the performance for each component of the proposed WSFCF. Besides, the limitations and transferability are presented in Section V. Finally, the conclusions are given in Section VI.

II. Related work

A. The application of deep learning and weakly supervised methods in land cover classification.

In recent years, deep learning has demonstrated great potential in feature extraction and has been applied by scholars to remote sensing information extraction [44-46]. It has exhibited superior performance across various objects [32, 47], such as impervious surfaces/built-up areas, crop, water and vegetation [30, 48]. Deep learning algorithms rely heavily on massive fine-labeled samples. However, the labeling of the samples is labor-intensive, time-consuming and cost. The pixel-level labels constraints the application of deep learning in remote sensing LULC classification [49-51]. The application of weakly supervised techniques [34] reduces the algorithm's demands for high-precision labels and makes breakthroughs in LULC mapping [52].

Weakly supervised deep learning algorithms show satisfactory performance in some remote sensing classification tasks, including single-target extraction such as water [53], roads [54], buildings [55, 56] and so on and multi-class classification [57, 58]. In these tasks, deep learning models commonly used in computer vision such as ResNet, VGGNet, Unet, DenseNet and HRNet are usually employed. These models tend to involve deeper network structures, with more flexible and powerful nonlinear fitting capabilities to extract more abstract and complex features. Imagery from unmanned aerial vehicles (UAVs), Google Earth and high-resolution satellites are the main data sources. The spatial resolution of these data is extremely high, particularly the UAV data that can reach centimeter-level resolution and can accurately characterize the spatial details of the land cover.

At present, there many studies applying weakly supervised techniques and deep learning algorithms to the classification task of medium-resolution remote sensing images [30]. Several studies have applied weakly supervised deep learning methods in LULC classification tasks [32, 33] with promising results, such as paddy rice mapping [59]. Moreover, some scholars have reported some global- or national-scale LULC products based on deep learning algorithms, such as Esri land cover and CRLC [39]. These works demonstrate the great potential of weakly supervised techniques and deep learning algorithms in large-scale medium resolution LULC classification.

B. The application of LULC products in remote sensing classification tasks.

The land cover classification is usually conducted using supervised classification. However, it is challenging to obtaining high-quality samples. There are many LULC products that provide pixel-level annotations. There is also crowdsourced data freely available such as OpenStreetMap (OSM) which provides substantial LULC information. Therefore, taking these data as pixel-level labels is a popular device in large-scale LULC classification nowadays [52]. With these data as classification labels in existing studies usually includes both direct and indirect ways.

Directly use of LULC products for labeling: Samples that directly use existing LULC products for classification often suffer from noisy labeling, such as mis-classification in the products. To address the problem of sample coarseness, Li et al. [40] propose a low-to-high network (L2HNet) for large-scale high-resolution LULC mapping, which ensures the extraction of high-resolution features while solving the noise problem caused by the resolution mismatch. Yang et al. [60] derived training samples directly from CLUDs, a time series product with high classification accuracy, for monitoring the dynamics of LULC in the Chinese region. Limited by the cost and difficulty of representative sample collection, Liu et al. [17] generated sample points in the middle Yangtze River basin area from crowdsourced OpenStreetMap (OSM) LULC data and updated the samples with LULC change detection results. In addition to specific LULC classification tasks, scholars have produced some public datasets using LULC products, such as BigEarthNet [61] and SEN2MS [62], which provide essential datasets for studying deep learning algorithms [24, 63-66].

Indirectly use of LULC products for labeling: The studies that indirectly use existing LULC products as samples usually start by selecting reliable sets based on these products and applying these samples into the classification tasks. For one thing, the accuracy of the labels can be improved by integrating multiple products through data fusion. Inconsistent predictions of the products with conflict labels can be eliminated by following a voting strategy which can keep the consistent pixels [41]. The Dempster-Shafer theory can bridge the uncertainty between products and obtain accurate labels [42]. The Dempster-Shafer evidence theory is usually a simple but effective way and so is voting. But the former requires massive prior knowledge. For another, products are first utilized as labels for pre-classification to obtain classification results, and the results or the LULC products optimized with the results are used as labels. To eliminate the influence of resolution mismatch and semantic errors in the labels, Liu et al. [39] modeled the spectral similarity and spatial adjacency of the training labels using a conditional random field to refine the training labels and used a class-conditional label correction methods to detect and correct the abnormal incorrect labels. To promote the accuracy of the labels, Chen et al. [32] trained a weakly supervised learning network using partially labeled hyper-pixels generated from the representative samples which are transferred from low-resolution LULC products; Zhang et al. [67] performed pre-classification based on a limited number of point samples and a support vector machine algorithm and took the high-confidence areas as training samples for the deep learning model; Liu et al. [41] create the sample by integrated existing LULC products and OSM.

The available LULC products provide valuable priori knowledge for land cover classification studies. Although previous studies provide a wealth of experience in utilizing those products, how to fully utilize this knowledge to guide subsequent studies remains a question that deserves in-depth exploration.

III. Study Area and Materials

A. Study area and remote sensing images.

Gaofen-1 wide field view (GF-1 WFV) satellite images serve as the data source for this study. The GF-1 optical satellite was launched on 26 April 2013. The WFV sensors exhibited a powerful acquisition capability with a 4-day revisit interval. The GF-1 WFV imagery has 4 bands with a spatial resolution of 16 m which is freely available. The GF-1 WFV satellite images downloaded from the China Centre for Resources Satellite Data and Application are geometrically corrected level-1 products and are not provided with cloud/cloud shadow masks. To enhance the usability and the user-friendliness of these images, we have pre-processed images, including ortho-rectification [68] and cloud detection.

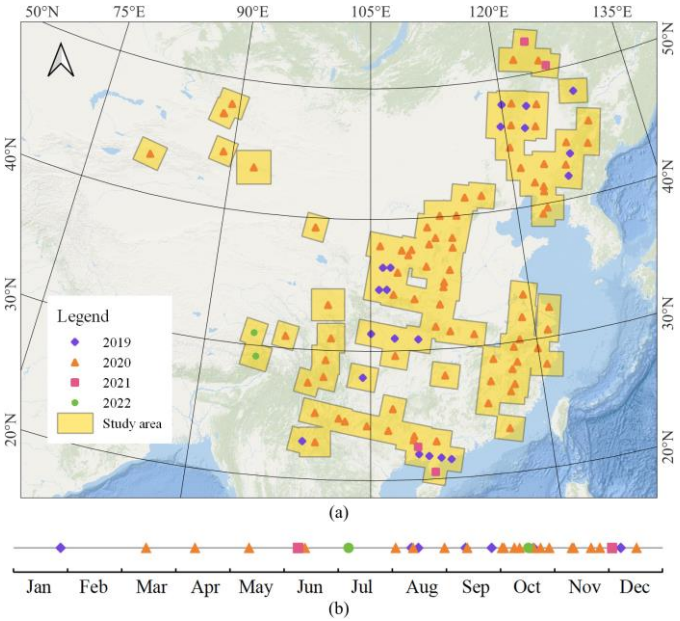

Fig. 1. The study areas and the acquisition time of the GF-1 WFV images. (a) indicates the spatial distribution of the study areas and the acquisition year of GF-1 WFV images. (b) indicates the temporal coverage of the collected images.

In the experiment, we select the area covered by 110 GF-1 WFV image scenes as the study area to validate the proposed method. The 110 images collected are from the year 2019-2022 and are centered in the months of August-November, as depicted in Fig. 1. The study area is mainly discretely distributed in the typical areas of China's forest with a total area of more than 3.6 million km^2, including the northeast, northwest, southwest and south China. The study area spans a wide range of latitudes (exceeding 34°) with significant variations in its coastal and terrestrial locations. The topography of the region is high in the west and low in the east, with a variety of terrain types and mountain ranges. Consequently, the variable combination of temperature and precipitation forms a rich climate, primarily encompassing subtropical monsoon climate, temperate monsoon climate, and temperate continental climate. The unique geographical environment, complex topography, and diverse climate have contributed to the variety of forest types [69] in the study area, including evergreen/deciduous broadleaf forests, evergreen/deciduous coniferous forests, and mixed forests.

B. Sample sets.

The sample comprises visual interpreted sample points and pseudo labels generated from forest-related data sets. The former is used for accuracy assessment of the results while the latter is designed for the model optimization.

1) The visual interpreted sample points

A total of 124,628 sample points is generated using the random sampling algorithm as displayed in Fig. 2. Generally, the forest and non-forest sample points comprised 39.2% and 60.8% of the total points respectively. The distribution of these sample points is carried out in two ways. One is distributed randomly within the study area, which consists of 45,780 forest points and 73,889 non-forest points. The other is randomly distributed in the forest edge and transition areas, with a total of 4959 points including 1891 non-forest points and 3068 forest points. The different distributions of the sample points are designed to verify the classification accuracy statistically in terms of the global study area and forest edges/transition areas. These points which correspond to 30m resolution pixels are visually interpreted by referring to several high-resolution satellite maps embedded in QGIS and historical images from Google Earth.

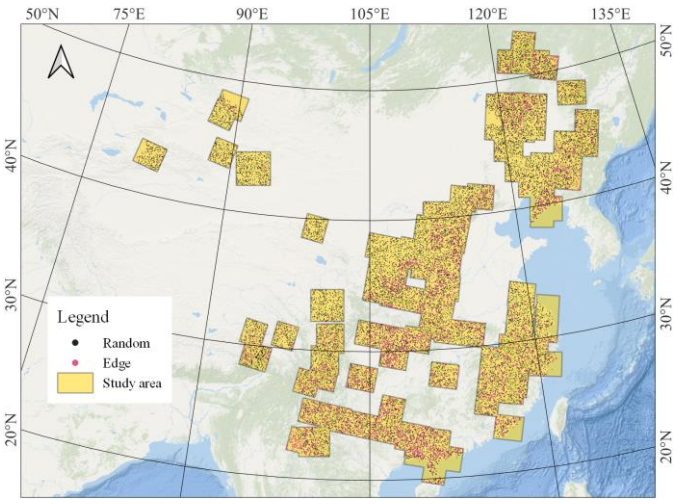

Fig. 2. Sample points (b) and (c) illustrates the spatial distribution of the first two part and third part point in section 3.2.3(1) respectively

2) Pseudo label (PL)

Acquisition of reliable labels is time-consuming and labor-intensive, but forest-related data sets with pixel-level

annotations provide an effective solution. However, it is challenging to use these data sets directly due to the limitations of accuracy and classification schemes. It has been shown [13, 70] that the higher voting results of different data sets eh higher accuracy of the region.

Therefore, we designed a simple rule to obtain LULC maps and forest/non-forest (FNF) labels with higher accuracy by referencing some LULC products including WordCover, Esri land cover, FROM-GLC10, GLC-FCS30, GlobeLand30, and TreeCover. However, the FNF labels still contain some errors, such as incomplete and inaccurate label. The detailed scheme is shown in Section 4.2.

IV. Methodology

A. Overall framework

The weakly supervised forest classification framework (WSFCF) proposed in this paper is implemented by noise label learning and weak supervision. This section provides the details of WSFCF. The forest-related datasets are initially employed to generate high-quality but noisy labels, called pseudo labels (PL). Then a weakly supervised classification scheme is designed, which realizes the robust training of the model by introducing a label correction module and a sample selection module. Specifically, the proposed WSFCF consists of 4 components, comprising data preparation, sample selection and optimization, dynamic correction of the PL and prediction, as illustrated in Fig. 3.

Prior to introducing each component, we would like to first introduce the forest classification model which is of great importance in the WSFCF. Given the characteristics of forests and medium-resolution satellite imagery, we adopt a simple model named Spectral-Spatial network (SSNet) as the main method for forest extraction, which is described in detail in Section 4.2.

