A Review of Individual Tree Crown Detection and Delineation From Optical Remote Sensing Images

Current progress and future

JUEPENG ZHENG, SHUAI YUAN, WEIJIA LI, HAOHUAN FU, LE YU, AND JIANXI HUANG

©SHUTTERSTOCK.COM/CINEMANIKOR

Trees in forest ecosystems contribute extensively to ecology, environmental, economic, and society domains in both global and local regions [1]. Therefore, better forest management not only provides renewable resources for human activities but also makes great contributions to ecological conservation and the global energy circle. In the traditional investigation period, surveyors regularly measure the parameters for individual trees by field surveys, zonal sampling, or manual aerial imagery interpretation, which cost a large amount of human labor, work time, and expense. Fortunately, with the development of commercial satellites with high resolutions, along with the rapid progress in computer techniques [2], [3], [4], especially automatically detecting objects from digital image processing [5], [6], [7], [8], researchers have opportunities for automatic ITCD through high-resolution remote sensing images [9].

There is a variety of reviews about trees [20], including tree species classification [21], fruit detection [22], and yield estimation [23]. Most of them emphasize lidar [24] or thermal imaging [25]. Some surveys review only one specific tree species, such as oil palm [26]. We list almost all the surveys in Table 1 with the comparisons among reviewed methods, topics, tasks, and data. Hyyppä et al. [10] and Wulder et al. [12] focus only on lidar data and vertically distributed

Digital Object Identifier 10.1109/MGRS.2024.3479871
Date of publication 12 November 2024; date of current version 27 March 2025.

forest attributes estimation. Yin and Wang [14] review the available techniques for evaluating detected individual tree locations and crown delineation maps using remote sensing data. They mainly discussed ITCD assessment rather than ITCD methods. Zhen et al. [13] conduct a comprehensive survey for two of the ITCD topics (detecting and delineating) using traditional image processing-based ITCD methods, while they focus only on lidar data. Zhao et al. [15] review the convolutional neural network (CNN)-based applications in ITCD, which focus only on deep learning-based methods while ignoring other traditional methods. Walker and Dahle [18] survey all kinds of methods but neglect the ITCD tasks that cannot provide whole insights of the ITCD domain. Ke et al. [11] review ITCD methods using passive remote sensing imagery, which only conducts a survey on traditional image processing-based ITCD methods.

We can tell that existing ITCD reviews cannot completely comprise ITCD development from all aspects, especially recent developments under the continuously rapid progress of machine learning and deep learning algorithms [see Figure 1(b)]. Furthermore, various important ITCD-related tasks and applications (e.g., counting the number of trees, health monitoring, parameter estimation, and so on) are paid rare attention in existing reviews. On the other hand, as shown in Figures 1(a) and 10, we can observe that the use of optical data and deep learning-based methods in the ITCD domain has increased significantly, especially in the previous 10 years, accounting for nearly 80% and 70% of ITCD-related publications, respectively.

Therefore, it is necessary to summarize the overall trends from ITCD-related research during the past 10 years to help readers comprehend the past, present, and future of the ITCD domain. Notably, this review mainly surveys ITCD from optical remote sensing images as well as combines optical remote sensing images and lidar data. Research that only adopts lidar data on ITCD is outside the scope of this review. The contributions of this article include the following three points:

1) We conduct a review of ITCD, including a meta-analysis of the literature, a thorough review and comparison of the methodology, an in-depth discussion, extensive related applications, and potential prospects. This article provides a systematic review of ITCD development in the most recent two decades.
2) We cater to the rapid progress in computer science and its usage in ITCD and discuss the advantages and disadvantages of all kinds of existing ITCD approaches from three aspects: traditional image processing, traditional machine learning, and deep learning. We conduct comparisons between general deep learning models and their applications.
3) We conduct in-depth discussions on the multisensor data in ITCD, dataset construction in ITCD, and comparisons among different ITCD algorithms, and analyze the criteria of choosing the proper methods. Also, we list extensive ITCD-related applications and tasks and envision promising future works in the ITCD domain. We point out that optical remote sensing data will continue to be a key driver of future ITCD-related studies.

The remainder of the article is organized as follows. We present the meta-analysis of related literature in the "Meta-Analysis of Related Literature" section. Following that, we conduct a thorough review of the methodology of ITCD in the "Methodology Review" section. After that, we provide an in-depth discussion of the comparison of different ITCD methods, the characteristics of ITCD research, an assessment of the accuracy, and so forth, followed by extensive ITCD related applications, such as tree parameters, forest monitoring, and so on, in the "ITCD-Related Applications" section. We envision our promising prospects in the ITCD domain in the "Prospects" section. Finally, we conclude this article in the "Conclusions" section.

META-ANALYSIS OF RELATED LITERATURE

As shown in Figure 1(a), the number of ITCD-related articles using optical remote sensing data has exponentially increased since 2017, which, for those who are involved in the ITCD domain, is notoriously difficult to keep track of ITCD-related research. To this end, it is essential to periodically conduct a review to summarize recently implemented ITCD methods, study areas, tree species, and the types of optical remote sensing data. In this section, we conduct a meta-analysis regarding the ITCD domain to investigate these subjects.

OVERALL TREND OF ITCD DEVELOPMENT

Figure 2 displays the representative ITCD-related research from 2000 to 2023. The circles, triangles, and rectangles denote traditional image processing-based ITCD methods, traditional machine learning methods, and deep learning-based ITCD methods, respectively. We can observe some tendencies in the ITCD field: 1) the number of ITCD-related research projects has exponentially increased recently, 2) deep learning-based methods have emerged with high-accuracy results, 3) more large-scale ITCD research projects have been proposed, and 4) most of the large-scale studies utilize remote sensing data with 0.5–1 m.

QUANTITATIVE ANALYSIS

Along with the collection of ITCD-related publications, quantitative data are presented through figures in the next sections, including tree species, study sites and area, the types of optical remote sensing data, and so forth.

TREE SPECIES

Figure 3 displays the statistics of tree species in ITCD-related publications. In the specific tree, we show only the species that have been studied at least two times in ITCD-related articles. According to existing ITCD publications, 44.21% of them take mixed forest as the study objective and the rest take only specific tree species as their study objective. Traditional image processing-based ITCD methods have been adopted for most of the times when the study object is mixed forest (63.72%). The palm tree is the most popular study species among other single tree species (37 times). The most probable reasons include the benefit of positive economics and the impact of a negative environment as the increasing expansion of oil palm plantation areas in tropical developing countries [27]. The urban tree is another popular objective in the ITCD domain (29 times). Other popular study species include pine tree (16 times) and citrus tree (13 times).

STUDY SITES

Figure 4 displays the spatial distribution of study sites according to our database. Study sites of countries where the number is over 20 times are the USA (38 times), China (34 times) and Canada (30 times). Similar to the spatial distribution of research institutions, most of them are principally located in North America, East Asia, and North Europe. Meanwhile, tropical forest areas (such as Brazil) are a hot study site for ITCD research because of their substantial

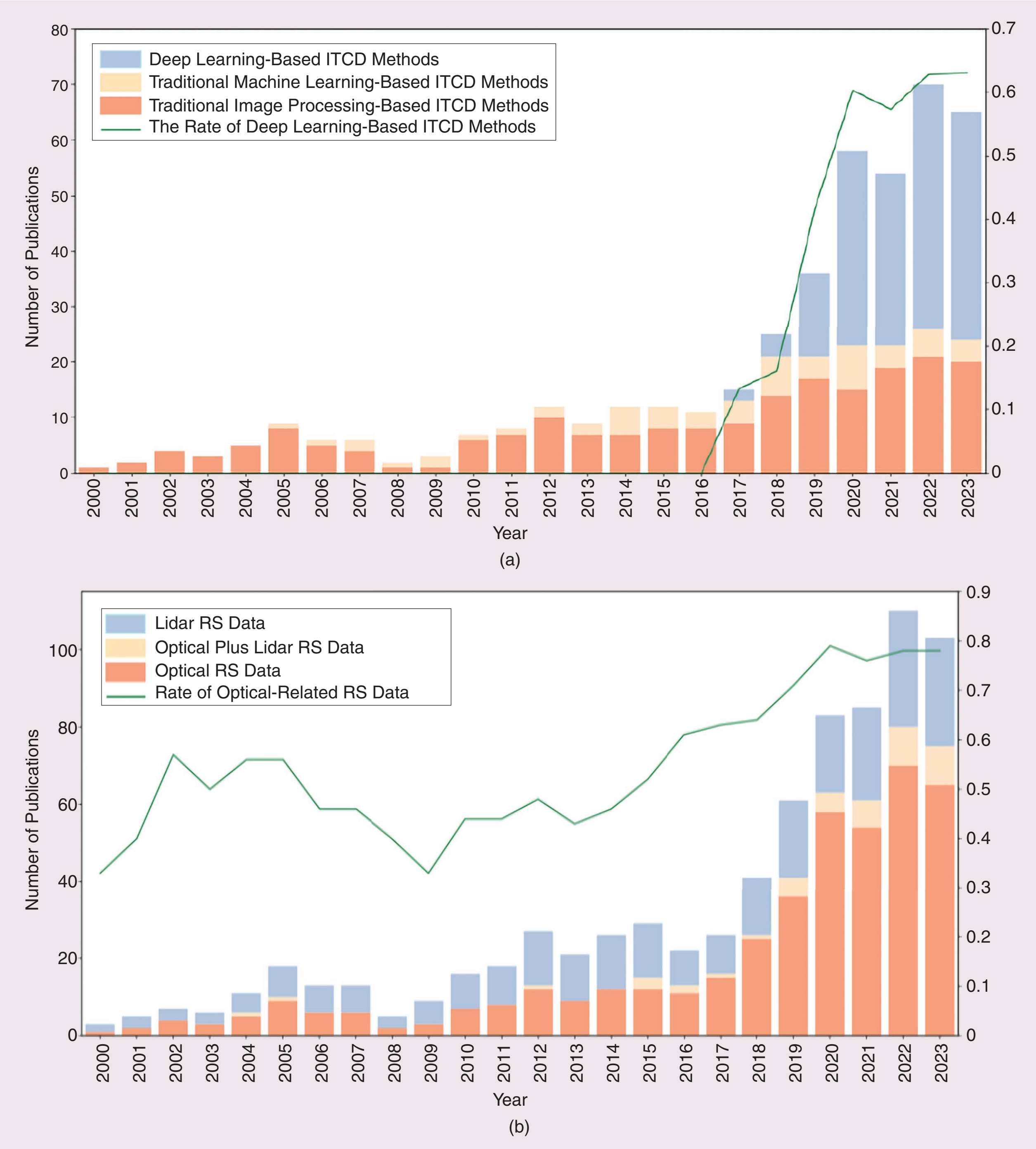

FIGURE 1. (a) The number of ITCDs from lidar data and optical remote sensing (RS) data-related publications from 2000 to 2023. (b) The number of ITCDs from optical RS image-related publications from 2000 to 2022.

impact and significance on global climate change. For others, such as Africa, although it has significant research value and a large distribution of tropical forests, the number of research times is quite low because of its complicated topography, lots of clouds, scarce fieldwork, and poor photograph conditions.

STUDY AREA

As Figure 5 displays, we count the different study areas in our collected articles. Different textures denote different study area and different colors denote different ITCD methods. The red line represents the rate of study area less than or equal to 10 ha and the gray line represents the rate of study area greater than or equal to 1,000 ha. It can be seen that before 2010, the majority of the study areas were smaller than 10 ha. Although, after 2010, the percentage of study areas larger than 1,000 ha are steadily increasing. Only 5.8% of studies are beyond 10,000 ha, and more than 60% of those adopt deep learning-based ITCD methods. We can observe that traditional image processing-based ITCD methods are mainly applied to study areas smaller than 100 ha (at the bottom of Figure 5), and deep learning-based ITCD methods are more applied in larger study areas (≥ 100 ha) (at the top left of Figure 5). We can also infer that the larger the study area is, the more that deep learning-based methods are adopted, which demonstrates that deep learning-based ITCD methods generally have a stronger capacity for efficiency, generalization, and robustness.

SENSOR TYPE

Figure 6 shows the number of sensor types used in ITCD-related publications, displaying the kinds of satellite images that have been used at least two times in ITCD-related articles. More than half of ITCD-related publications adopt aerial images (54%). Recently, spherical cameras have begun to be applied in the ITCD domain, such as cameras with fisheye [28], Google Street images [29], and so on. As for satellite images, it can be distinguished that WorldView and QuickBird data are adopted 23 and 20 times, respectively, for ITCD applications, and present the top two places among other satellite sensor types. The total number of the studies illustrated in Figure 6 is larger than the number of articles examined through satellite images, indicating

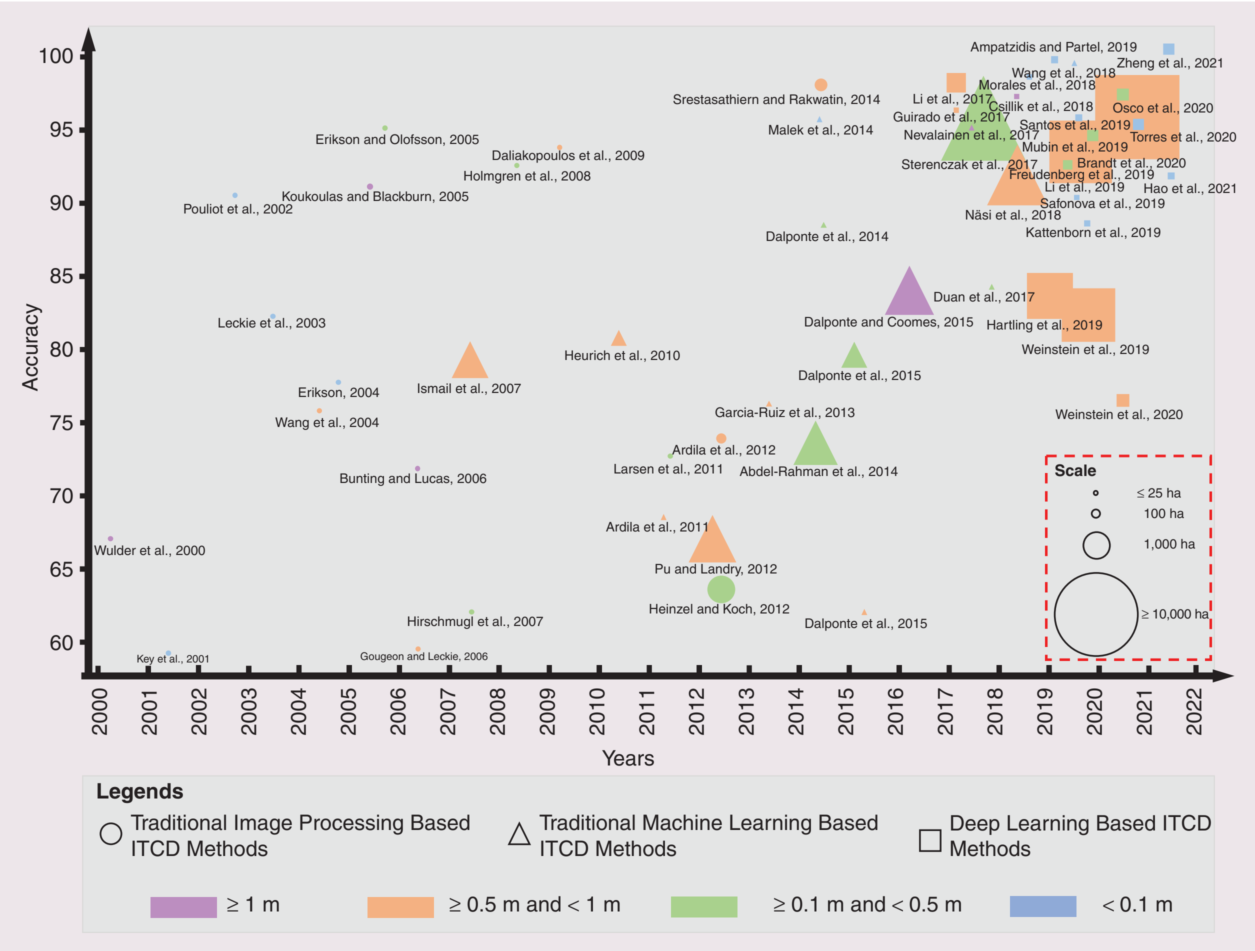

FIGURE 2. The overall trend of ITCD development from some typical examples since 2000. Different shapes represent different ITCD methods, and different colors represent the different spatial resolution of optical images. The larger the size, the larger the study area.

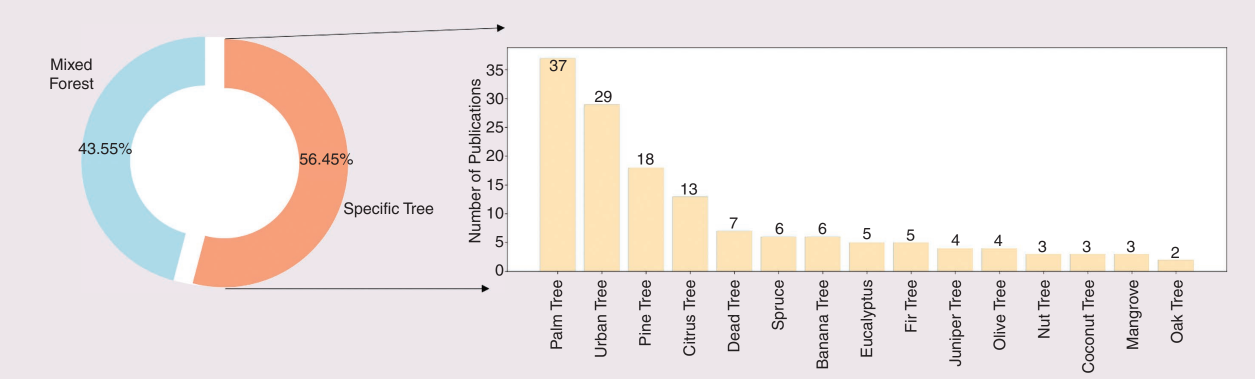

FIGURE 3. The statistics of tree species in ITCD-related publications. In specific trees, we display only the species that have been studied at least two times in ITCD-related articles.

that data from more than one type of sensor are utilized in some publications.

