Remote Sensing of Environment 323 (2025) 114654

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Influence of data and methods on high-resolution imagery-based tree species recognition considering phenology: The case of temperate forests

Xinlian Liang [a,*], Jianchang Chen [a], Weishu Gong [b], Eetu Puttonen [c], Yunsheng Wang [c]

[a] The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 430070 Wuhan, China
[b] Department of Geographical Sciences, University of Maryland, 20742 College Park, MD, United States of America
[c] Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute (National Land Survey of Finland), Vuorimiehentie 5, 02150 Espoo, Finland

ARTICLE INFO

Editor: Jing M. Chen

Keywords:
Phenology
Multi-temporal
Tree species
Recognition
Deep learning
Forest
Remote sensing
Close-range sensing

ABSTRACT

Seasonal phenological transformations alter tree appearances, notably by influencing the size and color of the foliage. It has long been anticipated that such phenology induced characteristics can help address the tree-species recognition problem, a fundamental challenge in forest science. Yet, studies on tree-species recognition using remote sensing and phenological characteristics have been rare, due to the very limited availability of high spatiotemporal resolution observations. Moreover, the interactions between the effectiveness of phenological characteristics, remote sensing data, and the analytical methodologies have not yet been sufficiently explored. The understanding of how to integrate multi-temporal observations and phenological characteristics in tree-species recognition has been lacking. This study aims to identify principles for optimizing species recognition by combining data, methods, and phenological dynamics. This involves understanding the impact factors of various methodologies, and how they interact with phenological characteristics and datasets at different times and/or frequencies. The study was carried out using multi-temporal high-resolution optical images of a temperate forest, which were collected in 2021 during leaf growth and senescence periods between May and October, i.e., three leaf growth (May–August) and three leaf senescence (September–October) periods. The test site comprised 14 different tree classes, including 11 species, 2 genera, and 1 dead tree class. The experimental results showed that, for deep learning approaches, the current main limitations in the tree species recognition lie in sample imbalance as the targeted species number increases. With the state-of-the-art data and methods, distinguishing between species within a same genus is much more challenging than differentiating between species from different genera or families. It is also revealed that the best timing for tree species classification is early autumn (September) or late spring (May) when a single-temporal (one-timepoint) data is applied; all-temporal (six-timepoint) data improves the recognition results in comparison with single-temporal observations; however, the improvements from adding additional timepoints became marginal after two timepoint are used with one from late spring and other from early autumn. Furthermore, prior knowledge of individual crown boundaries, typically obtained through individual tree crown delineation, is essential for efficiently incorporating phenological variations into species recognition.

1. Introduction

Tree species information is indispensable in understanding forest conditions and functions, as an independent variable (Gamfeldt et al., 2013; Wessely et al., 2024) or as an input of species-dependent models (Vorster et al., 2020; Wang et al., 2019b). However, to collect tree-level species information over large areas is a long-lasting challenge. Up to now, conventional field inventory via in situ manual recognition remain as the most trusted sources of tree-species information. Forest species composition is regularly estimated in national forest inventories (Tomppo et al., 2010). However, considering the vast area and restricted accessibility of forests, the spatial and temporal coverage of such manual collection is obviously limited. The sample-based inventories are conducted only every several years, which do not give spatially explicit wall-to-wall and temporally up-to-date tree information and far beyond enough to meet the needs of regular monitoring of forest ecosystems over large areas. The recent advances in close-range sensing improve the automation level in the forest inventories (Liang et al., 2015, 2024b;

* Corresponding author.
E-mail address: xinlian.liang@whu.edu.cn (X. Liang).

https://doi.org/10.1016/j.rse.2025.114654
Received 7 December 2024; Received in revised form 21 January 2025; Accepted 11 February 2025
Available online 2 April 2025

Hyyppä et al., 2020b; Hyyppä et al., 2020a; Wang et al., 2021; Mokroš et al., 2021), yet to meet the requirement of continuous coverage over large areas is still challenging. Consequently, the lack of reliable tree species information remains a significant and persistent challenge, e.g., for biodiversity conservation (Boonman et al., 2024; van Tiel et al., 2024) and forest management (Kahl and Bauhus, 2014; Oettel and Lapin, 2021).

Facilitating the acquisition of tree species information in forests has been a central goal of remote sensing (RS). However, most existing studies have concentrated on patch-level assessments, while tree-level investigations have been relatively limited, e.g., using high-resolution satellite imagery such as the WorldView (Jiang et al., 2021; Qiu et al., 2019). Bridging the gap between RS methodologies and tree-level species inventories across extensive geographic areas remains one of the most significant challenges in the field (Fassnacht et al., 2016).

Over the past decade, rapid advancements in close-range sensing technologies have significantly enhanced the effectiveness and efficiency of forest investigations (Liang et al., 2022). These developments have introduced convenient, flexible, and even fully automated solutions for tree-level studies at plot and area scales (Wang et al., 2019a; Jurjević et al., 2020; Balenović et al., 2021), e.g., utilizing terrestrial and low-altitude aerial imagery as well as laser scanning systems and at a high level of details (Pyörälä et al., 2018a, 2018b; Liang et al., 2019, 2024a). Driven by these advancements, interests in acquiring tree-level species information using close-range sensing are rapidly growing (Chen et al., 2024). In particular, close-range sensing enables high-frequency temporal observations, opening new opportunities for in-situ phenology studies and allowing the linkage of phenological characteristics with species recognition (Shcherbacheva et al., 2024).

Phenology varies across species, resulting in distinct color and texture dynamics in multi-temporal images for each species. It is known as a useful trait for RS-based species classification, e.g., if the time of RS data collection can be aligned with the phenological cycles of the targeted species (Gaertner et al., 2016). However, research on the application of multi-temporal RS in tree species recognition have been limited, as the availability of high-resolution time-series RS data has only increased recently. Specifically, close-range sensing using unmanned aerial vehicle (UAV) platforms provide a handy solution to acquire high-resolution imagery data with detailed tree-canopy characteristics. Time-series UAV observations enable frequent data collections at interannual, seasonal, and even daily paces, provide rich spectral, structural, and temporal information, and facilitate a comprehensive documentation of the growth cycle, phenological changes, and health status. The key question now is how to integrate the rich dynamics captured from UAV image time series into species-specific traits and to enhance the accuracy of tree species recognition.

One direction to explore the benefits of multi-temporal observations is to study the influence of phenology on species classification. The best data-acquisition timing for an application can be identified by comparing the performances of data collected at different timepoints. For example, data at seven timepoints were collected over a temperate forest in Quebec, Canada, i.e., one timepoint per month from May to August, and three in September and October (Cloutier et al., 2024). The study suggested that the early autumn (e.g., early September) images gave the highest overall accuracy (OA), instead of the peak autumn images (e.g., during October when the fall foliage is at its most vibrant and colorful period) as suggested by previous studies (Hill et al., 2010; Key et al., 2001). The study also revealed that variations in phenological characteristics influence the effectiveness of tree-species recognition, making it related with the timing of data acquisition. In (Miyoshi et al., 2020), hyperspectral (HS) images were acquired in June and July over three years in an Atlantic Forest located in São Paulo, Brazil. The study utilized three-year data to classify eight tree species, achieving a 50 % OA using statistical machine learning (ML), e.g., Random Forest (RF). According to the study, due to the lack of phenological variations during June and July, the contribution of multiple timepoints was considerably less compared to that of multiple spectral channels.

An alternative approach to leveraging the advantages of multi-temporal observations is to treat the phenological dynamics across multiple time points as a unified trait for tree-species classification. In (Grybas and Congalton, 2021), UAV RGB images were collected at five timepoints in a sample plot in New Hampshire, USA. The effectiveness of tree species recognition using multi-temporal images was assessed across single, two, three, four, and five timepoints. Thirteen tree species were classified using RF, with the highest OA of 61.1 % achieved using all five timepoints. However, the improvement in OA became marginal when more than three timepoints were incorporated. Images from mid to late spring yielded the highest classification accuracy, attributed to the significant spectral heterogeneity among different species during this period.

In (Shcherbacheva et al., 2024), the reflectance patterns of trees were demonstrated to be able to capture annual phenological characteristics of three main boreal species (Norway spruce, Scots pine, and silver Birch), at a single wavelength (1550 nm) from LiDAR (Light detection and ranging) data. When all key phenological events were captured, e.g., using bi-weekly observations from April to November, the OA of species classification can reach close to 100 % with a few exceptions raised by erroneous individual tree segmentation. In (Veras et al., 2022), deep learning (DL) method and multi-temporal UAV RGB images were used to recognize tree species in the Brazilian Amazon rainforest. The experimental results showed that the highest classification accuracy was 90.5 % for eight tree species using images from four timepoints. The improvements brought by multi-temporal data were notable when compared to the single timepoint, where the accuracies in the rainy season were 83.5 % in November and 81.9 % in February, and in the dry season were 69.3 % in May and 78.8 % in August.

Previous studies have suggested that multi-temporal data, or phenological information, enhances the ability to recognize tree species. However, the effects of phenology on tree crown appearance are complex, and their influence on species recognition may not always align with assumptions based solely on phenological patterns. For example, Cloutier et al. (2024) reported that among seven time points from different seasons, the peak autumn data presented the worst accuracy, contradicting the initial hypothesis. The same study suggested that the phenological dynamics influence the performance of species classification not only due to the changes on leaf coloring, but also due to the changes in the contrast between the foreground, i.e., tree crowns, and the background, e.g., forest floor, under growth, and other objects, which are attributed to variations in leaf sizes and densities, and the accuracy of individual tree crown (ITC) delineation.

The key question to address is whether the phenological trends of different species can be captured with a limited number of timepoints, how the temporal context between images enhances species recognition, and how to effectively collect and process multi-temporal data to maximize the benefits of phenological dynamics in overcoming the challenging task of species recognition.

However, previous studies have often blended the effects of phenological characteristics with those of data processing methods. As a result, it becomes difficult to pinpoint the primary factor driving performance changes, i.e., whether it is the data processing techniques or the phenological characteristics themselves. In particular, the role of data analysis methods in influencing the final outcomes has been insufficiently addressed.

This study aims to clarify the impact of phenological dynamics and the role of processing methodologies in the species recognition using multi-temporal images. More specifically, the study investigates 1) the impact of the processing methodology to the performances of tree-species classification; 2) the effectiveness of the data acquired at individual and multiple timepoints; and 3) any potential interactions between processing methodologies and phenological characteristics.

The study focuses on DL approaches, considering the overall profound reported performances of DL approaches in comparison with

classic ML approaches (Chen et al., 2024). Four network architectures are benchmarked to classify 14 tree classes using UAV-RGB images from six timepoints, i.e., the current most comprehensive multi-temporal datasets. Altogether, 72 tree-species classification experiments were carried out in two groups. The first category, which includes 48 experiments, uses data from a single timepoint to demonstrate the influence of methodology and species-specific characteristics related to preferred timing. The second category, consisting of 24 experiments, presents outcomes from multiple timepoints to highlight the effectiveness of phenological characteristics and explore the interactions between methodology and phenology.

2. Materials and methods

This section presents the dataset, and experimental designs used in the study.

2.1. Dataset

The dataset used in this study was published in (Cloutier et al., 2024), which is so far the most comprehensive dataset to support studies on species classification using high resolution multi-temporal imagery. The images were collected from a temperate mixed forest in the Montressor University Research Station in St. Hypolite (Quebec, Canada) using a DJI Phantom 4 RTK aircraft (DJI Science and Technology Co. Ltd., Shenzhen, China) equipped with an RGB camera. The flight was 60 m above the canopy, and the image was with a 0.02 m spatial resolution.

The published data included six timepoints from the growing season in 2021, i.e., on May 28 in spring, Jul. 21 and Aug. 18 in summer, and Sep. 2, Sep. 28, and Oct. 7 in autumn. The original image was cropped into images of size 256 × 256 pixels. Manual annotations of individual tree crowns were carried out by combining the field identified stem locations and tree species information with manual crown delineation on acquired UAV images. In total, 22,139 annotated individual tree crowns in 14 classes were available as vector polygons, i.e., 11 species, 2 genera, and 1 dead tree class. Table 1 summarizes the class allocation of the ITC annotations in the published dataset. For more information about the test site and the data, readers are referred to (Cloutier et al., 2024).