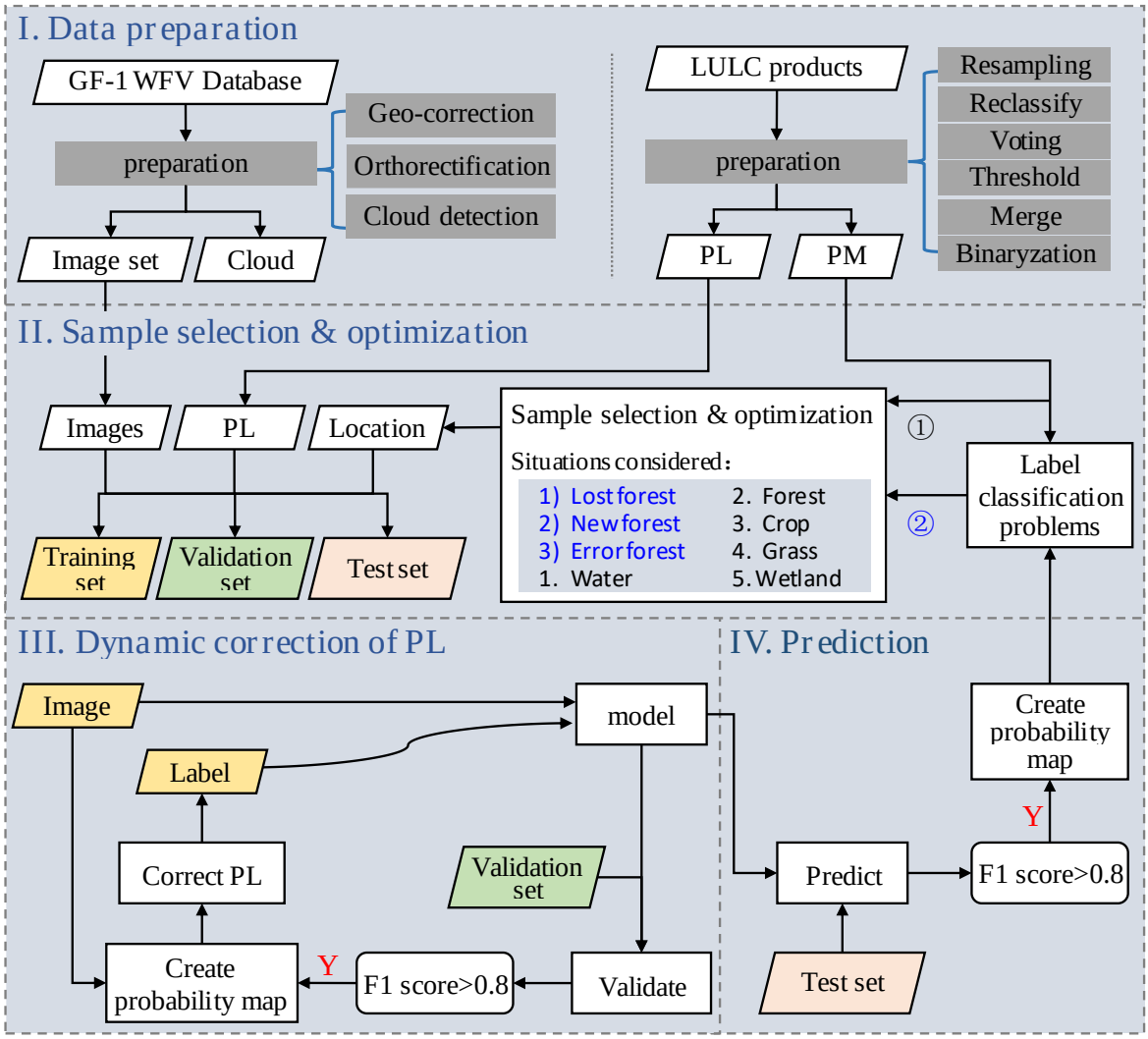

Fig. 3. Flow chart of proposed weakly supervised forest classification framework. ① represents the initial determination of sample position which accounts for land cover such as water, forest, crop, grass and wetland; ② represents optimization of sample position which considers not only the land cover but also areas prone to be misclassified.

The flowchart starts with data preparation, including satellite data preparation and a priori knowledge retrieval. **(1) Satellite data preparation**. The GF-1 WFV satellite images downloaded from the China Centre for Resources Satellite Data and Application are geometrically corrected level-1 products. To obtain a more accurate geometric localization, these images are orthorectified with ground control points and the Digital elevation model (DEM) data. In addition, the GF-1 WFV level-1 products do not provide quality assessment (QA) or cloud mask, which are necessary for forest monitoring [71]. Therefore, a threshold segmentation method is conducted to roughly extract clouds. Although it might result in the wrong extraction of some highlighted objects, it will not influence the forest classification because forests are usually dark objects in images. **(2) Priori knowledge retrieval.** The priori knowledge mainly refers to the freely available LULC products to create the PL and pseudo-LULC map (PM) required for the subsequent process, which is described in Section IV(C). There are some regions in the PL and the PM without annotations called uncertain areas which are often located in the boundaries and transition areas. These areas in the PL are not involved in the calculation of the loss function during model training.

Next, the satellite images, PL and PM matched based on location are used to create the training, validation and test sets for model inputs in the sample location selection and optimization. In the process of initial sample location determination and location optimization, both the informativeness of land cover and the difficulties of forest classification are considered. The test areas include all study area, while the training and validation areas are randomly selected from the study area with reference to the PM to ensure sample diversity. All training area or validation area is small than 5% of the total area. The sample location optimization might introduce new training areas, but the gross training area is less than 10% of the total area. The training area might partially or completely duplicate the initial training area after optimization, yet the training area is completely independent of validation area. In particular, the position optimization changes the training area only, not the validation area. The details are explained in Section IV(D).

Subsequently, the model is optimized using the training set in the third step and the PL is corrected dynamically (Section IV(E)). The experimental settings is presented in Section IV(F). During the model training, the PL would be dynamically corrected according to the prediction probability once the model reaches some fitting. And the PL correction employs a combination [40, 72] of a threshold and predicted probabilities, as detailed in Section IV(E). Because of the large thresholds used in create PM and PL (as shown in Section IV(C)), the forest annotations remained in the PL tend to be areas that are accurately annotated and easily classified correctly. Moreover, the forest area that satisfy the thresholds is much larger than those do not satisfy the threshold. However, the areas prone to misclassification, such as forest boundaries and transition areas of different land cover, are regarded as uncertain areas. The accuracy appears to be very high when the PL is used as the label to calculate the accuracy. In the experiments, it is observed that the model possessed some discrimination capability when the F1 score of predictions reaches 0.8 for comparison with the PL.

Finally, we predict the test set using the trained model. And the program proceeds to the next iteration until the condition do not satisfy and the program terminated. The purpose of the loop is to identify more representative and informative training areas. And the max loop number is set to 2.

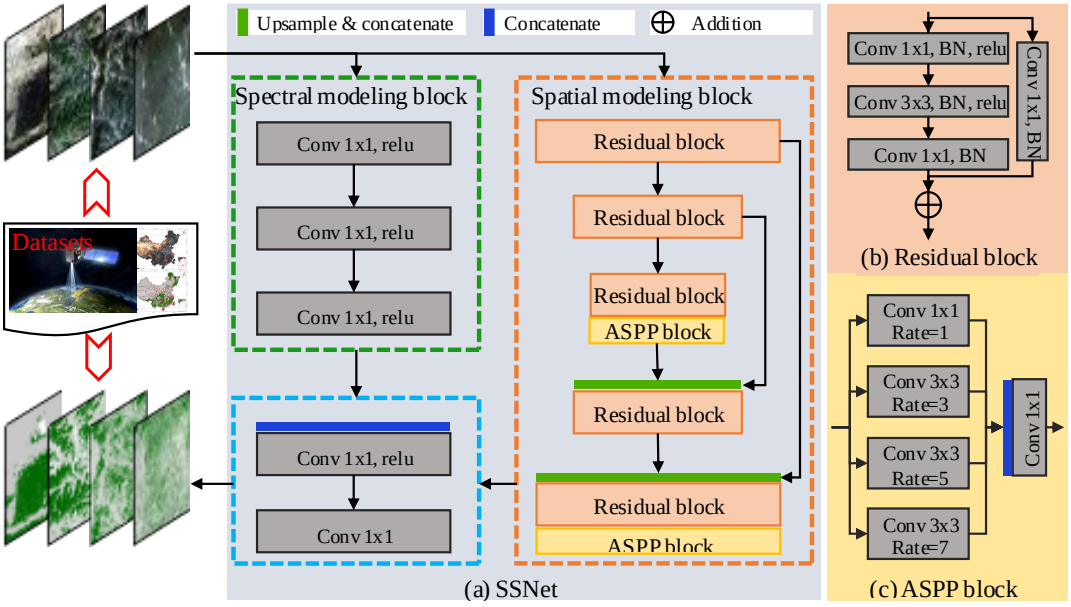

Fig. 4. The structure of the SSNet.

B. A network structure for capturing the spectral and spatial contextual information.

Moderate-resolution multispectral images exhibit spectral and spatial heterogeneity and are sensitive to atmosphere and light. These images can provide spectral information but suffer a lot from mixed pixels. Forests are usually patchy with irregular geometric boundaries, except some planted forests. Moreover, moderate-resolution images have disadvantages in depict spatial details. Therefore, the spatial texture of the forest is relatively blurred in moderate-resolution images.

Following the characteristics of medium-resolution images and forests, we designed the Spectral-Spatial network (SSNet) for forest classification, which can capture the spectral and spatial contextual information at the same time. As shown in Fig. 4, SSNet consists of a spectral modeling block (SeMB) and a spatial modeling block (SaMB). The acquired spectral and spatial features are fused and serve as input for classification.

SeM: Multi-layer perceptron constitutes this module for mining the spectral differences between forest and non-forest features. The SeM which consists of three linear layers with output dimensions of 64, 128 and 256 is designed for exploring the spectrum of FNF regardless of the spatial context of the scenes. It not only emphasizes the spectral differences of FNF and enhances the boundaries of FNF in transition and edge areas, but also compensates for the blurring and smoothing effect of the SaM.

SaM: This block comprises of a shallow U-shaped structure for exploiting the spatial contextual relationship and textures. As illustrated in Fig. 4, the SaM, with the residual structure as fundamental units, undergoes two downsampling and two upsampling operations. The residual structure adopts the classical bottleneck structure [73], as shown in Fig. 4(b). In addition, the standard atrous spatial pyramid pooling (ASPP) [74], which is designed to expand the reception field (as illustrated in Fig. 4(c)), is employed to enhance the feature extraction capability of SaM. Two primary factors influence the design of the shallow U-shaped structure. On the one hand, forests lack well-defined textures in moderate-resolution images. On the other hand, the deeper network diminishes resolution as downsampling progresses, leading to increased blurring of details in less prominent features.

C. The pseudo label (PL) generation.

Pixel-level FNF labels are generated with forest-related products. To rationalize the distribution of the samples, different LULC classes are considered, especially those easily confused with forests such as water and grass. Consequently, we created a pseudo LULC map (PM) alongside the PL.

Given the inconsistency of classification schemes, our first step involves reclassifying these products to ensure the consistency of classes. Then, we obtain voting results for different classes from the six products. Studies [75-79] indicate that classes such as crop, grass, water, built area and forest are more accurate, but classes like shrub, wetland and tundra are inaccurate in these products. To ensure the precision of PL and the diversity of the classes in PM, we establish distinct thresholds for forest and different non-forest classes. Following existing studies and visual comparison, we set a forest vote threshold of 3. This means that areas receiving more than 3 votes are categorized as forest areas. For non-forest classes with high accuracy, the threshold is set at 2, whereas for those with lower accuracy, it is set at 1. A PM featuring uncertain areas is created by consolidating the results of threshold segmentation of different classes and subsequently masking out the regions with overlap classes. Other than that, a PL with uncertain areas is obtained by aggregating the non-forest classes of the PM.

D. Sample location selection and optimization.

Drawing on the characteristics of sample selection in active learning, we use the informativeness and representativeness of the samples as the basis for determining sample position. Moreover, we also expect to select easily misclassified scenarios as hard samples to optimize the network. As a result, the determination of sample location includes initial determination and optimization of sample position.

In the process of training and validation area selection, we apply a non-overlapping sliding window to divide all images into patches. By setting some constraints, some patches are filtered out that satisfy the requirements as candidates of training and validation areas. Then, 5% of these candidate areas are randomly selected as training and validation areas respectively.

