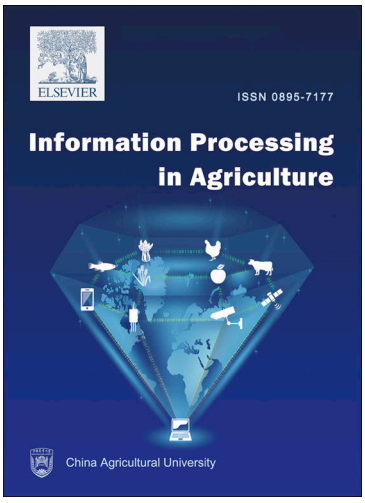
Journal Pre-proofs



Optimization of Sassafras tzumu leaves color quantification with UAV RGB imaging and Sassafras-net

Qingwei Meng, Wei Qi Yan, Cong Xu, Zhaoxu Zhang, Xia Hao, Hui Chen, Wei Liu, Yanjie Li

PII: S2214-3173(25)00001-0
DOI: https://doi.org/10.1016/j.inpa.2025.02.001
Reference: INPA 458

To appear in: *Information Processing in Agriculture*

Received Date: 1 July 2024
Revised Date: 24 January 2025
Accepted Date: 1 February 2025

Please cite this article as: Q. Meng, W. Qi Yan, C. Xu, Z. Zhang, X. Hao, H. Chen, W. Liu, Y. Li, Optimization of Sassafras tzumu leaves color quantification with UAV RGB imaging and Sassafras-net, *Information Processing in Agriculture* (2025), doi: https://doi.org/10.1016/j.inpa.2025.02.001

■ **Title Page:**

Title: Optimization of Sassafras tzumu leaves color quantification with UAV RGB imaging and Sassafras-net

Authors: Qingwei Meng1,2, Wei Qi Yan3, Cong Xu4, Zhaoxu Zhang5, Xia Hao5, Hui Chen6, Wei Liu7, Yanjie Li1*

Affiliations:

1. State Key Laboratory of Tree Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou, 311400, Zhejiang China

2. College of Horticulture and Gardening, Yangtze University, Jingzhou, 434025, Hubei, China

3. Auckland University of Technology, Auckland 1010, New Zealand

4. School of Forestry, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

5. College of Information Science and Engineering, Shandong Agricultural University, Taian, 271018, Shandong China

6. Forest Pest Control and Quarantine Station of Shaoxing City, Shaoxing, 312000, Zhejiang, China

7. College of Optical, Mechanical and Electrical Engineering, Zhejiang A&F University, Hangzhou 311300, Zhejiang China

Corresponding author: *Yanjie Li: aj7105@gmail.com*

Keywords: UAV RGB images; Color leaves; YOLOX; CCTrans network; Tree phenotypes

- UAV RGB imagery model outperforms in leaf detection/counting
- Sassafras detection and leaf counting accuracy were enhanced by YOLOX and CCTrans
- Sassafras-net model significantly reduces MAE and MSE in leaf counting

- YOLOX minimizes false/missed detections effectively

Optimization of *Sassafras tzumu* leaf color quantification with UAV RGB imaging and Sassafras-net

Qingwei Meng[1,2], Wei Qi Yan[3], Cong Xu[4], Zhaoxu Zhang[5], Xia Hao[5], Hui Chen[6], Wei Liu[7], Yanjie Li[1]*

8. State Key Laboratory of Tree Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou, 311400, Zhejiang China

9. College of Horticulture and Gardening, Yangtze University, Jingzhou, 434025, Hubei, China

10. Auckland University of Technology, Auckland 1010, New Zealand

11. School of Forestry, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

12. College of Information Science and Engineering, Shandong Agricultural University, Taian, 271018, Shandong China

13. Forest Pest Control and Quarantine Station of Shaoxing City, Shaoxing, 312000, Zhejiang, China

14. College of Optical, Mechanical and Electrical Engineering, Zhejiang A&F University, Hangzhou 311300, Zhejiang China

***Corresponding author: Yanjie Li:** *aj7105@gmail.com*

Abstract

Quantifying the leaf density and coloration of trees is critical for assessing landscape esthetics and photosynthetic efficiency; however, traditional leaf-counting methods are labor-intensive and potentially harmful to trees, making accurate measurements challenging. To address these issues, we present "Sassafras-net," an advanced model specifically designed to detect and count colored leaves on *Sassafras tzumu* trees.

The methodology consists of two steps. First, we used an improved model termed YOLOX-CBAM to accurately detect and isolate individual trees. This model proved to be more effective than alternatives, such as YOLOX, YOLOv8, YOLOv7, YOLOv5, and Fater-RCNN. Second, the Sassafras-net model, which is based on the CCTrans network, counts the number of colored leaves per tree. Compared with the original CCTrans model of 52.30 and 84.90, the Sassafras-net model achieved significantly lower mean absolute error and mean squared error values of 27.29 and 39.00, respectively. These results confirm the ability of the model to accurately and efficiently quantify colored leaves.

To the best of our knowledge, this is the first study to quantify colored leaves in trees. Our method provides forestry researchers with an effective and economical tool for selecting and breeding *S. tzumu* trees with enhanced color traits. In addition, this study opens new avenues for studying tree traits related to leaf coloration.

Keywords: UAV RGB images; Color leaves; YOLOX; CCTrans network; Tree phenotypes

1 Introduction

Sassafras tzumu, which is predominantly found in the subtropical regions of southern China [1], is a deciduous tree of great ecological and economic value. Renowned for its rapid growth and high tolerance to diverse soil conditions [2-4], it is used for various purposes ranging from shipbuilding to interior decoration. Its wood, which is known for its durability and fine-grained texture, is highly sought after for furniture and shipbuilding [5]. The medicinal properties of this tree, particularly in the treatment of rheumatism and trauma, and its use for extraction of essential oils such as safrole, underscore its importance in traditional medicine and commercial applications [6]. In addition, *S. tzumu* is characterized by sympodial branching, and its leaves are known for their variable lobed structures and unique gray-green coloration which changes to bright shades of red or yellow in the fall, making it a popular choice for ornamental horticulture [7]. This species also plays an important role in genetic diversity studies because of its moderate genetic differentiation and unique environmental adaptations [8].

Quantifying the number and density of colored leaves in trees such as *S. tzumu* is crucial for understanding various physiological and ecological aspects, including nutrient transport, crown size, and overall tree growth. These factors serve as important indicators for the selection of species that enhance landscape coloration and contribute to biodiversity [9, 10]. Leaf color changes, especially in deciduous trees, are not only visually striking but also indicate important biochemical and physiological processes, such as chlorophyll degradation and resource reallocation during leaf senescence [11]. Monitoring these changes is essential for assessing tree health and adaptive responses to environmental conditions.

Conventional methods for quantifying leaf color and density frequently lack accuracy and efficiency. For instance, manual counting can result in error rates ranging from 15% to 25% owing to human fatigue and difficulties in distinguishing overlapping leaves in dense canopies. In addition, the process is extremely time-consuming and often requires several hours of worktime per tree, making it impractical for large-scale studies [12, 13]. These methods can be invasive or harmful to trees, particularly when dealing with thick branches and dense canopies. Spectrophotometry and colorimetric analysis provide accurate color measurements but are limited to small samples and cannot capture the spatial distribution across entire trees or forests [14].

Advances in remote sensing and image analyses have introduced new methods for leaf quantification in forestry. Unmanned aerial vehicles (UAVs) equipped with high-resolution RGB cameras have revolutionized forest monitoring and assessment [15, 16].

In forestry, UAVs are used for tasks such as tree species classification [17], canopy structure analysis [18], biomass estimation [19], and leaf area index measurement [20]. High-resolution imagery from UAVs allows the precise delineation of individual tree crowns in mixed forests [21], thereby enhancing species identification and forest inventory accuracy. UAV applications have also expanded to include tree health assessments and pest detection, highlighting the versatility and value of this technology in managing forest ecosystems [22, 23]. Specifically related to leaf counting and phenotypic trait analysis, UAV-based imagery has been used to estimate leaf area and assess defoliation in boreal forests, demonstrating the potential for detailed leaf-level analysis from aerial perspectives [24, 25]. Additionally, the integration of deep-learning (DL) models with UAV data has facilitated the extraction of canopy information and the estimation of leaf biomass, contributing significantly to precision forestry practices [26, 27]. These developments highlight the efficacy of combining UAV technology with advanced image analysis techniques to improve the accuracy and efficiency of forestry management and research.

These advanced methods offer significant advantages over traditional techniques. Remote sensing and image analysis provide nondestructive, efficient, and scalable solutions for leaf quantification [28]. Unlike manual counting and visual assessments, which are labor-intensive and time-consuming, UAVs can quickly capture detailed images of large forest areas without physical contact, thereby minimizing the risk of damaging trees. DL algorithms process these images to analyze complex patterns and variations in leaf color and shape, thereby providing greater accuracy and consistency than human observers [29]. Additionally, these technologies can adapt to various environmental conditions and tree species, thus increasing their applicability in diverse forestry contexts [30].

Machine learning (ML) methods, such as support vector machines (SVMs) [31-33], random forests (RFs) [34], and recurrent neural networks (RNNs) [35-37] can be used to extract various categorical features; however, these methods are limited in terms of the recognition rate and generalization for complex recognition tasks. While methods

such as RFs offer some robustness against overfitting, SVMs and RNNs frequently require manual feature selection, which becomes increasingly cumbersome and less effective as dataset sizes increase. For example, SVM can suffer from high computational complexity and long training times when applied to large datasets and tasks involving nonlinear feature selection [38]. In contrast, DL has significant advantages in image processing because it can learn complex features autonomously through multilayer abstraction, thereby reducing the reliance on manual feature design. DL overcomes the challenges of traditional ML methods in image recognition by autonomously extracting critical features from data across multiple layers, which traditional methods struggle with, especially when scaling to large datasets or complex feature spaces [39].

DL, a sophisticated branch of ML, utilizes complex neural architectures to intelligently compute and autonomously recognize targets in large datasets. These methods significantly improve model accuracy and generalization, facilitating the extraction of key features for tasks such as plant phenotyping [40-42]. DL methods are generally divided into segmentation-, object detection-, and density map-based approaches [43]. Segmentation-based models, such as U-net [44], are highly effective in image segmentation tasks, facilitating the precise delineation of leaf and plant structures. The encoder–decoder architecture of U-net allows it to capture both spatial and contextual information, making it suitable for detailed leaf segmentation and phenotypic analyses. Graph Convolutional Networks (GCNs) [45] are a further class of DL algorithms that operate on graph-structured datasets. GCNs are adept at modeling relationships and interactions in complex structures, such as the arrangement of leaves in a canopy or the branching patterns of trees. This capability makes them valuable for tasks that require an understanding of relational data in forestry applications. Object detection-based methods, such as You Only Look Once X (YOLOX) [46], have shown remarkable speed and accuracy in detecting objects within images, making them suitable for real-time applications. Density map-based methods, including the Multi-column Convolutional Neural Network (MCNN) [47], Congested Scene Recognition Network (CSRNet) [48], and Distribution Matching for Crowd Counting (DM-Count) [49], utilize convolutional neural networks to estimate the object density for counting tasks. These methods perform well in scenarios involving high-density objects and occlusions.