2.2. Methods

A clear understanding of the role of methodologies and phenological dynamics in species classification can only be achieved when the impacts of other factors in the processing pipeline are thoroughly studied. First, the studies focused on single timepoints to clarify the methodology and phenology influences. Secondly, different timepoints were combined to explore the interaction between the species-specific phenology and processing methodology, as well as the phenological impacts on the tree-species recognition.

2.2.1. Experiment design

The study comprised a total of 72 experiments that were designed to comprehensively examine the factors affecting the performance of image-based species recognition, considering the influences of methodology, preprocessing, data collection timing, and the integration of multiple timepoints.

From the methodology aspect, this study focused on DL approaches over statistical ML methods due to the overall superior performance of DL in tree species analysis (Chen et al., 2024). Two types of classic DL methodologies were investigated, i.e., the object-based instance classification and the pixel-based semantic segmentation. For each methodology, two network architectures were employed to assess the performance, i.e., ResNet (He et al., 2016) and Swin Transformer (Liu et al., 2021) for the object-based instance classification approach and UNet (Ronneberger et al., 2015) and DeepLab V3+ (Chen et al., 2018) for the pixel-based semantic segmentation approach. These architectures were selected because they are the most commonly applied state-of-the-art approaches for image-based species classification (Chen et al., 2024).

Another key factor was preprocessing, which involved excluding background information through the ITC delineation. The background information in the images significantly influence the performance of individual-tree-level species classification, e.g., forest floor, undergrowth vegetation, entangled crown parts from neighboring trees, and other non-crown objects (Cloutier et al., 2024). Reliable ITC delineation was essential for employing object-based methods, and it also served to reduce background information in subsequent analyses as a side benefit. However, ITC delineation is a challenging task, specifically in complex forest stands where trees from multiple canopy layers have overlapped and/or entangled crowns. Considering the difficulty of automated ITC delineation from images, the experiments explored the significance of ITC delineation for species recognition, particularly in terms of background removal, by utilizing exist ITC masks. Therefore, each DL model is implemented in two scenarios: with and without background information.

From the phenology perspective, it is worth noting that the phenological dynamics are species specific, and variations in color and texture over time of a tree crown can be regarded as an integrated trait representing its species (Shcherbacheva et al., 2024). Therefore, when only a single data collection is possible, it is crucial to determine the optimal timing for data collection to maximize the contribution of phenological

Table 1
The statistics of the tree species in the experiment.

No.	Name	Abbreviation	Conifers or deciduous	Sample size			
				Tree		Pixel	
				Number	%	Million	%
1	*Abies balsamea*	ABBA	Conifers	2878	13.00	26.92	4.82
2	*Acer pensylvanicum*	ACPE	Deciduous	751	3.39	9.99	1.79
3	*Acer rubrum*	ACRU	Deciduous	5829	26.30	129.75	23.20
4	*Acer saccharum*	ACSA	Deciduous	1004	4.53	41.60	7.45
5	*Betula alleghaniensis*	BEAL	Deciduous	282	1.27	15.28	2.74
6	*Betula papyrifera*	BEPA	Deciduous	5861	26.50	176.32	31.60
7	*Fagus grandifolia*	FAGR	Deciduous	220	0.99	7.59	1.36
8	*Larix laricina*	LALA	Conifers	185	0.84	3.03	0.54
9	Picea spp.	Picea	Conifers	1020	4.61	11.26	2.02
10	*Pinus strobus*	PIST	Conifers	564	2.55	35.02	6.27
11	Populus spp.	Populus	Deciduous	1107	5.00	78.54	14.10
12	*Thuja occidentalis*	THOC	Conifers	1508	6.81	15.28	2.74
13	*Tsuga canadensis*	TSCA	Conifers	59	0.27	2.00	0.36
14	Dead tree	Mort	\	871	3.93	5.95	1.06

characteristics of tree crowns to species recognition. On the other hand, when multi-temporal data collection is feasible, it is important to learn how many timepoints are needed and how they should be distributed across seasons to fully capture the phenological characteristics of different species.

To explore the interactions between methodology, preprocessing, and phenology, the 72 experiments were divided into two main groups: 48 experiments (4*2*6) focusing on single - timepoint analyses, which consisted of the combinations of four architectures (ResNet, Swin Transformer, UNet, DeepLab V3+), two scenarios (include and exclude background), and six single timepoints (May 28, Jul. 21, Aug. 18, Sep. 2, Sep. 28, and Oct. 7); and 24 (4*2*3) experiments focusing on multi-timepoint analyses, which consisted of combinations of the four methods, two scenarios, and three multi-temporal combinations (two, three, and six timepoints).

2.2.2. Methodologies

Object-based instance classification and pixel-based semantic segmentation comprised two major methodology categories in DL image processing. The object-based methodology tackles images that only contain one targeted object instance, and the purpose is to classify the images based on the object, e.g., in the case of classic cat and dog classification. The pixel-based methodology is commonly applied to images that contains various objects, and the target is to segment the image according to the types or classes of the objects.

In the context of tree species classification, both methodologies are applicable. This study tested two most applied architectures for each methodology. In training, the ratio of training and validation datasets was 4:1. To expand the training dataset, the image samples of the training dataset were rotated for 90°, 180°, and 270° following a standard data expansion process that was commonly applied in DL approaches.

ResNet

ResNet is currently the most widely used object-based NN architecture in tree species recognition. ResNet introduced the residual module to solve the performance degradation problem of NNs during training, where the training error becomes larger as the depth of the network increases (He et al., 2016). This improvement allows the network to be deeper and improve its performance. In this study, the ResNet18 model is employed. Its main architecture consists of an input convolutional and pooling layer, four residual block groups where each group consists of multiple residual blocks, a global average pooling layer, and a fully connected layer.

Swin transformer

Swin Transformer is one of the most advanced object-based image instance classification architectures that represent the state-of-the-art that is based on Vision Transformer. The standard Vision Transformer (Dosovitskiy et al., 2021) computed self-attention directly on the entire image, which significantly increase the computation and memory cost, especially with increasing image size. Swin Transformer strikes a balance between computational cost and model performance by introducing a layered structure and a local window attention mechanism. Its core idea is to reduce the size of the feature map through a hierarchical architecture (Liu et al., 2021).

UNet

UNet (Ronneberger et al., 2015) is a classic NN architecture that has been often used in the pixel-based tree species classification. UNet is based on convolutional neural network (CNN) for image semantic segmentation. It is composed of down-sampling encoder and up-sampled decoder. The down-sampling encoder extracts the abstraction features of the input image using a series of convolutional and pooling layers and reduces the spatial size of the feature map. The up-sampled decoder constructs tree species semantic information through skip connections that fuse the feature maps of the corresponding layers in the encoder with the feature maps in the decoder.

DeepLab V3+

DeepLab V3+ is an image segmentation architecture based on DeepLab V3 (Chen et al., 2017). It is another commonly used architecture to train DL model for pixel-based tree species recognition. DeepLab V3+ combines the multi-scale feature extraction and dilated convolution technique of DeepLab V3 to capture rich contextual information through the atrous spatial pyramid pooling (ASPP) module. The encoder-decoder architecture was introduced in DeepLab V3+. The encoder utilized a deep CNN to extract multi-scale features and extended the receptive field without increasing the computational effort through dilated convolution. The decoder part fused the low-level features from the encoder and the output of ASPP to generate high-resolution segmentation results. The model generally performed well in segmentation tasks (Chen et al., 2018).

2.2.3. Individual tree crown delineation as preprocessing

To enable object-based instance classification approaches for tree species recognition using images, ITC delineation was required. Specifically, the ITC delineations were used to crop tree crowns from the image and to generate a sub-image for each individual tree crown, as shown in Fig. 1.

For pixel-based approaches, the ITC delineation outcomes can be used to label the pixels and to generate reference semantic masks for tree species classes, thus, to exclude the background information surrounding each individual crown, as shown in Fig. 2.

In this study, the impacts of the ITC delineation and the background removal are investigated by comparing the outcomes using different types of input data for object- and pixel-based approaches, where foreground and background information are mixed, as illustrated in Fig. 1 (a) and Fig. 2 (a), respectively, and where the background information are excluded, for object-based and pixel-based approaches, as illustrated in Fig. 1 (b) and Fig. 2 (b), respectively.

2.2.4. Phenological dynamics as an integrated species trait

The contribution of phenology to the species recognition is investigated through two approaches. First, each timepoint is processed separately assuming that data from multiple timepoints are independent. The analyses using single timepoints provide hints for the preferable individual data-acquisition timing.

Secondly, spectral variations are treated as a connected sequence rather than as independent events. The series of consecutive variations within this sequence are considered as the key trait for species classification. The phenological trends of different species are expected to become more apparent by integrating various timepoints into a cohesive flow. Thus, the capability to distinguish species is expected to be strengthened by incorporating the temporal context.

The multi-temporal data were chronologically assembled into a new dataset as the inputs of the DL models. Three combinations are formed using the six valid time points. The first combination consists of two time points: one from spring (May 28) and one from autumn (Sep.2), as these

(a) (b)

Fig. 1. Examples of input images for the object-based methodology. (a) an input image using coarse ITC bounding box, where foreground (the tree crown) and background (forest floor, under growth, or neighboring crown) are mixed. (b) an input image using accurate ITC delineation, where the background elements are excluded.

Fig. 2. Examples of input images for the pixel-based methodology. (a) An input image without applying ITC outcomes, where foreground and background are mixed. (b) An input image with the aid of ITC delineations, i.e., the background pixels are excluded. (c) The reference semantic mask generated with the aid of ITC delineations, each color represents a tree species class, and background is set to white.

are the most favored timepoints for nearly all processing approaches (object- and pixel-based, with and without background) in the single-timepoint analyses in section 3.1. The second combination includes three time points by adding a summer timepoint (Aug. 18) that is relatively effective. This represents the optimal combination of three seasons, i.e., spring, summer, and autumn. The third combination encompasses all six valid timepoints across the three seasons: one in spring, two in summer, and three in autumn.

Fig. 3 illustrates the compilation of data from six timepoints using inter-band link. The same approach is applied for other experimental scenarios where two and three timepoints were assembled.

2.2.5. Evaluation metrics

The performances of the tree species recognition results are evaluated using OA, Average Accuracy (AA), and Recall, where OA is the ratio of the total number of correctly recognized samples to the total number of samples; AA is the mean of correctly recognized samples for each class; and Recall is the ratio between correctly recognized samples of each class and the total number of samples in each class. The relevant equations are listed as following:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$AA = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

where TP represents the number of true positive cases, TN represents the number of true negative cases, FP represents the number of false positive cases, FN represents the number of false negative cases, and N represents the total number of classes.

It should be noted that the evaluation metrics, e.g., OA, AA, and Recall, for species classification from object- and pixel-based approaches need to be interpreted differently, even though the equations used are identical.

For the object- and pixel-based approaches, the unit for evaluation is the individual tree and pixel, respectively. The pixel-based approach does not require ITC delineation. Consequently, the results of pixel-based species classification only indicate the proportions of species occupancy at the area level and cannot be directly applied to deductions at the individual tree level. On the other hand, the outcomes of the object-based approaches are directly connected to individual trees, which can then be easily aggregated to the area level. Because object-based outcomes can be interpreted at the area level to represent species composition that is similar to pixel-based outcomes, the evaluation metrics derived from both methodologies are comparable for demonstrating overall performance.

The differences in evaluating outcomes at the object- versus pixel-level, i.e., using individual trees or pixels as the basic unit for calculating evaluation metrics, are discussed further in Section 4.1.

3. Results

The experimental results from the 48 experiments using single timepoint are provided in Section 3.1. Results of the other 24 experiments using multiple timepoints are reported in Section 3.2. Figures and tables in this section highlighted the most important results.

The complete outcomes of all experiments were provided in detailed tables in an appended document of this publication, which can be find through the link below: https://drive.google.com/file/d/1sF6Wvs7y9poU0I95yhah5rQHIs3lCbDj/view?usp=sharing

3.1. Results based on single timepoint observation

In single timepoint based evaluations, each timepoint is regarded as an independent event. The results reveal the influences of methodology and the species-specific phenological variance.

3.1.1. Performance for all species

The performances of the tree-species recognition are firstly reported as the average values across all species based on the method, the timing of data collection, and the existence of background.