Initial determination of sample position. In the initial stage of the experiment, the representative patches are selected based on PM, as illustrated in Fig. 3(II-①). There are 3 rules for patch selection; (1) sequentially determine whether a patch contains objects list in Fig. 3 (water, forest, crop, grass and wetland). The order is finalized by the area and the likelihood of being misclassified as forest. (2) There are at least 2 classes in the patch and the pixel number of target class greater than the minimum value (here set to 20 pixels, about 0.5 hectares). (3) The patches used for training and testing are randomly selected from those patches that satisfy the requirements (1) and (2). The number of non-forest patches is no more than that of forests.

Optimization of sample position. It is challenging to define the distribution of the hard samples at the beginning. Therefore, the model is first trained with the initial samples. The results predicted by the trained model are checked against the PL to find the potential hard sample area, which are marked in the PM.

There are three potential areas considered. The first, loss forest, donates the forest area in the PL that is predicted to be non-forest area. The second is error forest which indicates that the non-forest area in the PL is regarded as forest area. And the third is new forest which indicates that the uncertain area in the PL is classified to be forest area. It should be noted that loss/error/new forest is just a name for convenience. Meanwhile, it ensures that sufficient samples for objects that are prone to confuse with forests. That is, the classes that determine the sample initial position are taken into account,

including water, forest, crop, grass and wetland.

The samples are re-selected and the problematic areas are prioritized, as presented in Fig. 3(II-②). Sample position optimization probably bring about an increase in training time and unnecessarily difficult samples (especially in forest sample-rich regions), so the conditions for sample location optimization are restrict via the prediction accuracy and positions.

E. Dynamic correction of pseudo labels.

Based on the knowledge that the deep model tends to learn the correct samples at the early stage [31, 80] when there are noisy labels, we leverage the PL obtained from the freely available products for model training. Inaccurate labels remain in the PL which is not conductive to model fitting, even though some noise has been filtered out by voting threshold. Therefore, the PL will be dynamically corrected using predicted probability according to (1) once the model shows some fitted.

$$UPL_{p<\theta_1}^{H\times W}=0;\ UPL_{p>\theta_2}^{H\times W}=1. \tag{1}$$

where UPL means the updated PL and p is the probability of forest. θ_1 and θ_2 are the maximum probability of non-forest and minimum probability of forest respectively. Areas where the predicted probability of forest is less than θ_1 are updated to non-forest. In contrast, the areas with a predicted probability greater than θ_2 are updated to forest. In the experiment, we compare different combinations of θ_1 and θ_2 and set them to 0.2 and 0.8 respectively.

F. Experimental settings.

In the experiments, the cross entropy is employed to optimize the models. The Adam with initial learning rate 0.005 weight decay 1e-4. The batch size used in the experiments is 4 and the size of input image is 64x64. The maximum epoch of training is set to 20. All 110 images are served as the test set whereas the training and testing sets are less than 5% of the total data. The experiments are implemented based on PyTorch framework, and a Linux platform with an NVIDIA TITAN V GPU with 12 G memory is used for training and testing.

All models adopt the same settings for the sake of fairness. In this study, we choose the overall accuracy (OA), producer accuracy (PA), user accuracy (UA), F1 score, and intersection over union (IoU), to quantitatively evaluate the performance of the proposed method. These metrics are calculated as (2-6). PA shares the same formula as recall (R), so does UA and precision (P). We focus more on the F1 score and IoU.

$$OA=\frac{TP+TN}{TP+TN+FP+FN}. \tag{2}$$

$$PA=R=\frac{TP}{TP+FN}. \tag{3}$$

$$UA=P=\frac{TP}{TP+FP}. \tag{4}$$

$$F1=\frac{2\times P\times R}{P+R}. \tag{5}$$

$$IoU=\frac{TP}{TP+FP+FN}. \tag{6}$$

V. Results and Discussions

In this section, we demonstrate the effectiveness of each component of the proposed framework. An appropriate classification model is crucial for improving the accuracy of the results. As a core component of the WSFCF, the performance of the SSNet is our primary concern. Therefore, we first conduct several experiments to compare the SSNet with some deep learning methods commonly used for remote sensing classification. Then, we perform a series of ablation experiments to verify the effectiveness of each part of the proposed WSFCF in improving the classification accuracy as well as the potential uncertainty. Finally, we transfer the classification framework to other object classification and explore the transferability and limitations of the proposed framework.

TABLE I

Quantitatively comparison of different methods

Methods	UA (P, %)	PA (R, %)	F1 (%)	IoU (%)	OA (%)
U-Net	92.95 | 78.35	79.34 | 41.81	85.61 | 54.52	74.84 | 37.48	90.37 | 59.66
ResUNet	92.51 | **78.91**	80.91 | 46.35	86.33 | 58.40	75.94 | 41.24	90.75 | 61.81
ResUnet-a	**92.97** | 75.90	73.27 | 35.98	81.95 | 48.82	69.42 | 32.29	88.35 | 56.37
SwinUnet	89.48 | 72.16	79.61 | 46.77	84.26 | 56.75	72.80 | 39.62	89.27 | 58.78
HRNet	91.98 | 74.49	78.67 | 41.35	84.80 | 53.18	73.62 | 36.22	89.83 | 57.89
Deeplab V3	90.38 | 70.89	77.93 | 44.42	83.70 | 54.62	71.97 | 37.57	89.04 | 57.30
DeeplabV3+	91.42 | 74.19	80.87 | 47.47	85.82 | 57.90	75.17 | 40.74	90.36 | 60.07
SSNet	92.42 | 78.77	**83.35 | 52.40**	**87.65 | 62.94**	**78.02 | 45.92**	**91.52 | 64.30**
RF [1]	92.35 | 77.71	81.07 | 47.09	86.34 | 58.65	75.97 | 41.49	90.75 | 61.59
SSNet [1]	91.93 | 76.45	82.63 | 51.33	87.03 | 61.42	77.04 | 44.32	91.11 | 62.71

[1] donates the classification results based on point training samples. The bold values are the best accuracies, and the underline means the second best. The accuracy for indicator X is (x1 | x2), where x1 represents the accuracy of the study area while x2 represent the accuracy of the forest edge region. x1 is calculated based on P1 while x2 is based on P2.

A. Comparison with commonly used classification methods.

The SSNet compares with networks commonly used for remote sensing image classification [39], including U-Net [81], ResUNet [82], ResUneta [83], SwinUnet [84], HRNet [85], DeepLab v3 [74] and Deeplab V3+ [86]. These networks embody the classical structures of convolution networks, such as U-shaped structure, residual structure, transformer and ASPP. Besides, we select RF as a comparison method, which is the most widely used method in forest study. The samples for RF are randomly picked from the training set. These points are employed as training set with sparse labeling for SSNet. This section compares the results of different models both quantitatively and qualitatively.

1) Quantitative comparison

The accuracy metrics calculated based on the visually interpreted samples are illustrated in Table I. The metrics indicate that these algorithms exhibit high classification accuracy, although they exhibit some variations. Among the compared deep learning methods, ResUNet achieves the highest classification accuracy with 86.33% and 75.94% for F1 score and IoU respectively, followed by Unet with 85.61% for F1 score and 74.84% for IoU. However, the RF algorithm yields a higher classification accuracy than the deep learning-based methods, with F1 score and IoU reaching 86.34% and 75.97% respectively.

We perform experiments on the SSNet with point-based sparsely labeled samples and patch samples. The results reveal that both point- and patch-based SSNet outperform the other methods, especially the patch-based. The F1 score and IoU of patch-based SSNet are 87.65% and 78.02% respectively, which are superior to other methods. Especially in the edge/transition area, the patch-based SSNet significantly surpassed the comparison methods (RF), with F1 score and IoU 4.29% and 4.43% higher than the optimal comparison method respectively.

McNemar's test was adopted to evaluate the significance in the difference of the classification accuracy. And the results reveal that the differences of the results between the proposed method and compared methods are significant with p=0.000 (p is the significance level), which indicates the proposed methods outperform the comparing methods.

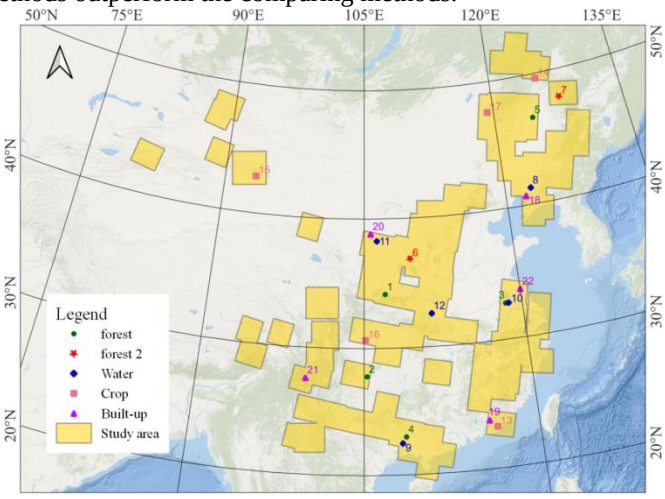

Fig. 5. Spatial distribution of the following displayed results.

2) Qualitative comparison.

We make extensive visual comparison of the classification results of these algorithms. Some regions are selected for visualization, and the spatial distribution is shown in Fig. 5. In general, these methods pose robust classification capabilities and discrepancies in fine object differentiation such as roads and rivers. There are two sites selected to demonstrate the details of the classification algorithms.

As presented in Fig. 6-7, non-forest areas can be accurately recognized in broad places, but different algorithms perform great discrepancies for small non-forest areas. Fig. 6 illustrate that Region 1 which is located in the Qinling Mountain is widely forested and traversed by several roads. U-Net, SwinUnet and RF show better results visually. And the patch-based SSNet achieves more complete non-forest areas than compared methods.

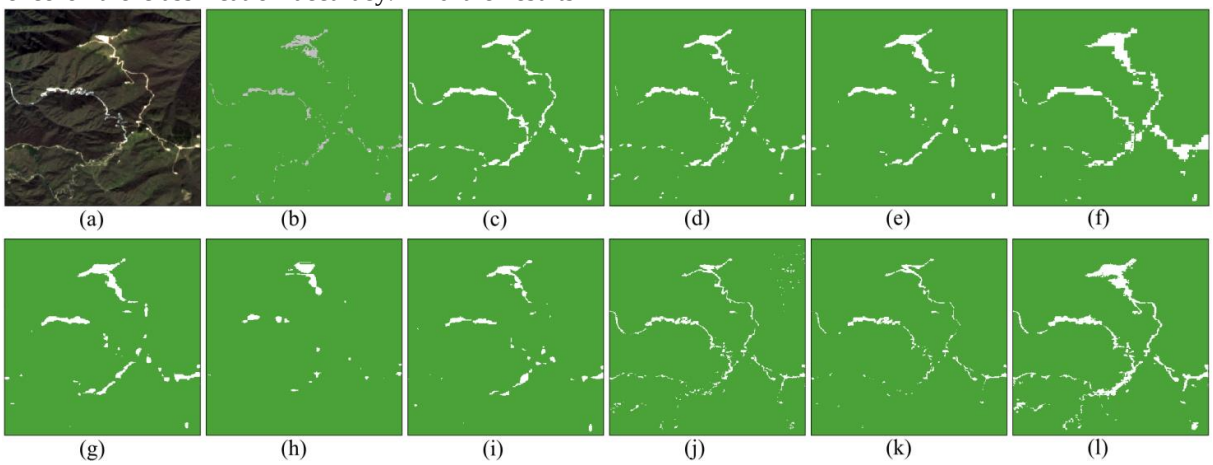

Fig. 6. Details of the results in Region 1. (a) is GF-1 WFV image (RGB); (b) is the PL; (c)-(j) represent the classification results of U-Net, ResUnet, ResUneta, SwinUnet, HRNet, Deeplab V3 and Deeplab V3+ respectively; (k) and (l) are indicates the results of point- and patch-based SSNet. Green in (b)-(c) donates forests and white donates non-forest; gray in (b) donate the unlabeled area, i.e. uncertain area.