SPATIAL RESOLUTION OF DATA

It can be seen from Figure 7 that researchers use a very high spatial resolution of data in the ITCD domain. Different textures denote different spatial resolutions, and different colors denote different ITCD methods. The red line represents the rate of spatial resolution greater than or equal to 1 m, and the gray line represents the rate of spatial resolution less than or equal to 0.1 m. We can observe that before 2010, the data with spatial resolution greater than or equal to 1 m were widely used in many ITCD-related articles, while the rate of spatial resolution less than or equal to 0.1 m was exponentially increasing, especially after 2016. The most probable reason is that unmanned aerial vehicle (UAV) images have been extensively used in forest inventory. As we can see, ITCD-related articles focus mainly on individual tree detection using high-resolution images. Furthermore, although traditional image processing-based ITCD methods are still the majority, with the increase of spatial resolution, more deep learning-based ITCD methods have been employed, and that can be summarized as deep learning-based ITCD methods own more advantages in very high-resolution image-based individual tree detection with stronger feature extraction and higher accuracy.

METHODOLOGY REVIEW

This section reviews the development and summary of ITCD methodology. We first separate ITCD into individual tree crown detection and individual tree crown delineation, and then we categorize existing ITCD methods into three classes: traditional image processing-based ITCD, traditional machine learning-based ITCD, and deep learning-based ITCD methods. We further categorize existing deep learning-based ITCD methods into two subclasses: object detection-based ITCD and semantic segmentation-based ITCD methods.

ITCD

In this review, ITCD includes ITCD. Individual tree crown detection is mainly oriented to the location of individual trees, such as the center or the coordinates of four corners of the tree crown. Individual tree delineation focuses mostly on sketching the contour and shape of the tree crown or the area of tree crown canopy volume [31]. Table 2 lists detailed ITCD functions for different ITCD methods. In traditional image processing-based ITCD methods, local maximum filtering is the best at tree crown detection, while image segmentation is the best at tree crown delineation. Although image segmentation and image binarization can achieve tree crown counting, they probably require some postprocessing steps. As for traditional machine learning-based ITCD methods, patch-based methods are similar to sliding-window-based ITCD methods; they usually need coordinates to merge after image classification by pixel-based distance [32] or intersection-of-union (IoU) metric [33]. Although pixel-based methods are similar to semantic segmentation-based ITCD ones, they are experts in individual tree crown delineation. On the contrary, as they have the

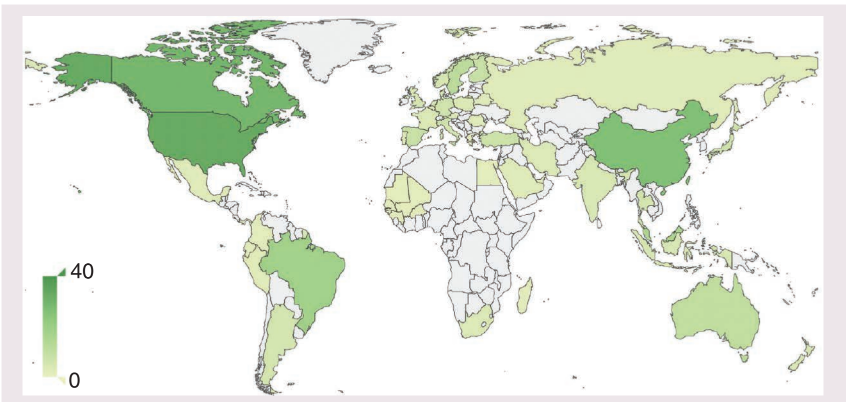

FIGURE 4. The number of study sites around the world according to our database.

common appearance of trees overlapping with each other, machine learning pixel-based and semantic segmentation-based ITCD methods require a postprocessing procedure to produce the final location and contours of individual tree crowns, such as the local maximum detection [34]. Most of the object detection methods can completely accomplish tree crown detection, while they are unable to conduct tree crown delineation except for mask region-based CNNs (R-CNNs) [35]. As seen in Figure 8, a mask R-CNN is an extension algorithm of a faster R-CNN [2], combining both object detection and instance segmentation functions. To this end, a mask R-CNN is capable of individual tree detection and delineation.

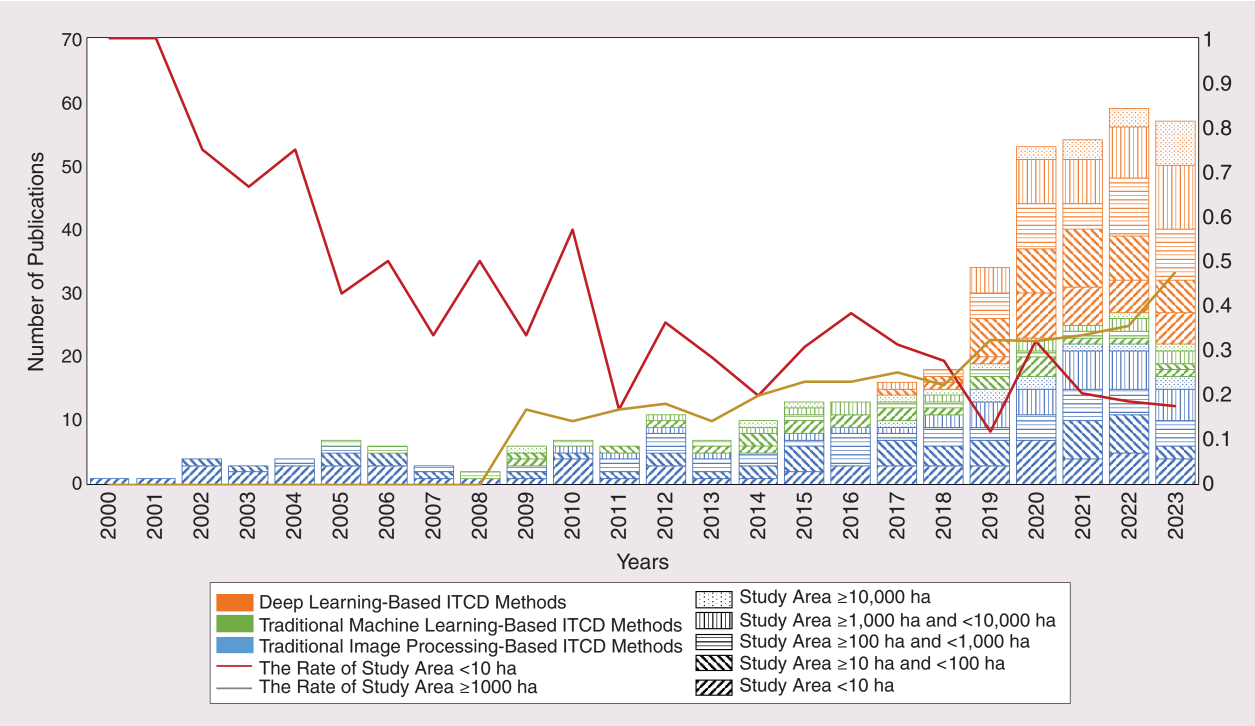

FIGURE 5. The statistics of study areas in ITCD-related publications. Different textures denote different study areas, and different colors denote different ITCD methods. The red line represents a rate-of-study area less than or equal to10 ha and the gray line represents a rate-of-study area greater than or equal to 1,000 ha.

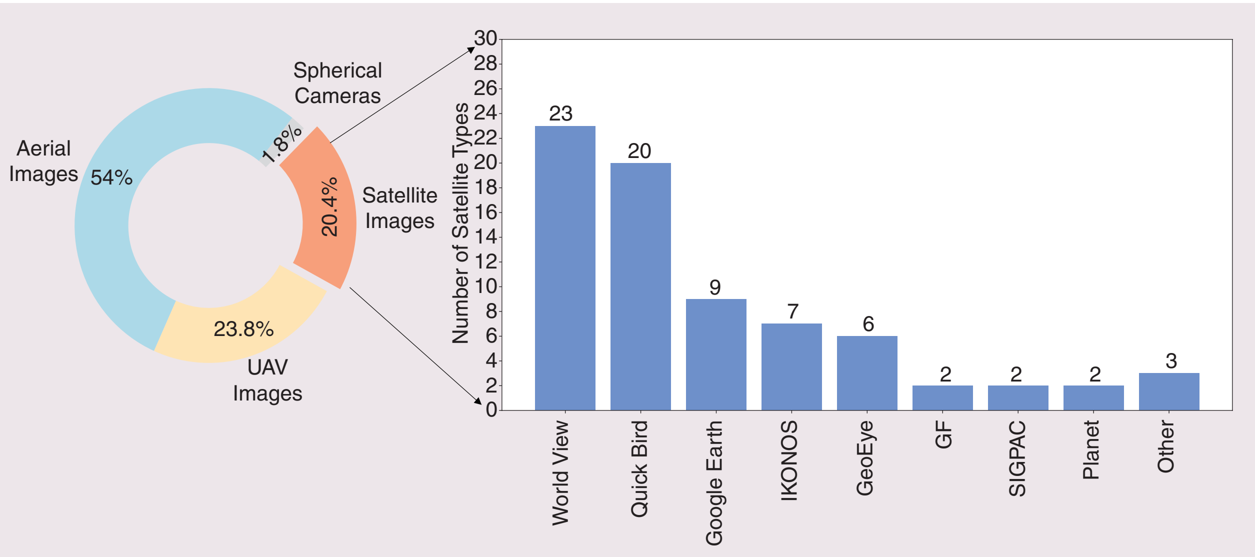

FIGURE 6. The number of sensor types used in ITCD-related publications, displaying the kinds of satellite images that have been used at least two times in ITCD-related articles. GF: GaoFen; UAV: unmanned aerial vehicle.

TRADITIONAL IMAGE PROCESSING-BASED ITCD METHODS

Traditional image processing-based ITCD methods include mainly local maximum filtering, image binarization, template matching, object-based image analysis, image segmentation, and so forth. According to a previous review [11] and its tasks, they can be categorized into two major types: tree crown detection and tree crown delineation (see the "Methodology Review" section). The former four methods major in tree crown detection tasks, while image segmentation majors in tree crown delineation tasks. Table 3 lists traditional image processing-based ITCD methods and the collected examples. Figure 9 displays some typical examples of traditional image processing-based ITCD methods.

TREE DETECTION

The local maximum filtering premise is that the presence of tree crown centers is located at the local maximum reflectance. This simple and efficient method soon became

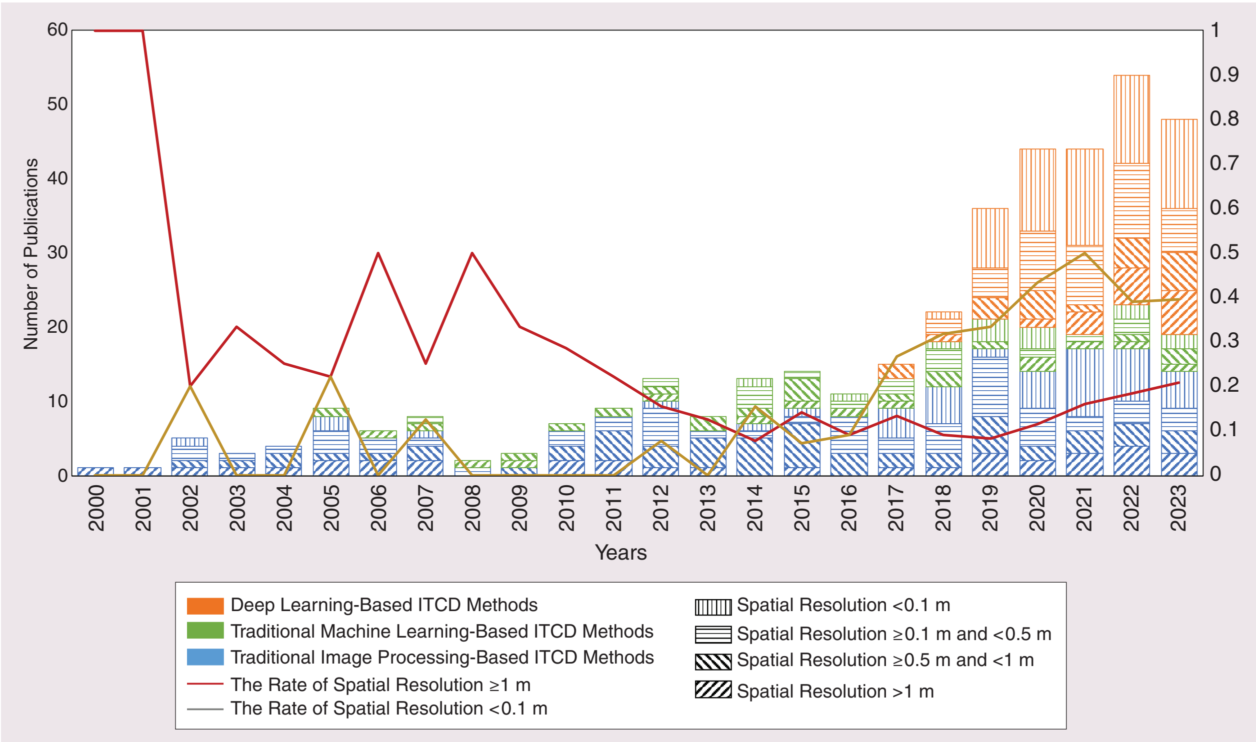

FIGURE 7. The statistics of spatial resolution of images used in publications. Different textures denote different spatial resolutions, and different colors denote different ITCD methods. The red line represents a rate of spatial resolution greater than 1 m, and the gray line represents a rate of spatial resolution less than or equal to 0.1 m.

TABLE 1. A SUMMARY OF EXISTING ITCD-RELATED REVIEWS.

	REVIEWED ITCD METHODS			REVIEWED ITCD TOPICS		REVIEWED ITCD TASKS		
PUBLICATIONS	TIP	TML	DL	DETECTING	DELINEATING	COUNTING	APPLICATIONS	REVIEWED DATA
[10]	x	x	x	✓	✓	x	x	Lidar data
[11]	✓	x	x	✓	✓	x	x	Optical data
[12]	x	x	x	✓	✓	x	x	Lidar data
[13]	✓	x	x	✓	✓	x	x	Lidar data
[14]	x	x	x	✓	✓	x	x	Lidar data
[15]	x	x	✓	✓	✓	x	x	Optical data
[16]	✓	x	x	✓	✓	✓	✓	Optical data
[17]	x	x	✓	✓	✓	x	x	Optical data
[18]	✓	✓	✓	✓	✓	x	x	Optical/lidar data
[19]	✓	✓	x	✓	✓	✓	x	Optical/lidar data
Ours	✓	✓	✓	✓	✓	✓	✓	Optical data

TIP: traditional image processing-based ITCD methods; TML: traditional machine learning-based ITCD methods; DL: deep learning-based ITCD methods; applications: ITCD-related applications such as tree species classification, health monitoring, tree parameter estimation, and so forth.

TABLE 2. DETAILED ITCD FUNCTIONS FOR DIFFERENT ITCD METHODS.

METHOD			DETECTION	DELINEATION	APPLICATIONS
Traditional image processing-based ITCD methods	Local maximum filtering		✓	x	✓
	Image segmentation		x	✓	✓$^{\#}$
	Template matching		✓$^{\#}$	x	✓$^{\#}$
	Image binarization		✓$^{\#}$	✓	✓$^{\#}$
Traditional machine learning-based ITCD methods	Patch based		✓$^{\#}$	x	✓$^{\#}$
	Pixel based		✓$^{\#}$	✓	✓$^{\#}$
Deep learning-based ITCD methods	Semantic segmentation		✓$^{\#}$	✓	✓$^{\#}$
	Object detection	Mask R-CNN	✓	✓	✓
		Others	✓	x	✓

✓ and x: the ITCD method was completely implemented and failed to implement the corresponding functions, respectively; ✓$^{\#}$: the ITCD method can implement corresponding functions through other preprocessing or postprocessing procedures.

the most common treetop detection approach among the traditional image processing-based ITCD methods [40], [41]. Another important branch is template matching, which recognizes trees by calculating the similarity between the templates (ground-truth trees) and the image patches that probably contain tree crowns [43]. The position where the similarity score is highest corresponding to the location where the template best matches the image patch can be recognized as the target tree [57]. Image binarization classifies mainly the image patches into two types, i.e., tree crown and background, through threshold or filtering, which is also named *image thresholding* [58]. Object-based image analysis is also widely because of the improved performance in complex scenarios [59]. During the initial segmentation and low-level feature extraction, object-based image analysis detects trees through the segmented images.

TREE DELINEATION

Image segmentation methods refer mainly to morphological approaches, which basically comprise two major operations: dilation and erosion. Dilation is used to expand the tree regions, making them more connected and complete. Erosion can be applied to refine the tree boundaries by removing small, isolated pixels or noise. To delineate tree crowns, works have developed morphological-based methods, including watershed segmentation [60], region growing [61], valley following [62], and so on.

Note that some research proves that combining methods, even with machine learning or deep learning methods, may achieve better results. For instance, Weinstein et al. [63] use local maximum filtering to create a big set of noisy training samples for training deep learning models, which are fine-tuned by handcrafted labels. Pu et al. [64] design a new combination method that involves watershed segmentation to first segment individual trees, and a k-nearest neighbor (KNN) classifier that refines the final outputs.