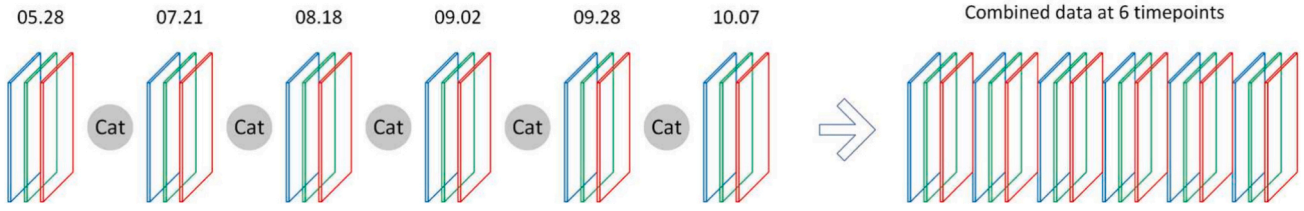


Fig. 3. The compilation of images from multiple timepoints as input for DL models. "Cat" represents the inter-band link. Blue, green, and red colors represent three image channels. Numbers represent the dates of each timepoint when the image data were collected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

General trends in accuracy variations across data collection timepoints

Fig. 4 illustrates the OA and AA of tree species recognition, and the achieved OA are generally higher than AA regardless the methodology and the timepoints.

The variations of OA and AA values among different timepoints are clear regardless the applied methodology. These results confirm that phenological characteristics of tree species affect the effectiveness of species classification due to changes on crown appearances in images.

For object-based approaches, the difference between the highest and lowest AAs for ResNet is 11.55 % with background (66.49 % on May. 28 vs. 54.94 % on Sep. 28) and 10.69 % without background (68.04 % on Sep. 2 vs. 57.35 % on Oct. 7), while the difference between the highest and lowest AAs for Swin Transformer is 10.00 % with background (59.21 % on May. 2 vs. 49.21 % on Jul. 21) and 12.89 % without background (63.12 % on Sep. 2 vs. 50.23 % on Oct. 7).

For pixel-based approaches, the difference between the highest and lowest AAs for UNet is 10.58 % with background (54.2 % on Sep. 2 vs. 43.62 % on Jul. 21) and 14.24 % without background (64.88 % on Sep. 2 vs. 50.64 % on Jul. 21), while the difference for DeepLab V3+ is 9.89 % with background (50.07 % on Sep. 2 vs. 40.5 % on Jul. 21) and 13.27 % without background (56.65 % on Sep. 2 vs. 43.38 % on Sep. 28).

The object-based approaches, i.e., ResNet and Swin Transformer, perform best with the data from May 28 and Sep. 2, with May 28 being preferred when the background is included and Sep. 2 when the background is excluded. The least favorable timepoint for these approaches is Oct. 7, where all lowest AA values came from, followed by Sep. 28, which produced the second-lowest AA values. For the pixel-based approaches, i.e., UNet and DeepLab V3+, Sep. 2 was the most favorable timepoint, regardless of background inclusion. The least favorable timepoint was Jul. 21, followed by Sep. 28 and Oct. 7.

Overall, Sep. 2 is the most favorable timepoint for all processing approaches, providing either the best or the 2nd best outcomes. On the other hand, Sep. 28 is the least favorable timepoint for most of the approaches. These results suggested that the autumn time window, during which phenological characteristics most effectively contribute to species classification, is quite narrow.

Average accuracies across six timepoints

Table 2 reports average OA and AA across all six time points. The object-based approaches outperform the pixel-based approaches in

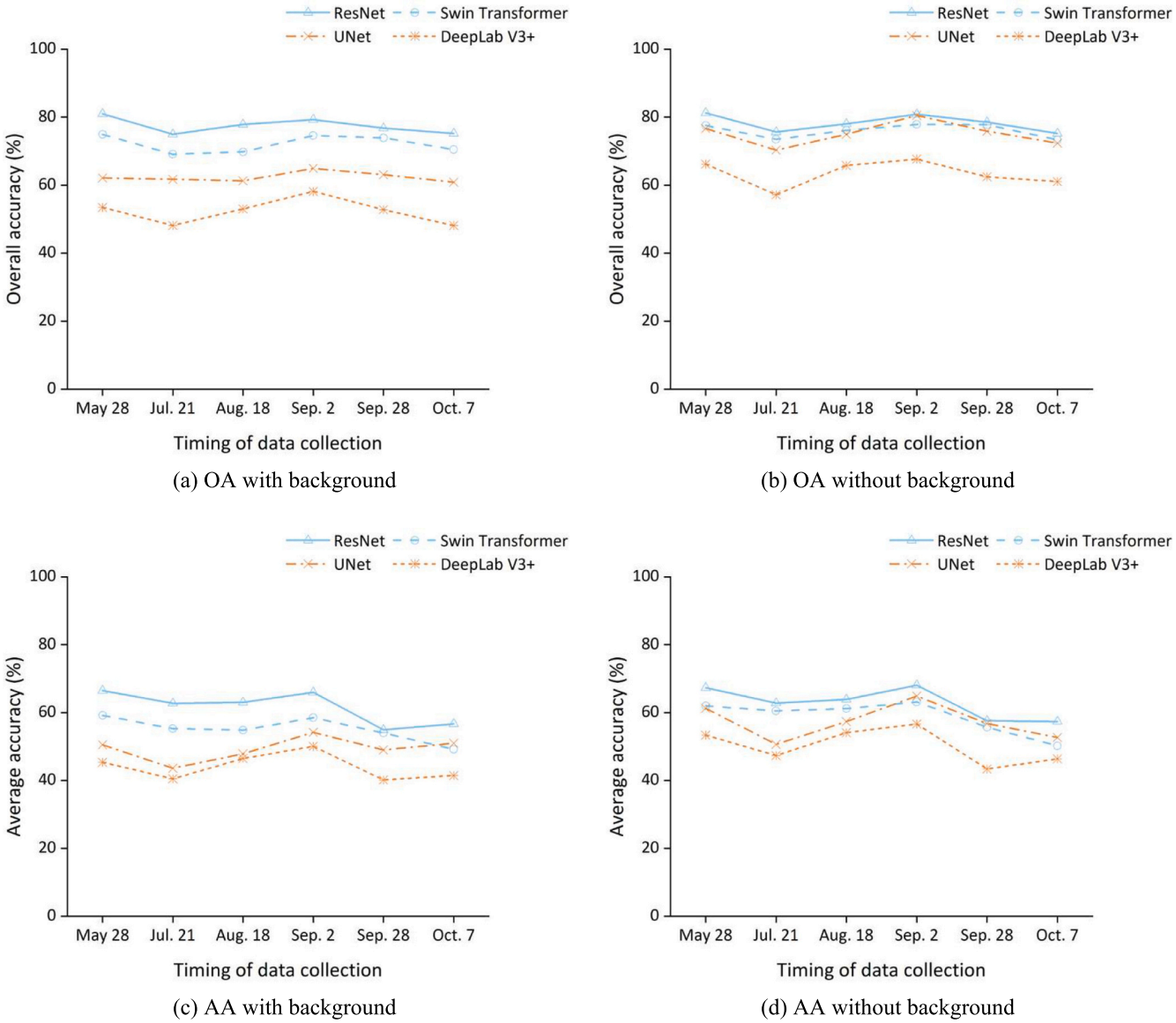


Fig. 4. The overall accuracy (OA) and average accuracy (AA) of tree species recognition using datasets at different timepoints and (a, c) with and (b, d) without background.

terms of the overall and the average accuracies.

For pixel-based approaches, background removal noticeably improved both OA and AA, as shown in Table 2. Specifically, for UNet and DeepLab V3+, the OAs increase by 12.77 % and 11.13 % on average, respectively; and the AAs increase by 7.90 % and 6.19 % on average, respectively. The impact of removing background information is less significant for object-based approaches, i.e., ResNet and Swin Transformer, with less than a 1 % increase in both OA and AA for ResNet, and less than a 3.4 % increase for both metrics for Swin Transformer. These results highlight the contribution and significance of the background removal facilitated by the ITC delineation, and the advantageous of object-based approaches.

3.1.2. Species-specific performance

The performances of the tree-species recognition are reported in this section across individual species with respect to the method and timepoint.

Specie-specific trends in accuracy variations across timepoints

Fig. 5 illustrates the recall values of the species recognition across different methods and timepoints. Results of each species are illustrated in a single cell. In each cell, the results of two input data scenarios are illustrated, i.e., with and without background information in the upper and lower subfigure, respectively.

A strong correlation was observed between sample size and recall value, regardless of species or DL models, i.e., larger sample sizes lead to higher recall values. For object-based approaches, a larger tree-wise sample size, indicated by the first number next to the species name, generally corresponds to higher recall values. Similarly, for pixel-based approaches, high recall values generally correspond to large pixel-scale sample sizes of the species, indicated by the second number next to the species name in the subtitle.

Overall, when a species accounts for more than 10 % of the total sample size across all species, either at the object- or pixel-scale, its recall values typically reach around 80 % from the corresponding object- or pixel-based methods, as shown by the results from the *Abies balsamea* (13 % recall at the object-scale) using ResNet and Swin Transformer and *Populus* spp. (14.1 % recall at the pixel scale) using UNet. When a species represents around 5 % of the total sample size, its recall value can generally be expected to be around 60 % across all DL models. However, for species with less than 2 % of the total sample size, recall values tend to be more variable and can be randomly high or low due to unpredictable reasons, e.g., Fig. 5. (h), (n).

The recall value of a given species is strongly correlated with its sample size during the training process, namely, larger sample sizes result in higher recall values. For example, coniferous species, such as *Abies balsamea*, has narrow tight crowns, where a relatively large object-scale sample size (i.e., 13.0 %) only takes a relatively small pixel-scale sample size (i.e., 4.8 %). Consequently, the object-based methods outperform the pixel-based methods, e.g., for recognizing *Abies balsamea*. On the contrary, deciduous species has wide open crowns, such as *Populus* spp., where a small object-scale sample size (i.e., 5.0 %) corresponds to a large pixel-scale sample size (i.e., 14.1 %). Consequently, the pixel-base methods outperform the object-based methods, e.g., for recognizing *Populus* spp. Thus, the varying performance of species recognition across different methodologies is largely attributed to the sample size available for each approach. Therefore, it is crucial to consider the crown geometries of the target species and their potential impact on pixel- and object-scale sample sizes when preparing training datasets for the respective processing approaches.

Pixel-based approaches exhibited greater sensitivity to background information and showed higher responsiveness to input data from different timepoints. This is because pixel-based approaches rely solely on the spectral information of individual pixels, without considering the spatial context among neighboring pixels within a single crown, such as structure, texture, and internal spectral variations. Meanwhile, background elements often exhibit a wide range of spectral randomness that can easily mislead the recognition task. Therefore, removing background information has a greater impact on pixel-based approaches than on object-based methods, as it more effectively minimizes disturbances from non-species-related elements in pixel-based processing. Similarly, since the spectral response of each pixel to changing data-collection timepoints can be influenced by various factors beyond phenology, such as lighting conditions and shadows, pixel-based classification methods that rely solely on the spectral characteristics of individual pixels without considering spatial context at the crown level are more susceptible to randomness, especially for deciduous trees, e.g., Fig. 5 (b), (d), and (g), where the recall values can be markedly high at some timepoints and markedly low at other time points.

Confusion between species

Overall, confusion between deciduous and conifer trees is insignificant across all processing approaches and timepoints, demonstrating that the DL models effectively learn key features to distinguish between these two groups. However, at the species or genera level, the misclassification patterns varied depending on the processing approach applied.

Fig. 6 illustrates the confusion matrix, to study the factors contributing to misclassifications among species. The results were from two representative approaches, i.e., ResNet and UNet for the objective- and pixel-based methods, respectively, and from the observations on Sep. 2 that gave the best results across all timepoints.

In general, the confusions are largely concentrated among species within the same genera, such as Acer genus (ACPE, ACRU, and ACSA) and Betula genus (BEAL and BEPA). According to the confusion matrixes in Fig. 6, the overall recall values for recognizing ACPE, ACRU, and ACSA as Acer genus are quite high, for example, being 94.15 %, 86.02 %, and 91.46 %, respectively, using ResNet with background, and 93.03 %, 87.75 %, and 89.57 %, respectively, using UNet with background. If the recall values are calculated at the genera-levels, the overall AA value for 11 genera-level classes (10 genera + dead tree) is 74.50 % and 64.40 % using ResNet and UNet with background, respectively, which are approximate 10 % higher than those for 14 classes where species-level classes are considered, i.e., 65.97 %.