Innovative approaches, such as Crowd Counting with Transformer (CCTrans) [50], employing the Twins backbone network [51], and a pyramid feature fusion module, have been developed to calculate the number of colored leaves through density map regression. This method excels in distinguishing between overlapping and occluded colored leaves, demonstrating strong generalization capabilities and robustness. Originally designed for crowd density estimation and management in congested settings, these algorithms have been adapted for forestry target detection [52, 53]. Although still in the exploratory stage in forestry, they utilize high-spatial-resolution images to estimate the number of trees and phenotypic traits in specific areas, showing great potential for precise forest resource management and environmental monitoring [53]. The application of these advanced DL models underscores the efficacy of

combining UAV technology with sophisticated image analysis techniques to improve the accuracy and efficiency of forestry management and research [54].

Building on these advancements, there remains a need for specialized models that can accurately quantify phenotypic traits, such as colored leaves in specific tree species, which existing methods have yet to address comprehensively. This study introduces a novel colored-leaf-counting model for *S. tzumu* using UAV RGB images by integrating the YOLOX and CCTrans frameworks. By doing so, we aimed to bridge the gap between advanced DL techniques and practical forestry applications. The objectives of this study were:

1) To develop a single-tree identification model (Sassafras-net) to segment individual *S. tzumu* trees, facilitating precise isolation and analysis of each tree within a forested landscape.

2) To implement a counting method that can accurately quantify the number of colored leaves on individual *S. tzumu* trees using advanced image processing and DL algorithms for detailed phenotypic trait analysis.

2 Materials and methods

We developed "Sassafras-net," a model designed to detect and count the colored leaves of *S. tzumu* trees using UAV-captured RGB images. Fig. 1 illustrates the overall framework of this model, which comprises two main components. The first component, a single *S. tzumu* extraction segment, utilizes an improved YOLOX DL trained on a dataset of manually annotated *S. tzumu* bounding boxes. This facilitates the precise detection and extraction of individual *S. tzumu* trees while effectively filtering out complex background elements such as concrete roads, other tree species, and weeds. The second component focuses on the detection and counting of colored leaves. This segment employs a training model developed with manually labeled coordinates of colored *S. tzumu* leaves, guiding the learning process of the counting model to accurately enumerate and generate a density map of the colored leaves. The integration of these components ensures accurate quantification of colored leaves on individual *S. tzumu* trees, thus enhancing the utility of the model for ecological and botanical studies.

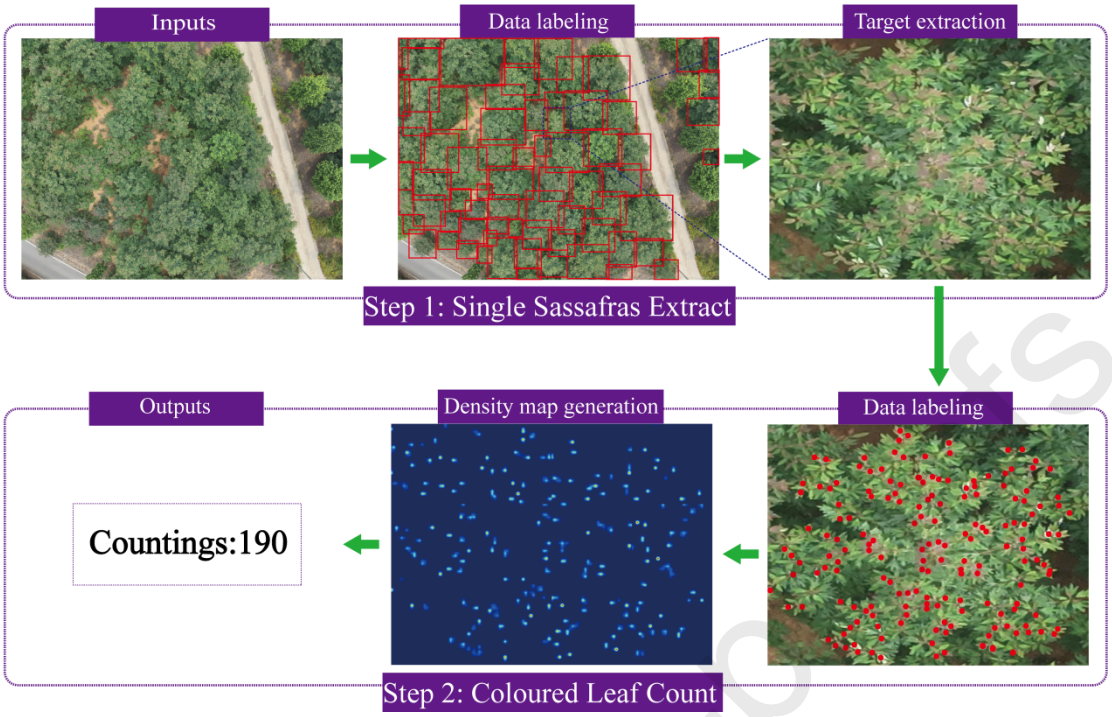


Fig. 1. Experimental and Technical Design Specifications. (Original Image: scale of 4000 × 3000, approximate size of 9216 KB, single *S. tzumu* tree image: scale of 1024 × 768, approximate size of 120 KB).

2.1 Experimental sites

The study area, a *S. tzumu* family plantation, is situated within the Longshan State Forest in Anji County, Huzhou City, Zhejiang province, China, with geographical coordinates of 30°50' N and 119°42' E (Fig. 2). This region is characterized by a subtropical climate and features a varied landscape of rolling hills with elevations ranging from 80 to 230 m. The climatic conditions of the area are defined by an annual precipitation of 1286.5 mm, a frost-free period lasting approximately 120 days, and an average annual temperature of 15.3 °C [55].

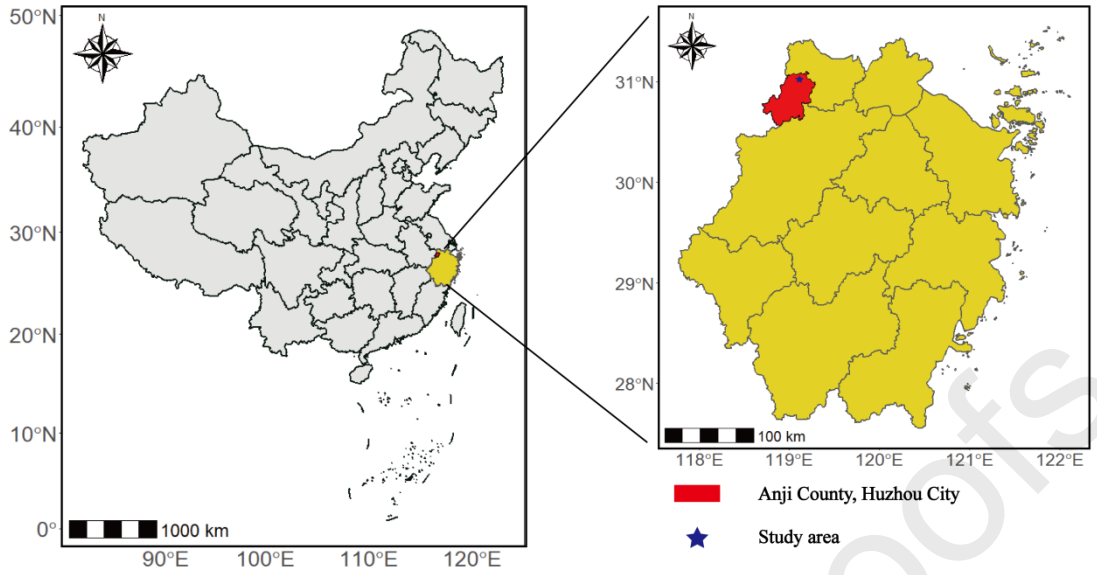


Fig. 2. Location of Image Collection Site.

2.2 Collection and preprocessing of data sets

Images of *S. tzumu* plantations were acquired using a DJI Matrice 300 RTK drone equipped with a Zenmuse P1 sensor. To minimize the effects of sunlight reflection and to ensure consistent lighting conditions, image acquisition was conducted on clear, cloudless days at approximately 11 a.m. An altitude of 40 m was selected as the optimal height to balance image resolution and coverage area. At this altitude, the UAV was able to capture high-resolution images with sufficient detail to detect individual colored leaves while covering a reasonable area to increase data collection efficiency. Lower flying would result in higher resolution but cover a smaller area per image, thus increasing the flight time and data processing load. Conversely, flying higher may reduce image resolution, which would render it difficult to accurately detect and count colored leaves. The Zenmuse P1 sensor captured images at a resolution of 4000 × 3000 pixels, which was chosen to ensure that the spatial resolution was sufficient to distinguish individual leaves and their color variations. This high-resolution setting allows for the accurate detection and counting of colored leaves while keeping file sizes manageable for data storage and processing. Higher resolutions would generate excessively large files, thus complicating data handling without significant gains in detection accuracy. Data collection included two monthly flights from August to November 2023 to maintain consistent conditions across time points. The camera settings provided substantial image overlap with side and front overlaps of 80% and 85%, respectively, to facilitate accurate stitching and analysis during data processing.

A total of 3,560 images were captured and stored in JPG format on the SD card of the UAV before being transferred to a computer for analysis (example images are shown in Fig. 3). For data processing, the original data set was resized to 640 × 640 pixels and divided into training, test, and validation sets in a 6:2:2 ratio. Specifically:

- For the single *S. tzumu* detection and extraction task, 1420 images were selected and divided into 852 images for training, 284 for testing, and 284 for validation.

- For the task of detecting and counting the colored leaves of individual *S. tzumu* trees, 310 images were selected, with 224 in the training set and 43 each in the test and validation sets.