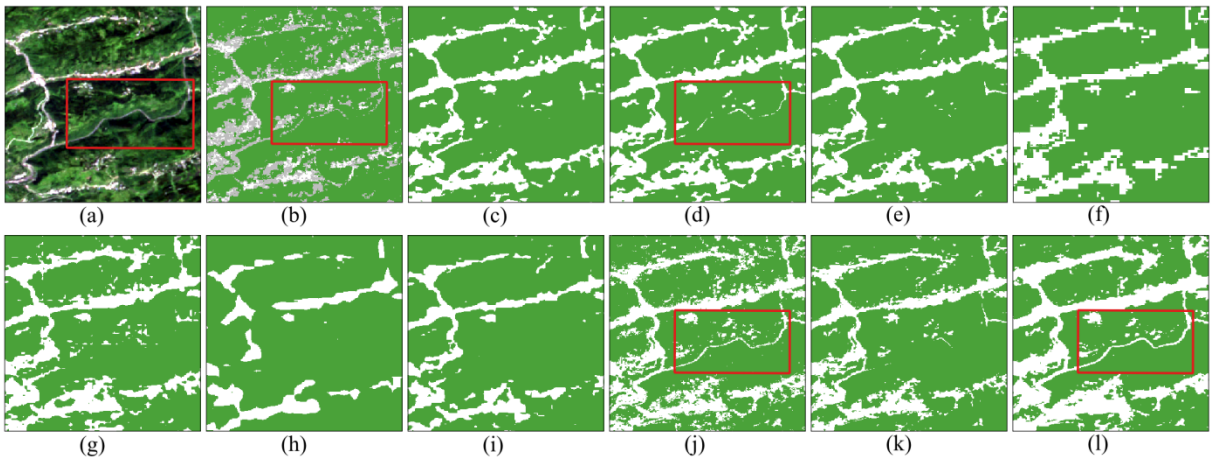

Fig. 7. Details of the results in Region 2. (a) is GF-1 WFV image (RGB); (b) is the PL; (c)-(j) represent the classification results of U-Net, ResUnet, ResUneta, SwinUnet, HRNet, Deeplab V3 and Deeplab V3+ respectively; (k) and (l) are indicates the results of point- and patch-based SSNet. Green in (b)-(c) donates forests and white donates non-forest; gray in (b) donate the unlabeled area, i.e. uncertain area.

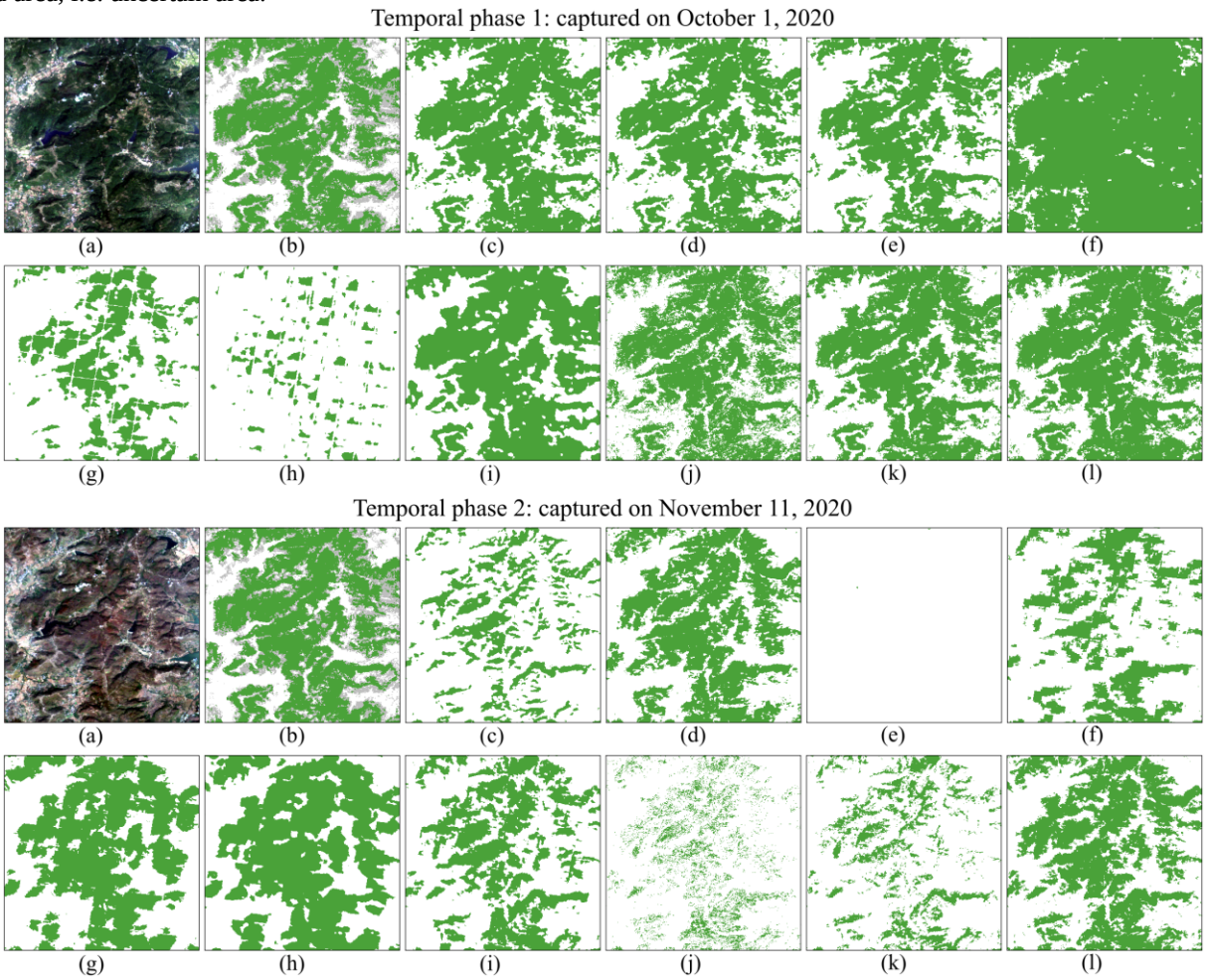

Fig. 8. The influence of seasonal shifts on the results. (1) and (2) donate the classification result of images acquired on October 1, 2020 and November 11, 2020. (a) is GF-1 WFV image (RGB); (b) is the PL; (c)-(j) represent the classification results of U-Net, ResUnet, ResUneta, SwinUnet, HRNet, Deeplab V3, Deeplab V3+ and RF; (k) and (l) are indicates the results of point- and patch-based SSNet. Green in (b)-(c) donates forests and white donates non-forest; gray in (b) donate the unlabeled area, i.e. uncertain area.

The Fig. 7 donates the results for region 2 which locates in the northern part of the Yunnan-Guizhou Plateau (Region 2). The red boxes indicate the river in the valley. Patch-based SSNet extracts the clearest river boundaries, followed by RF

and ResUnet. Other contrasting methods show huge errors in distinguishing non-forests (river) in this area.

Furthermore, the robustness of these methods varies significantly across different image and sample qualities, particularly when the quality of the sample or image is unsatisfactory. In the following section, we compared the classification result of these models under different conditions.

1) Comparing the influence of different factors on the models.

The study area spans a large space, and there are differences in forest properties, image temporal and quality and sample quality. The models are sensitive to these differences. To intuitively demonstrate the disparity of the methods, some typical phenomenon and regions are selected to be visualized, as shown in Fig. 8-10.

Seasonal shifts. Vegetation is strongly influenced by phenology. Therefore, the forest classification results suffer a lot from seasonal changes, especially in forest areas with pronounced seasonal changes. Region 3 sits in the central Jiangsu province with distinct seasonal variations.

The first image was captured on October 1, 2020, when the forest was in the growing season and the forest was flush. SwinUnet overestimate the forested area of Region 3 by misclassifying a large number of non-forested areas as forested. In contrast, HRNet and Deeplab V3 underestimate the forested area of this region by misclassifying vast forested areas as non-forested areas. In particular, Deeplab V3 exhibited catastrophic classification errors.

The second image was captured on November 11, 2020 when the forest was dormant the green decreased. The results indicate that U-Net, ResUneta and point-based methods (RF and SSNet) fail to overcome the negative effects of seasonal changes on forest classification and fail to distinguish the forest and non-forest areas in the non-growing season.

As displayed in Fig. 8, most of these methods can accurately classify the growing season images. However, it is challenging to perform on dormant period images. The results suggest that ResUnet and patch-based SSNet are more robust to the forest seasonal changes. It is observed that HRNet and Deeplab V3 yield better results for dormant season images than for growing season images. Forests in the dormant season probably exhibit more texture features.

Inaccurate labels. It is inevitable that there are inaccuracies in the forest samples due to the following factors. For one thing the acquisition time of the images used in forest-related products may not necessarily align with the acquisition time of the images used in this study, potentially leading to changes in the forest during this time interval. For another thing, existing products may suffer from inaccuracies stemming from various factors, including data quality and classification algorithms.

The Region 4 is situated in Nanning, Guangxi Province, which serves as a crucial timber resource area. Historical images reveal a history of frequent forestry activities in this region. As depicted in Fig. 9, extensive forest harvesting has been undertaken in this area. Nevertheless, there are serious errors in the PL and wide range of post-harvest forested areas are still labeled as forest. In this case, the SSNet are able to get rid of the effects of inaccurate labels and precisely distinguish forest/non-forest areas.

Conversely, the performance of these comparative methods is less satisfactory. Most comparative methods are able to distinguish between deforested areas to some extent, as reflected in Fig. 9 (c)-(j). The HRNet performs poorly and seems to confuse the forest and non-forest features in the Region 4. RF and SwinUnet perform particularly well among the contrasting methods, with both methods able to distinguish the harvest forest areas. However, their results are coarse and imprecise. In particular, the RF algorithm misclassify a significant number of forested areas as non-forested, even while it is able to identify some deforested areas.

Incomplete labels. The forest labels in this study area areas with high voting for existing products, thus some areas contain incomplete forest labels.

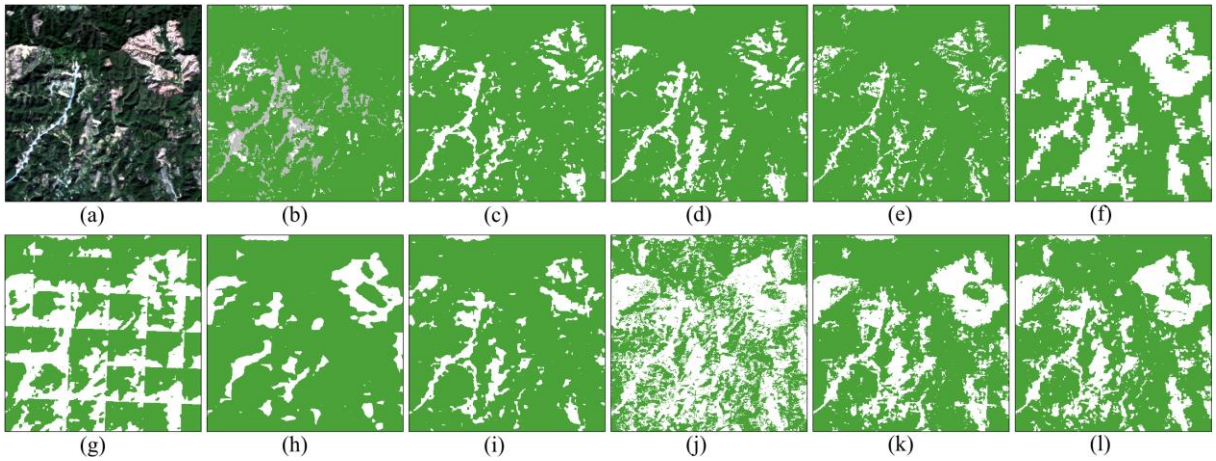

Fig. 9. The effects of inaccuracy labels on the results. (a) is GF-1 WFV image (RGB); (b) is the PL; (c)-(j) represent the classification results of U-Net, ResUnet, ResUneta, SwinUnet, HRNet, Deeplab V3, Deeplab V3+ and RF; (k) and (l) are indicates the results of point- and patch-based SSNet. Green in (b)-(c) donates forests and white donates non-forest; gray in (b) donate the unlabeled area, i.e. uncertain area.