The background information, i.e., including or excluding in the input data, did not significantly change the misclassification patterns, for the object-based approaches. Within deciduous or coniferous species groups, the misclassification tends to assign species with smaller sample sizes to those with larger sample sizes. When sample size is significantly small, the confusion between conifer and deciduous is possible. For instance, *Tsuga canadensis* (TSCA) that has the lowest recall value due to significant small sample size is confused with both coniferous species like *Thuja occidentalis* (THOC) and deciduous species like Populus spp. (Populus).

On the other hand, the background information significantly affects the misclassification patterns in pixel-based approaches, as the background is included as an additional class that can be randomly confused with other species due to the complex spectral characteristics of the background. Once the background is excluded, the overall confusion patterns in the pixel-based approach become similar to those observed in the object-based approach. The confusion matrices of UNet, as illustrated in Fig. 6 (c) and (d), further confirm that pixel-based methods are more significantly impacted by sample sizes. This is evident from the 0

Table 2
The average values of the overall accuracy (OA) and average accuracy (AA) across all six timepoints.

Type	Model	With background (%)		Without background (%)	
		OA	AA	OA	AA
Object-based	ResNet	**77.51**	**61.65**	**78.24**	**62.83**
	Swin Transformer	72.15	55.18	76.02	58.79
Pixel-based	UNet	62.35	49.37	75.12	57.27
	DeepLab V3+	52.28	44.02	63.41	50.21

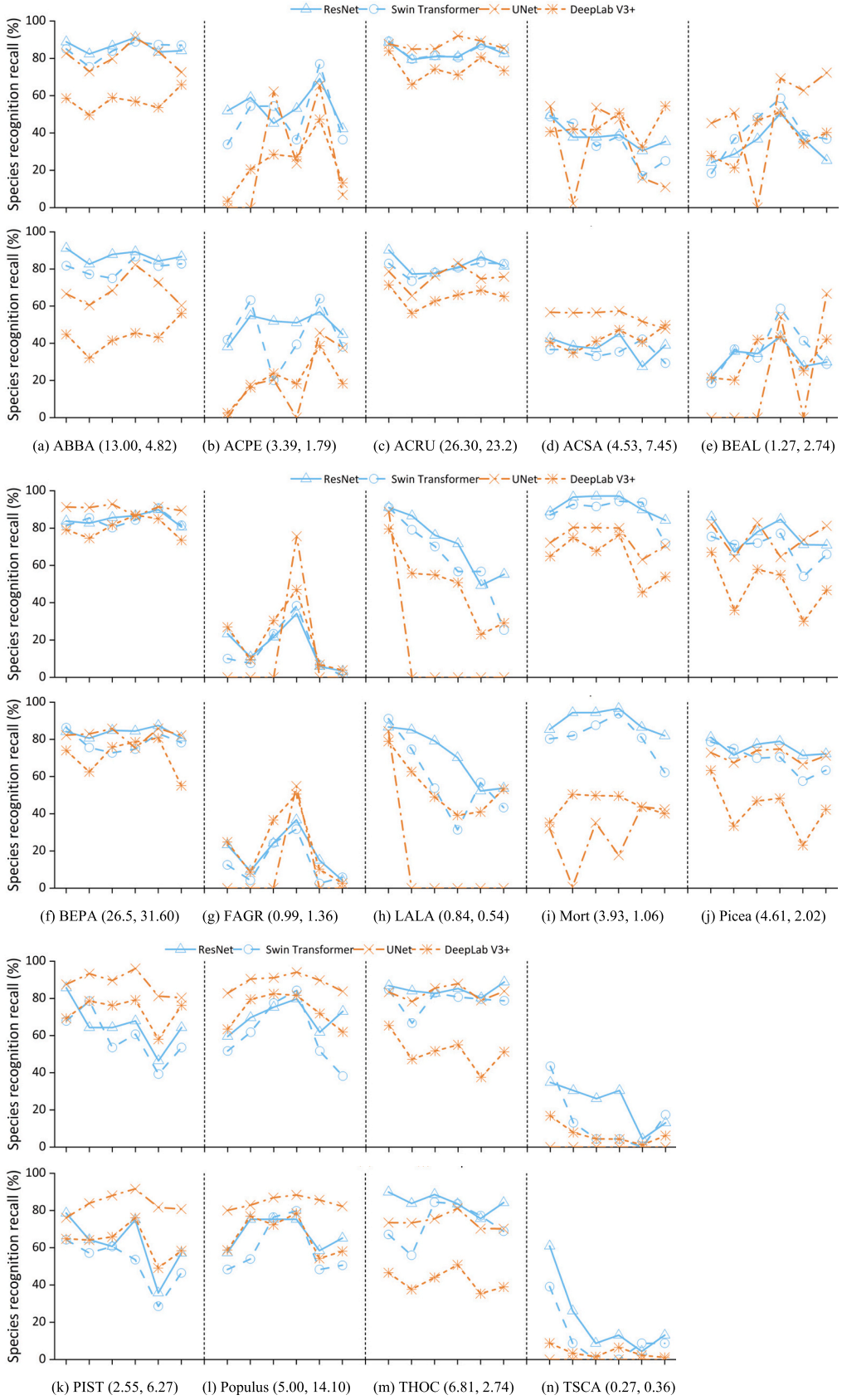


Fig. 5. The recall values of each tree species from each DL model across different timepoints, where in each cell the upper and bottom subfigures are the one without and with background, respectively. The numbers after each specie name in the braces are the relative sample sizes (%) of the species, i.e., the proportion of its own population in total population of all species, in tree number and pixel, respectively. The results of the object- and pixel-based methods are in blue and orange, respectively. The intervals on the x-axis represent different timepoints, in order from left to right: May 28, Jul. 21, Aug. 18, Sep. 2, Sep. 28, and Oct. 7. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

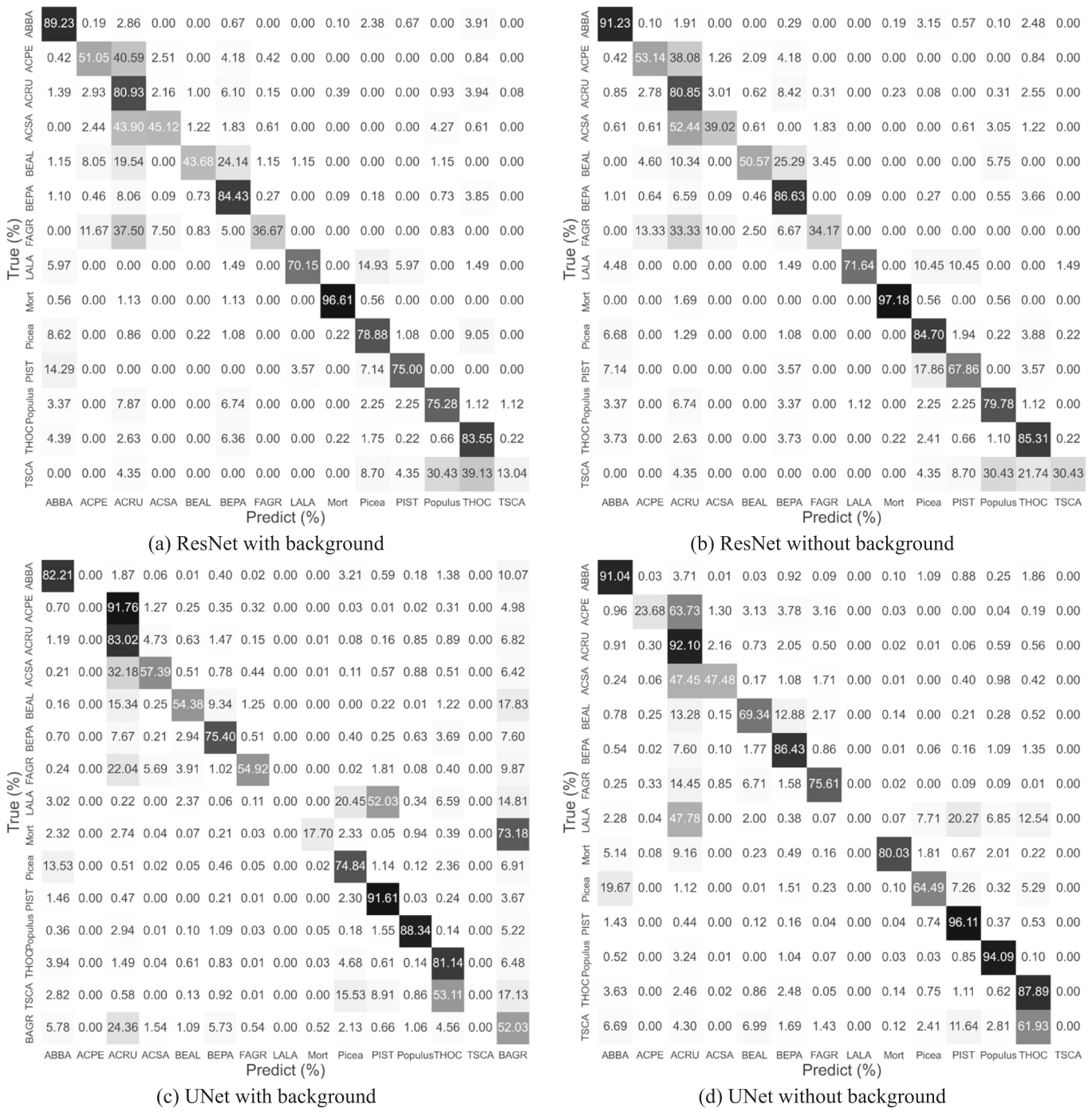


Fig. 6. The confusion matrix of tree species recognition from observations on Sep. 2 for the ResNet and UNet that represents the object- and pixel-based approach, respectively.

% recall values for *Larix laricina* (LALA) and *Tsuga canadensis* (TSCA), which represent only 0.54 % and 0.36 % of the total sample size (at the pixel scale) across all species, respectively.

3.2. Results based on multiple timepoints observations

The experiments using multi-temporal data revealed the interactions between the processing methodologies and effectiveness of phenological characteristics.

3.2.1. Overall performance for all species

The OA and AA for the multi-temporal combinations are presented in Fig. 7, alongside the results from Sep. 2 that represent the most favorable timepoint among all single timepoints and serves as a benchmark for the comparison between single- and multiple-timepoints.

As shown in Fig. 7, data at multiple timepoints outperformed that at a single timepoint, regardless of processing approaches, suggesting the advantage of extra timepoints. However, a more crucial finding is that the benefits of multi-temporal observations can only be realized when appropriate methodologies are applied. For object-based approaches ResNet and Swin Transformer, the performance improved consistently with the increasing number of timepoints. As indicated by the rising AA values, this improvement applied to almost all species, regardless of the sample sizes. Moreover, the advantage of adding more timepoints