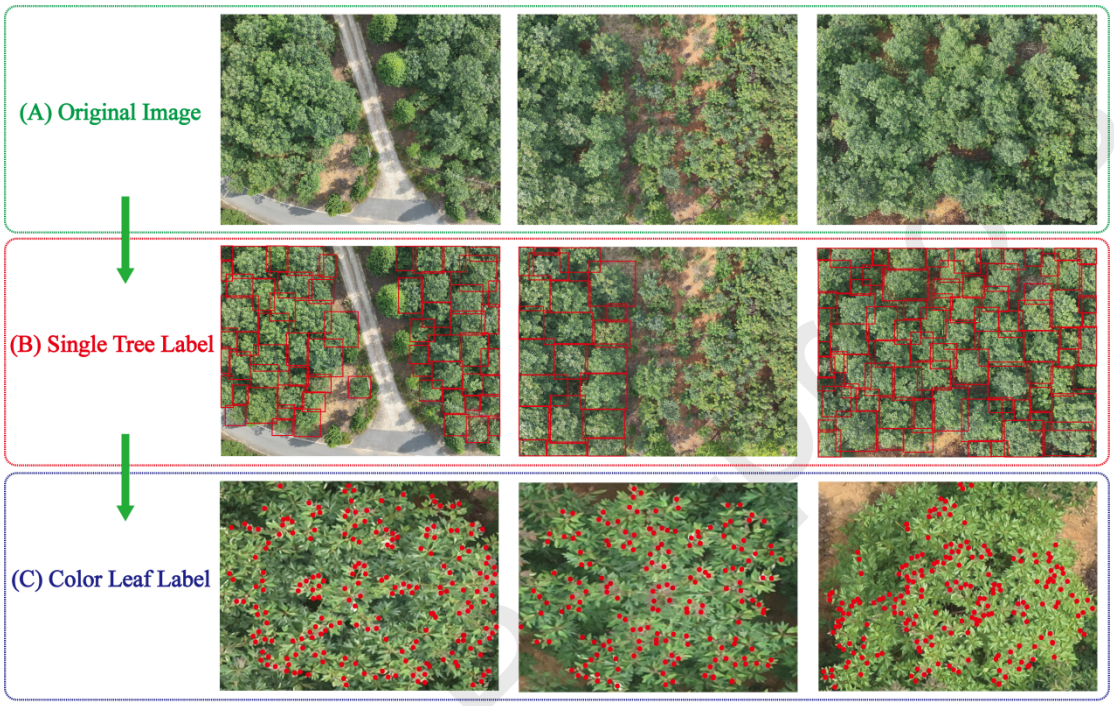


Fig. 3. Image Annotations from Drone-Captured Data. (A) Original image: captured by a DJI Matrice 300 RTK drone equipped with a Zenmuse P1 sensor, (B) single-tree label: the original image annotated using the graphical annotation tool Labelme (Version 5.3.1), (C) colored leaves label: individual colored *S. tzumu* leaves annotated using a custom MATLAB script.

2.3 Data set annotation

Annotation of the dataset involved two distinct tasks: delineating individual *S. tzumu* trees and marking the colored leaves of individual *S. tzumu* specimens. For tree annotation, we used the graphical annotation tool Labelme (version 5.3.1) [56] to draw bounding boxes, ensuring that each *S. tzumu* tree was completely enclosed within them. This process resulted in the annotation of 1420 original images and creation of 20,591 bounding boxes, as indicated in red in Fig. 3 (B). For leaf annotation, we used a custom MATLAB script to locate both the colored leaves and those on the cusp of the color transition. A total of 310 images were subjected to this process, resulting in 59,530 point annotations (shown as red dots in Fig. 3. (C)).

Branch shading poses a significant challenge for accurate data annotation in parts of *S. tzumu* plantations where tree health is declining. In addition, the inherent subjectivity

of human annotators and the limitations of existing annotation tools can contribute to mislabeling and omissions. To mitigate these issues, we implemented measures to improve the annotation accuracy. First, we repeatedly checked the labels against the *S. tzumu* family diagrams, which effectively reduced the incidence of mislabeling individual *S. tzumu* trees. Second, we reviewed the images of individual *S. tzumu* trees extracted by the model and examined the orientation of branches and leaves to verify the association of the correct *S. tzumu* tree, thereby reducing the mislabeling of colored leaves. Additionally, we implemented a quality control process in which a second annotator reviewed the annotations to check for errors or omissions. This iterative process allowed us to identify and correct errors, thus further improving the overall accuracy of the dataset. Although these strategies do not guarantee perfect accuracy, they significantly reduce the labeling error rate, resulting in a more reliable dataset for subsequent studies.

2.4 Individual *S. tzumu* tree extraction network: Convolutional Block Attention Module (CBAM)-YOLOX

We employed an enhanced YOLOX algorithm to detect and extract individual *S. tzumu* tree data. Based on the YOLOv3 framework [57], the algorithm comprises input, backbone network, neck network, and prediction stages. At the input stage, the integration of the mosaic and mixed-up data augmentation techniques enhances the diversity and complexity of the dataset, thus improving the robustness of the model.

The backbone network utilizes CSPDarknet to extract image features [58]. Considering potential occlusion among *S. tzumu* trees, we focused not only on the model's detection performance but also on the precise localization of the actual position of the trees. To this end, the CBAM module was incorporated into the output of the three feature layers using the backbone network, further enhancing the accuracy of tree detection and extraction.

The neck network utilizes a Feature Pyramid Network (FPN) and a Path Aggregation Network (PAN) [59]; the FPN propagates high-level semantic features from top to bottom, enriching the semantic content at all levels, whereas the PAN enhances precise localization features from the bottom up. Additionally, we replaced the traditional coupled detection heads with decoupled detection heads, thus significantly improving the speed and accuracy of detecting isolated and overlapping *S. tzumu* trees.

During the prediction phase, traditional predefined anchor boxes were abandoned, which substantially reduced the computational complexity and enhanced the efficiency of tree detection. Through these innovations and optimizations, our network structure (as shown in Fig.4) successfully achieved precise extraction of individual *S. tzumu* trees, thus fully validating the effectiveness of the enhanced YOLOX algorithm.

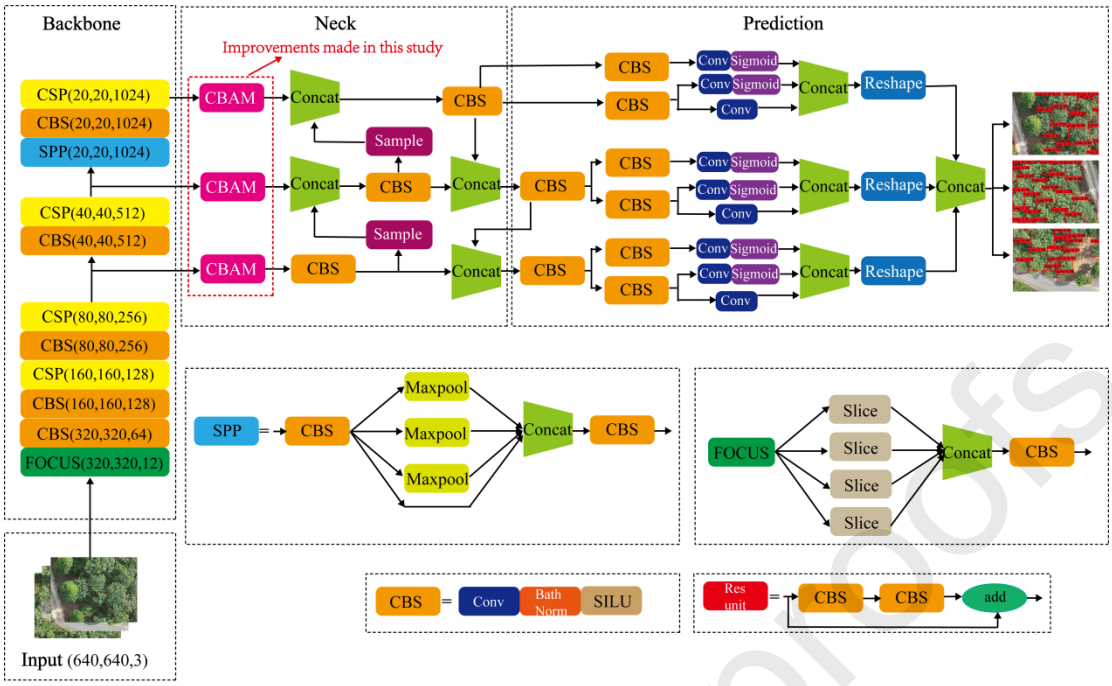


Fig. 4. Schematic Diagram of the YOLOX-CBAM Target Detection Network. CBS: Convolution, Batch normalization, and SILU (Swish-like activation function); CSP: Cross Stage Partial; SPP: Spatial Pyramid Pooling; CBAM: Convolutional Block Attention Module; Concat: Concatenation; Sigmoid: Sigmoid Activation Function; Max Pooling (Maxpool); Conv: Convolution; Batch Norm: Batch Normalization; Res: Residual Block.

2.5 CBAM attention mechanism

CBAM [60] consists of two complementary parts: a Channel Attention Module (CAM) and a Spatial Attention Module (SAM). The CAM begins by performing global max pooling and global average pooling on the input feature map and summarizing the feature information from different perspectives. The results of these pooling operations are then fed into a shared multilayer perceptron (MLP), whose output is activated by a sigmoid function to form the channel attention map M_c. This map weighs the importance of each channel to amplify useful information and suppress irrelevant information, thereby enhancing the model's ability to recognize critical details. The channel attention can be expressed using the following mathematical formula:

$$M_C(F) = \sigma((MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

$$F' = M_C(F) \otimes F \quad (2)$$

In equations (1) and (2), F represents the input feature map, AvgPool denotes the average pooling operation, MaxPool refers to the maximum pooling operation, MLP is

the multilayer perceptron, $M_c(F)$ is the channel attention weight factor, and F′ is the channel-attended feature map.

Next, the SAM takes the channel-attended feature map F′ as input and performs global maximum pooling and global average pooling on it. The pooled feature maps were then further processed using a 7 × 7 pixel convolutional layer. The output of this process is the spatial attention map M_s, which weighs the importance of each spatial location, emphasizes areas relevant to the target, and reduces the influence of irrelevant areas. The mathematical expression for spatial attention is

$$M_S(F) = \sigma f(f^{7\times7}([AvgPool(F)));(MaxPool(F)])) = \sigma(f^{7\times7}([F^S_{avg};F^S_{\max}])) \quad (3)$$

$$F'' = M_S(F') \otimes F' \quad (4)$$

In equations (3) and (4), F′ is the channel-attended feature map, AvgPool represents the average pooling operation, MaxPool denotes the maximum pooling operation, $f^{7\times7}$ is the 7×7 convolution operation, s is the sigmoid activation function, M_s (F) is the spatial attention weight coefficient, and F″ is the final mixed-domain attention feature map.