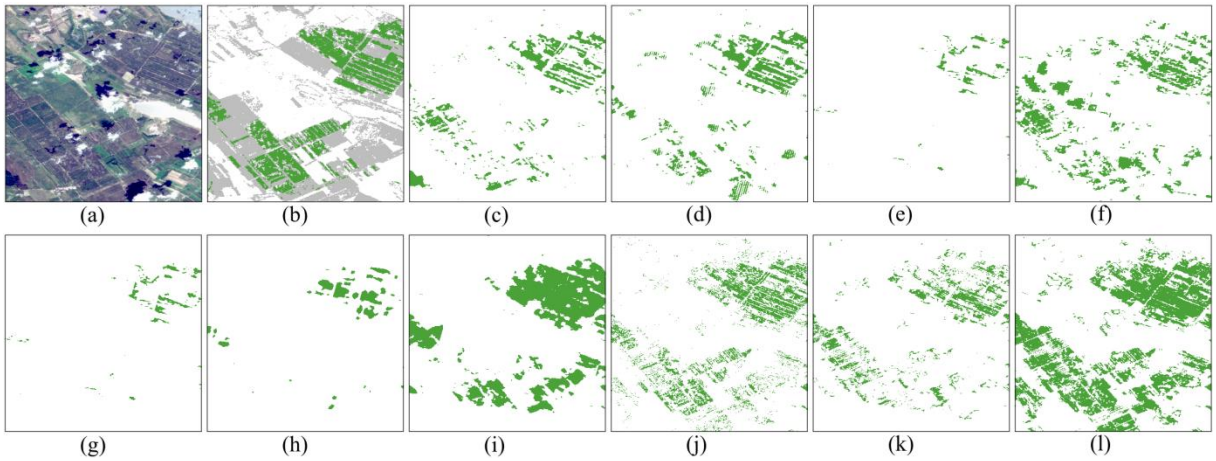

Fig. 10. The effects of incomplete labels on the results. (a) is GF-1 WFV image (RGB); (b) is the PL; (c)-(j) represent the classification results of U-Net, ResUnet, ResUneta, SwinUnet, HRNet, Deeplab V3, Deeplab V3+ and RF; (k) and (l) are indicates the results of point- and patch-based SSNet. Green in (b)-(c) donates forests and white donates non-forest; gray in (b) donate the unlabeled area, i.e. uncertain area.

The Region 5 is in Jilin Province. As shown in Fig. 10, part of the planted forest is in unlabeled in the PL. The comparison methods pose substantial uncertainties in this area. The patch-based SSNet demonstrates its ability to extract the most comprehensive results of planted forests, as depicted in Fig. 10. The results indicate that the patch-based SSNet outperforms both the point-based SSNet and the comparative methods in terms of completeness and accuracy.

Deep learning-based classification methods using patch-based labels yield results with reduced noise but exhibit pronounced omission errors. Among these, Deeplab V3+ produces relatively comprehensive results. When contrasted with patch-based labels, the method utilizing point-based samples generates results that are more fragmented, featuring a massive salt-and-pepper noise.

The unsatisfactory results are probably due to the following two aspects. On one hand, the plantation forests in Region 5 are young, and their features are not prominent in moderate-resolution images. This poses a challenge for the extraction of the plantation forests. On the other hand, the incomplete labels possibly lead to insufficient samples for model optimization.

Both the quantitative and qualitative results demonstrate that the patch-based SSNet performs better than the point-based SSNet. It suggests that optimizing the model parameters is challenging with sparse annotations. Deep network structures, such as HRNet and Deeplab V3, tend to smooth the edge and suffer from a disadvantage in discriminating fine targets. On the contrary, the U-shaped structure, which considers low-level features, excels in preserving spatial details and exhibits some capability to differentiate small targets. As well, RF can preserve spatial details in the result effectively. However, the outcomes present salt-and-pepper noise and fragmentation. Furthermore, when the samples contain a certain degree of error, precise classification fails to be realized by RF.

B. Validation of the proposed framework.

1) Ablation study.

We design a set of ablation experiments to verify the effectiveness of the proposed WSFCF with the setup described in Table II. Experiments 1-4 are used to assess the impact of each block of SSNet, while experiments 4-6 are conducted to verify the effectiveness of the framework. The quantitative results are presented in Table III.

TABLE II
THE SETTING OF ABLATION STUDY

	SSNet			Label correction	Location optimization
	SpeMB	SpaMB	ASPP		
1	√				
2		√			
3	√	√			
4	√	√	√		
5	√	√	√	√	
6	√	√	√	√	√

TABLE III
QUANTITATIVELY COMPARISON OF ABLATION STUDY

	UA (P, %)	PA (R, %)	F1 (%)	IoU (%)	OA (%)
1	92.52 | 78.49	82.11 | 48.48	87.00 | 59.93	77.00 | 42.79	91.15 | 62.52
2	**92.85** | 78.26	80.79 | 43.17	86.40 | 55.65	76.05 | 38.55	90.82 | 60.20
3	92.36 | 78.45	83.01 | 50.76	87.44 | 61.63	77.68 | 44.54	91.39 | 63.45
4	92.42 | **78.77**	83.35 | 52.40	87.65 | 62.94	78.02 | 45.92	91.52 | 64.30
5	91.45 | 76.45	84.64 | 56.30	87.91 | 64.85	78.43 | 47.98	91.60 | 64.70
6	90.73 | 75.19	**85.96** | **62.53**	**88.28** | **68.28**	**79.01** | **51.84**	**91.76** | **66.40**

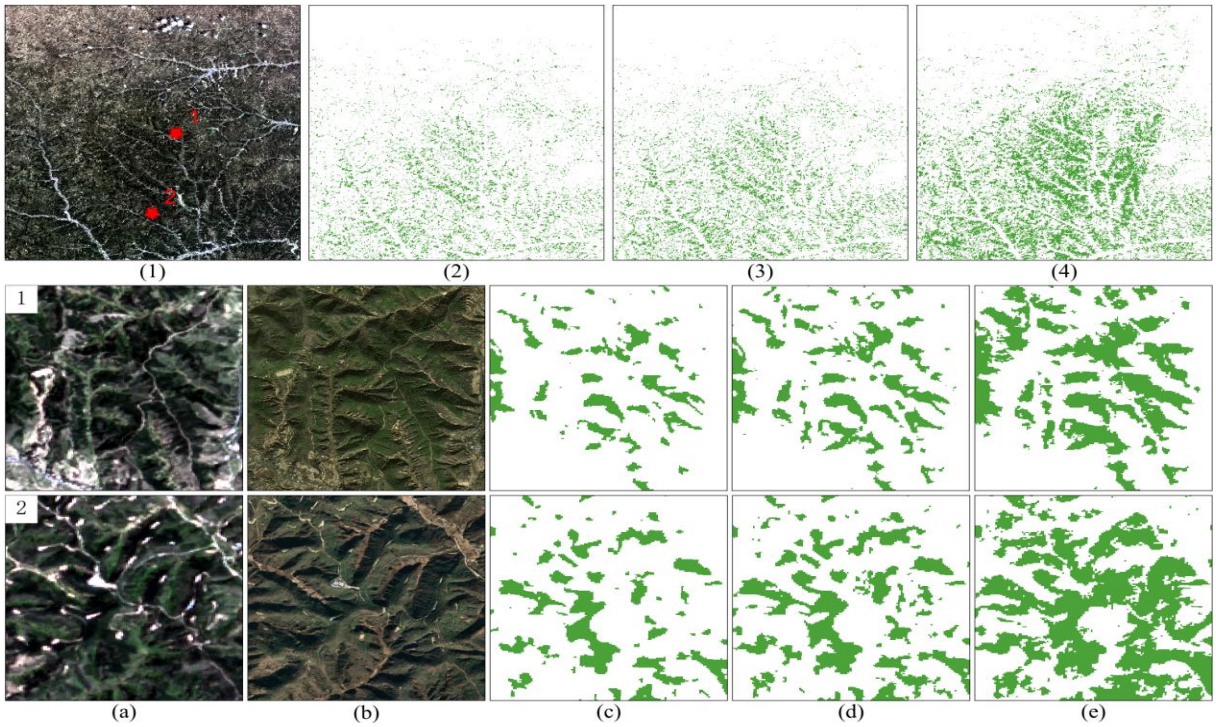

Fig. 11. The performance of sample location optimization in Region 6. In the first line (1) is the GF-1 WFV image, and (2) - (4) respectively represent the results for settings 4-6 in Table II. The second and third lines represent the details in (1), where (a) and (b) donate GF-1 WFV images and VHR image, and (c)-(e) donate the corresponding details in (2) - (4).

The Table III reveals that each block of SSNet enhance the model feature extraction capabilities. Specifically, the combination of SpeMB and SpaMB significantly improves the discriminative ability between forest and non-forest, especially in forest edge/transition areas, with IoU improvements of 1.75% and 5.99% compared to using SpeMB and SpaMB alone. The introduction of ASPP increases the model's F1 score and IoU by 1.31% and 1.38% respectively.

The quantitative results show that the sample correction and location optimization significantly boosted the model's discriminative ability in the hard-to-classify area, with F1 score and IoU reaching 88.28% and 79.0% respectively. In particular, the label correction improves the F1 score and IoU in the forest transition region from 62.94% and 45.92% to 64.85 % and 47.98%. the location optimization increases the F1 score and IoU by 5.34% and 5.92% respectively.

2) The effectiveness of label correction and sample position optimization.

The purpose of label correction is to make the samples more accurate and more favorable for model fitting. In the computer vision tasks, the PLs are usually created using predictive probability maps[31], class attention maps class attention map [72, 87] or class activation maps[31, 88, 89] that provide information about the foreground and background. Then threshold or clustering methods are employed to refine the PLs for more accurate samples. In this study, the reliable regions for optimizing the samples are obtained by applying thresholds to the prediction probabilities, and the diversity of samples is increased by optimizing the sample location.

In addition, an increased diversity of samples is identified as an instrument to improve the generalization ability of the model. In this study, the reliable regions for optimizing the samples are obtained by applying thresholds to the prediction probabilities, and the diversity and complexity of the samples are increased by augmenting the proportion of the hard samples. Experimental results indicate that dynamic label correction enhances label accuracy, yet sample location optimization sometimes interferes with the model. The influences of sample correction and location optimization on forest results in different environments are illustrated in Fig. 11 and 12.

There are noises in the original PL, but the CNN is still able to distinguish forests and non-forests, which reveals that the convolutional neural network can compensate for defective or noisy labels to a certain extent [90, 91]. Furthermore, PL correction enhances the model's ability to discriminate at details, realizing a more accurate distinction between forests and non-forests, as shown in Fig. 12(d).

The sample location optimization increases both sample size and diversity. It is particularly beneficial in regions lacking forest samples. As illustrated in Fig. 11, location optimization yields more comprehensive forest results for Region 9. However, in areas with an ample supply of forest samples, the inclusion of potentially confusable samples might compromise the model's learning capacity, especially concerning narrow line features, or even lead to an increase in errors, as evidenced in Fig. 12.