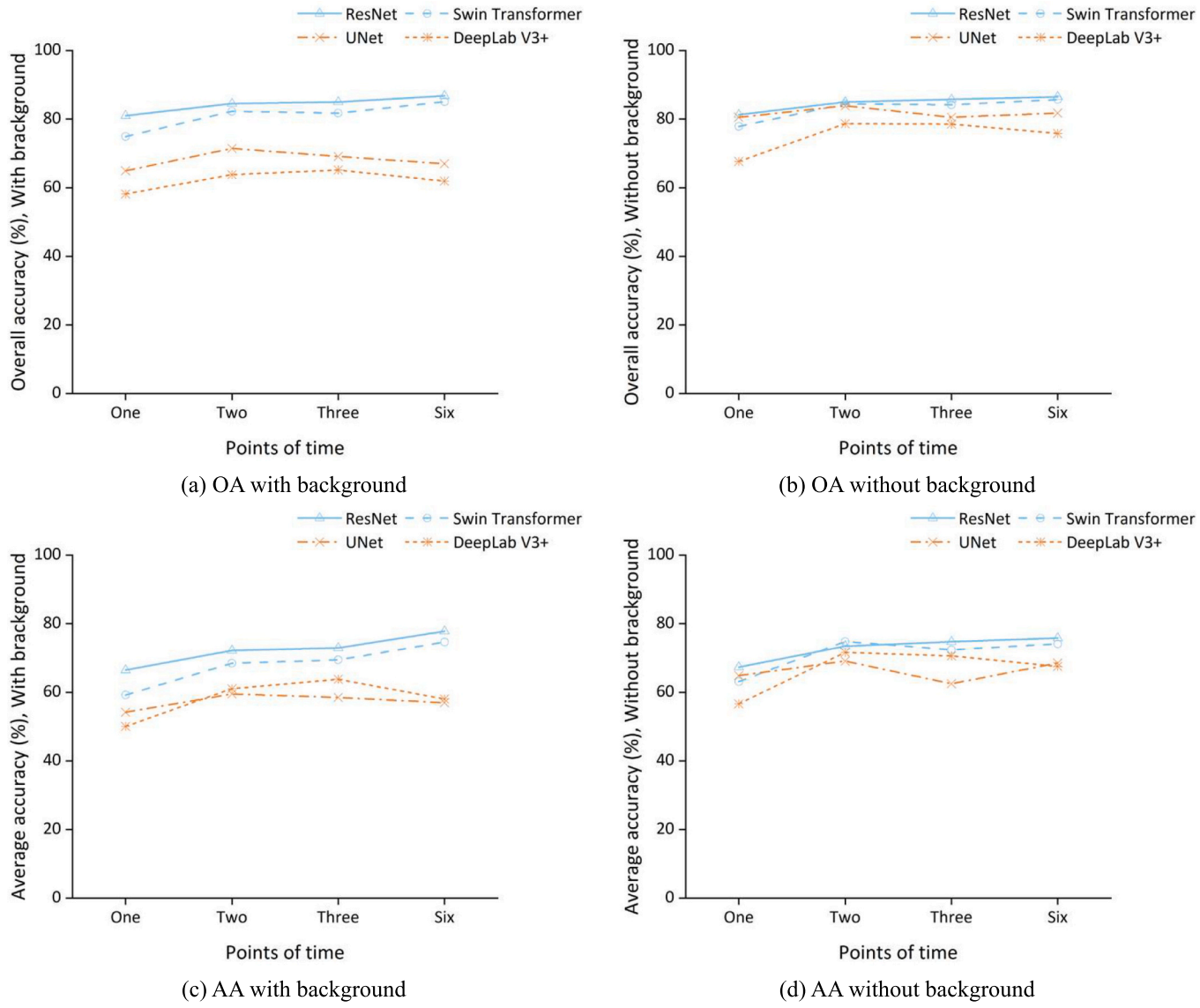


Fig. 7. The overall accuracy (OA) and average accuracy (AA) of tree species recognition using datasets from multiple points of time and (a, c) with and (b, d) without background.

becomes more pronounced in the scenarios when the prior background information was included. For both object-based approaches, the performances between the scenarios with and without background became comparable when all six timepoints were utilized.

On the contrary, for pixel-based approaches UNet and DeepLab V3+, the performance improvement did not persist once more than two timepoints were included. Both OA and AA values for UNet peaked when using two timepoints, regardless of the scenarios with and without background information. For DeepLab V3+, the best performance was achieved with two timepoints excluding background, whereas three timepoints performed the best including background. The overall performance of pixel-based approaches decreased markedly when all six timepoints were used, e.g., to the level where single timepoint was used, particularly when including background information.

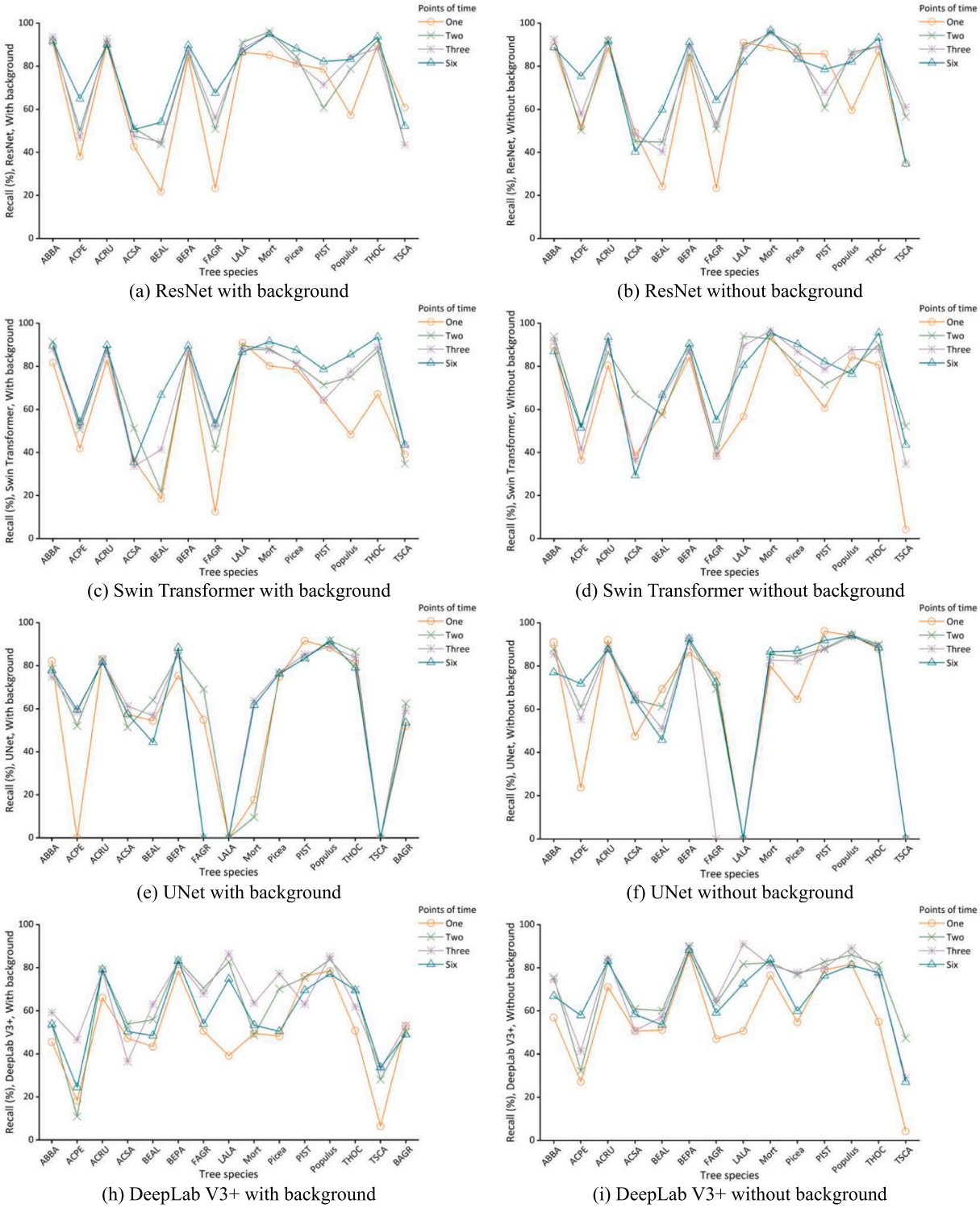
3.2.2. *Species-specific performance*

The recall values for each species, obtained from each processing approach using different combinations of timepoints, are illustrated in Fig. 8.

As mentioned before, for pixel-based approaches, the most favored time combination was two timepoints with one from late spring (May 28) and another one from early autumn (Sep 2). For those species that have small sample sizes in pixels, e.g., BEAL (2.74 %), FAGR (1.36 %), Picea (2.02 %), and TSCA (0.36 %), the recall values were most often the highest when combining data from two timepoints and then decline by adding more timepoints. This suggests that the significance of species-specific phenological characteristics becomes marginal after adding more than two timepoints, when local neighborhood context at the crown scale is absent.

On the other hand, for object-based approaches, the phenological characteristics of different species become more pronounced as the observation frequency increases. As illustrated in Fig. 8, the outcomes of ResNet and Swin Transformer suggest that adding timepoints can effectively mitigate the challenges posed by small sample sizes. This was evidenced by the consistent improvements in recall values for species with small object-scale sample sizes, such as BEAL (1.27 %), FAGR (0.99 %), PIST (2.55 %), and THOC (0.27 %). Additionally, the influence of background information was reduced when more timepoints are incorporated.

It is worth mentioning that adding more timepoints did not significantly alter the confusion patterns, regardless the applied processing approaches.



(a) ResNet with background
(b) ResNet without background
(c) Swin Transformer with background
(d) Swin Transformer without background
(e) UNet with background
(f) UNet without background
(h) DeepLab V3+ with background
(i) DeepLab V3+ without background

Fig. 8. The recall of tree species recognition using multi-timepoint data. The one point of time represents the highest overall accuracy using one-timepoint data. The two points of time are from Sep. 2 and May 28. The three points of time are from Sep. 2, May 28, and Aug. 18. The six points of time represent the combinate of all data.

3.3. Results based on input data scenarios

Table 3 reports the average of all accuracy values across nine timepoints, i.e., six single timepoints, and the combination of two, three, and six timepoints, which can be interpreted as the generally expectable accuracies under different input data scenarios.

Without existing prior knowledge, species recognition from images using pixel-based approaches produce on average approximate 60 % OA and 50 % AA, respectively. These values represent the lower performance limits of image-based species recognition, where species diversity is relatively high, e.g., over 10 species. Prior background information, i.e., significantly enhances the accuracy of pixel-based approaches, raising the OA and AA values to close to 80 % and 60 % at the best, e.g., using UNet, when the background removal was implemented. The

Table 3
The average accuracy (AA) overall accuracy (OA) across all timepoint combinations.

	Image							
	\		+ BR[1]		+ ITC[2]		+ BR + ITC	
	OA (%)	AA (%)	OA (%)	AA (%)	OA (%)	AA (%)	OA (%)	AA (%)
UNet	64.62	52.35	77.43	60.42	–	–	–	–
DeepLab V3+	56.07	49.67	68.16	56.78	–	–	–	–
ResNet	–	–	–	–	80.16	65.87	80.75	66.78
Swin-Transformer	–	–	–	–	75.77	60.40	78.94	63.78

[1] BR: background removal.
[2] ITC: individual tree crown.

performances became comparable to those of object-based approaches that require ITC delineation as a prerequisite, by reducing the accuracy gaps to around 5 %.

4. Discussion

This paper investigates the key factors that must be evaluated and balanced to achieve reliable species recognition outcomes within the constraints of available resources. These factors include three major aspects: the role of processing methodologies, the effectiveness of data, and the impact of phenological dynamics. For methodology, critical considerations include the feasibility of preprocessing approaches, such as ITC delineation or background removal, the achievable accuracy under various dataset conditions, and the availability of computational resources. Regarding data, maintaining consistency between the prepared data and the selected methodology is essential. This involves addressing variations in pixel- and object-level sample sizes influenced by the crown shapes of target species, determining the optimal timing for data collection, and assessing the benefit-cost ratio of acquiring multi-temporal data. Additionally, phenological dynamics significantly influence the effectiveness of the collected data. The success of integrating phenological variations over multiple time points depends on the chosen methodology and data. Properly considering these factors enhances the accuracy and reliability of species recognition while ensuring efficient resource utilization.

4.1. Considerations for methodology selection

This section discusses key considerations for selecting appropriate methods for tree-species recognition using images, emphasizing data requirements and feasible preprocessing approaches.

4.1.1. Characteristics of methodologies

As suggested by various benchmarking studies (Kaartinen et al., 2012; Vauhkonen et al., 2012; Wang et al., 2016; Liang et al., 2018; Wang et al., 2024), a clear understanding of the methodologies, specifically, when, where, and why certain algorithms outperform others in particular applications, is essential for setting realistic expectations about the final performance based on the available dataset and for developing effective strategies to improve the algorithms. Therefore, it is crucial to identify the key factors influencing the performance of the methods.

Object- vs. pixel- based approaches

Overall, object-based approaches outperform pixel-based methods, achieving higher recall values for over 70 % of tree species in most experiments. The advantage of the object-based method lies in that the model learns features from tree-scale instances (crowns) and therefore can leverage both the spectral and structural context (colors and textures) of pixels within each crown, which effectively mitigating random pixel-scale spectral variations due to factors like shading or lighting effects. In addition, the convenience of generating individual-tree-level species distribution products makes object-based approaches appealing in many practical applications. However, this advantage largely relies on prior knowledge of individual crown boundaries, which is typically facilitated by ITC delineation.