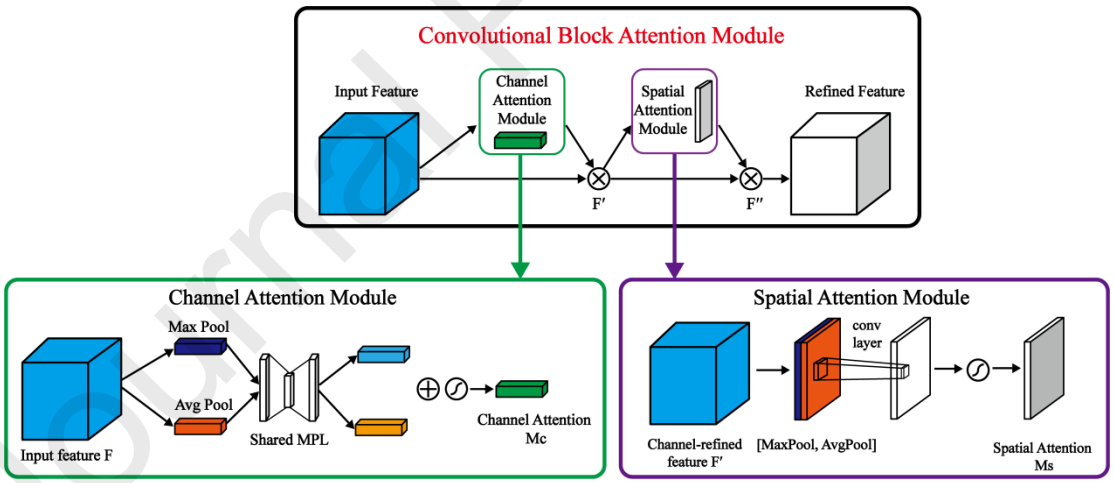


Fig. 5. CBAM Structure. (Channel Attention Module and Spatial Attention Module).

2.6 Colored leaf extraction network of Sassafras-net

We applied the crowd-counting model Sassafras-net to accurately enumerate the colored leaves of *S. tzumu* trees. Although convolutional neural networks (CNN) are a convolutional approach for crowd counting, their limited receptive field poses challenges in capturing global features, often necessitating additional mechanisms such

as attention layers, which complicates the model architecture [61]. To address these limitations, we developed Sassafras-net, a model specifically designed for precise quantification of colored leaves on individual *S. tzumu* trees. The model leverages the CCTrans network [50], incorporates a visual attention mechanism through a Vision Transformer [62], and utilizes a Twins architecture as its backbone [51]. By harnessing these advanced global context modeling capabilities, Sassafras-net not only simplifies the prediction process but also enhances the performance in densely populated leaf scenarios, offering significant improvements in the vision domain.

The architecture of the model illustrated in Fig. 6 is initiated by partitioning the input image into fixed-size segments and converting them into a one-dimensional sequence of vectors. Subsequent processing involves the extraction of global features from this sequence using an FPN, which transforms a one-dimensional sequence into a two-dimensional feature map for effective pyramid feature aggregation. Each resulting feature map is then up-sampled to a uniform resolution. To determine the final output, a straightforward multiscale regression header was employed that regressed the results across various scales. The total number of colored leaves was estimated in a weakly supervised manner, whereas the corresponding density maps of colored leaves were generated under a fully supervised framework, ensuring detailed and precise quantification.

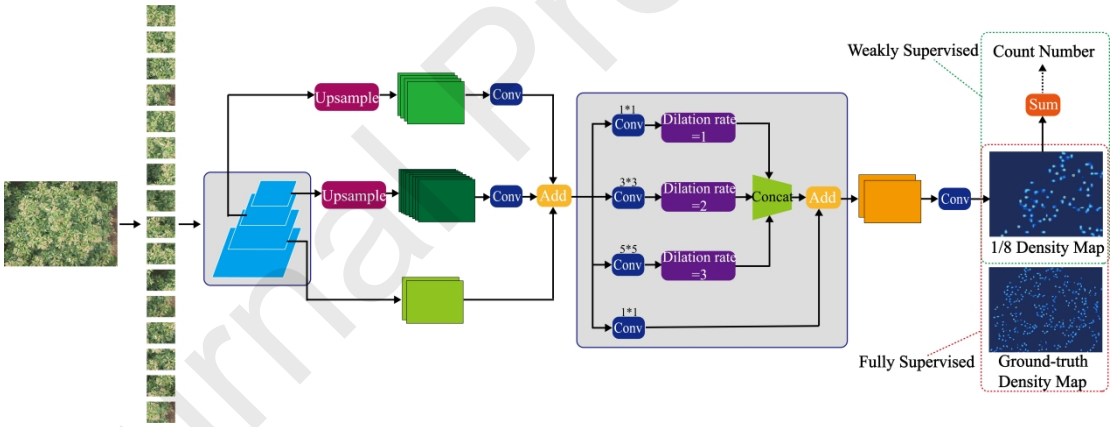


Fig. 6. Architecture of the CCTrans Network.

2.7 Programming environment

Python was used as the primary programming language throughout this project. The code editor PyCharm 2023.1.4 played a pivotal role in facilitating model training, prediction, and density map generation. The Integrated Development Environment Anaconda3 streamlined the installation of packages and management of the environment. A consistent programming environment was maintained across all algorithms using Python version 3.8.17, PyTorch 1.7.1, and CUDA 11.0. The computational resources were robust, featuring a system equipped with 128 GB of

RAM, powered by a 13th Generation Intel (R) Core (TM) i9-13900KF CPU, and an NVIDIA GeForce RTX 4090 graphics card with 24 GB of RAM.

2.8 Training data set

In addition to utilizing YOLOX weights for the single *S. tzumu* tree detection and extraction model, a comparative analysis was conducted with other notable models to benchmark their performance. Specifically, the analyses included YOLOv8 [63], YOLOv7 [64], YOLOv5 [65], and Faster-Convolutional Neural Networks for Regions (Faster-RCNN) [66]. The YOLOX model was trained with an input size of 640 × 640 pixels over 500 epochs and a batch size of 16, with a learning rate of 0.01. A weight decay of 0.0005 was implemented to prevent overfitting. The Twins-large architecture was employed as the backbone network for the detection and counting of colored leaves. It underwent an extended training period of 2000 epochs with a learning rate of 0.0001, enhancing the detection precision through prolonged adaptation to the dataset.

2.9 Evaluation indicators for the model

The performance of the model in extracting single *S. tzumu* trees was evaluated using several metrics, including Precision (P), Recall (R), F1-Score, Average Precision (AP), mean Average Precision (mAP), and Intersection over Union (IoU). Precision measures the proportion of correctly predicted positive samples out of all samples predicted as positive and is calculated as follows:

$$P = \frac{TP}{TP+FP} \quad (5)$$

R assesses the proportion of actual positive samples correctly identified by the model:

$$R = \frac{TP}{TP+FN} \quad (6)$$

The F1-Score, the harmonic mean of P and R, provides a balanced assessment of the model's performance, especially in cases of uneven class distribution:

$$\text{F1-Score} = 2 \times \frac{P \times R}{P+R} \quad (7)$$

The AP was obtained by integrating the P–R curve over all R values, offering a comprehensive evaluation of the model's ability to maintain high P across different R levels:

$$AP = \int_0^1 P(R)dR \quad (8)$$

As only one category (*S. tzumu*) was identified in this study (i=1, *N*=1), the mAP is equivalent to AP:

$$\text{mAP} = \frac{\sum_{i=1}^{N} AP_i}{N} \quad (9)$$

The IoU metric was used to assess the localization accuracy of the model. IoU quantifies the overlap between the predicted bounding box (A) and the ground truth bounding box (B) by calculating the ratio of the area of their intersection to the area of their union:

$$IoU = \frac{A \cap B}{A \cup B} \quad (10)$$

where true positives is the number of *S. tzumu* trees correctly identified by the model. False positives (FPs) are instances where non-*S. tzumu* elements were incorrectly identified as *S. tzumu*. False negatives were *S. tzumu* trees that the model failed to detect. *N* is the number of categories detected. Here, *N*=1 because only *S. tzumu* was considered.

By employing this suite of metrics—P, R, F1-Score, AP/mAP, and IoU—we conducted a comprehensive analysis and effective quantification of the model's efficacy in the task of single *S. tzumu* tree extraction. P and R measure the classification performance of the model, while the F1-Score balances these two metrics to provide a single performance indicator. The AP and mAP offer insights into the model's performance across different threshold settings, and IoU assesses the accuracy of the predicted bounding boxes in relation to the ground truth. Collectively, these metrics gauged both the robustness and accuracy of the model in detecting and localizing *S. tzumu* trees in the UAV images.

To evaluate the colored *S. tzumu* leaf identification and counting model, the Mean Square Error (MSE) and Mean Absolute Error (MAE) were used as evaluation metrics. MSE quantifies the closeness of the model's predicted values to the actual values, highlighting the accuracy of the model, whereas MAE measures the average deviation between the predicted and true values, reflecting the precision of the model. Additionally, the coefficient of determination, R^2, was used to assess the fit of the model to the ground truth data. Collectively, these metrics provided a robust framework for evaluating the performance and precision of the model.

$$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - G_i)^2} \quad (11)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - G_i| \quad (12)$$

$$R^2 = 1 - \frac{\sum_1^n |Y_i - G_i|^2}{\sum_1^n |Y_i - \overline{G}|^2} \quad (13)$$

In this formulae, n represents the total number of images in the test set, Y_i denotes the predicted number of colored leaves for the i_{th} image, and G_i represents the actual (ground truth) number of colored leaves in the i_{th} image (i≤n). The term $\overline{G}$ denotes the mean of the ground-truth values across all images in the test set.

3 Results

3.1 Identification and extraction of individual *S. tzumu* trees

We compared the detection performances of six target detection models using identical raw data and training parameters. Table 1 presents the performance metrics for individual *S. tzumu* tree extraction at a detection threshold of 0.5. Among the evaluated models, the improved YOLOX-CBAM model performed well, with a performance improvement of 3.4% compared with the YOLOX model, achieving P, R, and F1-scores above 93.8%. This performance surpassed those of YOLOv8 and YOLOv7, which displayed marginally lower metrics. When the threshold was 0.75, YOLOX-CBAM continued to lead with the highest performance metrics at approximately 70.5%, followed by YOLOv7 and YOLOX. In contrast, R of Faster-RCNN dropped significantly to 44.4%. The AP of the six models indicated that YOLOX-CBAM achieved the highest AP at both thresholds of 0.5 and 0.75, reaching 96.3% and 70.5%, respectively. Fig. 7 illustrates the IoU values for these models, and predicted boxes of YOLOX-CBAM demonstrate the closest alignment with the true-labeled box, resulting in the highest IoU value of 0.955.

Table 1. Comparative Detection Performance of the Various Models.