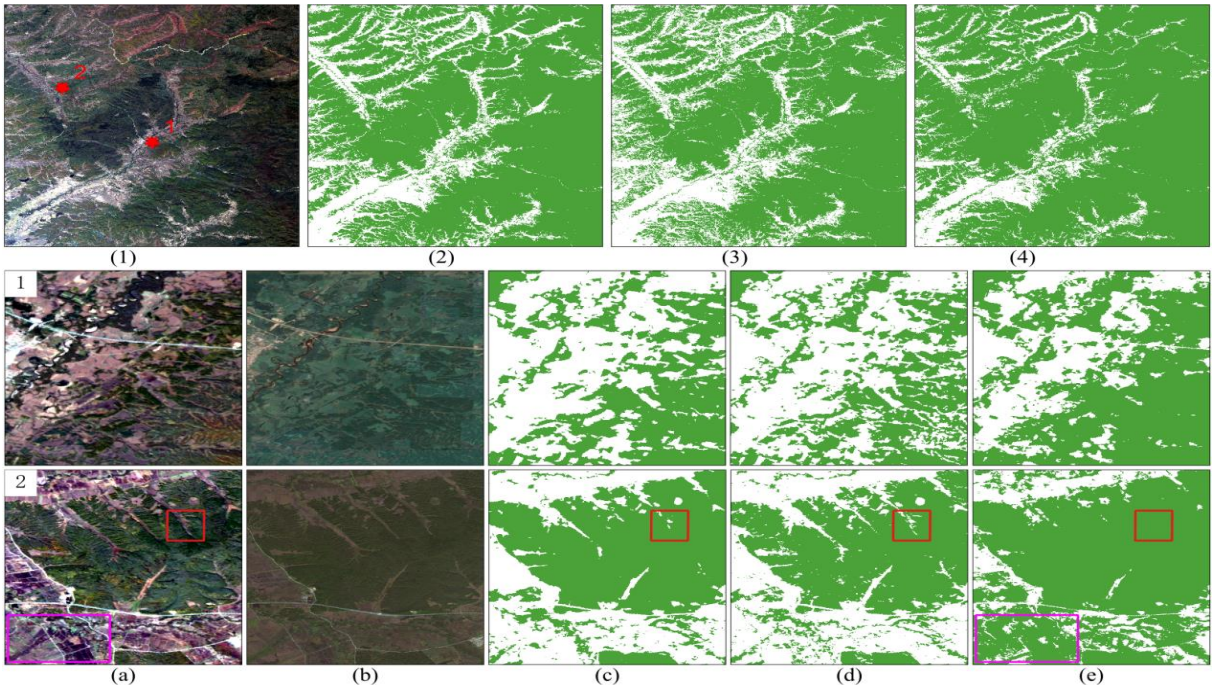

Fig. 12. The performance of sample location optimization in Region 7. In the first line (1) is the GF-1 WFV image, and (2) - (4) respectively represent the results for settings 4-6 in Table II. The second and third lines represent the details in (1), where (a) and (b) donate GF-1 WFV images and VHR image, and (c)-(e) donate the corresponding details in (2) - (4).

TABLE IV
QUANTITATIVELY COMPARISON OF TRANSFER LEARNING

	UA (P, %)	PA (R, %)	F1 (%)	IoU (%)	OA (%)
Water	94.71	90.06	92.30	85.83	89.18
Crop	74.62	85.35	79.42	66.20	78.89
Built-up	86.02	82.04	83.01	71.41	85.88

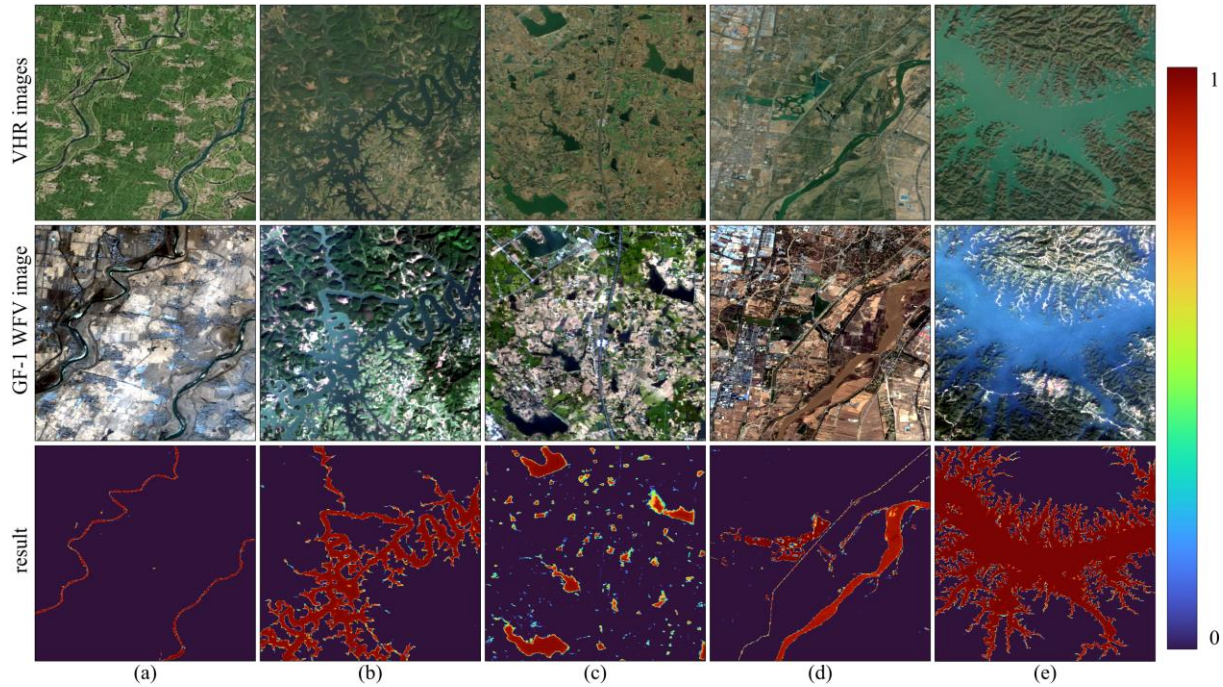

Fig. 13. The predicted probability map of water. (a)-(e) display the results of Regions 8-12 respectively.

Although increasing the diversity of samples is productive for model optimization, it is challenging to for forest

classification, especially adding samples that cannot be identified. The class of hard sample, that's lost/new/error forest area, are determined based on the predicted probabilities. Nevertheless, we practically do not know its true category. Therefore, it might introduce non-forest sample that are easily confused with forest. Under such circumstances, the hard samples tend to cause the model to be fitted in the wrong direction, resulting in inaccurate classification.

Furthermore, optimizing sample positions can result in an extended model training duration. In the worst-case, it might even double or further increase the original training time. Therefore, one should exercise caution when updating sample positions, despite the potential enhancement in model classification accuracy.

C. Validation of the proposed framework transferability.

The classification framework proposed in this study performs well on the forest. To explore the transferability and limitations of the framework, we perform experiments on other objects in the study area, such as water, crop, built-up. As illustrated in Table IV, the satisfying results have been achieved in these objects utilizing the proposed method, which suggest that the WSFCF manifests great potential on water, crop and built-up. The visualized results are presented in Fig. 13-15.

Water. Water bodies exhibit variations in size and spatial distribution. To facilitate a more comprehensive evaluation of the water body extraction capabilities presented in this paper, we have selected five distinct scenes to showcase the results of our methodology. These scenarios are situated in diverse geographical regions, as illustrated in Fig. 5, each possessing unique characteristics. Fig. 13 showcases the results for water. The visualization results show that whether for slender rivers (Region 8) and canals (Region 11), small and irregular ponds (Region 10), or tree-forked rivers (Region 9 and 12), SSNet shows a strong discriminative capability.

Crop. Due to the influence of climate and geography, there are significant variations in farmland size, irrigation methods, and crop types across different regions. These distinctions increase the complexity of farmland extraction. In this study, we have selected crop classification results of varying sizes in different regions to showcase the potential of our method in farmland classification, as illustrated in Fig. 14. SSNet demonstrates superior performance in regions characterized by extensive crop distribution. Nevertheless, it faces challenges in areas with intricate terrain and limited farmland (such as Region 16) as well as in regions with perplexing backgrounds (such as Region 17).

Built-up. The patterns of built-up areas are generally more intricate compared to those of forests and crops. In urban regions, built-up areas are both densely concentrated and wide spread, whereas in rural areas, they tend to be small and scattered. Fig. 15 depicts the results of various built-up area sizes. It is evident that the larger the built-up area, the more favorable the performance of the proposed method. Nevertheless, a notable capability decline is observed in rural areas, particularly along roads (Region 21 and 22). There are few or even no built-up labels in these regions, which poses a huge challenge for model learning. This might be attributed to the small and dispersed nature of built-up areas in rural regions. Small-scale built-up features also exhibit a tendency to yield large errors in existing products.

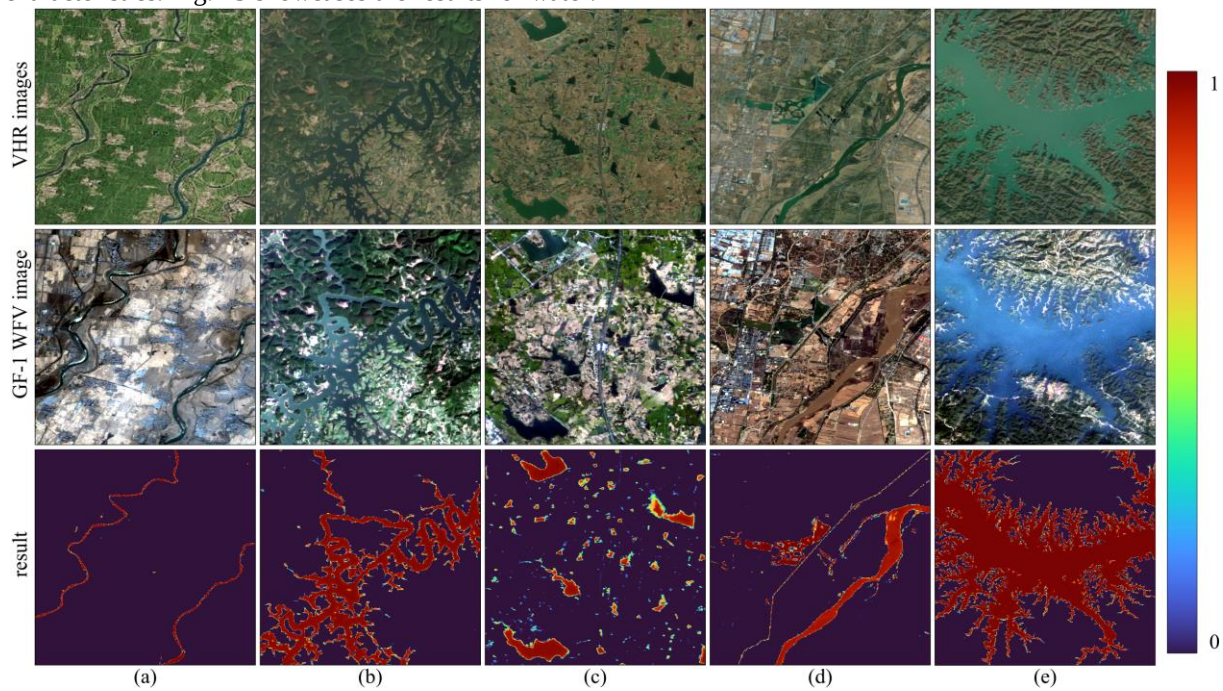

Fig. 14. The predicted probability map of crop. (a)-(e) display the results of Regions 13-17 respectively.