These traditional image processing-based methods are all dependent on manual threshold selection and have difficulties with noise images, whose poor generalization and loss of fine-grained information limit their applicability. Despite these drawbacks, these methods are still significant and popular for their simplicity and high efficiency in saving time and labor consumption. The combination of these traditional methods with other deep learning methods also brings new insights into providing fast end-to-end and convergence speed.

TRADITIONAL MACHINE LEARNING-BASED ITCD METHODS

The revolution in machine learning facilitates the development of ITCD by offering powerful, adaptable, and accurate solutions. Generally speaking, for both tree

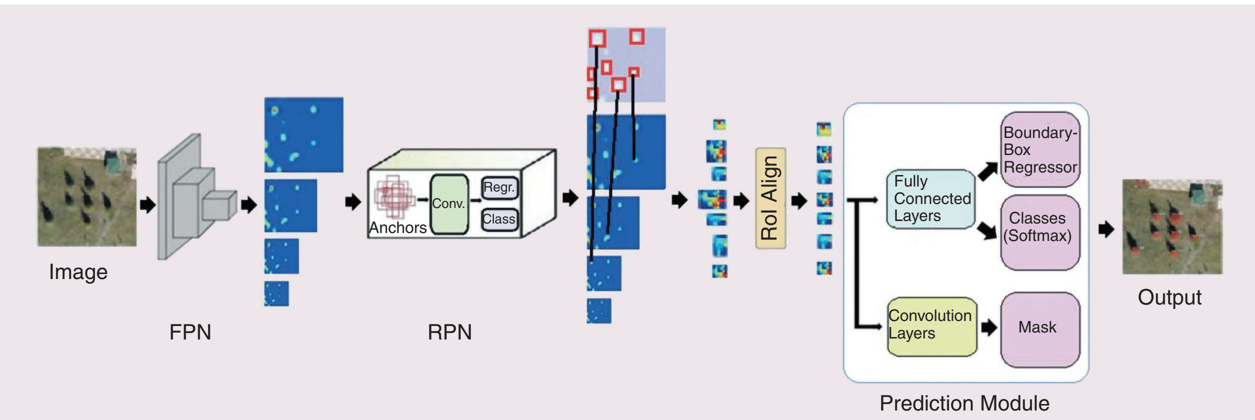

FIGURE 8. A typical ITCD example of an object detection-based ITCD method using the mask R-CNN proposed by [30]. Conv.: convolution; Regr.: regression; ROI: region of interest; RPN: Region Proposal Network; FPN: Feature Pyramid Network.

detection and tree delineation, there are four steps in traditional machine learning-based ITCD methods: 1) image preprocessing, 2) feature extraction, 3) classifier training, and 4) model prediction. Here we focus more on the nature of ITCD, which is progressing in feature extraction and classifier training. Table 4 lists traditional machine learning-based ITCD methods and the collected examples. This section does not separate detection and delineation, because for tree detection and tree delineation, feature extraction and classifier training are both necessary and the employed methods are similar.

FEATURE EXTRACTION

There is a variety of feature extraction methods, which can be simply classified into two types, i.e., nonhandcrafted features and handcrafted features. Nonhandcrafted features utilize mainly the obvious inner features of images themselves, such as spectral information, vegetation index [69], texture characteristics (e.g., the Gray-Level Co-occurrence Matrix) [70], structure characteristics [83], and so forth. Some studies also take spectral reflectance [65], canopy height models [82], and point cloud data [84] into consideration. On the other hand, handcrafted features are specific image representations that are crafted by domain knowledge and prior understanding of the data. These features are created by specific methods (e.g., the principal component transform [69], scale-invariant feature transform [71], histogram of oriented gradient [75], and so on) to capture relevant information that is deemed important for ITCD tasks.

The interpretability of these features makes them useful for understanding and reasoning about the content as they are explicitly designed to capture certain visual attributes like tree shapes and edges. Compared to nonhandcrafted features, handcrafted features present more data-driven characteristics. Still, due to the requirements of manual design and expert understanding, these features lack scalability and transferring ability to new scenarios. In a nutshell, a full understanding of the characteristics and the specific demands of specific ITCD tasks is essential to harness the full potential of these features and is beneficial for later classifier training.

CLASSIFIER TRAINING

Classifier training is the most important part of traditional machine learning ITCD methods. Potential classifiers contain DTs, Gaussian maximum likelihood, linear discriminant analysis, support vector machines (SVMs), extreme learning machines (ELMs), RFs, multilayer perceptrons (MLPs), *k*-means, KNNs, logistic regression, and so forth. Nevalainen [67] compares the ITCD

TABLE 3. A SUMMARY OF TRADITIONAL IMAGE PROCESSING-BASED ITCD METHODS.

TASKS	METHODS		EXAMPLES
Detection	Local maximum filtering		[40], [41]
	Template matching		[42], [43]
	Image binarization		[38], [44]
	Others	Scale-space filtering	[45], [46]
		Object-based image analysis	[47], [48]
		Marked point process	[49], [50]
Delineation	Image segmentation	Watershed segmentation	[51], [52]
		Region growing	[53], [54]
		Valley following	[55], [56]

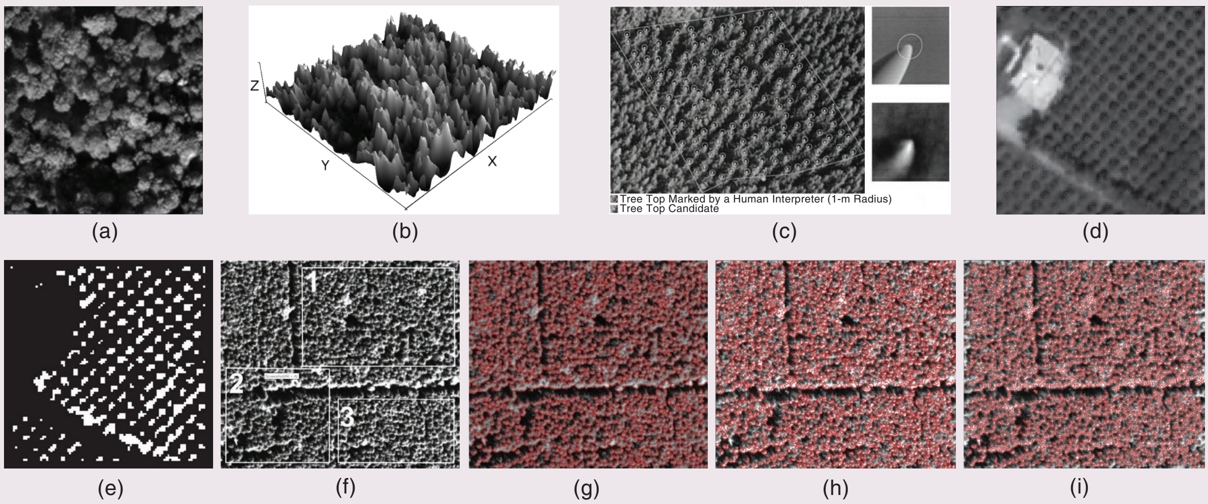

FIGURE 9. Some typical examples of traditional image processing-based ITCD methods. (a) The original image from [36]. (b) The local maxima appearing in the third dimension are associated with the presence of trees [36]. (c) Identification of trees through template matching from [37]. (d) Red band thresholding from [38]. (e) Tree detection results through image binarization [38]. (f) The original QuickBird image from [39]. (g)–(i) Tree delineation results using valley-following, region-growing, and watershed segmentation [39].

performance of five different classifiers: KNN, Bayes classifier, DT, MLP, and RF. They build a high-resolution dataset based on hyperspectral and point cloud data and extract approximately 350 features. The experimental results indicate that MLP achieves the best accuracy, with 95.4%, followed by the KNN, RF, DT, and Bayes classifiers.

In summary, the performance of traditional machine learning-based methods in ITCD relies on efficient feature extraction and powerful classifier training.

TABLE 4. A SUMMARY OF FEATURE EXTRACTION AND ADOPTED CLASSIFIERS IN TRADITIONAL MACHINE LEARNING-BASED ITCD METHODS.

ITEMS	METHODS	EXAMPLES
Feature extraction	Nonhandcrafted features	[65], [66] [67], [68]
	Handcrafted features	[69], [70]; [71], [72]
Adopted classifiers	DT	[65], [73]
	Gaussian maximum likelihood	[74]
	Linear discriminant analysis	[70]
	Support vector machine	[75], [76]
	Extreme learning machine	[71]
	RF	[77], [68]
	Multilayer perceptron	[67]
	k-means	[78], [79]
	k-nearest neighborhood	[80], [81]
	Logistic regression	[82]

TABLE 5. A SUMMARY OF DEEP LEARNING-BASED ITCD METHODS.

METHODS	NETWORKS	EXAMPLES
Sliding window	LeNet [86]	[32], [87]
	VGG [88]	[89], [90], [91]
	ResNet [92]	[93], [94]
	Inception [95]	[96]
	DenseNet [97]	[98]
Object detection	YOLO [5]	[99], [100]
	SSD [101]	[102]
	RetinaNet [103]	[104], [105]
	EfficientDet [106]	[107]
	Faster R-CNN [2]	[28], [108]
	Mask R-CNN [35]	[30], [29]
	MMDetection [3]	[109]
	DetectNet [110]	[111]
	Transformer [112]	[113], [114]
Semantic segmentation	DeepLabV3+ [115]	[116], [117]
	U-Net [118]	[119], [9]
	FCN [120]	[121], [122]
	SegNet [123]	[124]
	FC-DenseNet [125]	[126]
	ResNet like [92]	[34], [127]

SSD: single shot detector.

Compared to traditional image processing-based methods, the ability to automatically learn and extract relevant features from raw data that involve specific expert understanding of traditional machine learning-based methods makes them more adaptable to more different scenarios for ITCD. However, it is important that a set of high-quality and high-quantity input data are a sufficient condition for the promising performance of traditional machine learning-based methods [85]. For example, if the study area is a small region with simple tree targets and landscape invariance and the images are full of noise, traditional image processing-based methods may have better performance. Therefore, choosing or comparing traditional image processing-based and traditional machine learning-based methods, depends on the specific scenarios and data conditions.

DEEP LEARNING-BASED ITCD METHODS

As successful cases emerge in various applications, today, many ITCD methods adopt neural networks, achieving high-accuracy and real-time ITCD results in complex and large-scale regions. Here we review deep learning-based ITCD methods by an extended taxonomy: object detection-based methods for tree detection, and semantic segmentation-based methods for tree delineation. Table 5 lists deep learning-based ITCD methods. Figure 10 displays the number of deep learning-based ITCD methods-related publications since 2017.

TREE DETECTION

A rich line of object detection approaches has been applied to detect a variety of ground objects in the remote sensing field in the past few decades, including tree detection using high-resolution remote sensing data. Object detection algorithms can generally be categorized into two classes: sliding-window-based and end-to-end methods (i.e., two- and one-stage object detection methods, respectively). Sliding-window-based object detection methods were the first deep learning-based tree detection methods and were proposed in 2017 [32] (see Figure 11). According to the characteristics of tree crowns, many scholars have designed or modified new neural network architectures to improve the performance of tree detection. For example, Wu et al.[128] present a two-stage CNN architecture that detects and counts oil palms in Malaysia. The first stage classifies the land cover type, and the second stage classifies the object. The experimental results demonstrate that a two-stage CNN has much fewer confusions with other land cover types (such as other vegetation and buildings) in the whole QuickBird image and achieves the most improvement at 21.27%, compared to a traditional one-stage CNN with respect to the F1 score. Furthermore, some researchers propose novel approaches to reduce time-consuming label interpretation work. For instance, He et al. [90] propose feature learning from image markers to mostly decrease the number of training images in fully connected layers; the accuracy has a slight improvement of 0.3% compared to fine-tuning the VGG structure for coconut detection. These methods are derived from natural image processing, but researchers in the ITCD field re-create or redesign the structures to make the methods appropriate to the features of ITCD.

On the other hand, end-to-end methods contain two- and one-stage object detection frameworks. A two-stage object detection framework like a faster R-CNN (Figure 12), to some extent, consists of the mechanism of the human brain, first giving a coarse scan of the whole image and then focusing on areas of interest. It contains several correlated stages, such as generating region proposals, feature extraction, bounding-box regression, and classification. Driven by the promising performance of a two-stage framework for natural image object detection, many ITCD researchers have adopted a two-stage pipeline for tree detection tasks [129], [130], [131]. For example, Zheng et al. [108] propose the Multiclass Oil PAlm Detection approach (MOPAD) based on the faster R-CNN model to automatically detect five fine-grained oil palm growing statuses in Indonesia. MOPAD is achieved by designing the refined pyramid feature and a hybrid class-balanced loss module to achieve satisfying observation of the growing status for individual oil palms. Their work proves the considerable potential for not only individual oil palm tree detection but also for monitoring growing statuses. Recently, some works have utilized transformers to extract stronger feature representations. For example, Chen et al. [113] propose a semisupervised transformer-based framework for tree counting that designs a pyramid tree representation module based on transformer blocks to extract multiscale features during the encoding stage.

Although the aforementioned methods have achieved good performance, they are still time consuming because

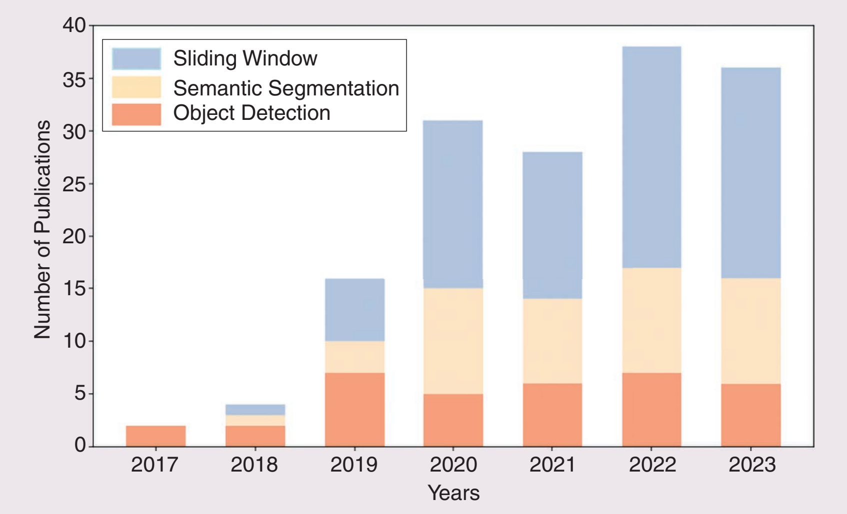

FIGURE 10. The number of deep learning-based ITCD methods-related publications since 2017.

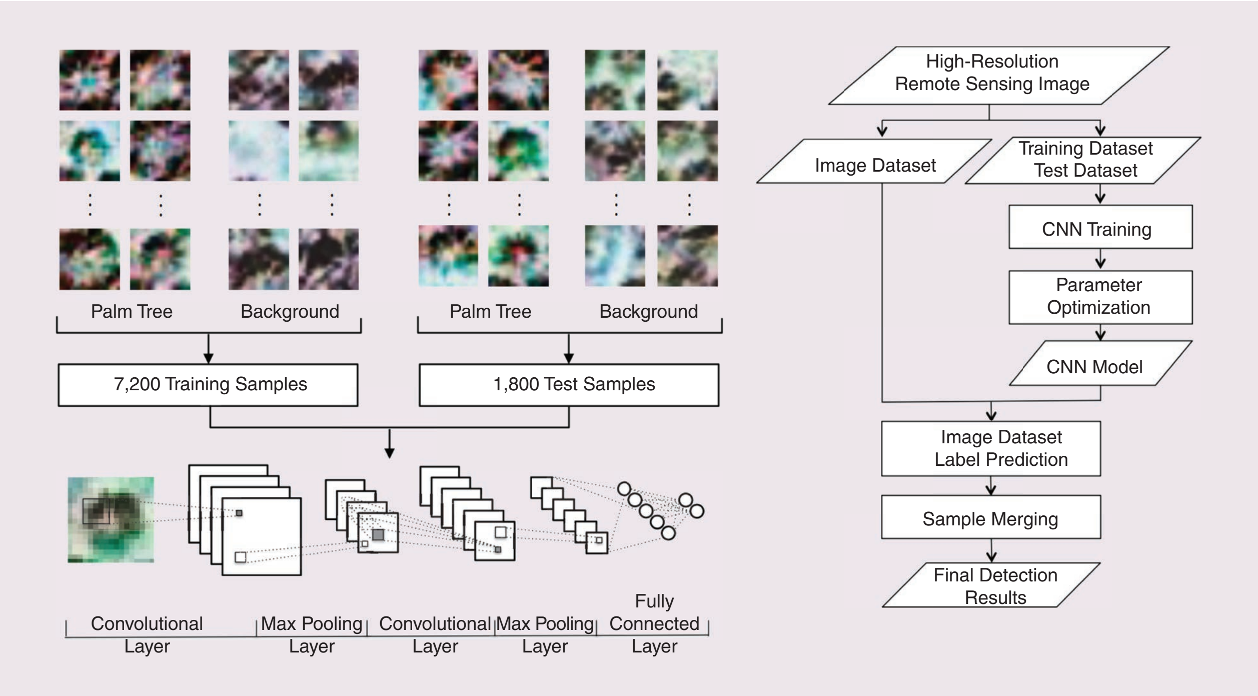

FIGURE 11. A typical example of a CNN classification-based ITCD method proposed by [32]. max: maximum.

of the original proposal detection two-stage scheme. La Rosa et al. [132] conduct an analysis of the efficiency and accuracy of the aforementioned algorithms applied on ITCD applications, indicating that two-stage object detection approaches (such as mask R-CNNs and faster R-CNNs) generally attain higher accuracy than one-stage object detection approaches [such as RetinaNet and YOLO (you only live once) v2/v3], but one-stage methods potentially accelerate the ITCD speed. To achieve faster and real-time tree detection, some researchers adopt one-stage approaches. Based on global regression/classification, a one-stage object detection framework like YOLO and single shot detector (SSD) straightly maps from image pixels to class probabilities and bounding-box coordinates. For example, Albuquerque et al. [133] propose a novel lightweight architecture for dead-tree detection based on the YOLO framework. This architecture includes a specially designed feature extraction module that reuses the features from previous layers for dense connectivity and reduced dependence and a depthwise separable convolution with a small number of parameters to reduce the number of model parameters to achieve real-time detection.