Model	Threshold=0.5				Threshold=0.75			
	P /%	R /%	F 1/%	A P/%	P /%	R /%	F 1/%	A P/%
YOLOX -CBAM	94.3 ↑	93.8 ↑	94.0 ↑	96.3 ↑	76.2 ↑	75.8 ↑	76.0 ↑	70.5 ↑

YOLOX	9	9	9	9	7	7	7	6				
	1.7	2.1	2.0	3.1	0.2	0.5	0.0	2.7				
YOLOv8	9	9	9	9	6	7	7	6				
	3.4	1.5	2.0	4.1	8.8	0.3	0.0	0.1				
YOLOv7	9	9	9	9	7	6	7	6				
	4.0	0.0	2.0	5.3	2.5	9.4	1.0	3.8				
YOLOv5	9	8	8	9	6	5	6	5				
	2.0	4.6	8.0	2.9	4.7	9.6	2.0	3.0				
Faster-RCNN	8	8	8	8	4	4	4	3				
	2.7	7.7	5.0	6.9	1.9	4.4	3.0	1.0				

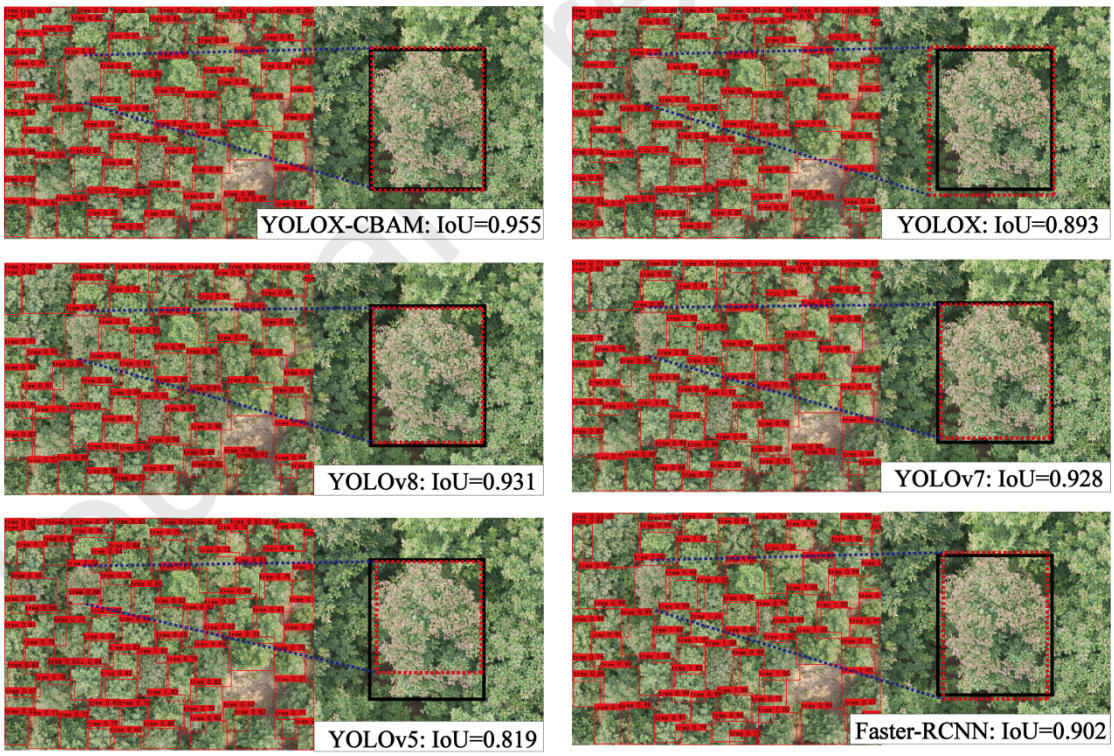


Fig. 7. Comparison of IoU Values for Six Models in Identifying an Individual *S. tzumu* Tree. (red dashed lines represent the predicted boxes by the models, and black solid lines represent the true location of the *S. tzumu* tree. A higher IoU value indicates greater precision in the model's prediction of the tree's true location).

Fig. 8 provides a comparative analysis of the detection capabilities of the six models for identifying individual *S. tzumu* trees. False detections and missed targets are highlighted by yellow and blue circles, respectively. The findings illustrate that while models such as YOLOX, YOLOv8, YOLOv7, YOLOv5, and Faster R-CNN exhibited varying levels of FPs and omissions in complex settings, the enhanced YOLOX-CBAM model considerably outperformed its counterparts. Considering the density and complexity of *S. tzumu* forests, the YOLOX-CBAM model proved to be the most effective solution for this particular task.

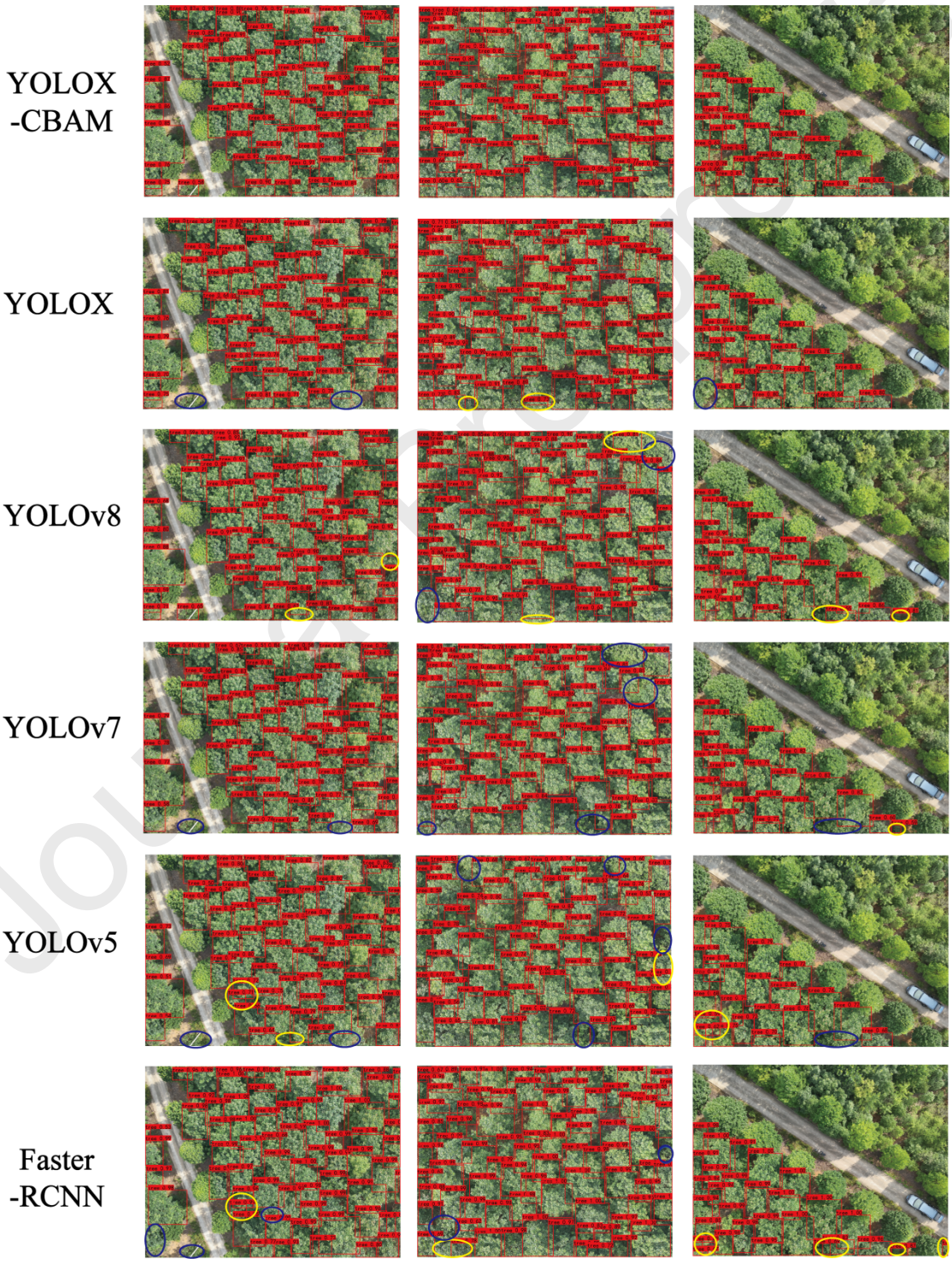


Fig. 8. Detection Outcomes of Six Models for Identifying Individual *S. tzumu* Trees. (Yellow circles indicate false detections, and blue circles denote missed detections.).

3.2 Detection and counting of colored leaves

A plot generated by the Sassafras-net model, depicted in Fig. 9, illustrates the density of colored leaves on a single *S. tzumu* tree. In this visualization, the locations and counts of the colored leaves predicted by the model are represented by light-colored dots against a complex background and shaded in dark blue. The model's prediction accuracy was demonstrated to be exceptional, with an MAE of 27.29 and an MSE of 39.00. These metrics signify a notable reduction in both the MAE and MSE compared with the baseline model (as detailed in Table 2). Additionally, Table 3 presents the Sassafras-net prediction results for 20 randomly selected images of a single *S. tzumu* tree, showing high accuracy and robust generalization capability across diverse leaf count scenarios with an average error detection rate of 4.73%. Fig. 10 illustrates the confusion matrix for the model's colored leaf detection of three random *S. tzumu* trees.

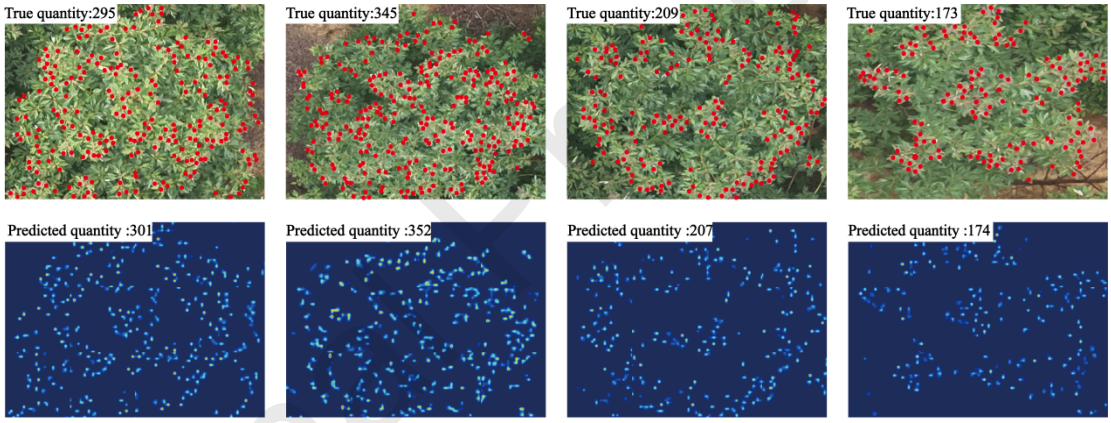


Fig. 9. Comparison of Actual and Predicted Colored Leaves Quantities by the Sassafras-net Model. (Red points denote the actual quantity and location of colored leaves; light-colored points represent the model-predicted quantity and location of colored leaves).