When multi-temporal data were used as a consequent event flow, the object- and pixel-based approaches gave clearly different results, as illustrated in Fig. 7. The divergent reactions between object- and pixel-based approaches towards increased timepoints suggest that species-specific phenological dynamics become more recognizable in multi-temporal image series when spatial context is provided by the boundaries of individual crowns. In contrast, when crown-level spatial context is missing, e.g., without crown delineation and without background removal, the pixel-scale phenological dynamics in the time series can be overshadowed by random pixel-scale variations caused by changing lighting conditions and shadow effects at different timepoints. In fact, adding more timepoints might even hinder the analysis due to a reduced signal-to-noise ratio among these spectral variations overtime. In other words, the contribution of the phenological characteristics is methodology dependent.

Moreover, for species that exhibit distinctive textural patterns on images due to their crown structures, object-based approaches significantly alleviate the challenges of small sample sizes. For instance, LALA, with an only 0.84 % sample size at the object level, achieved over 70 % recall with a single timepoint (averaged across all six timepoints) from both object-based approaches. Meanwhile, for the pixel-based method, LALA has a very small pixel-scale sample size at 0.54 %. The UNet yielded 0 % recall for LALA in five out of the six timepoints, and DeepLab V3+ achieved less than 50 % recall on average across all timepoints. The confusion matrices from data on Sep. 2 (the best timing among all six timepoints) in Fig. 6 indicates that, without background interference, LALA is frequently misclassified with other coniferous species like Picea spp. and PIST when using object-based approach such as ResNet. Conversely, LALA is mostly confused with the deciduous species ACRU when using pixel-based approach such as UNet. Fig. 9 illustrates the examples of these species, and suggests that the reason for such varied confusion pattern is that the texture plays a more influential role in object-based methods, whereas color has greater impacts in pixel-based methods.

Nevertheless, the value of pixel-based approaches should not be undermined, as direct ITC delineation from aerial images remains a highly challenging task. The latest ISPRS international contest on ITC detection and delineation using high-resolution images demonstrated that state-of-the-art approaches achieve accuracies ranging from 25 % to 55 %, depending on forest types and stand conditions (*ISPRS ITC Segmentation Contest [WWW Document]*, 2024). In other words, without reliable 3D data (e.g., from LiDAR systems or photogrammetry point clouds) to support ITC delineation, directly delineating ITCs from images remains a challenge, significantly restricting the use of object-based approaches for species recognition. From a practical point view, if the targeted area is rather large, and the primary goal of species recognition is to broadly assess species composition across the area, a pixel-based approach may be the preferred solution because it skips the requirement for the ITC segmentation and thus is much easier to implement.

Characteristics of the deep learning models

Considering the model efficiency, the object-based models are

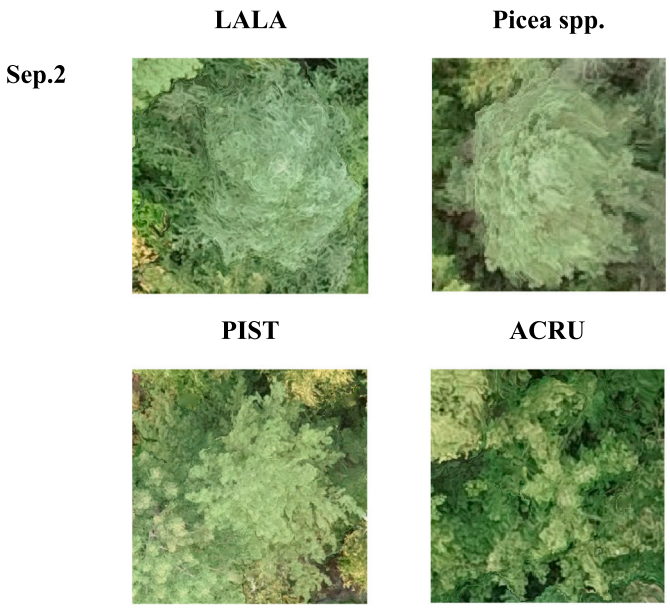


Fig. 9. Example images of LALA, Picea spp., PIST, ACRU on Sep.2.

clearly compact in size and less costly for computation, as shown in Table 4. Among the object-based approaches, the Swin Transformer presented similar overall performance as ResNet, suggesting that the contribution of the attention mechanism in the Transformer model is insignificant in the object-based species recognition. This may stem from the uniformity in texture and spectral features presented within the tree crowns, which can hinder the model from effectively attribute learning. Among the two object-based methods, ResNet can be preferable for its simplicity and reliability.

For pixel-based approaches, DeepLab V3+ demonstrated great robustness across different species, e.g., sample sizes, as well as time-points and their combinations (Fig. 8), although it did not achieve the highest OA or AA values in most experiments. The multi-scale feature extraction mechanism allows the model to concentrate on robust features across various scales, making it resilient to noise, but insensitive to subtle features that distinguish different species. Conversely, UNet is more susceptible to randomness, e.g., in shading and lighting dynamics. It may provide accurate results with appropriately allocated sample sizes and well-timed data collection, but can also provide unsatisfactory outcome when dataset is not ideal. Therefore, a multi-scale learning approach like what is in DeepLab V3+ could be tested first when the quality of the dataset is uncertain.

4.1.2. Background removal and individual tree crown delineation

According to the experimental results, preprocessing steps that provide prior knowledge, such as background exclusion and crown boundary delineation, are critical for enhancing the performance of tree-level species recognition. Furthermore, the connection between prior knowledge and the chosen methodology helps identify optimal solutions for the task.

It is important to note that background removal and ITC delineation represent two distinct types of prior knowledge. While ITC delineation often produces background removal as a byproduct, the removal of background does not necessarily depend on ITC delineation. Compared to ITC delineation, background removal is relatively less demanding, e.g., fusing spectral and 3D spatial data (Balestra et al., 2024). Typical background elements, such as open forest floors visible through canopy gaps, can be effectively identified using auxiliary datasets like low-density airborne LiDAR (a.k.a., airborne laser scanning - ALS) data or georeferenced photogrammetry point cloud data (Silva et al., 2019; Yilmaz and Gungor, 2018). Therefore, for pixel-based approaches, background removal can serve as a practical solution for improving species recognition accuracy without the need for more complex ITC delineation processes. Moreover, it is also important to highlight that forest structures, including canopy closure, vertical canopy stratification, and the type of subordinate vegetation, can significantly impact the complexity and effectiveness of background removal.

Table 4
The model size and computational complexity.

Type	Model	Parameters (Million)	FLOPs * (Giga)
Object-based	ResNet	11.19	2.38
	Swin Transformer	27.51	6.66
Pixel-based	UNet	31.04	54.80
	DeepLab V3+	54.61	20.77

* Floating Point Operations Per Second (FLOPs).

Another point worth noting is that the original outcomes of pixel-based approach only indicate the proportions of species occupancy at the area level and cannot be directly applied to applications at the individual tree level. Moreover, the initial outcomes of pixel-based approaches can be noisy due to the nature of the methods that takes pixels as the individual instances. As a result, denoising procedures are generally required when a species distribution map is the final product of the recognition process. Such denoising approaches affect the initial species recognition accuracy, either positively or negatively. Under the condition that the crown boundaries are not known, the denoising can create an "over-smoothing" effect, which may diminish the representation of species with smaller populations or crown sizes in the final map. This may obscure important biodiversity details by losing local species variations. Additionally, in cases where a small number of correctly identified pixels are surrounded by misclassified pixels, the correct pixels can be overridden by incorrect classifications, thereby reducing the accuracy of the original outcomes.

Therefore, if reliable ITC delineation is feasible, object-based approaches remain as the most optimal choice for image-based individual-tree-level species recognition, because the tree-level studies facilitate high-quality species mapping and more in-depth species-related analyses. When the aerial images are acquired with sufficient overlaps that facilitate the generation of a point cloud, the ITC delineation, and consequently, object-based species recognition would become practical because both 2D crown height model (CHM) and 3D point cloud-based approaches for individual tree delineation have been well studied and are much more reliable (Wang et al., 2016).

4.2. Considerations for data configuration

In high-resolution image-based tree species recognition, data preparation primarily involves image acquisition and data annotation. Both conventional statistical learning methods, such as Random Forest (RF) and Support Vector Machines (SVM), and DL approaches require annotated reference data for training, making data annotation essential for utilizing DL approaches. Recent advancements in DL have introduced various options to reduce the required sample sizes for reference data (Safonova et al., 2023), making data annotation increasingly manageable and less being a barrier.

4.2.1. Sample size balancing

The sample size balancing directly impacts the results. Thus, it is crucial to provide a relatively balanced population of samples across the targeted species for the species recognition using DL methods. According to the experiment results, a relative sample size at 5 % is generally required to achieve approximately 60 % recall for the species recognition. When the relative sample size reaches 10 %, the recall typically improves to around 80 %. For species that accounts for less than 2 % of all targeted species, their recall values become unpredictable because their presence in the dataset resembles that of background or noise, making them prone to randomness in the used dataset that can lead to

markedly low or high recall values. Above all, all DL models tend to allocate species with smaller samples sizes to those with larger sample sizes, thus, the gaps among the proportional sample sizes should be minimized to achieve unbiased results.

The unevenly distributed sample sizes across different species classes also introduce divergence between accuracy evaluator OA and AA. As shown in section 3.1.1, the recognition accuracy (recall) of a species with a markedly larger sample size has a greater impact on OA than on AA. For instance, as illustrated in Fig. 4 (b), when the background is excluded, the OA values of the UNet approach were similar to that of the two object-based approaches ResNet and Swin Transformer; however, the advantages of the object-based approaches become more apparent when AA values were considered. This is because pixel-based UNet is more sensitive to sample size and presents marked higher recall values for species with larger sample sizes (Fig. 5), which leads to higher OA. In other words, AA is a relatively more robust evaluation metric if all tree species are considered equally important, regardless of their sample sizes.

The impacts of the sample size are method- and crown-geometry-dependent. Therefore, it is important to assess sample sizes in alignment with the processing approach, along with a fundamental understanding of the crown geometries of the target species in the study area. The specific crown geometry of a species can lead to significantly different proportional sample sizes at the object- and pixel-scales, as demonstrated by *Populus* spp. and *Abies balsamea* in this study. These differences contributed to variations in recognition accuracy between object- and pixel-based approaches. Thus, a thorough understanding of the crown geometries of the target species can effectively aid in balancing sample sizes for the chosen processing approach.

4.2.2. Optimal timing for single data collection

This study compared datasets from six timepoints, i.e., from May 28, Jul. 21, Aug.18, Sep. 2, and Sep. 28, and Oct. 7 for species recognition in a temperate forest. The experiment results suggested that when only one timepoint was used, Sep. 2 yielded the highest OA and AA across all four DL models when background information was excluded. This timing also resulted in the highest accuracy for the two pixel-based DL models when background information was included. The second most effective timing was May 28, where the object-based approaches achieved their highest accuracies when background was included. Additionally, for experiments where Sep. 2 performed best, May 28 consistently delivered the second-highest accuracy values. For the object-based approaches, the accuracy results on Sep. 2 and May 28 were nearly equivalent, with the difference in OA and AA being less than 1.2 %. For the pixel-based approaches, the difference between these dates was more pronounced, ranging from 2.5 % to 4.9 %, suggesting that pixel-based approaches that rely on spectral characteristics of pixels is more sensitive to the timing of data collection.

Nevertheless, all four DL models agreed on the preferred timepoints, i.e., either Sep. 2 or May 28, which highlighted a strong and consistent influence of autumn and spring phenology on species recognition. This also aligns with other previous studies that showed the optimal timing for data collection in species classification within boreal and temperate forests is typically mid-to-late spring, e.g., using UAV RGB images in (Grybas and Congalton, 2021), or late autumn, e.g., using laser reflectance in (Shcherbacheva et al., 2024).

The fact that even pixel-based approaches are in favor of early autumn and late spring suggested that the variance of the phenological processes among different species is at the peak at the secondary morphogenesis period in spring (Bar and Ori, 2014) and at the early foliar senescence period in autumn (Dox et al., 2020), making the spectral features of the crowns at these two stages more distinctive among different species.