The first row of Table 5 summarizes the algorithms and the collected examples for object detection-based methods for tree detection. In general, sliding-window-based methods are time-consuming approaches due to them producing a considerably large number of latent candidates spanning a variety of sizes. Therefore, sliding-window-based methods are inflexible and inefficient to detect trees with various crown sizes because the patch size of the subimage is required to be predefined through prior human knowledge. As for end-to-end object detection-based methods such as faster R-CNNs, they are more robust and faster, which greatly alleviates the performance drop caused by confusion with other vegetation or complex topography, and so on. Compared to traditional machine learning-based ITCD and sliding-window-based methods, end-to-end object detection-based ones have improved considerably in accuracy and efficiency [108]. Today, end-to-end object detection-based algorithms are increasingly popular and common among all the ITCD methods, representing 25%, 33.3%, 50%, and 46.4% of deep learning-based ITCD methods in 2018, 2019, 2020, and 2021, respectively (see details in Figure 10). Although most of the existing deep learning-based methods have achieved significant success in tree detection within the ITCD domain, they are primarily adapted from techniques (e.g., R-CNNs, faster R-CNNs, SSD, and so on) that were originally designed for natural scene images. However, remote sensing images differ significantly from natural scene images, particularly in terms of rotation, scale variation, and complex, cluttered backgrounds. Although some of these challenges have been partially addressed by incorporating prior knowledge or developing specialized models, ITCD remains an open problem that warrants further research.

TREE DELINEATION

Without requiring the time-consuming sliding-window scheme, the semantic segmentation-based tree delineation method is an end-to-end algorithm. Dissimilar to the object detection-based methods that produce one label for a patch of an image, semantic segmentation

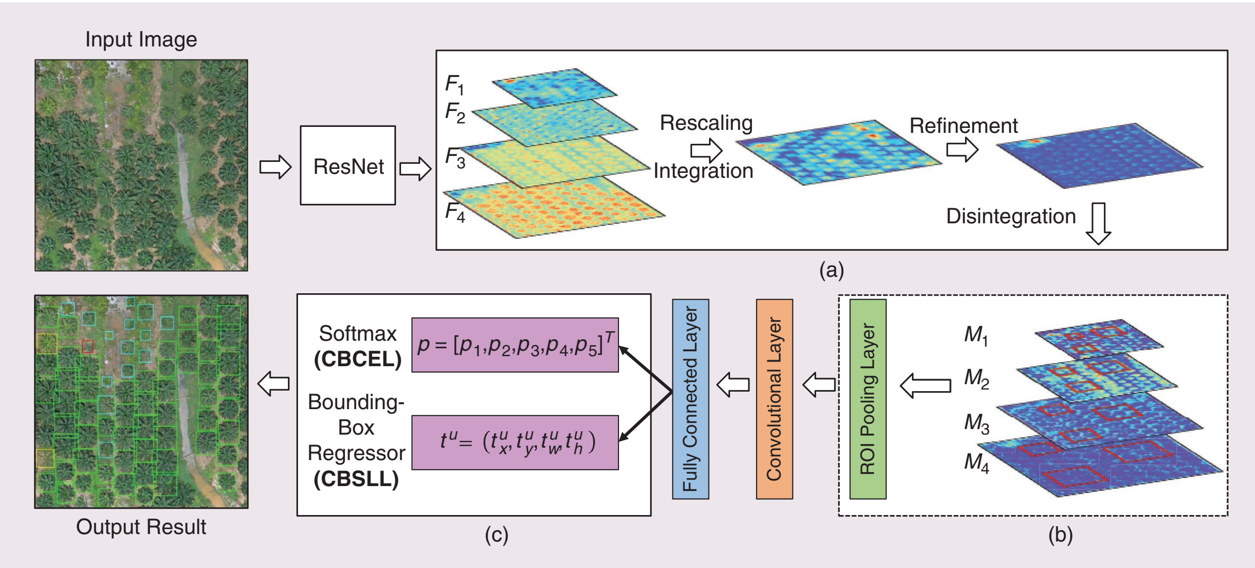

FIGURE 12. A typical example of an object detection-based ITCD method proposed by [108]. (a) RPFs. (b) Multilevel RPN. (c) Hybrid Class-Balanced Loss Module. RPFs: refined pyramid features; ROI: region of interest; ResNet: residual network; CBCEL: class-balanced cross-entropy loss; CBSLL: class-balanced smooth L1 loss.

methods aim at generating dense classes for each pixel in the whole image (see Figure 13). Similar to tree detection, deep learning-based tree delineation also derives from semantic segmentation methods for natural images. Some state-of-the-art semantic segmentation architectures, such as DeepLabV3+, U-Net, fully connected networks (FCNs), PointNet and SegNet, and so forth, have been applied to the ITCD domain in recent years. Some researchers have employed semantic segmentation-like models to generate confidence maps for tree crown extraction [142], where pixels with high confidence indicate the locations of tree crowns. Others have developed new algorithms to mitigate the challenge of requiring large volumes of labeled data for training. For instance, Miraki et al. [143] propose a weakly supervised deep learning pipeline using class activation mapping to detect individual, red-attacked trees, utilizing only image-level labeled remote sensing data. Some studies have also proposed modified semantic segmentation models for delineation tasks. For example, Özcan et al. [144] integrate a set of residual U-Nets and a sequence of automatically derived input scales to introduce a new scale-sequence residual U-Net-based deep learning algorithm. This approach adapts to variations among different types of trees and consistently achieves the highest detection accuracy (DA) (91.67% on average) compared to four other ITCD methods. For comparison among different semantic segmentation methods, Ochoa and Guo [122] evaluate five state-of-the-art tree delineation methods for segmenting citrus trees from UAV multispectral images: DeepLabV3+, dynamic dilated convolution network (DDCN), SegNet, U-Net, and FCN. The experimental results showed comparable F1 scores, with DDCN achieving the highest F1 score: 94.42%. However, DDCN exhibited the lowest detection efficiency, taking 1.02 min per hectare, while other algorithms processed each hectare in approximately 15 s.

The second row of Table 5 summarizes the algorithms and the collected examples for semantic segmentation-based tree delineation methods. In general, semantic segmentation-based tree delineation methods are more efficient than sliding-window-based ones because they generate the detection results of the whole image at once. For example, Brandt et al. [9] extract more than 1.8 billion individual trees over a land area that covers 1.3 million km^2 in the West African Sahara, Sahel, and subhumid zones, with only 5% of the annotated tree crowns overlooked in the final results. Semantic segmentation-based tree delineation methods need to use an overlapping partition way for large-scale remote sensing image prediction into several image patches. In the meantime, every two adjacent patches in the whole image have an overlapping height (width) to make sure that corners are not missed by the algorithm (see Figure 13). However, the performance of semantic segmentation-based tree delineation methods cause worse results for regions with tree crowns that appear to torch or overlap with each other, leading to segmenting some touching or overlapping tree crowns as only one tree crown. Besides that, the output of semantic segmentation-based tree delineation methods is a "confidence" or a "probability" map, meaning the probability that a pixel belongs to the type of tree crown. These methods usually need a postprocessing step to produce the final maps of an individual tree crown and segment overlapping tree crowns, for example, local maximum detection [119].

DATASET IN ITCD FOR DEEP LEARNING METHODS

The bottleneck for deep learning applications regarding ITCD is the difficulty in collecting high-quality training samples. We list representative articles in Table 6 to show how these articles address this bottleneck and achieve promising ITCD results. Also, we list all public ITCD-related datasets and calculate the usage frequency of these datasets.

DATASET CONSTRUCTION

There are four important indexes when constructing a useful dataset. First, the data source is the base for building high-quality datasets. As shown, these selected publications use either high-resolution satellite/aerial images

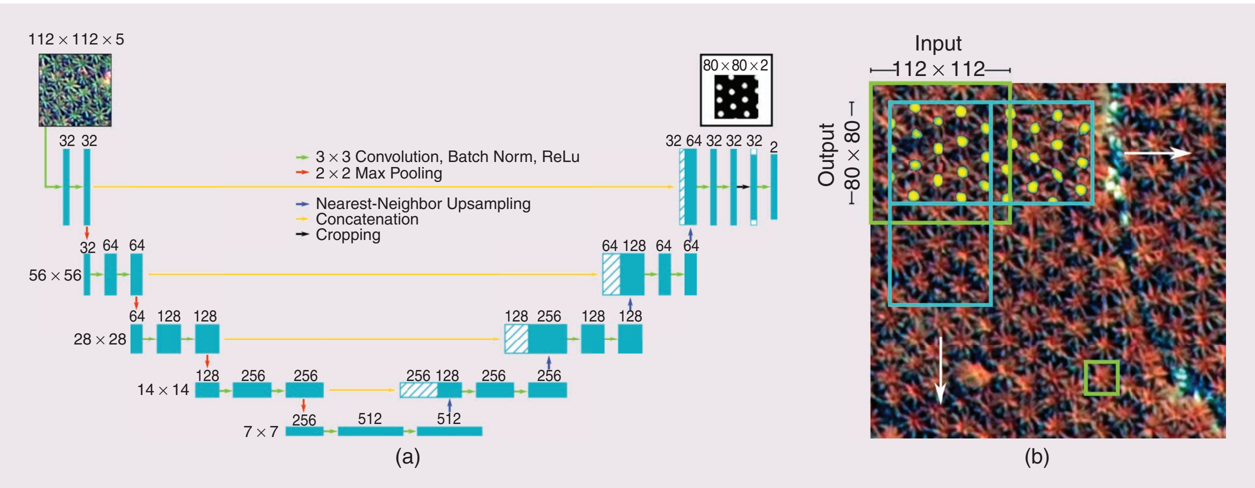

FIGURE 13. A typical example of a semantic segmentation-based ITCD method proposed by [119]. ReLu: rectified linear unit.

or UAV images. High-resolution satellite/aerial images are from mainly high-resolution commercial satellites, such as QuickBird, WorldView, and so on, while most of the UAV images are collected by researchers themselves, with postprocessing on the fly. The second thing is the resolution. ITCD tasks require high-resolution images and these publications achieve resolutions below 1 m. The highest resolution is 0.02 m. The third thing is the image number. These publications all first manually annotate sufficient target samples for training, validating, and testing, regarding the application areas and the number of model parameters. Normally, more than 2,000 images for training are a necessity, which largely prevents models from overfitting. When deep learning models have more than 100 million parameters (i.e., transformers), the basic number of annotation images should increase. The fourth thing is the image size. Sometimes, image size selection is a tradeoff between accuracy and efficiency. To fully dig into the model potential, presurveying the tree size is necessary for predefining the proper image size.

EXISTING DATASETS

There are six public ITCD-related datasets in total. For example, Zheng et al. [33] propose a satellite image dataset for oil palm trees with 0.6 m resolution and 40,000 images. Qiao et al. [141] propose a mixed-tree dataset with aerial images and 104,675,304 annotated instances. We can see that all the public datasets are designed for deep learning-based methods with high resolutions and over thousands of training images. Moreover, we should note that only a few publications publish their annotated datasets for reproduction and re-creation. We should encourage researchers to publish their datasets to contribute to the whole community. Moreover, the usage frequency is low compared to public datasets in other fields (e.g., building extraction and ship detection), which indicates that ITCD researchers tend to construct specific datasets for specific scenarios, rather than using existing ones. The strong heterogeneity of different locations, tree species, and resolutions may contribute to this phenomenon.

EVALUATION OF ITCD RESULTS

TREE CROWN DETECTION

For tree crown detection, we usually adopt true positive, false positive (FP), and false negative to describe the number of trees that are detected correctly, number of others that are detected as trees by model fault, and the amount of ground-truth trees that are overlooked in detection results. According to these three indexes, we can calculate precision, recall, overall accuracy (OA), and F1 score. Precision and recall evaluate the algorithm's capability of correctly detecting trees and the algorithm's capability of completely detecting ground-truth trees, respectively. The OA and F1 scores depict the overall results of the algorithm [145]

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100\% \\ \text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100\% \end{aligned} \tag{1}$$

$$\begin{aligned} \text{OA} &= \frac{\text{Precision}+\text{Recall}}{2} \\ \text{F1 score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} 100\%. \end{aligned} \tag{2}$$

Actually, precision and recall are also named *user's accuracy* and *producer's accuracy*, respectively [70], or *correctness* and *completeness*, respectively [146]. In addition, recall is sometimes considered as DA [121]. Other researchers

TABLE 6. STATISTICS OF REPRESENTATIVE PUBLICATIONS USING DEEP LEARNING-BASED METHODS TO COLLECT DATA.

PUBLICATION	SOURCE	RESOLUTION	SPECIES	IMAGE NUMBER	IMAGE SIZE	INSTANCE NUMBER	AREA	AVAILABILITY	USAGE FREQUENCY
[34]	UAV	0.129 m	Citrus	2,389 for training	256	37,353	70 ha	x	–
[9]	Aerial image	0.5 m	mixed	334,000	256	89,899	5,000 ha	x	–
[134]	UAV	0.04 m	Mixed	14,000 for training	128	–	30 ha	x	–
[89]	UAV	0.04 m	Fir	3,520 for training	150	–	10 ha	x	–
[7]	Satellite image	0.6 m	Coconut	2,000 for training	512	136,500	1,475 ha	x	–
[135]	UAV	0.02 m	mixed	903 for training	1,024	–	36 ha	x	–
[136]	UAV	0.04 m	Mixed	1,603 for training	1,024	–	200 ha	x	–
[137]	UAV	0.02 m	Fir	25,446	128	197,922	4 ha	x	–
[138]	Aerial image	0.3 m	Mangrove	18,000 for training	256	–	645 ha	x	–
[33]	Satellite image	0.6 m	Palm	40,000	17	431,101	7,429 ha	✓	3
[108]	UAV	0.1 m	Palm	6,000 for training	1,024	363,877	3,700 ha	✓	3
[32]	Satellite image	0.6 m	Palm	20,000	17	100,000	7,500 ha	✓	3
[139]	UAV	0.02 m	Mixed	7,521 for training	1024	27,160	105 ha	✓	–
[140]	Aerial image	0.2 m	Mixed	45,343 for training	300	–	4,771,000 ha	✓	11
[141]	Aerial image	–	Mixed	10,000 for training	400	104,675,304	–	✓	8

may adopt both omission and commission errors to evaluate detection results

$$\begin{aligned}\mathrm{OE} &= \frac{\mathrm{FN}}{\mathrm{TP}+\mathrm{FN}} \times 100\% \\ \mathrm{CE} &= \frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}} \times 100\%.\end{aligned} \tag{3}$$

Other overall tree detection accuracy metrics include accuracy index (AI) [49] and matching score (M score) [147], which can be calculated as

$$\begin{aligned}\mathrm{AI} &= \frac{\mathrm{TP}-\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}} \times 100\% \\ \mathrm{M\ score} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}} \times 100\%.\end{aligned} \tag{4}$$

Yin and Wang [14] conclude that OA is the most commonly used measurement in individual tree crown detection assessment. However, according to our collected ITCD publications, F1 score has become the most popular overall tree crown detection accuracy metric (F1 score, M score, AI, and DA). Specially, more than half of the articles that adopt CNN classification and object detection-based ITCD methods use F1 score to quantitatively describe the overall performance of their ITCD algorithms. Furthermore, mean average precision (mAP) has started to gain more attention in the accuracy evaluation of tree crown detection [148]. mAP both combines both recall and precision into a single metric by calculating the area under the precision-recall curve, resulting in a score ranging from zero to one, which is defined as the mean precision at a set of eleven equally spaced recall levels (from zero to one with a step-size of 0.1) by the Pascal VOC Challenge [149]. mAP can be formulated as

$$\mathrm{mAP} = \frac{1}{11} \sum_{\mathrm{Recall} \in \{0, 0.1, \ldots, 1\}} \mathrm{Precision}(\mathrm{Recall}). \tag{5}$$

TREE CROWN DELINEATION

Crown delineation segments an image into multiple parts, each of which is required to be one tree crown. To this end, crown delineation performance can be assessed by segmentation accuracy evaluation. Similar to tree crown detection assessment, most of its evaluation metrics are also available for tree crown delineation evaluation, while we adopt pixel-based rather than object-based (tree-based) evaluation. Some research considers adopting the matching rate to describe the performance of tree crown delineation according to the over- or under-segmentation rate of the segments [51], the overlapping rate of the segments [150], or to other self-defined segmentation criteria [47]. Recently, mean IoU (mIoU) has been adopted in tree crown delineation evaluation [151], which computes the tree crown area overlapped by manual delineation (A^{ref}) and generated delineation (A^{est}) (intersection area) divided by the sum of the tree crown area from both the manual and generated delineations (union area). mIoU can be calculated by the following equation. Some articles use other similar criteria, such as Jaccard score (J score) [152] and area error ratio [153].

$$\mathrm{mIoU} = \frac{1}{S} \sum_{i=1}^{S} \frac{A_i^{\mathrm{ref}} \cap A_i^{\mathrm{est}}}{A_i^{\mathrm{ref}} \cup A_i^{\mathrm{est}}}. \tag{6}$$

DISCUSSIONS

ITCD is of utmost importance for a comprehensive understanding of the ecological environment on both global and local scales. The meta-analysis presented is convenient to outline the past, current, and potential future of ITCD for those who want to know about this specific domain. A thorough introduction of ITCD algorithms in this review article may be of interest to them. In this section, we discuss three ITCD-related issues to further comprehend the ITCD domain.