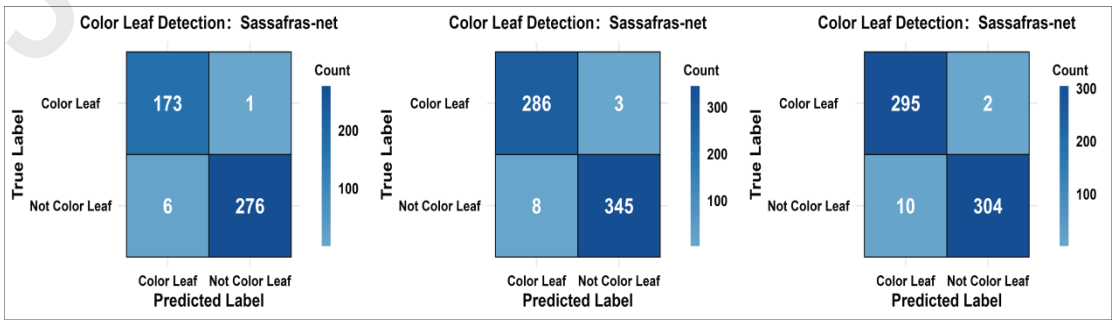


Fig. 10. Comparison of Colored Leaf Detection Performance by the Sassafras-net

Model.

Table 2. Comparative Analysis of MAE and MSE: Original CCTrans Model vs. Sassafras-net Model.

Model	MAE	MSE
CCTrans	52.30	84.90
Sassafras-net	27.29	39.00

Table 3. Comparison of Actual vs. Predicted Colored Leaf Counts by the Sassafras-net Model Across 20 Randomly Selected Images of a Single *S. tzumu* Tree.

Image ID	Actual number	Correct detection	Image ID	Actual number	Correct detection
1	301	294	11	169	172
2	192	199	12	165	172
3	99	105	13	176	166
4	133	143	14	295	301
5	185	196	15	184	176
6	345	352	16	61	58
7	247	252	17	147	136
8	153	148	18	116	122
9	209	207	19	228	232

10	209	226	20	173	174

3.3 Projected results

R^2 values were calculated for each model to quantify the accuracy of the respective leaf counts. As shown in Fig. 11, for single *S. tzumu* tree extraction, the YOLOX-CBAM model had the highest correlation with the actual ground truth values, with an R^2 value of 0.994. By contrast, the YOLOX, YOLOv8, YOLOv7, and YOLOv5, and the Faster R-CNN models were less accurate because of the different degrees of omissions and misdetections, with R^2 values of 0.989, 0.983, 0.984, 0.973, and 0.961, respectively.

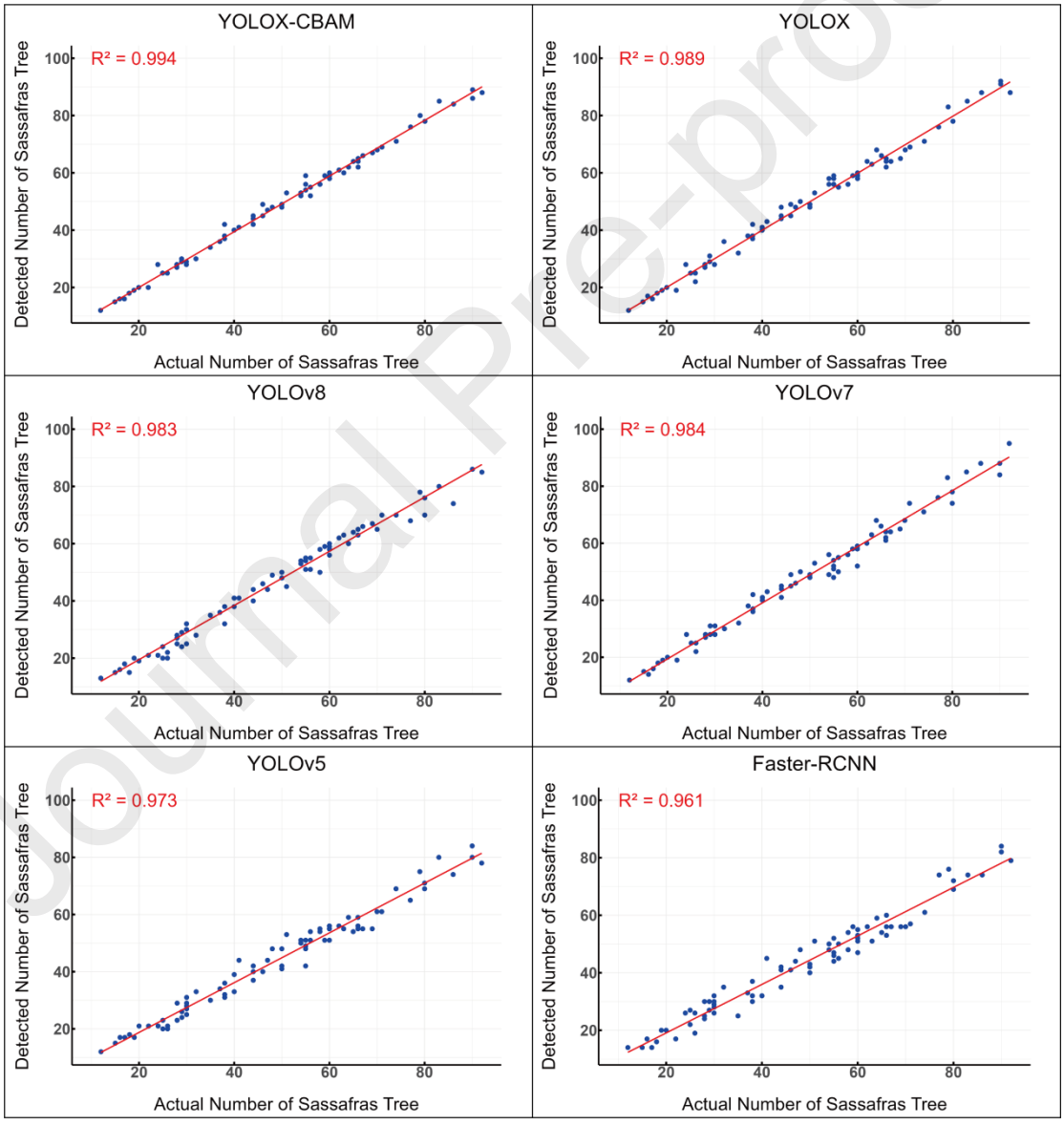


Fig. 11. Scatter Plot of True vs. Predicted Values by Multiple Models on the Test Set. (X-axis: true value, Y-axis: predicted value).

As shown in Fig. 12, the accuracy of the Sassafras-net model for counting single colored *S. tzumu* leaves decreased as the number of predicted leaves increased. Despite this, the model maintained high accuracy with an R^2 value of 0.964.

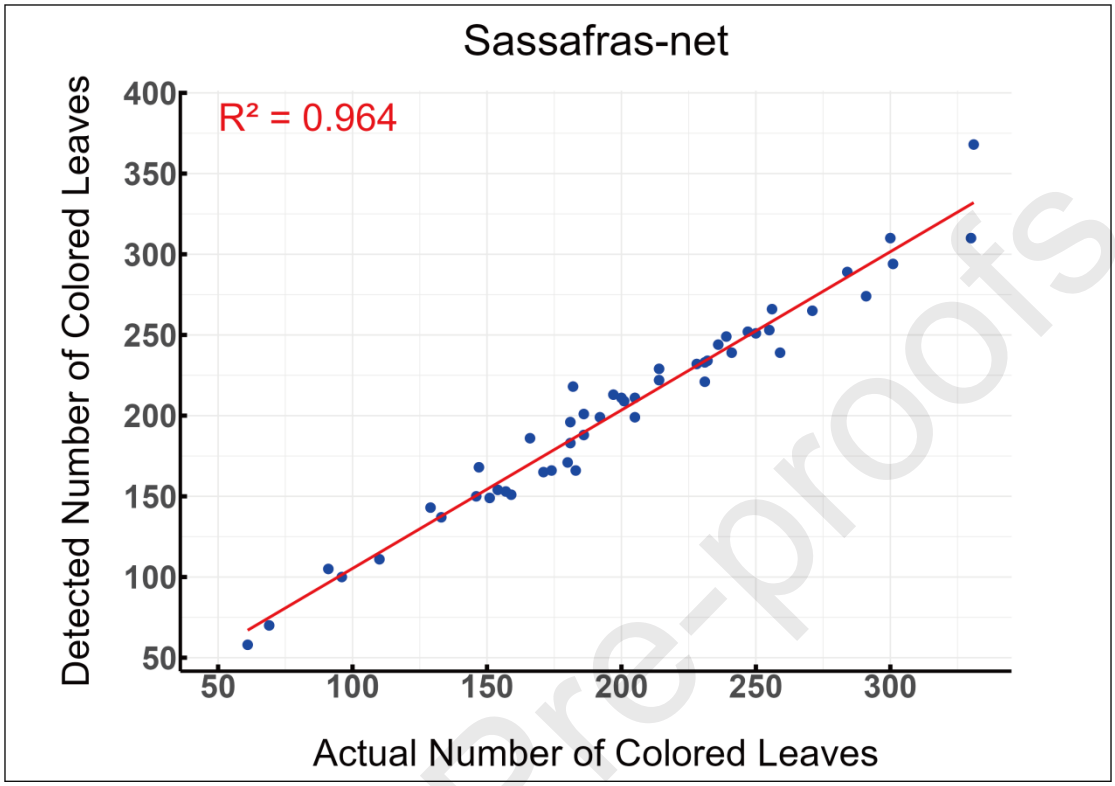


Fig. 12. Detection Results of Sassafras-net on the Test Set. (X-axis: true value, Y-axis: predicted value).

4 Discussion

The accurate quantification of phenotypic traits, such as the number of colored leaves on forest trees, is critical for advancing smart forestry. This not only improves our understanding of plant health and vigor but also has important ecological implications [67]. Colored leaves, a characteristic of *S. tzumu*, provide esthetic value and play a central role in ecological assessments, helping identify and classify tree species in different habitats [4]. Recent studies have highlighted the efficacy of using DL techniques alongside RGB images from UAVs for counting tasks, a method heavily influenced by advances in crowd-counting methods [68-71].

In the present study, we used a UAV equipped with RGB cameras alongside DL models to quantify colored leaves on individual *S. tzumu* trees and compared this approach with traditional ground-based measurements and satellite remote sensing. We found that ground-based measurements, while highly accurate, are inefficient for large-scale applications owing to significant labor and time costs, especially in inaccessible forest areas. By contrast, satellite remote sensing can cover large areas and is suitable for long-

term ecological monitoring; however, its resolution limitations prevent it from capturing minute phenotypic variations, such as changes in individual leaves. UAV technology offers greater flexibility and precision, providing high-resolution imagery that is particularly suited for accurate measurements of phenotypic traits in complex forest environments [72-75].