This study also agrees with what reported in (Cloutier et al., 2024) that the species recognition accuracy was highest at the onset of senescence (i.e., Sep. 2) and lowest for the peak autumn (i.e., Sep. 28). Such result emphasizes that image data should be collected during periods when species exhibit distinct phenological stages, e.g., morphogenesis period in spring and early senescence period in autumn, rather than when they all converge to a similar state, e.g., matured or peaked.

It is also worth highlighting that the results of this study indicate a narrow time window for effective image collection. Considering that, for all processing approaches, the best and worst outcomes were observed at two closely spaced timepoints, i.e., Sep. 2 and Sep. 28, reflecting the rapid pace of phenological changes, it is advisable to conduct data collection within a two-week window during critical spring and autumn phenological periods for the targeted forest.

4.2.3. Considerations for multi-temporal data approach

Adding extra timepoints clearly improves the reliability of species recognition, specifically, by tackling the challenges posed by the significantly uneven distribution of sample sizes across species, as seen in the experiments in this study. However, the effectiveness of using more than two timepoints depends significantly on the chosen processing approach.

According to the outcomes of this study, when using pixel-based approaches, i.e., without knowing crown boundaries, using two timepoints from late spring and early autumn, offers the best return in improved accuracy relative to the additional efforts for extra data collections. Compared to DeepLab V3+, UNet shows greater responsiveness to the benefits of adding an extra timepoint, which leveraged AA to 59.52 % and OA to 71.46 % with background, and AA to 69.09 % and OA to 83.88 % without background. Given the highly uneven distribution of sample sizes across species in the dataset, the results, i.e., 69.09 % AA and 83.88 % OA in recognizing 14 species from a temperate forest without the aid of ITC delineation, demonstrate a strong performance of the bi-temporal data for species recognition. Meanwhile, it is important to note that for pixel-based approaches, adding additional timepoints beyond bi-temporal data may risk reducing accuracy due to the limitations in pixel-scale analysis.

When ITC delineation is feasible, enabling the use of object-based approaches, adding observations from additional timepoints steadily improves species recognition accuracy, suggesting that subtle phenological variations among different species enhance species recognition, particularly using crown-scale analysis. Background removal is often a byproduct of ITC delineation; thus, for object-based approaches, it is more meaningful to look at the scenarios where backgrounds are removed. Using six timepoints and excluding the background, ResNet and Swin Transformer models showed accuracy improvements of 8.5 % and 11.0 % in AA, and 5.3 % and 7.8 % in OA, respectively, compared to the highest accuracy achieved with a single timepoint, which yielded AA and OA values of 75.8 % and 86.5 % for ResNet, and 74.1 % and 85.7 % for Swin Transformer, respectively.

Despite the general trend of accuracy gains with additional timepoints, the largest improvement for both object-based approaches occurred when a second timepoint was added. When background was excluded, adding a second timepoint increased AA by 6.1 % for ResNet and 11.7 % for Swin Transformer, and OA by 3.7 % and 6.6 %, respectively. This accounts for more than 70 % of the total accuracy improvement observed from using one to six timepoints. The marginal effectiveness of additional timepoints beyond bi-temporal data highlights the complexity of species recognition tasks, where further accuracy gains become difficult once AA surpasses 70 % and OA exceeds 80 %, especially under the complex scenarios such as in this study where differentiation among 14 species classes is required.

In (Grybas and Congalton, 2021) where species recognition was studied using RF approaches and multi-temporal UAV images, it was also reported that, although combining five timepoints yielded the highest OA, the improvement became marginal once three timepoints were utilized. Meanwhile, in (Shcherbacheva et al., 2024), it was suggested that it is possible to reach close to 100 % species recognition accuracy using LiDAR reflectance when observations were densified to

twice a week throughout a year, thus, the exact spectral patterns of species-specific phenology can be recorded. However, without proper infrastructural setup such as the permanent observation station for multifold research tasks employed in (Shcherbacheva et al., 2024), it is unrealistic to exhaustively collecting high-resolution image data at such high frequencies only for species recognition. If solely for species recognition, a more practical solution would be to collect bi-temporal image data with proper timing, e.g., one from late spring and one from early autumn, which could safeguard a generally plausible accuracy with properly selected method, e.g., transformer DL architectures for pixel-based approaches, or classic image segmentation architectures for object-based approaches.

4.3. Applicability of the findings

This study reveals the preferable processing designs of using multi-temporal data for tree species recognition to maximize the effectiveness of the phenological characteristics captured in data, through the comparison of different combinations of DL models and timepoint observations.

From a practical standpoint, UAV systems have become a cost-effective data source for tree-scale studies. However, the limited coverage of UAV systems presents a significant constraint, making them insufficient for characterizing large forest areas. Nonetheless, as long as high-resolution images are used for species recognition, the core relationships between methodology and data preparation identified in this study are applicable regardless of the specific study site or data source. Recent advancements in high-resolution satellite imaging, such as 0.3-m spatial resolution and beyond, offer an alternative source for high-resolution images with much broader data coverage. Therefore, the findings of this study can serve as a valuable reference for tree-scale species studies over large areas using high-resolution satellite imagery.

Conversely, the relationship between phenology and optimal timing for data collection identified in this study is more specific to temperate forests. The exact optimal period or time window, as well as the ideal number of timepoints, is dependent on forest type and site conditions, and can even vary from year to year due to the changing climate. However, the core methodological principles for effectively utilizing multiple timepoints to enhance species recognition performance remain consistent.

5. Conclusion

This study discussed how datasets and methodologies should be efficiently combined to effectively incorporate species-specific phenological variations in optimizing species recognition. The study utilized an open dataset consisting of high-resolution UAV RGB images collected across six different timepoints in a growth season, encompassing 22,139 annotated individual tree crowns representing 14 classes (including 11 specie-level, two genus-level, and one dead tree classes) in a typical temperate mixed forest in Quebec, Canada. Four state-of-the-art deep learning (DL) architectures, including two pixel- and two object-based approaches, were employed to study the species recognition performances using data from different times and different combination of timepoints and with different preprocessing approaches such as background removal through individual tree crown (ITC) delineation. Altogether, 72 experiments were implemented.

The experimental results also indicate that a balanced distribution of sample sizes across target species is crucial for achieving satisfactory species recognition performance regardless the methods. This requires accounting for the effects of crown geometry on sample sizes at pixel- and object-scales, which correspond to pixel-based and object-based DL approaches, respectively.

Although object-based approaches outperform pixel-based methods by achieving higher recall values for over 70 % of tree species in most experiments, the value of pixel-based approaches should not be undermined, because object-based approaches require prior knowledge of individual crown boundaries, and direct ITC delineation to derive these boundaries from aerial images remains a highly challenging task. With the aid from a background removal approach, the accuracy of pixel-based approaches can be improved to be comparable with object-based approaches. Such outcomes encourage the species recognition using high-resolution satellite images (e.g., with meter or sub-meter spatial resolution) and pixel-based approach for the potential to provide plausible species mapping outcomes over large areas with the aid from multi-temporal observations and background (e.g. terrain) removal supported by global terrain model products.

Overall, incorporating both spring and autumn phenology into bi-temporal dataset has the most significant benefit for species recognition by showing the capability of mitigate the challenge of markedly small sample sizes across all four DL architectures. However, pixel-based approaches struggle to extract meaningful features from datasets with more than two timepoints, and their performance declines with the addition of further timepoints. In contrast, object-based approaches that are typically enabled by prior ITC delineation can effectively leverage species-specific phenological features from multiple timepoints, thus, benefiting from multi-temporal datasets. Nevertheless, the gains from adding additional timepoints become marginal beyond two timepoints, unless more frequent observations (e.g., daily, bi-weekly, or weekly) can be implemented to capture the species-specific phenological processes during the key development periods.

Specifically, for temperate forests, when single-temporal (one-timepoint) data is used, the experimental results indicate that the optimal timing for species recognition is early autumn, followed by late spring. It is also important to note that the effective window for optimal timing can be narrow, often within a two-week period, due to the rapid phenological changes during the morphogenesis and senescence phases. Overall, research on tree-species recognition using multi-temporal observations is still in its early stages. Further studies are needed to explore this approach in different forest environments, with other data types, or across large forest management areas.

CRediT authorship contribution statement

Xinlian Liang: Conceptualization, Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition. **Jianchang Chen:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Weishu Gong:** Formal Analysis, Writing – review & editing. **Eetu Puttonen:** Investigation, Writing – review & editing. **Yunsheng Wang:** Conceptualization, Writing – review & editing, Writing – original draft, Methodology.

Declaration of competing interest

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

Acknowledgment

The authors would like to acknowledge the financial supports from the National Key Research and Development Program of China (2023YFF1303901), the National Natural Science Foundation of China (32171789, 12411530088, 32211530031), Research Council of Finland (RCF) (334060, 359204, 356137, 337656/357908), EU Horizon 2020 project “TRANSFORMIT” (101135263).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2025.114654.

Data availability

The authors do not have permission to share data.

References

Balenović, I., Liang, X., Jurjević, L., Hyyppä, J., Seletković, A., Kukko, A., 2021. Hand-held personal laser scanning: current status and perspectives for forest inventory application. Croat. J. For. Eng. 42, 163–174. https://doi.org/10.5552/crojfe.2021.858.

Balestra, M., Marselis, S., Sankey, T.T., Cabo, C., Liang, X., Mokroš, M., Peng, X., Singh, A., Stereńczak, K., Vega, C., 2024. LiDAR data fusion to improve Forest attribute estimates: a review. Curr. For. Rep. 10, 281–297. https://doi.org/10.1007/s40725-024-00223-7.

Bar, M., Ori, N., 2014. Leaf development and morphogenesis. Development 141, 4219–4230. https://doi.org/10.1242/dev.106195.

Boonman, C.C.F., Serra-Diaz, J.M., Hoeks, S., Guo, W.-Y., Enquist, B.J., Maitner, B., Malhi, Y., Merow, C., Buitenwerf, R., Svenning, J.-C., 2024. More than 17,000 tree species are at risk from rapid global change. Nat. Commun. 15, 166. https://doi.org/10.1038/s41467-023-44321-9.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. https://doi.org/10.48550/arXiv.1706.05587.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with Atrous separable convolution for semantic image segmentation. In: Presented at the Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.

Chen, J., Liang, X., Liu, Z., Gong, W., Chen, Y., Hyyppä, J., Kukko, A., Wang, Y., 2024. Tree species recognition from close-range sensing: a review. Remote Sens. Environ. 313, 114337. https://doi.org/10.1016/j.rse.2024.114337.

Cloutier, M., Germain, M., Laliberte, E., 2024. Influence of temperate forest autumn leaf phenology on segmentation of tree species from UAV imagery using deep learning. Remote Sens. Environ. 311, 114283. https://doi.org/10.1016/j.rse.2024.114283.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929.

Dox, I., Gricar, J., Marchand, L.J., Leys, S., Zuccarini, P., Geron, C., Prislan, P., Marien, B., Fonti, P., Lange, H., Penuelas, J., Van den Bulcke, J., Campioli, M., 2020. Timeline of autumn phenology in temperate deciduous trees. Tree Physiol. 40, 1001–1013. https://doi.org/10.1093/treephys/tpaa058.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87. https://doi.org/10.1016/j.rse.2016.08.013.

Gaertner, P., Foerster, M., Kleinschmit, B., 2016. The benefit of synthetically generated RapidEye and Landsat 8 data fusion time series for riparian forest disturbance monitoring. Remote Sens. Environ. 177, 237–247. https://doi.org/10.1016/j.rse.2016.01.028.

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M.C., Fröberg, M., Stendahl, J., Philipson, C.D., Mikusiński, G., Andersson, E., Westerlund, B., Andrén, H., Moberg, F., Moen, J., Bengtsson, J., 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. Nat. Commun. 4, 1340. https://doi.org/10.1038/ncomms2328.

Grybas, H., Congalton, R.G., 2021. A comparison of multi-temporal RGB and multispectral UAS imagery for tree species classification in heterogeneous New Hampshire forests. Remote Sens. 13, 2631. https://doi.org/10.3390/rs13132631.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: In: 2016 Ieee Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, New York, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hill, R.A., Wilson, A.K., George, M., Hinsley, S.A., 2010. Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. Appl. Veg. Sci. 13, 86–99. https://doi.org/10.1111/j.1654-109X.2009.01053.x.