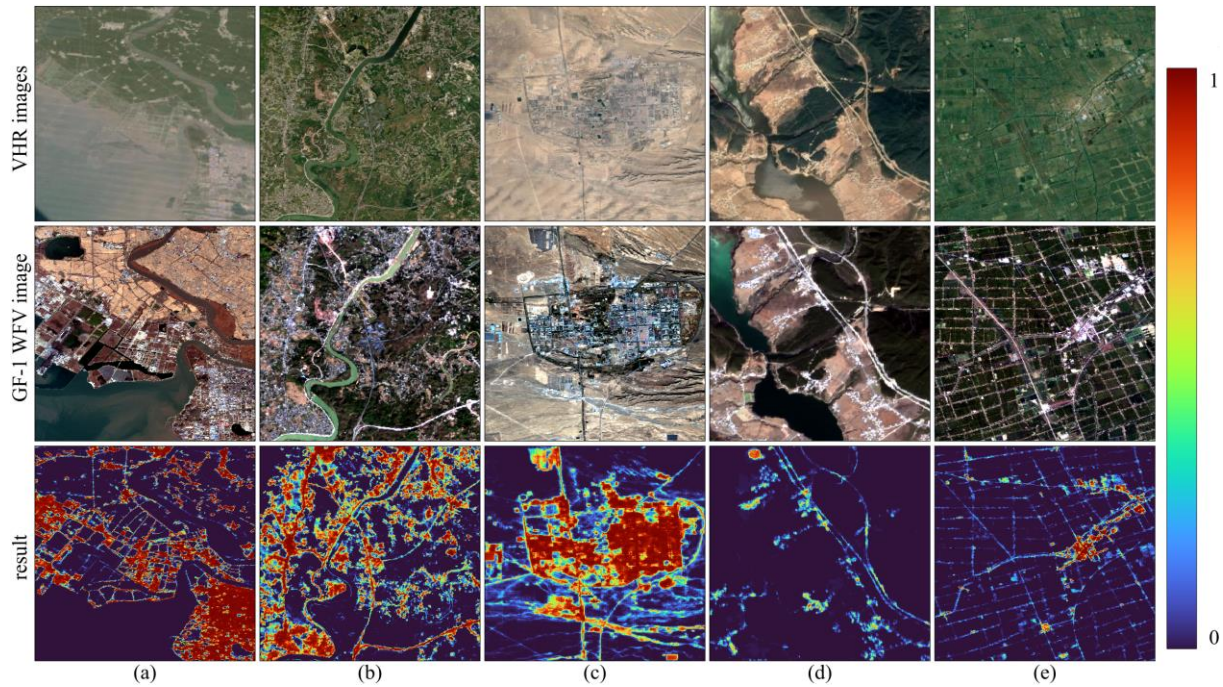

Fig. 15. The predicted probability map of built-up. (a)-(e) display the results of Regions 18-22 respectively.

We have conducted extensive experiments on the transferability of the algorithm, including single- and multi-class classification. The consequences suggest that the proposed method performs impressively in single-class classification, but less satisfactory in multi-class classification. We argue that this phenomenon is caused by an imbalance in the sample size of the LULC classes and the generation rules of the PM. The unbalanced sample inclined the model to favor the numerically dominant classes [92, 93], resulting in inaccurate classification results [94]. Therefore, it is inferred that improvements can be made in the following two aspects in future surface cover categorization studies. One is to improve the generation rules of the PM and the strategy of sample selection to narrow the sample size gap between classes. The other is to adopt reasonable classification strategies, such as a hierarchical classification scheme [95] and a reasonable loss function [96], to reduce the impact of LULC classes imbalance on the results.

Additionally, there is a shortage of samples representing small objects in the PM. In the experiments, the extraction of small objects is suboptimal. This is doubtless attributed to the initial intention which aimed at achieving accuracy in non-forest area as much as possible. The completeness, diversity and fine details of specific objects were overlooked. Furthermore, small objects commonly occur as mixed pixels, which are inherently pose challenges for separation. Therefore, the proposed WSFCF provides the potential to handle other objects classifications. However, it is essential to take into account the fineness and richness of target objects labels.

Last but not least, it is noticed that the probability distribution of the targets is uneven in the experiments with different objects but are significantly differentiated from the background as presented in Fig. 13-15. The prediction probability tends to be very high in areas with dense targets but tends to be lower in areas with dispersed targets. On the on hand, this might be attributed to the representativeness and the spatial distribution of the samples. On the other hand, it might be revealing that we might create more accurate result through some post-processing such as local adaptive threshold segmentation.

VI. Conclusions

In this paper, we have focused on the method for large-scale forest extraction using existing products and medium-resolution imagery, which is of significant importance for sustainable forest monitoring and management. We propose a weakly supervised forest classification framework (WSFCF) including model design, sample generation and training strategies. Further, SSNet, a network that takes into account spectral-texture, was designed based on medium-resolution imagery and forest characteristics. Label correction and sample location play positive roles in forest classification. However, the positive impact of sample location is conditional. In areas where forest samples are sparse or insufficient, updating sample locations can significantly improve the accuracy and completeness of classification results. However, in regions with an abundance of forest samples, this might potentially confuse the model's discriminative ability somewhat.

We perform experiments on 110 GF-1 WFV images to test the proposed method. The results suggest that the proposed method achieves the best classification in the study area, with the F1 score and IoU reaching 88.28% and 79.0% respectively. Especially in the edge and transition area of the forest, the method in this paper outperforms the other methods, with F1 score and IoU higher than the best comparative methods by

9.63% and 10.35% respectively. It reveals that the classification framework proposed in this study is effective in the task of extracting medium-resolution forest classification. This provides a feasible solution for mapping the subsequent large-scale (such as national, continental or global scale) forest cover. So far, we have created a forest cover map in China in 2020 using the proposed method and GF-1 WFV images. We also validated the effectiveness of the proposed method on Landsat data, and we will exploit the scalability and feasibility of the method for time-series forest monitoring in subsequent studies. In addition, the proposed method exhibits excellent transferability and achieves excellent classification results in water, crop, and built-up areas. The diversity and fineness of the samples may contribute to the transfer of the model to other classes.

REFERENCES

[1] M. C. Hansen *et al.*, "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, no. 6160, pp. 850-3, Nov 15 2013, doi: 10.1126/science.1244693.

[2] "Global Forest Resources Assessment 2020." https://www.fao.org/forest-resources-assessment/2020/en/ (accessed on 22 May 2024).

[3] D. Zhang, "China's forest expansion in the last three plus decades: Why and how?," *Forest Policy and Economics*, vol. 98, pp. 75-81, 2019, doi: 10.1016/j.forpol.2018.07.006.

[4] J. C. White, N. C. Coops, M. A. Wulder, M. Vastaranta, T. Hilker, and P. Tompalski, "Remote Sensing Technologies for Enhancing Forest Inventories: A Review," *Canadian Journal of Remote Sensing*, vol. 42, no. 5, pp. 619-641, 2016, doi: 10.1080/07038992.2016.1207484.

[5] M. Shimada *et al.*, "New global forest/non-forest maps from ALOS PALSAR data (2007–2010)," *Remote Sensing of Environment*, vol. 155, pp. 13-31, 2014, doi: 10.1016/j.rse.2014.04.014.

[6] X. Zhang *et al.*, "Rapid generation of global forest cover map using Landsat based on the forest ecological zones," *Journal of Applied Remote Sensing*, no. 2, 2020.

[7] C. Jun, Y. Ban, and S. Li, "Open access to Earth land-cover map," *Nature*, vol. 514, no. 7523, pp. 434-434, 2014.

[8] P. Gong *et al.*, "Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data," *International Journal of Remote Sensing*, vol. 34, no. 7, pp. 2607-2654, 2012, doi: 10.1080/01431161.2012.748992.

[9] P. Gong *et al.*, "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," *Science Bulletin*, vol. 64, no. 6, pp. 370-373, 2019, doi: 10.1016/j.scib.2019.03.002.

[10] X. Zhang, L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi, "GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery," *Earth System Science Data*, vol. 13, no. 6, pp. 2753-2776, 2021, doi: 10.5194/essd-13-2753-2021.

[11] D. Zanaga, Van De Kerchove, Ruben, De Keersmaecker, Wanda, Souverijns, Niels, Brockmann, Carsten, Quast, Ralf, Wevers, Jan, Grosu, Alex, Paccini, Audrey, Vergnaud, Sylvain, Cartus, Oliver, Santoro, Maurizio, Fritz, Steffen, Georgieva, Ivelina, Lesiv, Myroslava, Carter, Sarah, Herold, Martin, Li, Linlin, Tsendbazar, Nandin-Erdene, … Arino, Olivier. *ESA WorldCover 10 m 2020 v100*. [Online]. Available: https://doi.org/10.5281/zenodo.5571936

[12] K. Karra, C. Kontgis, Z. Statman-Weil, J. C. Mazzariello, M. Mathis, and S. P. Brumby, "Global land use / land cover with Sentinel 2 and deep learning," presented at the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021.

[13] X. Peng, G. He, G. Wang, T. Long, X. Zhang, and R. Yin, "User-Aware Evaluation for Medium-Resolution Forest-Related Datasets in China: Reliability and Spatial Consistency," *Remote Sensing*, vol. 15, no. 10, 2023, doi: 10.3390/rs15102557.

[14] J. R. Townshend *et al.*, "Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges," *International Journal of Digital Earth*, vol. 5, no. 5, pp. 373-397, 2012, doi: 10.1080/17538947.2012.713190.

[15] S. Talukdar *et al.*, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sensing*, vol. 12, no. 7, 2020, doi: 10.3390/rs12071135.

[16] J. Koskinen *et al.*, "Participatory mapping of forest plantations with Open Foris and Google Earth Engine," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 148, pp. 63-74, 2019, doi: 10.1016/j.isprsjprs.2018.12.011.

[17] D. Liu, N. Chen, X. Zhang, C. Wang, and W. Du, "Annual large-scale urban land mapping based on Landsat time series in Google Earth Engine and OpenStreetMap data: A case study in the middle Yangtze River basin," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 337-351, 2020, doi: 10.1016/j.isprsjprs.2019.11.021.

[18] K. Cheng *et al.*, "Mapping China's planted forests using high resolution imagery and massive amounts of crowdsourced samples," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 356-371, 2023, doi: 10.1016/j.isprsjprs.2023.01.005.

[19] Y. Qin *et al.*, "Forest cover maps of China in 2010 from multiple approaches and data sources: PALSAR, Landsat, MODIS, FRA, and NFI," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 109, pp. 1-16, 2015, doi: 10.1016/j.isprsjprs.2015.08.010.

[20] M. C. Hansen *et al.*, "Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD)," *Remote Sensing of Environment*, vol. 140, pp. 466-484, 2014, doi: 10.1016/j.rse.2013.08.014.

[21] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24-31, 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[22] L. Collins, G. McCarthy, A. Mellor, G. Newell, and L. Smith, "Training data requirements for fire severity mapping using Landsat imagery and random forest," *Remote Sensing of Environment*, vol. 245, 2020, doi: 10.1016/j.rse.2020.111839.

[23] G. Lassalle, M. P. Ferreira, L. E. C. La Rosa, and C. R. de Souza Filho, "Deep learning-based individual tree crown delineation in mangrove forests using very-high-resolution satellite imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 189, pp. 220-235, 2022, doi: 10.1016/j.isprsjprs.2022.05.002.

[24] N. Ahmed, S. Saha, M. Shahzad, M. M. Fraz, and X. X. Zhu, "Progressive Unsupervised Deep Transfer Learning for Forest Mapping in Satellite Image," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 752-761.

[25] F. Zhao *et al.*, "Monthly mapping of forest harvesting using dense time series Sentinel-1 SAR imagery and deep learning," *Remote Sensing of Environment*, vol. 269, 2022, doi: 10.1016/j.rse.2021.112822.

[26] R. Dalagnol *et al.*, "Mapping tropical forest degradation with deep learning and Planet NICFI data," *Remote Sensing of Environment*, vol. 298, 2023, doi: 10.1016/j.rse.2023.113798.

[27] G. Song *et al.*, "Monitoring leaf phenology in moist tropical forests by applying a superpixel-based deep learning method to time-series images of tree canopies," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 183, pp. 19-33, 2022, doi: 10.1016/j.isprsjprs.2021.10.023.

[28] F. Schiefer *et al.*, "Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 170, pp. 205-215, 2020, doi: 10.1016/j.isprsjprs.2020.10.015.

[29] L. E. C. La Rosa, C. Sothe, R. Q. Feitosa, C. M. de Almeida, M. B. Schimalski, and D. A. B. Oliveira, "Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 35-49, 2021, doi: 10.1016/j.isprsjprs.2021.07.001.

[30] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24-49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[31] Y. Cao and X. Huang, "A coarse-to-fine weakly supervised learning method for green plastic cover segmentation using high-resolution remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 188, pp. 157-176, 2022, doi: 10.1016/j.isprsjprs.2022.04.012.

[32] Y. Chen *et al.*, "A novel weakly supervised semantic segmentation framework to improve the resolution of land cover product," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 73-92, 2023, doi: 10.1016/j.isprsjprs.2022.12.027.