For example, Xu et al. [76] used YOLOv5 to detect fully expanded maize leaves with a P of 92.0%, R of 84.4%, and AP of 89.6%, which were lower than our results for colored leaves on individual *S. tzumu* trees with a P of 94.3%, R of 93.8%, and AP of 96.3%. Similarly, Zhang et al. [77] achieved an F-score of > 90% when counting canola leaves, whereas our study achieved an F-score of 94.0%. Importantly, current plant-counting methods primarily focus on large-scale crop assessments in smart agriculture, with few studies dedicated to the phenotypic traits of individual trees, especially with regard to the number of colored leaves. These traits are critical for assessing the health and growth performance of specific tree species, as changes in leaf color often indicate physiological responses or adaptability to environmental changes [78, 79]. In forestry, the ability to accurately count and analyze colored leaves on individual trees provides deeper insights into the biodiversity and functionality of forest ecosystems [80].

We have successfully developed two models focused on quantifying colored leaves on individual *S. tzumu* trees: one for extracting individual trees from UAV-collected RGB images and another called "Sassafras-net" for counting colored leaves on these extracted trees. This approach represents an advancement of current technological methods.

Initially, the YOLOX-CBAM network was selected for the single-tree extraction model because of its impressive performance metrics, achieving the highest P, R, and AP values among the evaluated models. This model aligns with the latest enhancements in the YOLO series, which have transitioned to an anchor-free detection method, improving the detection efficacy across various model scales [81, 82]. The success of the YOLOX model in this study reflects its broader applicability in diverse object detection tasks, renowned for its precision and speed, particularly in complex and dynamic environments, such as forestry and agriculture [52, 83, 84].

In the subsequent phase, we developed the "Sassafras-net" model based on the CCTrans framework specifically for counting colored leaves. This CCTrans-based network demonstrated considerable improvements over traditional counting methods, such as MCNN, CSRNet, and DM-Count, as evidenced by significantly lower MAE and MSE values. The accuracy of the Sassafras-net model surpasses that of the YOLOv5 model used for maize leaves counting [76] and models for rice leaf disease detection [85], highlighting the potential of advanced DL models for enhancing phenotypic trait analysis in smart forestry.

Although our study highlights the efficacy of UAV-based RGB imaging and DL models for phenotypic trait extraction, we must acknowledge the potential biases inherent in

these methods. In complex forest environments, UAV imaging is affected by varying lighting conditions, shadow effects, and overlapping canopies [86]. These factors can lead to uneven image quality, which affects the accuracy of the tree and leaf detection models. To minimize these imagery biases, we acquired images under consistent lighting conditions and at slower flight speeds whenever possible. In addition, we screened the original flight images to remove unclear or repetitive images, thereby improving the uniformity of our dataset.

Building on the CCTrans framework, our work leverages a transformative approach to crowd counting by applying it to phenotypic trait analyses in forestry. The capacity of CCTrans to capture detailed global context information through a pyramid vision transformer backbone and a pyramid feature aggregation model sets it apart from conventional methods [50], offering a sophisticated tool for forestry applications. The performance of the Sassafras-net model represents not only an improvement but also a significant leap in computational efficiency. Unlike conventional CNN approaches that often grapple with context modeling and are constrained by local receptive fields, CCTrans integrates local and global contexts to form a more comprehensive understanding of an image [87]. This integration enables the Sassafras-net model to significantly outperform the existing models. Our findings highlight the potential of the Sassafras-net model as a pivotal tool in smart forestry, facilitating accurate and detailed monitoring of phenotypic traits. This advancement opens up avenues for exploring new applications in forestry research, ranging from genetic diversity studies to the development of new species with desirable traits such as optimal leaf coloration.

The quantification of colored leaves on *S. tzumu* trees, while esthetically and ecologically important, also plays a vital role in understanding plant health and adaptive behaviors across varied habitats. The exploration of *S. tzumu* through advanced DL techniques and UAV RGB imagery not only deepens our understanding of this species but also provides valuable insights into the broader field of phenotypic trait analysis in forestry.

5 Limitations and future work

Despite these significant advances, it is important to recognize the inherent limitations of our model. The prevalence of declining *S. tzumu* forest regions, compounded by the physical obstruction of branches, often leads to the overestimation of colored leaf counts. In addition, the broad and large *S. tzumu* leaves may have obscured the smaller underlying colored leaves, hampering their detection by the UAV and subsequent prediction by the model. Nevertheless, the accuracy of the Sassafras-net model in counting colored leaves remains largely unaffected by these challenges, underscoring its value for accurate phenotypic quantification despite the noted limitations.

The development of the Sassafras-net model required careful labeling and training, focusing exclusively on the colored leaves of individual *S. tzumu* trees. During the

predictive phase, the model demonstrated its ability to recognize colored leaves, as evidenced by its ability to include both labeled and previously unlabeled leaves in the overall count. This consistent production of accurate leaf counts, even with the expansion of the dataset and the ongoing refinement of the model, suggests promising directions for future research. Future iterations of this model may incorporate advanced image-processing techniques or alternative data collection methods to overcome these challenges. Additionally, exploring the integration of multispectral imagery may provide a solution to the challenge of detecting obscured foliage, thereby enhancing the utility of the model. Moreover, although the Sassafras-net model has demonstrated high accuracy in quantifying colored leaves of *S. tzumu*, its applicability to other tree species and different forest types has not been tested. This model was developed and trained specifically for *S. tzumu* in a particular ecological setting, which may limit its generalizability to other species or environments with different characteristics. Future work should include pilot studies on various forest types and ecological regions to validate the practical utility of the model in diverse settings. In addition, exploring the adaptation of the model to multispecies tree identification tasks in complex forest environments may enhance its versatility and contribute to broader applications in forestry research and management.

6 Conclusion

This study established a robust methodology for the precise quantification of colored leaves of *S. tzumu* trees by developing the Sassafras-net model, which integrates an improved YOLOX network architecture. This model provides a solid foundation for advancing research on the phenotypic traits associated with foliar coloration and supports the cultivation and utilization of ornamental leaf varieties. The Sassafras-net model represents not only a technological advancement in colored leaf quantification but also a significant methodological leap in smart forestry. The ability to accurately count colored leaves on *S. tzumu* trees provides a new standard for phenotypic trait analysis and paves the way for exploring novel applications in genetic diversity studies and the development of species with optimal leaf coloration.

Data Availability

The data mentioned in this paper are available on request from the corresponding author.

Authors' contributions

Qingwei Meng conducted the experiment and wrote the manuscript. **Yanjie Li** designed the study, supported the data collection and field experiments and performed revisions of the manuscript. **Wei Qi Yan, Zhaoxu Zhang and Xia Hao** supervised experiments and performed revisions of the manuscript. **Cong Xu**, **Hui Chen** and **Wei Liu** supervised experiments and performed revisions of the manuscript, all authors read and approved the final manuscript.

Conflicts of interest

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

Acknowledgments

This work was funded by Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (2021C02070-7-3) and the Fundamental Research Funds of CAF, No. CAFYBB2022QA001.

References

[1] S. Wang, et al., Genetic diversity and population structure of an arctic tertiary relict tree endemic to China (Sassafras tzumu) revealed by novel nuclear microsatellite (nSSR) markers, Plants 11(20) (2022) 2706.

[2] H. Zhao, et al., Effects of cadmium stress on growth and physiological characteristics of sassafras seedlings, Scientific reports 11(1) (2021) 9913.

[3] K.F. Chung, et al., Isolation and characterization of microsatellite loci in Sassafras randaiense (Lauraceae), American journal of botany 98(11) (2011) e326-e329.

[4] K. Zhang, et al., Species distribution modeling of Sassafras tzumu and implications for forest management, Sustainability 12(10) (2020) 4132.

[5] A.G. De Soyza, D.T. Kincaid, Patterns in leaf morphology and photosynthesis in shoots of Sassafras albidum (Lauraceae), American Journal of Botany 78(1) (1991) 89-98.

[6] J. Sun, et al., Composition and environmental interpretation of the communities of Sassafras tzumu, a protected species, at Zhejiang province in eastern China, Global Ecology and Conservation 24 (2020) e01218.

[7] X. Zhou, Comparative analysis of the effect of wood mixed with wood-load afforestation, For. By-Prod. Spec. China 4 (2022) 22-24.

[8] Y. Li, et al., Spectroscopic determination of leaf chlorophyll content and color for genetic selection on Sassafras tzumu, Plant methods 15 (2019) 1-11.

[9] R.C. Kuijken, et al., Root phenotyping: from component trait in the lab to breeding, Journal of experimental botany 66(18) (2015) 5389-5401.

[10] S. Kisvarga, et al., Plant responses to global climate change and urbanization: Implications for sustainable urban landscapes, Horticulturae 9(9) (2023) 1051.

[11] X. Hu, et al., Research progress in the interconversion, turnover and degradation of chlorophyll, Cells 10(11) (2021) 3134.

[12] V.S.d. SILVA, LiDAR-derived methods for volume estimation and individual tree detection in Eucalyptus spp. plantations, (2019).

[13] X. Liang, et al., Close-range remote sensing of forests: The state of the art, challenges, and opportunities for systems and data acquisitions, IEEE geoscience and remote sensing magazine 10(3) (2022) 32-71.

[14] J. Ubbens, et al., The use of plant models in deep learning: an application to leaf counting in rosette plants, Plant methods 14 (2018) 1-10.

[15] A.P. Dalla Corte, et al., Measuring individual tree diameter and height using GatorEye High-Density UAV-Lidar in an integrated crop-livestock-forest system, Remote Sensing 12(5) (2020) 863.

[16] F. López-Granados, et al., An efficient RGB-UAV-based platform for field almond tree phenotyping: 3-D architecture and flowering traits, Plant Methods 15(1) (2019) 1-16.

[17] M. Miraki, et al., Individual tree crown delineation from high-resolution UAV images in broadleaf forest, Ecological Informatics 61 (2021) 101207.

[18] C. Torresan, et al., Forestry applications of UAVs in Europe: A review, International journal of remote sensing 38(8-10) (2017) 2427-2447.

[19] D.J. Kachamba, et al., Biomass estimation using 3D data from unmanned aerial vehicle imagery in a tropical woodland, Remote Sensing 8(11) (2016) 968.

[20] L. Lin, et al., UAV based estimation of forest leaf area index (LAI) through oblique photogrammetry, Remote Sensing 13(4) (2021) 803.

[21] H. Zhong, et al., Individual Tree Species Identification for Complex Coniferous and Broad-Leaved Mixed Forests Based on Deep Learning Combined with UAV LiDAR Data and RGB Images, Forests 15(2) (2024) 293.