MULTISENSOR DATA IN THE ITCD DOMAIN

COMPARISON BETWEEN LIDAR DATA AND OPTICAL REMOTE SENSING DATA IN THE ITCD DOMAIN

Lidar is a critical data source for forestry inventory and ecological analysis [154] and has been increasingly adopted in individual tree crown detection and tree parameters estimation [155], such as diameter breast height, leaf area index, above-ground biomass (AGB), and so forth. However, regardless whether it's terrestrial laser scanning (TLS) or airborne laser scanning (ALS), most of the existing forestry inventory concentrates on region scales because of their difficulties and high cost for data collection. Although a UAV equipped with laser scanning is a low-range, low-cost lidar system, its study area is even smaller than TLS and ALS systems. On the other hand, lidar measurements (such as the European Space Agency's BIOMASS and NASA's Global Ecosystem Dynamics Investigation) from satellites that cover larger-scale areas do not satisfy research of the individual tree scale [156], and they focus mainly on some tree parameters' retrieval at a coarser scale. On the contrary, optical data capture tree crown reflectance with more spectral, texture, and semantic information using passive remote sensing instruments. This rich information is beneficial to represent the intrinsic features of vegetation and observe conditions and status (such as disease). Furthermore, high-resolution and global optical data are much easier to acquire than lidar and have a large number of storage data from the last two decades that are waiting for us to use. Actually, with high-resolution optical satellite data, we could soon map every tree on Earth [157]. Individual tree crown detection over a large area in West Africa [9] suggests that it is possible to detect the location and size of every individual tree worldwide according to existing optical satellite data. Although it is unable to provide 3D information, some researchers have explored the potential of side-view optical data (such as fish-eye cameras and Google Street images) to better describe the tree trunk and branches (see details in the "Multisensor Data in the ITCD Domain" section).

MULTISENSOR FUSION IN THE ITCD DOMAIN

Besides the revolution in algorithms, the prosperity of multisensor data also provides strong support for the development of the ITCD domain. Multisensor data fusion plays a pivotal role in advancing ITCD in various environmental monitoring and remote sensing applications [158]. The primary aim of multisensor data fusion is to integrate information from diverse sensors such as optical, lidar, radar, and multispectral to improve the accuracy, completeness, and robustness of tree detection and delineation processes. Hakkenberg et al. [159] fuse lidar and hyperspectral data to map 15 urban tree species, which could provide both vertical and horizontal information, and have shown great potential in improving tree species identification. Multisensor data fusion in ITCD enables researchers to combine complementary data sources that capture different aspects of tree characteristics. For example, optical sensors provide valuable color and texture information, while lidar offers detailed 3D structural data. Radar sensors are proficient at penetrating vegetation, especially in adverse weather conditions. Multispectral sensors provide spectral signatures that are useful for discriminating among tree species. As a result of multisensor data fusion, ITCD algorithms benefit from enhanced spatial and spectral information. The fusion process aids in distinguishing between trees and other objects, accurately estimating tree height and crown diameter and identifying changes in tree cover over time. The outcomes include a more precise forest inventory, better forest management, and informed decision making regarding environmental conservation [66].

COMPARISON OF DIFFERENT ITCD METHODS

Table 7 displays a qualitative assessment for different ITCD methods in three aspects: annotations, efficiency, and accuracy. Here we conduct in-depth discussions on them.

ANNOTATION

It is necessary and fundamental to conduct annotation work in supervised learning. Traditional image processing-based ITCD methods have the least cost, and most of them are unsupervised learning methods and do not require any annotation work, except template matching. Annotation work of semantic segmentation-based ITCD methods is the most difficult and complex among all the methods because it is a pixel-level annotation and has to carefully outline all kinds of fine-grained tree crown shapes. As for traditional machine learning-based ITCD and CNN classification methods, we not only have to annotate the samples of tree crowns but also have to annotate the samples of other land cover types, such as croplands, bare land, water, impervious areas, and so on. As for object detection-based methods, we have to annotate the location of the four corners of a tree crown and generate a bounding box for each tree crown. Of course, for a mask R-CNN, we further have to annotate the thorough shape of tree crowns to conduct tree segmentation. The annotation works of the three aforementioned ITCD methods are more difficult than traditional image processing-based ITCD methods, while they are easier than semantic segmentation-based ITCD methods.

EFFICIENCY

Algorithm efficiency is a crucial and key factor in ITCD applications, especially when applied to large-scale study areas. As most of them are unsupervised learning algorithms, traditional image processing-based ITCD methods cost the most time in simple and basic image operations, usually with low computation complexity and fewer iteration times. Traditional machine learning-based ITCD and CNN classification methods have the worst performance on algorithm and implementation efficiency, given that they require the time-consuming sliding-window scheme to achieve the location and recognition of tree crowns. In addition, classifiers or neural network training and parameter tuning phases worsen their efficiency. Although semantic segmentation-based and object detection-based methods have time-consuming neural network training work, they belong to end-to-end algorithms that allow the detection of several trees in the whole patch image. To this end, these two algorithms are moderately efficient: higher than traditional machine learning-based methods, while lower than machine learning-based ones.

TABLE 7. A QUALITATIVE COMPARISON OF DIFFERENT ITCD METHODS IN ANNOTATION, EFFICIENCY, AND ACCURACY.

METHOD		ANNOTATION	EFFICIENCY	ACCURACY
Traditional image processing-based ITCD methods		+++	+++	+
Traditional machine learning-based ITCD methods		++	+	+
Deep learning-based ITCD methods	CNN classification	++	+	+++
	Semantic segmentation	+	++	+++
	Object detection	++	++	+++

+++: the method that performs best in this respect; +: the method that performs worst in this respect.

ACCURACY

ITCD accuracy is the most important evaluation used to judge whether the ITCD algorithm is successfully applied to practical tree inventory. In general, deep learning-based methods perform the best in accuracy, with a high capacity of robustness and generalization. In particular, deep learning-based ITCD methods achieve more convincing and satisfying results in complex areas. Notably, both semantic segmentation-based and object detection-based ITCD methods achieve a slightly better performance than CNN classification methods. As for traditional image processing-based ITCD methods, their accuracy is generally the lowest among different algorithms and only has satisfactory performance in simple areas or under specific parameters or specific regions. When the study sites turn to a varied topography, complicated environment, or regions with overlapping crowns, the accuracy may incur a terrible deterioration. The accuracy of traditional machine learning-based methods is between that of deep learning-based and traditional image processing-based ones.

COMPARISONS OF GENERAL DEEP LEARNING MODELS AND THEIR APPLICATIONS IN THE ITCD DOMAIN

In Figure 14, the top of the timeline shows the development of general deep learning architectures, and the bottom shows the years that these deep learning models were first used in the ITCD domain. LeNet and AlexNet were proposed in 1998 and 2012, respectively, and were applied in the ITCD domain until 2017. After 2017, in general, novel deep learning models would be adopted in the ITCD domain within three years. For example, faster R-CNNs, Inception, FC-DenseNet, and mask R-CNNs were applied to the ITCD domain three years after they were proposed. YOLOv3, RetinaNet, and MMDetection were adopted in ITCD applications two years after they were first proposed. It took only one year for ResNet and EfficientDet to be utilized in the ITCD domain. Furthermore, DeepLavV3+ was applied to the ITCD domain at almost the same time as when it was proposed. To this end, the time gap between general deep learning models and their applications in the ITCD domain has become increasingly closer. The progress of deep learning architectures plays a vital role in the development of deep learning-based ITCD methods.

According to different deep learning models, we can complete different ITCD tasks (see Table 2). General deep learning models can be directly applied to ITCD scenarios. However, the following differences and modifications should be considered:

- As the size of a remote sensing image is too large to be data that are input for a general deep learning model, we need to utilize an overlapping partition method for a large-scale remote sensing image to be divided into several subimages in the inference phase (see Figure 13). After that, we apply coordinates' transformation and merge the results of all the subimages to achieve the final ITCD results [119].
- As for the design of deep learning architectures, we need to modify the sizes and ratios of candidate anchor boxes in object detection-based ITCD methods because the size of tree crown is usually different from general objects [108]. Furthermore, the number of channels in the first layer usually needs to be modified because of multispectral bands for remote sensing images, rather than only three bands for general images [99].
- Many deep learning methods need to have some postprocessing procedures. For example, the results of semantic segmentation-based ITCD methods are a "confidence map," meaning that they usually require local

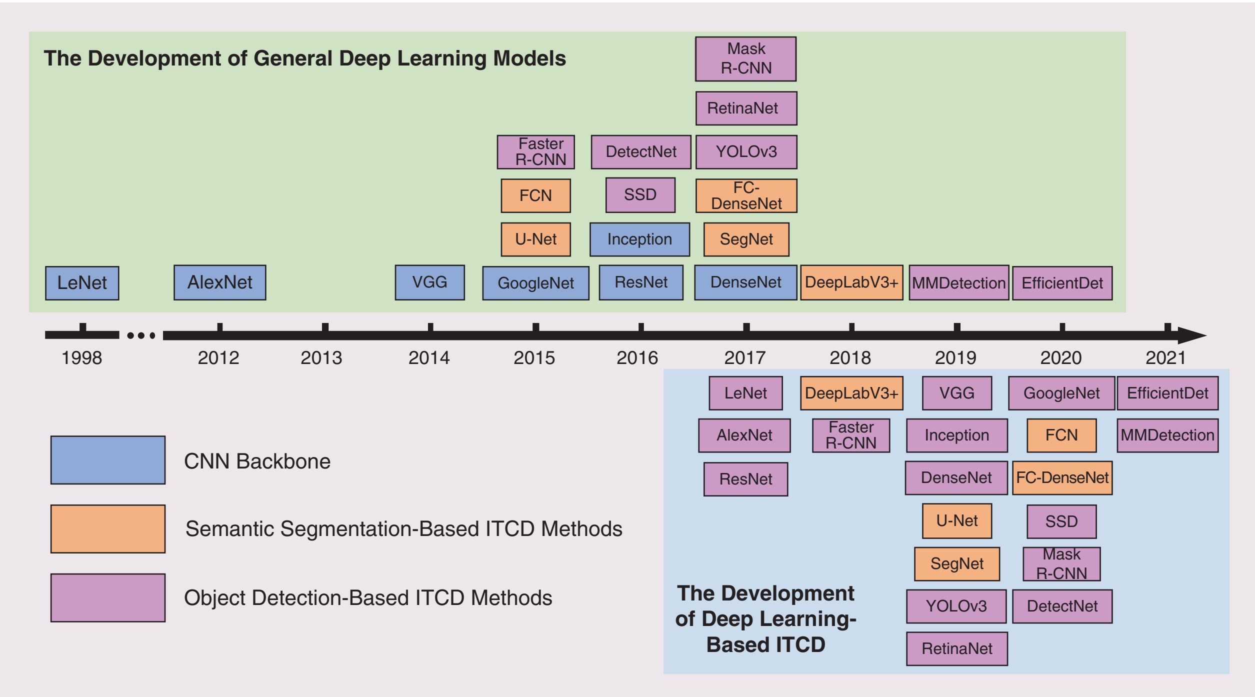

FIGURE 14. Comparisons of general deep learning models and their applications in the ITCD domain. FCN: fully convolutional network.

maximum detection to produce the final locations and contours of individual trees [34]. Some studies design a specific postprocessing regularization to reduce FPs [129] and improve the accuracy of the ITCD model.

CRITERIA FOR CHOOSING APPROPRIATE ITCD METHODS

Due to the complexity of different research subjects (e.g., mixed forests, specific tree species, and so forth) with different attributes (e.g., areas, density, locations, and so on), it is desirable to design or select appropriate ITCD approaches that address ITCD tasks under different scenarios. In this section, we discuss multiple influencing factors of ITCD approaches.

TREE SPECIES

We list the statistics of tree species in ITCD-related publications in Figure 3. There are hardly any traditional machine learning-based or semantic segmentation-based methods that address ITCD tasks within mixed forests. Instead, traditional machine learning-based methods are adopted more for specific tree detection (i.e., urban or dead trees) whose distribution is sparse, and semantic segmentation-based ones are utilized more on trees that gather together. This is not to say traditional machine learning-based or deep learning-based methods cannot handle mixed-tree scenarios. Mixed forests often involve a large number of trees with different species, and machine learning- or deep learning-based methods are data driven, so it is not cost-effective to annotate the mixed forests for training samples, especially when areas are small. Moreover, each method has its own advantages and disadvantages. Taking sliding-window-based methods as an example, the sliding-window scheme is straightforward and easy to understand and implement. Moreover, these schemes can produce results with high recall, which is very important in tree counting and management applications. However, the computation cost is high, which contributes to poor real-time detection and scalability. The fixed window size also limits their transferability. We can see that most of the sliding-window-based articles study one specific area with a fixed tree species, which makes full use of its advantages and avoids its disadvantages. This indicates that when the study area is comparatively small and the tree size is somehow fixed, sliding-window-based methods are a proper selection for ITCD.

TREE DENSITY

Tree density also influences the choice and performance of ITCD approaches. From a vertical comparison, even though 63.72% of traditional image processing-based method publications handle mixed-forest scenarios that own high tree density, their average performance is lower than that of traditional image processing-based methods on scenarios with low tree density. It is the same with traditional machine learning- and deep learning-based methods. From a horizontal comparison, deep learning-based methods have outstanding average performance compared to traditional machine learning-based and traditional image processing-based ones when the tree density is at the same level. However, it is important to note that when the density is sparse, it is not desirable to use semantic segmentation-based methods to map individual trees. In a nutshell, the performance is negatively correlated with tree density. Data-driven methods may have a better performance at the same tree density scenario compared to other methods, only if the training samples are sufficient.

FOREST TYPE AND STRUCTURE

We divide the forest into three types: boreal, temperate, and tropical; and two structures: forest and plantation. Boreal forests consist of mainly regular pine trees and spruce trees, which are suitable for both traditional image processing-based and traditional machine learning-based methods because of their feature invariance. Temperate forests are complex for multiple tree species and high tree density. All the methods' performance drops slightly in this scenario, so it is the application area that determines the choice of the utilized method. If the area is smaller than 1,000 ha, it is better to use traditional image processing-based or unsupervised traditional machine learning-based methods due to their efficiency. Otherwise, data-driven methods show their superiority. There are few ITCD studies that focus on tropical forests because of the severe overlapping of trees. For plantations like oil palm tree and olive, because of human management, there is barely any overlapping or tree shading. The proper distance between trees allows for all kinds of methods to perform well. However, when the scenarios are complex (i.e., mixed with dead or growing trees), and the areas are large (i.e., larger than 1,000 ha), it is ideal to utilize deep learning-based methods.

ITCD-RELATED APPLICATIONS

In this section, we introduce some practical ITCD-related applications. Other ITCD applications include wildfire potential estimation [160], plant heterogeneity [161], wildlife protection [162] and biodiversity research [163], and so forth. Most of them first conduct individual tree crown detection and then further conduct other analyses on a single-tree scale. With the recent emergence of end-to-end deep learning techniques, we are able to achieve individual tree crown detection or delineation, along with individual tree species classification or health monitoring in the meantime.

TREE SPECIES CLASSIFICATION

Tree species classification is a valuable and important task in forest science, helping us to understand the role of trees' ecological functions [21]. Most of the previous individual tree species classifications are two-stage work, including individual tree crown detection or delineation, and then species recognition [164]. Researchers adopt common classifiers, such as RF [84], SVM [72], and so on, to classify the collected features for each detected tree crown. Today, CNN [165] and 3D CNN [166] are increasingly applied in tree species classification and achieve superior results under enormous input-extracted features. More recently, mask R-CNNs have been able to achieve both end-to-end individual tree detection and delineation, even along with individual tree species classification (see "Classes (Softmax)" in the "Prediction Module" in Figure 8) [136]. That is, individual tree species classification is no longer a two-stage workflow and has become a more simple but effective one-stage workflow through object detection-based ITCD methods [167].

HEALTH MONITORING

Multispectral information from optical remote sensing data, coupled with machine learning or deep learning techniques, plays a considerable role in tree health monitoring, including disease surveillance, growing status observation, tree mortality mapping, and so forth. Similar to tree species classification, previous popular individual tree's health monitoring is two-stage work [168], although, as deep learning-based ITCD methods emerge, existing tree's health assessment becomes a one-stage framework, accomplishing both individual tree crown detection and their status observation [68]. Compared to lidar data, deep learning may perform better on multispectral optical remote sensing data because of its rich semantic and texture information, which is also beneficial for capturing vegetation's intrinsic features. As a matter of fact, employing comprehensive health monitoring, especially for economical tree species, is beneficial to improve their productivity or yield, and further increase the economic effect [169].

TREE PARAMETERS ESTIMATION

Tree parameters are quite vital biophysical representations that influence water, energy, and carbon exchanges between the atmosphere and forest ecosystems. As aerial lidar and terrestrial lidar are able to acquire 3D point returns, most of the existing studies adopt them to estimate most of the tree-related parameters, including first-order properties (such as height, crown diameters, and so on) [170] and second-order properties (such as basal area, AGB, and so forth) [171]. On the other hand, optical remote sensing data are also widely applied in studies that link tree parameters from the field to observations through the sensitivity of optical reflectance to canopy structure variations. Optical remote sensing provides great potential for tree parameter estimation at a larger scale than lidar data [172]. For example, [9] analyze the canopy cover, tree density and tree crown size over 1.3 million km^2 in West Africa, after detecting trees by a semantic segmentation-based ITCD method. However, optical remote sensing data are poor in height-related parameters [173], which means that combining lidar or adopting side-view remote sensing data may address this issue. Furthermore, end-to-end forest attribute retrieving in the one-stage framework for individual trees still needs to be exploited and developed in the future.