[33] D. Ienco, Y. J. E. Gbodjo, R. Gaetano, and R. Interdonato, "Weakly Supervised Learning for Land Cover Mapping of Satellite Image Time Series via Attention-Based CNN," *IEEE Access*, vol. 8, pp. 179547-179560, 2020, doi: 10.1109/access.2020.3024133.

[34] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44-53, 2018, doi: 10.1093/nsr/nwx106.

[35] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. X. Zhu, "Weakly supervised semantic segmentation of satellite images for land cover mapping--challenges and opportunities," *arXiv preprint arXiv:2002.08254*, 2020.

[36] G. Foody and M. Arora, "An evaluation of some factors affecting the accuracy of classification by an artificial neural network," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 799-810, 1997.

[37] Z. Zhu *et al.*, "Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 122, pp. 206-221, 2016, doi: 10.1016/j.isprsjprs.2016.11.004.

[38] X. Peng *et al.*, "A Comparison of Random Forest Algorithm-Based Forest Extraction with GF-1 WFV, Landsat 8 and Sentinel-2 Images," *Remote Sensing*, vol. 14, no. 21, 2022, doi: 10.3390/rs14215296.

[39] Y. Liu, Y. Zhong, A. Ma, J. Zhao, and L. Zhang, "Cross-resolution national-scale land-cover mapping based on noisy label learning: A case study of China," *International Journal of Applied Earth Observation and Geoinformation*, vol. 118, 2023, doi: 10.1016/j.jag.2023.103265.

[40] Z. Li, H. Zhang, F. Lu, R. Xue, G. Yang, and L. Zhang, "Breaking the resolution barrier: A low-to-high network for large-scale high-resolution land-cover mapping using low-resolution labels," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 192, pp. 244-267, 2022, doi: 10.1016/j.isprsjprs.2022.08.008.

[41] S. Liu *et al.*, "Land Use and Land Cover Mapping in China Using Multimodal Fine-Grained Dual Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-19, 2023, doi: 10.1109/tgrs.2023.3285912.

[42] S. Schmitz, M. Weinmann, U. Weidner, H. Hammer, and A. Thiele, "Automatic generation of training data for land use and land cover classification by fusing heterogeneous data sets," *Publ. Der Dtsch. Ges. Für Photogramm. Fernerkund. Und Geoinf. EV*, vol. 29, pp. 73-86, 2020.

[43] Z. Li, W. He, M. Cheng, J. Hu, G. Yang, and H. Zhang, "SinoLC-1: the first 1-meter resolution national-scale land-cover map of China created with the deep learning framework and open-access data," *Earth System Science Data Discussions*, vol. 2023, pp. 1-38, 2023.

[44] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195-204, Feb 2019, doi: 10.1038/s41586-019-0912-1.

[45] A. Vali, S. Comai, and M. Matteucci, "Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review," *Remote Sensing*, vol. 12, no. 15, 2020, doi: 10.3390/rs12152495.

[46] Z. Sun *et al.*, "A review of Earth Artificial Intelligence," *Computers & Geosciences*, vol. 159, 2022, doi: 10.1016/j.cageo.2022.105034.

[47] L. Zhang and L. Zhang, "Artificial Intelligence for Remote Sensing Data Analysis: A review of challenges and opportunities," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 2, pp. 270-294, 2022, doi: 10.1109/mgrs.2022.3145854.

[48] X. Yang, S. Qiu, Z. Zhu, C. Rittenhouse, D. Riordan, and M. Cullerton, "Mapping understory plant communities in deciduous forests from Sentinel-2 time series," *Remote Sensing of Environment*, vol. 293, 2023, doi: 10.1016/j.rse.2023.113601.

[49] X. X. Zhu *et al.*, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8-36, 2017, doi: 10.1109/mgrs.2017.2762307.

[50] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539-556, 2017, doi: 10.1016/j.patcog.2016.07.001.

[51] P. Gong *et al.*, "A new research paradigm for global land cover mapping," *Annals of GIS*, vol. 22, no. 2, pp. 87-102, 2016, doi: 10.1080/19475683.2016.1164247.

[52] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. X. Zhu, "WEAKLY SUPERVISED SEMANTIC SEGMENTATION OF SATELLITE IMAGES FOR LAND COVER MAPPING – CHALLENGES AND OPPORTUNITIES," 2020.

[53] M. Lu, L. Fang, M. Li, B. Zhang, Y. Zhang, and P. Ghamisi, "NFANet: A Novel Method for Weakly Supervised Water Extraction From High-Resolution Remote-Sensing Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-14, 2022, doi: 10.1109/tgrs.2022.3140323.

[54] M. Zhou, H. Sui, S. Chen, J. Liu, W. Shi, and X. Chen, "Large-scale road extraction from high-resolution remote sensing images based on a weakly-supervised structural and orientational consistency constraint network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 193, pp. 234-251, 2022, doi: 10.1016/j.isprsjprs.2022.09.005.

[55] M. U. Ali, W. Sultani, and M. Ali, "Destruction from sky: Weakly supervised approach for destruction detection in satellite imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 115-124, 2020, doi: 10.1016/j.isprsjprs.2020.02.002.

[56] W. Qiao, L. Shen, J. Wang, X. Yang, and Z. Li, "A Weakly Supervised Semantic Segmentation Approach for Damaged Building Extraction From Postearthquake High-Resolution Remote-Sensing Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1-5, 2023, doi: 10.1109/lgrs.2023.3243575.

[57] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic Annotation of High-Resolution Satellite Images via Weakly Supervised Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3660-3671, 2016, doi: 10.1109/tgrs.2016.2523563.

[58] X. Zeng, T. Wang, Z. Dong, X. Zhang, and Y. Gu, "Superpixel Consistency Saliency Map Generation for Weakly Supervised Semantic Segmentation of Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-16, 2023, doi: 10.1109/tgrs.2023.3264232.

[59] M. Wang, J. Wang, and L. Chen, "Mapping Paddy Rice Using Weakly Supervised Long Short-Term Memory Network with Time Series Sentinel Optical and SAR Images," *Agriculture*, vol. 10, no. 10, 2020, doi: 10.3390/agriculture10100483.

[60] J. Yang and X. Huang, "The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019," *Earth System Science Data*, vol. 13, no. 8, pp. 3907-3925, 2021, doi: 10.5194/essd-13-3907-2021.

[61] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019: IEEE, pp. 5901-5904.

[62] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Sen12ms – a Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W7, pp. 153-160, 2019, doi: 10.5194/isprs-annals-IV-2-W7-153-2019.

[63] Y. Cong *et al.*, "Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery," *Advances in Neural Information Processing Systems*, vol. 35, pp. 197-211, 2022.

[64] M. Schmitt and Y.-L. Wu, "Remote sensing image classification with the SEN12MS dataset," *arXiv preprint arXiv:2104.00704*, 2021.

[65] D. Zhang, M. Gade, and J. Zhang, "SOFNet: SAR-Optical Fusion Network for Land Cover Classification," presented at the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021.

[66] M. Rußwurm, S. Wang, M. Korner, and D. Lobell, "Meta-learning for few-shot land cover classification," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition workshops*, 2020, pp. 200-201.

[67] W. Zhang, P. Tang, T. Corpetti, and L. Zhao, "WTS: A Weakly towards Strongly Supervised Learning Framework for Remote Sensing Land Cover Classification Using Segmentation Models," *Remote Sensing*, vol. 13, no. 3, 2021, doi: 10.3390/rs13030394.

[68] T. Long, W. Jiao, G. He, G. Wang, and Z. Zhang, "Digital orthophoto map products and automated generation algorithms of Chinese optical satellites," *National Remote Sensing Bulletin*, vol. 27, no. 03, pp. 635-650, 2023, doi: 10.11834/jrs.20232041.

[69] X. S. Zhang, "Vegetation Map of the People's Republic of China (1:1,000,000) and Its Illustration Put to Press," *Acta Ecologica Sinica*, 2007.

[70] K. Zheng, G. He, R. Yin, G. Wang, and T. Long, "A Comparison of Seven Medium Resolution Impervious Surface Products on the Qinghai–Tibet Plateau, China from a User's Perspective," *Remote Sensing*, vol. 15, no. 9, 2023, doi: 10.3390/rs15092366.

[71] C. Huang *et al.*, "Automated masking of cloud and cloud shadow for forest change analysis using Landsat images," *International Journal of Remote Sensing*, vol. 31, no. 20, pp. 5449-5464, 2010.

[72] A. Pardo, H. Alwassel, F. Caba, A. Thabet, and B. Ghanem, "Refineloc: Iterative refinement for weakly-supervised action localization," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 3319-3328.

[73] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[74] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[75] Z. S. Venter, D. N. Barton, T. Chakraborty, T. Simensen, and G. Singh, "Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover," *Remote Sensing*, vol. 14, no. 16, 2022, doi: 10.3390/rs14164101.

[76] J. Wang *et al.*, "Consistency Analysis and Accuracy Assessment of Three Global Ten-Meter Land Cover Products in Rocky Desertification Region—A Case Study of Southwest China," *ISPRS International Journal of Geo-Information*, vol. 11, no. 3, 2022, doi: 10.3390/ijgi11030202.

[77] Y. Ding *et al.*, "A Field-Data-Aided Comparison of Three 10 m Land Cover Products in Southeast Asia," *Remote Sensing*, vol. 14, no. 19, 2022, doi: 10.3390/rs14195053.

[78] H. Wang, H. Yan, Y. Hu, Y. Xi, and Y. Yang, "Consistency and Accuracy of Four High-Resolution LULC Datasets—Indochina Peninsula Case Study," *Land*, vol. 11, no. 5, 2022, doi: 10.3390/land11050758.

[79] L. Liu, X. Zhang, Y. Gao, X. Chen, X. Shuai, and J. Mi, "Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects," *Journal of Remote Sensing*, vol. 2021, pp. 1-38, 2021, doi: 10.34133/2021/5289697.

[80] E. Arazo, D. Ortego, P. Albert, N. O'Connor, and K. McGuinness, "Unsupervised label noise modeling and loss correction," in *International conference on machine learning*, 2019: PMLR, pp. 312-321.

[81] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (Lecture Notes in Computer Science, 2015, ch. Chapter 28, pp. 234-241.

[82] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted Res-UNet for High-Quality Retina Vessel Segmentation," presented at the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), 2018.

[83] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94-114, 2020, doi: 10.1016/j.isprsjprs.2020.01.013.

[84] H. Cao *et al.*, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*, 2022: Springer, pp. 205-218.

[85] J. Wang *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3349-3364, 2020.

[86] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801-818.

[87] W. Ge, S. Guo, W. Huang, and M. R. Scott, "Label-penet: Sequential label propagation and enhancement networks for weakly supervised instance segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3345-3354.

[88] B. Kim, Y. Yoo, C. E. Rhee, and J. Kim, "Beyond semantic to instance segmentation: Weakly-supervised instance segmentation via semantic knowledge transfer and self-refinement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4278-4287.

[89] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921-2929.

[90] T. Kattenborn, J. Eichel, S. Wiser, L. Burrows, F. E. Fassnacht, and S. Schmidtlein, "Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery," *Remote Sensing in Ecology and Conservation*, vol. 6, no. 4, pp. 472-486, 2020.

[91] Z. M. Hamdi, M. Brandmeier, and C. Straub, "Forest damage assessment using deep learning on high resolution remote sensing data," *Remote Sensing*, vol. 11, no. 17, p. 1976, 2019.

[92] W. Miao, J. Geng, and W. Jiang, "Multigranularity Decoupling Network With Pseudolabel Selection for Remote Sensing Image Scene Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-13, 2023, doi: 10.1109/tgrs.2023.3244565.

[93] Y. Li, Y. Zhou, Y. Zhang, L. Zhong, J. Wang, and J. Chen, "DKDFN: Domain Knowledge-Guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 186, pp. 170-189, 2022, doi: 10.1016/j.isprsjprs.2022.02.013.

[94] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Information Sciences*, vol. 513, pp. 429-441, 2020, doi: 10.1016/j.ins.2019.11.004.

[95] J. Chen *et al.*, "Global land cover mapping at 30m resolution: A POK-based operational approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 7-27, 2015, doi: 10.1016/j.isprsjprs.2014.09.002.

[96] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.