Hyyppä, E., Hyyppä, J., Hakala, T., Kukko, A., Wulder, M.A., White, J.C., Pyörälä, J., Yu, X., Wang, Y., Virtanen, J.-P., Pohjavirta, O., Liang, X., Holopainen, M., Kaartinen, H., 2020a. Under-canopy UAV laser scanning for accurate forest field measurements. ISPRS J. Photogramm. Remote Sens. 164, 41–60. https://doi.org/10.1016/j.isprsjprs.2020.03.021.

Hyyppä, E., Kukko, A., Kaijaluoto, R., White, J.C., Wulder, M.A., Pyörälä, J., Liang, X., Yu, X., Wang, Y., Kaartinen, H., Virtanen, J.-P., Hyyppä, J., 2020b. Accurate derivation of stem curve and volume using backpack mobile laser scanning. ISPRS J. Photogramm. Remote Sens. 161, 246–262. https://doi.org/10.1016/j.isprsjprs.2020.01.018.

ISPRS ITC Segmentation Contest [WWW Document]. URL. https://www2.isprs.org/commissions/comm3/wg1/news/contest/isprs-itc-segmentation-contest/ (accessed 11.12.24).

Jiang, Y., Zhang, L., Yan, M., Qi, J., Fu, T., Fan, S., Chen, B., 2021. High-resolution mangrove forests classification with machine learning using Worldview and UAV hyperspectral data. Remote Sens. 13, 1529. https://doi.org/10.3390/rs13081529.

Jurjević, L., Liang, X., Gašparović, M., Balenović, I., 2020. Is field-measured tree height as reliable as believed – Part II, a comparison study of tree height estimates from conventional field measurement and low-cost close-range remote sensing in a deciduous forest. ISPRS J. Photogramm. Remote Sens. 169, 227–241. https://doi.org/10.1016/j.isprsjprs.2020.09.014.

Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B.M., Wu, J.-C., 2012. An international comparison of individual tree detection and extraction using airborne laser scanning. Remote Sens. 4, 950–974. https://doi.org/10.3390/rs4040950.

Kahl, T., Bauhus, J., 2014. An index of forest management intensity based on assessment of harvested tree volume, tree species composition and dead wood origin. Nat. Conserv. 7, 15–27.

Key, T., Warner, T.A., McGraw, J.B., Fajvan, M.A., 2001. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood Forest. Remote Sens. Environ. 75, 100–112. https://doi.org/10.1016/S0034-4257(00)00159-0.

Liang, X., Wang, Y., Jaakkola, A., Kukko, A., Kaartinen, H., Hyyppa, J., Honkavaara, E., Liu, J., 2015. Forest data collection using terrestrial image-based point clouds from a handheld camera compared to terrestrial and personal laser scanning. IEEE Trans. Geosci. Remote Sens. 53, 5117–5132. https://doi.org/10.1109/TGRS.2015.2417316.

Liang, X., Hyyppä, J., Kaartinen, H., Lehtomäki, M., Pyörälä, J., Pfeifer, N., Holopainen, M., Brolly, G., Francesco, P., Hackenberg, J., Huang, H., Jo, H.-W., Katoh, M., Liu, L., Mokroš, M., Morel, J., Olofsson, K., Poveda-Lopez, J., Trochta, J., Wang, D., Wang, J., Xi, Z., Yang, B., Zheng, G., Kankare, V., Luoma, V., Yu, X., Chen, L., Vastaranta, M., Saarinen, N., Wang, Y., 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. ISPRS J. Photogramm. Remote Sens. 144, 137–179. https://doi.org/10.1016/j.isprsjprs.2018.06.021.

Liang, X., Kukko, A., Balenovic, I., Saarinen, N., Junttila, S., Kankare, V., Holopainen, M., Mokros, M., Surovy, P., Kaartinen, H., Jurjevic, L., Honkavaara, E., Nasi, R., Liu, J., Hollaus, M., Tian, J., Yu, X., Pan, J., Cai, S., Virtanen, J.-P., Wang, Y., Hyyppa, J., 2022. Close-range remote sensing of forests: the state of the art, challenges, and opportunities for systems and data acquisitions. IEEE Geosci. Remote Sens. Mag. 10, 32–71. https://doi.org/10.1109/MGRS.2022.3168135.

Liang, X., Qi, H., Deng, X., Chen, J., Cai, S., Zhang, Q., Wang, Y., Kukko, A., Hyyppä, J., 2024a. ForestSemantic: a dataset for semantic learning of forest from close-range sensing. Geo-spat. Inf. Sci. https://doi.org/10.1080/10095020.2024.2313325.

Liang, X., Wang, Y., Pyörälä, J., Lehtomäki, M., Yu, X., Kaartinen, H., Kukko, A., Honkavaara, E., Issaoui, A.E.I., Nevalainen, O., Vaaja, M., Virtanen, J.-P., Katoh, M., Deng, S., 2019. Forest in situ observations using unmanned aerial vehicle as an alternative of terrestrial measurements. Forest Ecosyst 6, 20. https://doi.org/10.1186/s40663-019-0173-3.

Liang, X., Yao, H., Qi, H., Wang, X., 2024b. Forest in situ observations through a fully automated under-canopy unmanned aerial vehicle. Geo-spat. Inf. Sci. https://doi.org/10.1080/10095020.2024.2322765.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986.

Miyoshi, G.T., Imai, N.N., Garcia Tommaselli, A.M., Antunes de Moraes, M.V., Honkavaara, E., 2020. Evaluation of hyperspectral multitemporal information to improve tree species identification in the highly diverse Atlantic Forest. Remote Sens. 12, 244. https://doi.org/10.3390/rs12020244.

Mokroš, M., Mikita, T., Singh, A., Tomaštík, J., Chudá, J., Wężyk, P., Kuželka, K., Surový, P., Klimánek, M., Zięba-Kulawik, K., Bobrowski, R., Liang, X., 2021. Novel low-cost mobile mapping systems for forest inventories as terrestrial laser scanning alternatives. Int. J. Appl. Earth Obs. Geoinf. 104, 102512. https://doi.org/10.1016/j.jag.2021.102512.

Oettel, J., Lapin, K., 2021. Linking forest management and biodiversity indicators to strengthen sustainable forest management in Europe. Ecol. Indic. 122, 107275. https://doi.org/10.1016/j.ecolind.2020.107275.

Pyörälä, J., Liang, X., Saarinen, N., Kankare, V., Wang, Y., Holopainen, M., Hyyppä, J., Vastaranta, M., 2018a. Assessing branching structure for biomass and wood quality estimation using terrestrial laser scanning point clouds. Can. J. Remote. Sens. 44, 462–475. https://doi.org/10.1080/07038992.2018.1557040.

Pyörälä, J., Liang, X., Vastaranta, M., Saarinen, N., Kankare, V., Wang, Y., Holopainen, M., Hyyppa, J., 2018b. Quantitative assessment of scots pine (*Pinus Sylvestris* L.) whorl structure in a Forest environment using terrestrial laser scanning. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 11, 3598–3607. https://doi.org/10.1109/JSTARS.2018.2819598.

Qiu, P., Wang, D., Zou, X., Yang, X., Xie, G., Xu, S., Zhong, Z., 2019. Finer resolution estimation and mapping of mangrove biomass using UAV LiDAR and WorldView-2 data. Forests 10, 871. https://doi.org/10.3390/f10100871.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention. Pt III. Springer International Publishing Ag, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., Ryo, M., 2023. Ten deep learning techniques to address small data problems with remote sensing. Int. J. Appl. Earth Obs. Geoinf. 125, 103569. https://doi.org/10.1016/j.jag.2023.103569.

Shcherbacheva, A., Campos, M.B., Wang, Y., Liang, X., Kukko, A., Hyyppä, J., Junttila, S., Lintunen, A., Korpela, I., Puttonen, E., 2024. A study of annual tree-wise LiDAR intensity patterns of boreal species observed using a hyper-temporal laser scanning time series. Remote Sens. Environ. 305, 114083. https://doi.org/10.1016/j.rse.2024.114083.

Silva, C.A., Valbuena, R., Pinage, E.R., Mohan, M., de Almeida, D.R.A., North, E., Jaafar, W.S.W.M., de Papa, D.A.M., Cardil, A., Klauberg, C., 2019. ForestGapR: an r package for forest gap analysis from canopy height models. Methods Ecol. Evol. 10, 1347–1356. https://doi.org/10.1111/2041-210X.13211.

Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E., Gabler, K., Schadauer, K., Vidal, C., Lanz, A., Ståhl, G., Cienciala, E., 2010. National forest inventories. Pathways for common reporting. Europ. Sci. Found. 1, 541–553.

van Tiel, N., Fopp, F., Brun, P., van den Hoogen, J., Karger, D.N., Casadei, C.M., Lyu, L., Tuia, D., Zimmermann, N.E., Crowther, T.W., Pellissier, L., 2024. Regional uniqueness of tree species composition and response to forest loss and climate change. Nat. Commun. 15, 4375. https://doi.org/10.1038/s41467-024-48276-3.

Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkanen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalen, P., Gobakken, T., Koch, B., Naesset, E., Tokola, T., Maltamo, M., 2012. Comparative testing of single-tree detection algorithms under different types of forest. Forestry 85, 27–40. https://doi.org/10.1093/forestry/cpr051.

Veras, H.F.P., Ferreira, M.P., da Neto, E.M.C., Figueiredo, E.O., Corte, A.P.D., Sanquetta, C.R., 2022. Fusing multi-season UAS images with convolutional neural networks to map tree species in Amazonian forests. Ecol. Inform. 71, 101815. https://doi.org/10.1016/j.ecoinf.2022.101815.

Vorster, A.G., Evangelista, P.H., Stovall, A.E.L., Ex, S., 2020. Variability and uncertainty in forest biomass estimates from the tree to landscape scale: the role of allometric equations. Carbon Balance Manag. 15, 8. https://doi.org/10.1186/s13021-020-00143-6.

Wang, Y., Hyyppä, J., Liang, X., Kaartinen, H., Yu, X., Lindberg, E., Holmgren, J., Qin, Y., Mallet, C., Ferraz, A., 2016. International benchmarking of the individual tree detection methods for modeling 3-D canopy structure for silviculture and forest ecology using airborne laser scanning. IEEE Trans. Geosci. Remote Sens. 54, 5011–5027. https://doi.org/10.1109/TGRS.2016.2543225.

Wang, Y., Lehtomäki, M., Liang, X., Pyörälä, J., Kukko, A., Jaakkola, A., Liu, J., Feng, Z., Chen, R., Hyyppä, J., 2019a. Is field-measured tree height as reliable as believed – a comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a boreal forest. ISPRS J. Photogramm. Remote Sens. 147, 132–145. https://doi.org/10.1016/j.isprsjprs.2018.11.008.

Wang, Y., Pyörälä, J., Liang, L., Lehtomäki, M., Kukko, A., Yu, X., Kaartinen, H., Hyyppä, J., 2019b. In situ biomass estimation at tree and plot levels: What did data record and what did algorithms derive from terrestrial and aerial point clouds in boreal forest. Remote Sensing of Environment 232, 111309. https://doi.org/10.1016/j.rse.2019.111309.

Wang, Y., Kukko, A., Hyyppä, E., Hakala, T., Pyörälä, J., Lehtomäki, M., El Issaoui, A., Yu, X., Kaartinen, H., Liang, X., Hyyppä, J., 2021. Seamless integration of above- and under-canopy unmanned aerial vehicle laser scanning for forest investigation. For. Ecosyst. 8, 10. https://doi.org/10.1186/s40663-021-00290-3.

Wang, X., Liang, X., Campos, M., Zhang, J., Wang, Y., 2024. Benchmarking of laser-based simultaneous localization and mapping methods in Forest environments. IEEE Trans. Geosci. Remote Sens. 62, 1–21. https://doi.org/10.1109/TGRS.2024.3439438.

Wessely, J., Essl, F., Fiedler, K., Gattringer, A., Hülber, B., Ignateva, O., Moser, D., Rammer, W., Dullinger, S., Seidl, R., 2024. A climate-induced tree species bottleneck for forest management in Europe. Nat. Ecol. Evol. 8, 1109–1117. https://doi.org/10.1038/s41559-024-02406-8.

Yilmaz, C.S., Gungor, O., 2018. Comparison of the performances of ground filtering algorithms and DTM generation from a UAV-based point cloud. Geocarto Int. 33, 522–537. https://doi.org/10.1080/10106049.2016.1265599.