MULTITEMPORAL CHANGE ANALYSIS

Multitemporal remote sensing data are not only able to conduct individual tree crown detection but also explore the changes of individual crown diameters, canopy cover, growth process, and so forth, evaluating the variants of ecological restoration [54] and carbon stock [174] or the impacts of tree species competition [175] and natural disasters [176]. Also, multitemporal data contribute to better accomplishing individual tree crown detection and ITCD-related applications through seasonal spectral and texture variations [177]. However, similar to tree parameters estimation, existing single tree-level change analyses are all two-stage works. Following the development of recurrent neural networks, we believe that coupling semantic segmentation-based and object detection-based ITCD methods with time-series analysis may achieve real-time, high-accuracy, and end-to-end single tree-level change analysis using multitemporal remote sensing data.

PROSPECTS

Based on the aforementioned literature analysis, methodology review, and in-depth discussion, ITCD-related prospects have emerged from this attempt, which concerns past, current, and future trends.

USING MULTISOURCE AND MULTIVIEW REMOTE SENSING DATA

Some ITCD researchers combine optical remote sensing data with other remote sensing data to extract high-dimension features, such as point clouds from lidar [72], digital topographic models [76], digital surface models [178], or geographic information system data [179]. Meanwhile, some articles simultaneously adopt satellite and UAV images to achieve ITCD. For instance, Tan et al. [104] first utilize multispectral satellite images (from *WorldView-2*, *Panet*, and *Sentinel-2*) to extract banana plantation regions, and then use UAV images to precisely locate each banana plant and recognize its health condition.

However, existing studies have not exploited the potential of fully fusing multisource remote sensing data. For example, most of the UAV images are unable to describe abundant spectral information because they have only three bands (red, green, and blue), rather than the multispectral images photographed by highly expensive multispectral-based cameras. If we make full use of high spatial-resolution UAV images and high spectral-resolution satellite images [180], it is considerably beneficial to precisely recognize tree crowns and classify fine-grained tree species or growing status with high spectral–spatial-resolution remote sensing data. On the other hand, remote sensing data acquired from a vertical view cannot perfectly extract the overlapping tree crowns or those that are sheltered from higher mature trees with larger crowns. If remote sensing data from the side view are available, we are capable of recognizing those trees that are easily missed from vertical-view remote sensing data. Some researchers have attempted to detect and delineate individual tree crowns using side-view remote sensing data [181]. We believe that integrating vertical- and side-view remote sensing data achieves better ITCD performance [182]. In addition, combining synthetic aperture radar (SAR) [183] or interferometric SAR [184] with lidar or optical remote sensing data also seems promising for the ITCD domain, and we may pay more attention to exploiting other observation platforms with optical remote sensing data in the future.

FINE-GRAINED TREE SPECIES OR GROWING STATUS CLASSIFICATION

Fine-grained individual tree classification includes both fine-grained tree species and fine-grained growing status classifications. The former is significant for understanding the distribution of forest species and protecting biodiversity. The latter not only observes damaged or diseased trees to prevent their proliferation but is also conducive to estimating yield and increasing the benefits for some economic trees. To this end, fine-grained tree classification is extremely valuable to both ecology and economy. Most of the existing individual tree species studies focus on small areas (smaller than 1,000 ha) [21]. Although Zhang et al. [185] estimate the number of tree species in tropical areas, they have not mapped the distribution of fine-grained tree species. As for growing status observation, most of them are classified into only two statuses: healthy and unhealthy trees [104]. By contrast, few studies are devoted to multiclass growing status classification [68]. It is highly demanded for plantations (such as oil palm, olive, and so on) to monitor more fine-grained healthy conditions, such as specific diseases. Until now, individual tree classification work has been a two-stage scheme, first detecting or delineating individual tree crowns, and then completing species or growing status recognition. Furthermore, fine-grained classification requires recognizing the slighted difference between similar classes, which may need high spatial- and high spectral-resolution remote sensing images. The data-fusing approaches mentioned in the "Using Multisource and Multiview Remote Sensing Data" section would be an effective solution.

LARGE-SCALE ITCD IN SPATIAL BIG DATA ERA

Undoubtedly, we are presently in the big data era and will continue to be so in the future. Massive remote sensing images acquired by satellites, aerial planes, UAVs, and even mobile phones create the spatial big data era. With these Earth-observation data, we have the opportunity to achieve large-scale ITCD and deeply comprehend global tree resources. Most of the study areas in existing ITCD research are smaller than 1,000 km^2, except for those in [9]. They extract more than 1.8 billion individual trees over a land area that spans 1.3 million km^2 in the West African Sahara, Sahel, and subhumid zones, with only 5% of labeled trees being overlooked in the final results. Despite the fact that Crowther et al. [31] estimate that there are roughly 3.04 trillion trees around the world, they only approximately map the global tree distribution and tree density. There are two major challenges in large-scale ITCD work. The first is the capacity of model generalization. As we have to prepare multitemporal, multisource, and multiregional remote sensing data to conduct large-scale ITCD [186], [187], [188], developing a more transferable, robust, and general model is a powerful foundation, using advanced algorithms such as domain generalization, domain adaptation [189], [190], and transfer learning. Another challenge is the capacity of computation performance to support the efficiency of ITCD in large-scale areas. At present, some studies adopt high-performance computation platforms (such as field-programmable gate arrays and GPUs) to accelerate ITCD algorithms [191], [192]. However, global-, continental-, or national-level ITCD research has not been completed in higher-performance computing platforms such as supercomputers, which may be a potential general platform for processing global observation issues.

FOUNDATION MODELS FOR ITCD

Large vision models (LVMs) that utilize transformers or contrastive learning have demonstrated remarkable success in various computer vision tasks, ranging from image classification [193], [194] to object detection and segmentation. Some remote sensing foundation models have been constructed to support remote sensing downstream tasks. Due to the heterogeneity in different areas, different tree species, and different scales, ITCD is often constrained to one region, small scales, and single-tree species. The transferability is poor. Moreover, the heterogeneity also increases dataset annotation consumption when transferring to other regions, enlarging the scale and classifying new tree species. We think that the application of LVMs in the field of ITCD holds significant promise. With billions of parameters and millions of input data, LVMs have more powerful visual representation ability to extract more useful deep features, leading to better performance on downstream tasks. For example, in ITCD, where distinguishing among different tree species or detecting individual tree crowns in dense forests can be challenging, the superior feature representation capabilities of LVMs can lead to extracting subtle differences in texture, color, and shape, which are critical for differentiating tree crowns from other objects or vegetation. The use of pretrained LVMs allow for transfer learning, where models trained on large general datasets can be fine-tuned for specific ITCD tasks. This approach reduces the need for extensive labeled data. In the meantime, because of its strong generalization ability, it allows the models to perform well across different geographical regions, forest types, and environmental conditions. Such scalability reduces the need for retraining models for specific datasets, making them more versatile and applicable to various ITCD scenarios globally.

CONCLUSIONS

ITCD using high-resolution optical remote sensing data is essential for forestry inventory and ecological analysis in an automated way. In this review article, a comprehensive overview of ITCD-related research was introduced. First, we conducted an investigation of scientific peer-reviewed journal articles over 20 years, building an available database and carrying out a meta-analysis. Second, intriguing and thorough ITCD methods that depict the trend and development of past years relating to this specific domain were presented. We classify ITCD methods into three types: traditional image processing based (such as local maximum filtering, image segmentation, and so on), traditional machine learning based (such as RF and DT and so forth), and deep learning based. In addition, we also categorized deep learning-based ITCD methods into three types (i.e., CNN classification, semantic segmentation, and object detection) and discussed their pros and cons. At the current pace that the methodology of ITCD research is conducted, such information is rather essential and truly valuable. In addition, we discussed three ITCD-related topics to further comprehend the ITCD domain, such as comparisons between lidar data and optical remote sensing data, comparisons among different algorithms, and different ITCD tasks. Finally, some ITCD-related applications and a few existing and emerging topics were presented, and we promise the significance and prosperity of ITCD in the future.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant 42401415 and Grant T2125006), Fundamental Research Funds for the Central Universities, Sun Yat-sen University (Project 24xkjc002), and Jiangsu Innovation Capacity Building Program (Project BM2022028). The cofirst authors are Juepeng Zheng and Shuai Yuan. The corresponding authors are Weijia Li and Haohuan Fu.

AUTHOR INFORMATION

Juepeng Zheng (zhengjp8@mail.sysu.edu.cn) is with the School of Artificial Intelligence, Sun Yat-Sen University, Zhuhai 519080, China, and National Supercomputing Center in Shenzhen, Shenzhen 518054, China. He is a Member of IEEE.

Shuai Yuan (shuai914@connect.hku.hk) is with the Department of Geography, The University of Hong Kong, Hong Kong 999077, China.

Weijia Li (liweij29@mail.sysu.edu.cn) is with the School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai 519082, China.

Haohuan Fu (haohuan@tsinghua.edu.cn) is with Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China, Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China, National Supercomputing Center in Shenzhen, Shenzhen 518054, China, and Tsinghua University (Department of Earth System Science) Xi'an Institute of Surveying and Mapping Joint Research Center for Next-Generation Smart Mapping, Beijing 100084, China. He is a Senior Member of IEEE.

Le Yu (leyu@tsinghua.edu.cn) is with the Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China. He is a Senior Member of IEEE.

Jianxi Huang (jxhuang@cau.edu.cn) is with the Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China, College of Land Science and Technology, China Agricultural University, Beijing 100083, China, and Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing 100083, China. He is a Senior Member of IEEE.

REFERENCES

[1] M. C. Hansen et al., "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, no. 6160, pp. 850–853, 2013, doi: 10.1126/science.1244693.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[3] K. Chen et al., "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*.

[4] S. Yuan et al., "FUSU: A multi-temporal-source land use change segmentation dataset for fine-grained urban semantic understanding," 2024, *arXiv:2405.19055*.

[5] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[6] S. Yuan et al., "Relational part-aware learning for complex composite object detection in high-resolution remote sensing images," *IEEE Trans. Cybern.*, vol. 54, no. 10, pp. 6118–6131, Oct. 2024, doi: 10.1109/TCYB.2024.3392474.

[7] J. Zheng et al., "Surveying coconut trees using high-resolution satellite imagery in remote atolls of the pacific ocean," *Remote Sens. Environ.*, vol. 287, Mar. 2023, Art. no. 113485, doi: 10.1016/j.rse.2023.113485.

[8] S. Yuan, J. Zheng, L. Zhang, R. Dong, R. C. Cheung, and H. Fu, "MUREN: Multistage recursive enhanced network for coal-fired power plant detection," *Remote Sens.*, vol. 15, no. 8, 2023, Art. no. 2200, doi: 10.3390/rs15082200.

[9] M. Brandt et al., "An unexpectedly large count of trees in the west African Sahara and Sahel," *Nature*, vol. 587, no. 7832, pp. 78–82, 2020, doi: 10.1038/s41586-020-2824-5.

[10] J. Hyyppä, H. Hyyppä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo, "Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1339–1366, 2008, doi: 10.1080/01431160701736489.

[11] Y. Ke and L. J. Quackenbush, "A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing," *Int. J. Remote Sens.*, vol. 32, no. 17, pp. 4725–4747, 2011, doi: 10.1080/01431161.2010.494184.

[12] M. A. Wulder et al., "Lidar sampling for large-area forest characterization: A review," *Remote Sens. Environ.*, vol. 121, pp. 196–209, Jun. 2012, doi: 10.1016/j.rse.2012.02.001.

[13] Z. Zhen, L. J. Quackenbush, and L. Zhang, "Trends in automatic individual tree crown detection and delineation—Evolution of LiDAR data," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 333, doi: 10.3390/rs8040333.

[14] D. Yin and L. Wang, "How to assess the accuracy of the individual tree-based forest inventory derived from remotely sensed data: A review," *Int. J. Remote Sens.*, vol. 37, no. 19, pp. 4521–4553, 2016, doi: 10.1080/01431161.2016.1214302.

[15] H. Zhao, J. Morgenroth, G. Pearse, and J. Schindler, "A systematic review of individual tree crown detection and delineation with convolutional neural networks (CNN)," *Current Forestry Rep.*, vol. 9, no. 3, pp. 149–170, 2023, doi: 10.1007/s40725-023-00184-3.

[16] E. Lindberg and J. Holmgren, "Individual tree crown methods for 3D data from remote sensing," *Current Forestry Rep.*, vol. 3, no. 1, pp. 19–31, 2017, doi: 10.1007/s40725-017-0051-6.

[17] T. Yun et al., "Status, advancements and prospects of deep learning methods applied in forest studies," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 131, Jul. 2024, Art. no. 103938, doi: 10.1016/j.jag.2024.103938.

[18] M. Walker and G. A. Dahle, "Literature review of unmanned aerial systems and LIDAR with application to distribution utility vegetation management," *Arboric. Urban For.*, vol. 48, no. 3, pp. 242–257, 2020.

[19] M. Mohan et al., "Individual tree detection using UAV-lidar and UAV-SFM data: A tutorial for beginners," *Open Geosci.*, vol. 13, no. 1, pp. 1028–1039, 2021, doi: 10.1515/geo-2020-0290.

[20] 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," *Can. J. Remote Sens.*, vol. 42, no. 5, pp. 619–641, 2016, doi: 10.1080/07038992.2016.1207484.

[21] F. E. Fassnacht et al., "Review of studies on tree species classification from remotely sensed data," *Remote Sens. Environ.*, vol. 186, pp. 64–87, Dec. 2016, doi: 10.1016/j.rse.2016.08.013.

[22] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning–method overview and review of use for fruit detection and yield estimation," *Comput. Electron. Agric.*, vol. 162, pp. 219–234, Jul. 2019, doi: 10.1016/j.compag.2019.04.017.

[23] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction," *IEEE Access*, vol. 9, pp. 63,406–63,439, 2021, doi: 10.1109/ACCESS.2021.3075159.

[24] M. Michałowska and J. Rapiński, "A review of tree species classification based on airborne LiDAR data and applied classifiers," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 353, doi: 10.3390/rs13030353.

[25] A. Ahmed, A. Ibrahim, and S. Hussein, "Detection of palm tree pests using thermal imaging: A review," in *Machine Learning Paradigms: Theory and Application*, A. E. Hassanien, Ed., Cham, Switzerland: Springer-Verlag, 2019, pp. 253–270.

[26] K. L. Chong, K. D. Kanniah, C. Pohl, and K. P. Tan, "A review of remote sensing applications for oil palm studies," *Geo-Spatial Inf. Sci.*, vol. 20, no. 2, pp. 184–200, 2017, doi: 10.1080/10095020.2017.1337317.

[27] T. Santika et al., "Impact of palm oil sustainability certification on village well-being and poverty in Indonesia," *Nature Sustainability*, vol. 4, no. 2, pp. 109–119, 2021, doi: 10.1038/s41893-020-00630-1.

[28] G. D. Pearse et al., "Detecting and mapping tree seedlings in UAV imagery using convolutional neural networks and field-verified data," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 156–169, Oct. 2020, doi: 10.1016/j.isprsjprs.2020.08.005.

[29] S. Lumnitz, T. Devisscher, J. R. Mayaud, V. Radic, N. C. Coops, and V. C. Griess, "Mapping trees along urban street networks with deep learning and street-level imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 144–157, May 2021, doi: 10.1016/j.isprsjprs.2021.01.016.

[30] N. E. Ocer, G. Kaplan, F. Erdem, D. Kucuk Matci, and U. Avdan, "Tree extraction from multi-scale UAV images using mask R-CNN with FPN," *Remote Sens. Lett.*, vol. 11, no. 9, pp. 847–856, 2020, doi: 10.1080/2150704X.2020.1784491.

[31] T. W. Crowther et al., "Mapping tree density at a global scale," *Nature*, vol. 525, no. 7568, pp. 201–205, 2015, doi: 10.1038/nature14967.

[32] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," *Remote Sens.*, vol. 9, no. 1, 2017, Art. no. 22, doi: 10.3390/rs9010022.

[33] J. Zheng et al., "Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 154–177, Sep. 2020, doi: 10.1016/j.isprsjprs.2020.07.002.

[34] L. P. Osco et al., "A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 97–106, Feb. 2020, doi: 10.1016/j.isprsjprs.2019.12.010.

[35] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, 2017, pp. 2961–2969, doi: 10.1109/ICCV.2017.322.

[36] M. F. Gomes and P. Maillard, "Detection of tree crowns in very high spatial resolution images," in *Environmental Applications of Remote Sensing*, M. Marghany, Ed., London, U.K.: IntechOpen, 2016, pp. 41–71.

[37] M. Larsen and M. Rudemo, "Optimizing templates for finding trees in aerial photographs," *Pattern Recognit. Lett.*, vol. 19, no. 12, pp. 1153–1162, 1998, doi: 10.1016/S0167-8655(98)00092-0.

[38] I. N. Daliakopoulos, E. G. Grillakis, A. G. Koutroulis, and I. K. Tsanis, "Tree crown detection on multispectral VHR satellite imagery," *Photogramm. Eng. Remote Sens.*, vol. 75, no. 10, pp. 1201–1211, 2009, doi: 10.14358/PERS.75.10.1201.

[39] Y. Ke and L. J. Quackenbush, "A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution imagery," *Int. J. Remote Sens.*, vol. 32, no. 13, pp. 3625–3647, 2011, doi: 10.1080/01431161003762355.

[40] M. Wulder, K. O. Niemann, and D. G. Goodenough, "Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery," *Remote Sens. Environ.*, vol. 73, no. 1, pp. 103–114, 2000, doi: 10.1016/S0034-4257(00)00101-2.

[41] X. Xu, Z. Zhou, Y. Tang, and Y. Qu, "Individual tree crown detection from high spatial resolution imagery using a revised local maximum filtering," *Remote Sens. Environ.*, vol. 258, Jun. 2021, Art. no. 112397, doi: 10.1016/j.rse.2021.112397.

[42] A. M. Niccolai, A. Hohl, M. Niccolai, and C. D. Oliver, "Decision rule-based approach to automatic tree crown detection and size classification," *Int. J. Remote Sens.*, vol. 31, no. 12, pp. 3089–3123, 2010, doi: 10.1080/01431160903154374.

[43] D. G. Leckie, N. Walsworth, and F. A. Gougeon, "Identifying tree crown delineation shapes and need for remediation on high resolution imagery using an evidence based approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 206–227, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.02.005.

[44] H. Z. Shafri, N. Hamdan, and M. I. Saripan, "Semi-automatic detection and counting of oil palm trees from high spatial resolution airborne imagery," *Int. J. Remote Sens.*, vol. 32, no. 8, pp. 2095–2115, 2011, doi: 10.1080/01431161003662928.

[45] T. Brandtberg, "Automatic individual tree based analysis of high spatial resolution aerial images on naturally regenerated boreal forests," *Can. J. For. Res.*, vol. 29, no. 10, pp. 1464–1478, 1999, doi: 10.1139/x99-150.

[46] L. Wang, "A multi-scale approach for delineating individual tree crowns with very high resolution imagery," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 4, pp. 371–378, 2010, doi: 10.14358/PERS.76.4.371.

[47] P. Bunting and R. Lucas, "The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data," *Remote Sens. Environ.*, vol. 101, no. 2, pp. 230–248, 2006, doi: 10.1016/j.rse.2005.12.015.

[48] A. Chemura, I. van Duren, and L. M. van Leeuwen, "Determination of the age of oil palm from crown projection area detected from WorldView-2 multispectral remote sensing data: The case of Ejisu-Juaben district, Ghana," *ISPRS J. Photogramm. Remote Sens.*, vol. 100, pp. 118–127, Feb. 2015, doi: 10.1016/j.isprsjprs.2014.07.013.

[49] J. Zhou et al., "Mapping local density of young eucalyptus plantations by individual tree detection in high spatial resolution satellite images," *For. Ecol. Manage.*, vol. 301, pp. 129–141, Aug. 2013, doi: 10.1016/j.foreco.2012.10.007.

[50] M. F. Gomes et al., "Individual tree crown detection in sub-meter satellite imagery using marked point processes and a geometrical-optical model," *Remote Sens. Environ.*, vol. 211, pp. 184–195, Jun. 2018, doi: 10.1016/j.rse.2018.04.002.

[51] L. Jing, B. Hu, T. Noland, and J. Li, "An individual tree crown delineation method based on multi-scale segmentation of imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 70, pp. 88–98, Jun. 2012, doi: 10.1016/j.isprsjprs.2012.04.003.

[52] Z. Zheng et al., "Mapping functional diversity using individual tree-based morphological and physiological traits in a subtropical forest," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112170, doi: 10.1016/j.rse.2020.112170.

[53] M. Erikson, "Species classification of individually segmented tree crowns in high-resolution aerial images using radiometric and morphologic image measures," *Remote Sens. Environ.*, vol. 91, nos. 3–4, pp. 469–477, 2004, doi: 10.1016/j.rse.2004.04.006.

[54] P. Gärtner, M. Förster, A. Kurban, and B. Kleinschmit, "Object based change detection of Central Asian Tugai vegetation with very high spatial resolution satellite imagery," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 31, pp. 110–121, Sep. 2014, doi: 10.1016/j.jag.2014.03.004.

[55] D. G. Leckie, F. A. Gougeon, N. Walsworth, and D. Paradine, "Stand delineation and composition estimation using semi-automated individual tree crown analysis," *Remote Sens. Environ.*, vol. 85, no. 3, pp. 355–369, May 2003, doi: 10.1016/S0034-4257(03)00013-0.

[56] D. G. Leckie, F. A. Gougeon, S. Tinis, T. Nelson, C. N. Burnett, and D. Paradine, "Automated tree recognition in old growth conifer stands with high resolution digital imagery," *Remote Sens. Environ.*, vol. 94, no. 3, pp. 311–326, 2005, doi: 10.1016/j.rse.2004.10.011.

[57] C. Hung, M. Bryson, and S. Sukkarieh, "Multi-class predictive template for tree crown detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 170–183, Mar. 2012, doi: 10.1016/j.isprsjprs.2012.01.009.

[58] D. Koc-San, S. Selim, N. Aslan, and B. T. San, "Automatic citrus tree extraction from UAV images and digital surface models using circular Hough transform," *Comput. Electron. Agric.*, vol. 150, pp. 289–301, Jul. 2018, doi: 10.1016/j.compag.2018.05.001.

[59] N. A. F. Nik Effendi et al., "Unlocking the potential of hyperspectral and LiDAR for above-ground biomass (AGB) and tree species classification in tropical forests," *Geocarto Int.*, vol. 37, no. 25, pp. 1–26, 2021, doi: 10.1080/10106049.2021.1990419.

[60] D. Pouliot, D. J. King, and D. G. Pitt, "Development and evaluation of an automated tree detection delineation algorithm for monitoring regenerating coniferous forests," *Can. J. For. Res.*, vol. 35, no. 10, pp. 2332–2345, 2005, doi: 10.1139/x05-145.

[61] J. Gu and R. G. Congalton, "Individual tree crown delineation from UAS imagery based on region growing by over-segments with a competitive mechanism," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2021, doi: 10.1109/TGRS.2021.3074289.

[62] D. Leckie, F. Gougeon, D. Hill, R. Quinn, L. Armstrong, and R. Shreenan, "Combined high-density lidar and multispectral imagery for individual tree crown analysis," *Can. J. Remote Sens.*, vol. 29, no. 5, pp. 633–649, 2003, doi: 10.5589/m03-024.

[63] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1309, doi: 10.3390/rs11111309.

[64] Y. Pu, D. Xu, H. Wang, X. Li, and X. Xu, "A new strategy for individual tree detection and segmentation from leaf-on and leaf-off UAV-LIDAR point clouds based on automatic detection of seed points," *Remote Sens.*, vol. 15, no. 6, 2023, Art. no. 1619, doi: 10.3390/rs15061619.

[65] T. R. Tooke, N. C. Coops, N. R. Goodwin, and J. A. Voogt, "Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications," *Remote Sens. Environ.*, vol. 113, no. 2, pp. 398–407, 2009, doi: 10.1016/j.rse.2008.10.005.

[66] M. Dalponte, H. O. Ørka, L. T. Ene, T. Gobakken, and E. Næsset, "Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data," *Remote Sens. Environ.*, vol. 140, pp. 306–317, Jan. 2014, doi: 10.1016/j.rse.2013.09.006.

[67] O. Nevalainen et al., "Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging," *Remote Sens.*, vol. 9, no. 3, 2017, Art. no. 185, doi: 10.3390/rs9030185.

[68] K. Johansen et al., "Mapping the condition of macadamia tree crops using multi-spectral UAV and worldview-3 imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 165, pp. 28–40, Jul. 2020, doi: 10.1016/j.isprsjprs.2020.04.017.

[69] Y. O. Ouma and R. Tateishi, "Urban-trees extraction from quickbird imagery using multiscale *spectex*-filtering and non-parametric classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 63, no. 3, pp. 333–351, 2008, doi: 10.1016/j.isprsjprs.2007.10.006.

[70] R. Pu and S. Landry, "A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species," *Remote Sens. Environ.*, vol. 124, pp. 516–533, Sep. 2012, doi: 10.1016/j.rse.2012.06.011.

[71] S. Malek, Y. Bazi, N. Alajlan, H. AlHichri, and F. Melgani, "Efficient framework for palm tree detection in UAV images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4692–4703, Dec. 2014, doi: 10.1109/JSTARS.2014.2331425.

[72] M. Dalponte, L. T. Ene, M. Marconcini, T. Gobakken, and E. Næsset, "Semi-supervised SVM for individual tree crown species classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 110, pp. 77–87, Dec. 2015, doi: 10.1016/j.isprsjprs.2015.10.010.

[73] G. Tochon et al., "On the use of binary partition trees for the tree crown segmentation of tropical rainforest hyperspectral images," *Remote Sens. Environ.*, vol. 159, pp. 318–331, Mar. 2015, doi: 10.1016/j.rse.2014.12.020.

[74] Y. Bai, N. Walsworth, B. Roddan, D. A. Hill, K. Broersma, and D. Thompson, "Quantifying tree cover in the forest–grassland ecotone of British Columbia using crown delineation and pattern detection," *For. Ecol. Manage.*, vol. 212, nos. 1–3, pp. 92–100, 2005, doi: 10.1016/j.foreco.2005.03.005.

[75] Y. Wang, X. Zhu, and B. Wu, "Automatic detection of individual oil palm trees from UAV images using hog features and an SVM classifier," *Int. J. Remote Sens.*, vol. 40, no. 19, pp. 7356–7370, 2019, doi: 10.1080/01431161.2018.1513669.

[76] L. Windrim, A. J. Carnegie, M. Webster, and M. Bryson, "Tree detection and health monitoring in multispectral aerial imagery and photogrammetric pointclouds using machine learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2554–2572, 2020, doi: 10.1109/JSTARS.2020.2995391.

[77] S. I. Roth, R. Leiterer, M. Volpi, E. Celio, M. E. Schaepman, and P. C. Joerg, "Automated detection of individual clove trees for yield quantification in northeastern Madagascar based on multi-spectral satellite data," *Remote Sens. Environ.*, vol. 221, pp. 144–156, Feb. 2019, doi: 10.1016/j.rse.2018.11.009.

[78] J. Recio, T. Hermosilla, L. Ruiz, and J. Palomar, "Automated extraction of tree and plot-based parameters in citrus orchards from aerial images," *Comput. Electron. Agric.*, vol. 90, pp. 24–34, Jan. 2013, doi: 10.1016/j.compag.2012.10.005.

[79] M. Dalponte, F. Reyes, K. Kandare, and D. Gianelle, "Delineation of individual tree crowns from ALS and hyperspectral data: A comparison among four methods," *Eur. J. Remote Sens.*, vol. 48, no. 1, pp. 365–382, 2015, doi: 10.5721/EuJRS20154821.

[80] M. Heurich, T. Ochs, T. Andresen, and T. Schneider, "Object-orientated image analysis for the semi-automatic detection of dead trees following a spruce bark beetle (Ips typographus) outbreak," *Eur. J. For. Res.*, vol. 129, no. 3, pp. 313–324, 2010, doi: 10.1007/s10342-009-0331-1.

[81] Y. Mollaei, A. Karamshahi, and S. Y. Erfanifard, "Detection of the dry trees result of oak borer beetle attack using world-view-2 satellite and UAV imagery an object-oriented approach," *J Remote Sens. GIS*, vol. 7, no. 1, 2018, Art. no. 232, doi: 10.4172/2469-4134.1000232.

[82] J. Wu, W. Yao, and P. Polewski, "Mapping individual tree species and vitality along urban road corridors with LiDAR and imaging sensors: Point density versus view perspective," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1403, doi: 10.3390/rs10091403.

[83] M. López-López, R. Calderón, V. González-Dugo, P. J. Zarco-Tejada, and E. Fereres, "Early detection and quantification of almond red leaf blotch using high-resolution hyperspectral and thermal imagery," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 276, doi: 10.3390/rs8040276.

[84] A. Kamińska, M. Lisiewicz, K. Stereńczak, B. Kraszewski, and R. Sadkowski, "Species-related single dead tree detection using multi-temporal ALS data and CIR imagery," *Remote Sens. Environ.*, vol. 219, pp. 31–43, Dec. 2018, doi: 10.1016/j.rse.2018.10.005.

[85] P. Gong, J. Wang, and H. Huang, "Stable classification with limited samples in global land cover mapping: Theory and experiments," *Sci. Bull.*, vol. 69, no. 12, pp. 1862–1865, 2024, doi: 10.1016/j.scib.2024.03.040.

[86] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[87] N. A. Mubin, E. Nadarajoo, H. Z. M. Shafri, and A. Hamedianfar, "Young and mature oil palm tree detection and counting using convolutional neural network deep learning method,"

Int. J. Remote Sens., vol. 40, no. 19, pp. 7500–7515, 2019, doi: 10.1080/01431161.2019.1569282.

[88] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[89] A. Safonova, S. Tabik, D. Alcaraz-Segura, A. Rubtsov, Y. Maglinets, and F. Herrera, "Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 643, doi: 10.3390/rs11060643.

[90] I. E. de Souza and A. X. Falcão, "Learning CNN filters from user-drawn image markers for coconut-tree image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2020.3020098.

[91] J. Zheng, W. Wu, S. Yuan, H. Fu, W. Li, and L. Yu, "Multisource-domain generalization-based oil palm tree detection using very-high-resolution (VHR) satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021, doi: 10.1109/LGRS.2021.3061726.

[92] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[93] Y. Sun, Q. Xin, J. Huang, B. Huang, and H. Zhang, "Characterizing tree species of a tropical wetland in southern China at the individual tree level based on convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4415–4425, Nov. 2019, doi: 10.1109/JSTARS.2019.2950721.

[94] M. Onishi and T. Ise, "Explainable identification and mapping of trees using UAV RGB image and deep learning," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, 2021, doi: 10.1038/s41598-020-79653-9.

[95] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[96] G. Hu, C. Yin, M. Wan, Y. Zhang, and Y. Fang, "Recognition of diseased Pinus trees in UAV images using deep learning and AdaBoost classifier," *Biosyst. Eng.*, vol. 194, pp. 138–151, Jun. 2020, doi: 10.1016/j.biosystemseng.2020.03.021.

[97] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[98] S. Hartling, V. Sagan, P. Sidike, M. Maimaitijiang, and J. Carron, "Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning," *Sensors*, vol. 19, no. 6, 2019, Art. no. 1284, doi: 10.3390/s19061284.

[99] Y. Ampatzidis, V. Partel, B. Meyering, and U. Albrecht, "Citrus rootstock evaluation utilizing UAV-based remote sensing and artificial intelligence," *Comput. Electron. Agric.*, vol. 164, Sep. 2019, Art. no. 104900, doi: 10.1016/j.compag.2019.104900.

[100] K. Itakura and F. Hosoi, "Automatic tree detection from three-dimensional images reconstructed from 360 spherical camera using YOLO v2," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 988, doi: 10.3390/rs12060988.

[101] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer-Verlag, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[102] A.-I. Pleşoianu, M.-S. Stupariu, I. Şandric, I. Pătru-Stupariu, and L. Drăguţ, "Individual tree-crown detection and species classification in very high-resolution remote sensing imagery using a deep learning ensemble model," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2426, doi: 10.3390/rs12152426.

[103] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, 2017, pp. 2999–3007, doi: 10.1109/ICCV.2017.324.

[104] M. G. Selvaraj et al., "Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 110–124, Nov. 2020, doi: 10.1016/j.isprsjprs.2020.08.025.

[105] B. G. Weinstein, S. Marconi, S. A. Bohlman, A. Zare, and E. P. White, "Cross-site learning in deep learning RGB tree crown detection," *Ecol. Inf.*, vol. 56, Mar. 2020, Art. no. 101061, doi: 10.1016/j.ecoinf.2020.101061.

[106] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2020, pp. 10,781–10,790, doi: 10.1109/CVPR42600.2020.01079.

[107] A. Ammar, A. Koubaa, and B. Benjdira, "Deep-learning-based automated palm tree counting and geolocation in large farms from aerial geotagged images," *Agronomy*, vol. 11, no. 8, 2021, Art. no. 1458, doi: 10.3390/agronomy11081458.

[108] J. Zheng et al., "Growing status observation for oil palm trees using unmanned aerial vehicle (UAV) images," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 95–121, Mar. 2021, doi: 10.1016/j.isprsjprs.2021.01.008.

[109] P. Zamboni et al., "Benchmarking anchor-based and anchor-free state-of-the-art deep learning methods for individual tree detection in RGB high-resolution images," *Remote Sens.*, vol. 13, no. 13, 2021, Art. no. 2482, doi: 10.3390/rs13132482.

[110] A. Tao, J. Barker, and S. Sarathy, "DetectNet: Deep neural network for object detection in digits," *Parallel Forall*, vol. 4, p. 424, Aug. 2016.

[111] D. Pulido, J. Salas, M. Rös, K. Puettmann, and S. Karaman, "Assessment of tree detection methods in multispectral aerial images," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2379, doi: 10.3390/rs12152379.

[112] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[113] H. A. Amirkolaee, M. Shi, and M. Mulligan, "TreeFormer: A semi-supervised transformer-based framework for tree counting from a single high resolution image," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, 2023, doi:10.1109/TGRS.2023.3295802.

[114] G. Chen and Y. Shang, "Transformer for tree counting in aerial images," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 476, doi: 10.3390/rs14030476.

[115] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[116] M. P. Ferreira et al., "Individual tree detection and species classification of amazonian palms using UAV images and deep learning," *For. Ecol. Manage.*, vol. 475, Nov. 2020, Art. no. 118397, doi: 10.1016/j.foreco.2020.118397.

[117] J. A. C. Martins et al., "Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3054, doi: 10.3390/rs13163054.

[118] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, Cham, Switzerland: Springer-Verlag, 2015, pp. 234–241.

[119] M. Freudenberg, N. Nölke, A. Agostini, K. Urban, F. Wörgötter, and C. Kleinn, "Large scale palm tree detection in high resolution satellite images using U-Net," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 312, doi: 10.3390/rs11030312.

[120] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[121] C. Xiao, R. Qin, and X. Huang, "Treetop detection using convolutional neural networks trained through automatically generated pseudo labels," *Int. J. Remote Sens.*, vol. 41, no. 8, pp. 3010–3030, 2020, doi: 10.1080/01431161.2019.1698075.

[122] L. P. Osco et al., "Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery," *Precis. Agric.*, vol. 22, no. 4, pp. 1–18, 2021, doi: 10.1007/s11119-020-09777-5.

[123] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[124] K. S. Ochoa and Z. Guo, "A framework for the management of agricultural resources with automated aerial imagery detection," *Comput. Electron. Agric.*, vol. 162, pp. 53–69, Jul. 2019, doi: 10.1016/j.compag.2019.03.028.

[125] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, 2017, pp. 11–19, doi: 10.1109/CVPRW.2017.156.

[126] D. Lobo Torres et al., "Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery," *Sensors*, vol. 20, no. 2, 2020, Art. no. 563, doi: 10.3390/s20020563.

[127] L. Yao, T. Liu, J. Qin, N. Lu, and C. Zhou, "Tree counting with high spatial-resolution satellite imagery based on deep neural networks," *Ecol. Indic.*, vol. 125, Jun. 2021, Art. no. 107591, doi: 10.1016/j.ecolind.2021.107591.

[128] W. Li, R. Dong, H. Fu, and L. Yu, "Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks," *Remote Sens.*, vol. 11, no. 1, 2019, Art. no. 11, doi: 10.3390/rs11010011.

[129] J. Zheng, W. Li, M. Xia, R. Dong, H. Fu, and S. Yuan, "Large-scale oil palm tree detection from high-resolution remote sensing images using faster-RCNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE Press, 2019, pp. 1422–1425, doi: 10.1109/IGARSS.2019.8898360.

[130] W. Wu, J. Zheng, H. Fu, W. Li, and L. Yu, "Cross-regional oil palm tree detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, 2020, pp. 248–257, doi: 10.1109/CVPRW50498.2020.00036.

[131] J. Zheng, W. Wu, L. Yu, and H. Fu, "Coconut trees detection on the Tenarunga using high-resolution satellite images and deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE Press, 2021, pp. 6512–6515, doi: 10.1109/IGARSS47720.2021.9555008.

[132] M. Xia, W. Li, H. Fu, L. Yu, R. Dong, and J. Zheng, "Fast and robust detection of oil palm trees using high-resolution remote sensing images," in *Proc. Autom. Target Recognit. XXIX*, vol. 10988, 2019, pp. 65–73, doi: 10.1117/12.2518352.

[133] X. Wang, Q. Zhao, P. Jiang, Y. Zheng, L. Yuan, and P. Yuan, "LDS-YOLO: A lightweight small object detection method for dead trees from shelter forest," *Comput. Electron. Agric.*, vol. 198, Jul. 2022, Art. no. 107035, doi: 10.1016/j.compag.2022.107035.

[134] 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 J. Photogramm. Remote Sens.*, vol. 179, pp. 35–49, Sep. 2021, doi: 10.1016/j.isprsjprs.2021.07.001.

[135] R. W. Albuquerque et al., "Mapping key indicators of forest restoration in the amazon using a low-cost drone and artificial intelligence," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 830, doi: 10.3390/rs14040830.

[136] C. Zhang et al., "Multi-species individual tree segmentation and identification based on improved mask R-CNN and UAV imagery in mixed forests," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 874, doi: 10.3390/rs14040874.

[137] Z. Hao et al., "Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask R-CNN)," *ISPRS J. Photogramm. Remote Sens.*, vol. 178, pp. 112–123, Aug. 2021, doi: 10.1016/j.isprsjprs.2021.06.003.

[138] 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 J. Photogramm. Remote Sens.*, vol. 189, no. 1, pp. 220–235, 2022, doi: 10.1016/j.isprsjprs.2022.05.002.

[139] J. Troles, U. Schmid, W. Fan, and J. Tian, "BAMFORESTS: Bamberg benchmark forest dataset of individual tree crowns in very-high-resolution UAV images," *Remote Sens.*, vol. 16, no. 11, 2024, Art. no. 1935, doi: 10.3390/rs16111935.

[140] S. Ahlswede et al., "*TreeSatAi benchmark archive:* A multi-sensor, multi-label dataset for tree species classification in remote sensing," *Earth Syst. Sci. Data Discuss.*, vol. 15, no. 2, pp. 681–695, 2022, doi: 10.5194/essd-15-681-2023.

[141] B. G. Weinstein et al., "A remote sensing derived data set of 100 million individual tree crowns for the national ecological observatory network," *Elife*, vol. 10, Feb. 2021, Art. no. e62922, doi: 10.7554/eLife.62922.

[142] G. T. Miyoshi et al., "A novel deep learning method to identify single tree species in UAV-based hyperspectral images," *Remote Sens.*, vol. 12, no. 8, 2020, Art. no. 1294, doi: 10.3390/rs12081294.

[143] R. Qiao, A. Ghodsi, H. Wu, Y. Chang, and C. Wang, "Simple weakly supervised deep learning pipeline for detecting individual red-attacked trees in VHR remote sensing images," *Remote Sens. Lett.*, vol. 11, no. 7, pp. 650–658, 2020, doi: 10.1080/2150704X.2020.1752410.

[144] C. Zhang, P. M. Atkinson, C. George, Z. Wen, M. Diazgranados, and F. Gerard, "Identifying and mapping individual plants in a highly diverse high-elevation ecosystem using UAV imagery and deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 280–291, Nov. 2020, doi: 10.1016/j.isprsjprs.2020.09.025.

[145] M. Miraki, H. Sohrabi, P. Fatehi, and M. Kneubuehler, "Individual tree crown delineation from high-resolution UAV images in broadleaf forest," *Ecol. Inf.*, vol. 61, Mar. 2021, Art. no. 101207, doi: 10.1016/j.ecoinf.2020.101207.

[146] A. H. Özcan, D. Hisar, Y. Sayar, and C. Ünsalan, "Tree crown detection and delineation in satellite images using probabilistic voting," *Remote Sens. Lett.*, vol. 8, no. 8, pp. 761–770, 2017, doi: 10.1080/2150704X.2017.1322733.

[147] M. Larsen, M. Eriksson, X. Descombes, G. Perrin, T. Brandtberg, and F. A. Gougeon, "Comparison of six individual tree crown detection algorithms evaluated under varying forest conditions," *Int. J. Remote Sens.*, vol. 32, no. 20, pp. 5827–5852, 2011, doi: 10.1080/01431161.2010.507790.

[148] B. G. Weinstein, S. Marconi, M. Aubry-Kientz, G. Vincent, H. Senyondo, and E. P. White, "DeepForest: A python package for RGB deep learning tree crown delineation," *Methods Ecol. Evol.*, vol. 11, no. 12, pp. 1743–1751, 2020, doi: 10.1111/2041-210X.13472.

[149] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010, doi: 10.1007/s11263-009-0275-4.

[150] F. H. Wagner et al., "Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 362–377, Nov. 2018, doi: 10.1016/j.isprsjprs.2018.09.013.

[151] J. R. G. Braga et al., "Tree crown delineation algorithm based on a convolutional neural network," *Remote Sens.*, vol. 12, no. 8, 2020, Art. no. 1288, doi: 10.3390/rs12081288.

[152] M. Dalponte, L. Frizzera, and D. Gianelle, "Individual tree crown delineation and tree species classification with hyperspectral and LiDAR data," *PeerJ*, vol. 6, Jan. 2019, Art. no. e6227, doi: 10.7717/peerj.6227.

[153] B. Wu, B. Yu, Q. Wu, Y. Huang, Z. Chen, and J. Wu, "Individual tree crown delineation using localized contour tree method and airborne LiDAR data in coniferous forests," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 52, pp. 82–94, Oct. 2016, doi: 10.1016/j.jag.2016.06.003.

[154] K. Calders et al., "Terrestrial laser scanning in forest ecology: Expanding the horizon," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112102, doi: 10.1016/j.rse.2020.112102.

[155] T. Yrttimaa et al., "Performance of terrestrial laser scanning to characterize managed Scots pine (*Pinus sylvestris L.*) stands is dependent on forest structural variation," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 277–287, Oct. 2020, doi: 10.1016/j.isprsjprs.2020.08.017.

[156] R. Dubayah et al., "The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography," *Sci. Remote Sens.*, vol. 1, Jun. 2020, Art. no. 100002, doi: 10.1016/j.srs.2020.100002.

[157] N. P. Hanan and J. Y. Anchang, "Satellites could soon map every tree on earth," *Nature*, vol. 587, no. 7832, pp. 42–43, 2020, doi: 10.1038/d41586-020-02830-3.

[158] Y. Quan, M. Li, Y. Hao, J. Liu, and B. Wang, "Tree species classification in a typical natural secondary forest using UAV-borne LiDAR and hyperspectral data," *GISci. Remote Sens.*, vol. 60, no. 1, 2023, Art. no. 2171706, doi: 10.1080/15481603.2023.2171706.

[159] L. Liu, N. C. Coops, N. W. Aven, and Y. Pang, "Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data," *Remote Sens. Environ.*, vol. 200, pp. 170–182, Oct. 2017, doi: 10.1016/j.rse.2017.08.010.

[160] M. A. Contreras and W. Chung, "Developing a computerized approach for optimizing individual tree removal to efficiently reduce crown fire potential," *For. Ecol. Manage.*, vol. 289, pp. 219–233, Feb. 2013, doi: 10.1016/j.foreco.2012.09.038.

[161] C. Hakkenberg, R. K. Peet, D. L. Urban, and C. Song, "Modeling plant composition as community continua in a forest landscape with LiDAR and hyperspectral remote sensing," *Ecol. Appl.*, vol. 28, no. 1, pp. 177–190, 2018, doi: 10.1002/eap.1638.

[162] K. Tochigi, T. Masaki, A. Nakajima, K. Yamazaki, A. Inagaki, and S. Koike, "Detection of arboreal feeding signs by Asiatic black bears: Effects of hard mast production at individual tree and regional scales," *J. Zool.*, vol. 305, no. 4, pp. 223–231, 2018, doi: 10.1111/jzo.12564.

[163] T. D. Stokely et al., "Experimental evaluation of herbicide use on biodiversity, ecosystem services and timber production trade-offs in forest plantations," *J. Appl. Ecol.*, vol. 59, no. 1, pp. 52–66, 2022, doi: 10.1111/1365-2664.13936.

[164] M. Alonzo et al., "Urban tree species mapping using hyperspectral and lidar data fusion," *Remote Sens. Environ.*, vol. 148, pp. 70–83, May 2014, doi: 10.1016/j.rse.2014.03.018.

[165] S. Nezami, E. Khoramshahi, O. Nevalainen, I. Pölönen, and E. Honkavaara, "Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1070, doi: 10.3390/rs12071070.

[166] J. Mäyrä et al., "Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks," *Remote Sens. Environ.*, vol. 256, Apr. 2021, Art. no. 112322, doi: 10.1016/j.rse.2021.112322.

[167] R. Rodriguez et al., "Comparing interpretation of high-resolution aerial imagery by humans and artificial intelligence to detect an invasive tree species," *Remote Sens.*, vol. 13, no. 17, 2021, Art. no. 3503, doi: 10.3390/rs13173503z.

[168] K. Einzmann et al., "Early detection of spruce vitality loss with hyperspectral data: Results of an experimental study in Bavaria, Germany," *Remote Sens. Environ.*, vol. 266, Dec. 2021, Art. no. 112676, doi: 10.1016/j.rse.2021.112676.

[169] M. M. Alam, M. N. Strandgard, M. W. Brown, and J. C. Fox, "Improving the productivity of mechanised harvesting systems using remote sensing," *Australian Forestry*, vol. 75, no. 4, pp. 238–245, 2012, doi: 10.1080/00049158.2012.10676408.

[170] D. Yin and L. Wang, "Individual mangrove tree measurement using UAV-based LiDAR data: Possibilities and challenges," *Remote Sens. Environ.*, vol. 223, pp. 34–49, Mar. 2019, doi: 10.1016/j.rse.2018.12.034.

[171] G. Goldbergs, S. R. Levick, M. Lawes, and A. Edwards, "Hierarchical integration of individual tree and area-based approaches for savanna biomass uncertainty estimation from airborne LiDAR," *Remote Sens. Environ.*, vol. 205, pp. 141–150, Feb. 2018, doi: 10.1016/j.rse.2017.11.010.

[172] Y. K. Karna et al., "Integration of WorldView-2 and airborne LiDAR data for tree species level carbon stock mapping in Kayar Khola watershed, Nepal," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 38, pp. 280–291, Jun. 2015, doi: 10.1016/j.jag.2015.01.011.

[173] C. Song, "Estimating tree crown size with spatial information of high resolution optical remotely sensed imagery," *Int. J. Remote Sens.*, vol. 28, no. 15, pp. 3305–3322, 2007, doi: 10.1080/01431160600993413.

[174] S. B. Turner, D. P. Turner, A. N. Gray, and W. Fellers, "An approach to estimating forest biomass change over a coniferous forest landscape based on tree-level analysis from repeated lidar surveys," *Int. J. Remote Sens.*, vol. 40, no. 7, pp. 2558–2575, 2019, doi: 10.1080/01431161.2018.1528401.

[175] Q. Ma, Y. Su, S. Tao, and Q. Guo, "Quantifying individual tree growth and tree competition using bi-temporal airborne laser scanning data: A case study in the Sierra Nevada Mountains, California," *Int. J. Digital Earth*, vol. 11, no. 5, pp. 485–503, 2018, doi: 10.1080/17538947.2017.1336578.

[176] M. Vastaranta, I. Korpela, A. Uotila, A. Hovi, and M. Holopainen, "Mapping of snow-damaged trees based on bitemporal airborne LiDAR data," *Eur. J. For. Res.*, vol. 131, no. 4, pp. 1217–1228, 2012, doi: 10.1007/s10342-011-0593-2.

[177] M. Papeş, R. Tupayachi, P. Martínez, A. T. Peterson, G. P. Asner, and G. V. N. Powell, "Seasonal variation in spectral signatures of five genera of rainforest trees," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 339–350, Apr. 2013, doi: 10.1109/JSTARS.2012.2228468.

[178] S. Chen, D. Liang, B. Ying, W. Zhu, G. Zhou, and Y. Wang, "Assessment of an improved individual tree detection method based on local-maximum algorithm from unmanned aerial vehicle RGB imagery in overlapping canopy mountain forests," *Int. J. Remote Sens.*, vol. 42, no. 1, pp. 106–125, 2021, doi: 10.1080/01431161.2020.1809024.

[179] J. Aval et al., "Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 197–210, Dec. 2018, doi: 10.1016/j.isprsjprs.2018.09.016.

[180] E. Alvarez-Vanhard, T. Houet, C. Mony, L. Lecoq, and T. Corpetti, "Can UAVs fill the gap between in situ surveys and satellites for habitat mapping?" *Remote Sens. Environ.*, vol. 243, Jun. 2020, Art. no. 111780, doi: 10.1016/j.rse.2020.111780.

[181] Z. Cheng, L. Qi, and Y. Cheng, "Cherry tree crown extraction from natural orchard images with complex backgrounds," *Agriculture*, vol. 11, no. 5, 2021, Art. no. 431, doi: 10.3390/agriculture11050431.

[182] L. Zhang et al., "DeepLight: Reconstructing high-resolution observations of nighttime light with multi-modal remote sensing data," 2024, *arXiv:2402.15659*.

[183] C. Magnard, F. Morsdorf, D. Small, U. Stilla, M. E. Schaepman, and E. Meier, "Single tree identification using airborne multibaseline SAR interferometry data," *Remote Sens. Environ.*, vol. 186, pp. 567–580, Dec. 2016, doi: 10.1016/j.rse.2016.09.018.

[184] M. Yazdani, S. S. Jouibary, J. Mohammadi, and Y. Maghsoudi, "Comparison of different machine learning and regression methods for estimation and mapping of forest stand attributes using ALOS/PALSAR data in complex Hyrcanian forests," *J. Appl. Remote Sens.*, vol. 14, no. 2, 2020, Art. no. 024509, doi: 10.1117/1.JRS.14.024509.

[185] J. F. Slik et al., "An estimate of the number of tropical tree species," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 112, no. 24, pp. 7472–7477, 2015, doi: 10.1073/pnas.1423147112.

[186] J. Zheng et al., "Unsupervised mixed multi-target domain adaptation for remote sensing images classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE Press, 2020, pp. 1381–1384, doi: 10.1109/IGARSS39084.2020.9323602.

[187] J. Zheng et al., "Open-set domain adaptation for scene classification using multi-adversarial learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 208, pp. 245–260, Feb. 2024, doi: 10.1016/j.isprsjprs.2024.01.015.

[188] Q. Li, Y. Wen, J. Zheng, Y. Zhang, and H. Fu, "HyUniDA: Breaking label set constraints for universal domain adaptation in cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–15, 2024, doi: 10.1109/TGRS.2024.3400959.

[189] J. Zheng, Y. Zhao, W. Wu, M. Chen, W. Li, and H. Fu, "Partial domain adaptation for scene classification from remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–17, 2022, doi: 10.1109/TGRS.2022.3229039.

[190] J. Zheng et al., "A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (S^2M^2T da) scenarios," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021, doi: 10.1109/TGRS.2021.3105302.

[191] H. Jiang, S. Chen, D. Li, C. Wang, and J. Yang, "Papaya tree detection with UAV images using a GPU-accelerated scale-space filtering method," *Remote Sens.*, vol. 9, no. 7, 2017, Art. no. 721, doi:10.3390/rs9070721.

[192] J. Zheng et al., "A parallel approach for oil palm tree detection on a SW26010 many-core processor," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Piscataway, NJ, USA: IEEE Press, 2022, pp. 1548–1551, doi: 10.1109/IGARSS46834.2022.9884785.

[193] J. Bai et al., "Qwen-VL: A frontier large vision-language model with versatile abilities," 2023, *arXiv:2308.12966*.

[194] P. Zhang et al., "Multi-scale vision longformer: A new vision transformer for high-resolution image encoding," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Montreal, QC, Canada, 2021, pp. 2998–3008, doi: 10.1109/ICCV48922.2021.00299.

GRS