[22] A. Duarte, et al., Recent advances in forest insect pests and diseases monitoring using UAV-based data: A systematic review, Forests 13(6) (2022) 911.

[23] S. Ecke, et al., UAV-based forest health monitoring: A systematic review, Remote Sensing 14(13) (2022) 3205.

[24] K. Otsu, et al., Calibrating the severity of forest defoliation by pine processionary moth with Landsat and UAV imagery, Sensors 18(10) (2018) 3278.

[25] S. Zhang, et al., Comparison of Semi-Physical and Empirical Models in the Estimation of Boreal Forest Leaf Area Index and Clumping With Airborne Laser Scanning Data, IEEE Transactions on Geoscience and Remote Sensing (2024).

[26] Y. Tian, et al., Aboveground mangrove biomass estimation in Beibu Gulf using

machine learning and UAV remote sensing, Science of the Total Environment 781 (2021) 146816.

[27] Z. Ye, et al., Extraction of olive crown based on UAV visible images and the U2-net deep learning model, Remote Sensing 14(6) (2022) 1523.

[28] J.A. Richards, J.A. Richards, Remote sensing digital image analysis, Springer2022.

[29] G. Yang, et al., Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives, Frontiers in plant science 8 (2017) 1111.

[30] S. Getzin, et al., Using unmanned aerial vehicles (UAV) to quantify spatial gap patterns in forests, Remote Sensing 6(8) (2014) 6988-7004.

[31] L. Bruzzone, L. Carlin, A multilevel context-based system for classification of very high spatial resolution images, IEEE transactions on Geoscience and Remote Sensing 44(9) (2006) 2587-2600.

[32] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Transactions on geoscience and remote sensing 42(8) (2004) 1778-1790.

[33] M. Pal, P.M. Mather, Support vector machines for classification in remote sensing, International journal of remote sensing 26(5) (2005) 1007-1011.

[34] M. Pal, Random forest classifier for remote sensing classification, International journal of remote sensing 26(1) (2005) 217-222.

[35] S.D. Kumar, et al., Design of disease prediction method based on whale optimization employed artificial neural network in tomato fruits, Materials Today: Proceedings 33 (2020) 4907-4918.

[36] B.S. Laurindo, et al., Optimization of the number of evaluations for early blight disease in tomato accessions using artificial neural networks, Scientia horticulturae 218 (2017) 171-176.

[37] C. Li, et al., Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection, Sensors and Actuators B: Chemical 125(1) (2007) 301-310.

[38] B. Schölkopf, Learning with kernels: support vector machines, regularization, optimization, and beyond, The MIT Press, 2002.

[39] Y. LeCun, et al., Deep learning, nature 521(7553) (2015) 436-444.

[40] M.P. Pound, et al., Deep machine learning provides state-of-the-art performance in image-based plant phenotyping, Gigascience 6(10) (2017) gix083.

[41] S. Taghavi Namin, et al., Deep phenotyping: deep learning for temporal phenotype/genotype classification, Plant methods 14 (2018) 1-14.

[42] Y. Jiang, C. Li, Convolutional neural networks for image-based high-throughput plant phenotyping: a review, Plant Phenomics (2020).

[43] A. Kamilaris, F.X. Prenafeta-Boldú, Deep learning in agriculture: A survey, Computers and electronics in agriculture 147 (2018) 70-90.

[44] O. Ronneberger, et al., U-net: Convolutional networks for biomedical image segmentation, Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234-241.

[45] F. Wu, et al., Simplifying graph convolutional networks, International conference on machine learning, PMLR, 2019, pp. 6861-6871.

[46] Y. Zhang, et al., Complete and accurate holly fruits counting using YOLOX object detection, Computers and Electronics in Agriculture 198 (2022) 107062.

[47] Y. Zhang, et al., Single-image crowd counting via multi-column convolutional neural network, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 589-597.

[48] Y. Li, et al., Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes, Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1091-1100.

[49] B. Wang, et al., Distribution matching for crowd counting, Advances in neural information processing systems 33 (2020) 1595-1607.

[50] Y. Tian, et al., Cctrans: Simplifying and improving crowd counting with transformer, arXiv preprint arXiv:2109.14483 (2021).

[51] X. Chu, et al., Twins: Revisiting the design of spatial attention in vision transformers, Advances in neural information processing systems 34 (2021) 9355-9366.

[52] J. Huang, et al., An improved YOLOX algorithm for forest insect pest detection, Computational Intelligence and Neuroscience 2022(1) (2022) 5787554.

[53] Y. Wang, et al., Recent advances in the application of deep learning methods to forestry, Wood science and technology 55(5) (2021) 1171-1202.

[54] R. Dainelli, et al., Recent advances in Unmanned Aerial Vehicles forest remote sensing—A systematic review. Part II: Research applications, Forests 12(4) (2021) 397.

[55] G.-z. CHU, The Avifauna in The Area Around the Longshan Forest Farm, Anji County, Zhejiang Province, Zoological Research 17(1) (1996) 45-51.

[56] B.C. Russell, et al., LabelMe: a database and web-based tool for image annotation, International journal of computer vision 77 (2008) 157-173.

[57] L. Zhao, S. Li, Object detection algorithm based on improved YOLOv3, Electronics 9(3) (2020) 537.

[58] P. Zhang, D. Li, CBAM+ ASFF-YOLOXs: An improved YOLOXs for guiding agronomic operation based on the identification of key growth stages of lettuce, Computers and Electronics in Agriculture 203 (2022) 107491.

[59] M. Zhu, et al., Dynamic feature pyramid networks for object detection, arXiv preprint arXiv:2012.00779 (2020).

[60] S. Woo, et al., Cbam: Convolutional block attention module, Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.

[61] J. Liang, et al., A detection approach for late-autumn shoots of litchi based on unmanned aerial vehicle (UAV) remote sensing, Computers and Electronics in Agriculture 204 (2023) 107535.

[62] A. Dosovitskiy, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929 (2020).

[63] B. Karthika, et al., Object Detection Using YOLO-V8, 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 2024, pp. 1-4.

[64] C.-Y. Wang, et al., YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 7464-7475.

[65] D. Thuan, Evolution of Yolo algorithm and Yolov5: The State-of-the-Art object detention algorithm, (2021).

[66] S. Ren, et al., Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE transactions on pattern analysis and machine intelligence 39(6) (2016) 1137-1149.

[67] M.M. Griggs, et al., Sassafras albidum (Nutt.) Nees, Burns, RM; Honkala, BH, Technical coordinators. Silvics of North America 2 (1990) 773-777.

[68] X. Bai, et al., Rice Plant Counting, Locating, and Sizing Method Based on High-Throughput UAV RGB Images, Plant Phenomics 5 (2023) 0020.

[69] Q. Wang, et al., NWPU-crowd: A large-scale benchmark for crowd counting and localization, IEEE transactions on pattern analysis and machine intelligence 43(6) (2020) 2141-2149.

[70] L. Boominathan, et al., Crowdnet: A deep convolutional network for dense crowd counting, Proceedings of the 24th ACM international conference on Multimedia, 2016, pp. 640-644.

[71] Z. Ma, et al., Bayesian loss for crowd count estimation with point supervision, Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6142-6151.

[72] M. Mohan, et al., Individual tree detection from unmanned aerial vehicle (UAV) derived canopy height model in an open canopy mixed conifer forest, Forests 8(9) (2017) 340.

[73] Y. Li, et al., Phenomic selection in slash pine multi-temporally using UAV-multispectral imagery, Frontiers in Plant Science 14 (2023) 1156430.

[74] W. Guo, et al., UAS-based plant phenotyping for research and breeding applications, Plant Phenomics (2021).

[75] Z. Song, et al., Enabling breeding selection for biomass in slash pine using UAV-based imaging, Plant Phenomics (2022).

[76] X. Xu, et al., Detection and Counting of Maize Leaves Based on Two-Stage Deep Learning with UAV-Based RGB Image, Remote Sensing 14(21) (2022) 5388.

[77] J. Zhang, et al., Rapeseed stand count estimation at leaf development stages with UAV imagery and convolutional neural networks, Frontiers in plant science 11 (2020) 617.

[78] L. Chalker - Scott, Environmental significance of anthocyanins in plant stress responses, Photochemistry and photobiology 70(1) (1999) 1-9.

[79] L. Li, et al., Physiological response and resistance of three cultivars of Acer rubrum L. to continuous drought stress, Acta Ecologica Sinica 35(6) (2015) 196-202.

[80] J.Y. Park, et al., Quantifying leaf phenology of individual trees and species in a tropical forest using unmanned aerial vehicle (UAV) images, Remote Sensing 11(13) (2019) 1534.

[81] P. Zamboni, et al., Benchmarking anchor-based and anchor-free state-of-the-art deep learning methods for individual tree detection in rgb high-resolution images, Remote Sensing 13(13) (2021) 2482.

[82] Y. Tian, et al., MD-YOLO: Multi-scale Dense YOLO for small target pest detection, Computers and Electronics in Agriculture 213 (2023) 108233.

[83] Z. Zheng, et al., Autonomous navigation method of jujube catch-and-shake harvesting robot based on convolutional neural networks, Computers and Electronics

in Agriculture 215 (2023) 108469.

[84] B. Wang, et al., Multiscale Maize Tassel Identification Based on Improved RetinaNet Model and UAV Images, Remote Sensing 15(10) (2023) 2530.

[85] H. Yang, et al., Disease Detection and Identification of Rice Leaf Based on Improved Detection Transformer, Agriculture 13(7) (2023) 1361.

[86] N. Guimarães, et al., Forestry remote sensing from unmanned aerial vehicles: A review focusing on the data, processing and potentialities, Remote Sensing 12(6) (2020) 1046.

[87] J. Li, et al., SoybeanNet: Transformer-Based Convolutional Neural Network for Soybean Pod Counting from Unmanned Aerial Vehicle (UAV) Images, arXiv preprint arXiv:2310.10861 (2023).

■ Authors' Biographies

As the first author, Meng Qingwei, male, received a Bachelor's degree in Agronomy from Inner Mongolia University for Nationalities in 2022. He is currently a second-year Master's student at the College of Horticulture and Forestry, Yangtze University. His research interests include tree genetic breeding and the intelligent extraction of tree phenotypic information. His email is 2022720926@yangtzeu.edu.cn.

Dr. Wei Qi Yan is a Professor in the Department of Computer and Information Sciences at Auckland University of Technology (AUT). His research expertise encompasses deep learning, computer vision, robotics, and multimedia computing. Dr. Yan serves as an Associate Editor for ACM Transactions on Multimedia Computing, Communications and Applications, Frontiers in Neuroscience, and Springer Nature Computer Science. He has also held the position of Editor-in-Chief for the International Journal of Digital Crime and Forensics (IJDCF) and worked as an exchange computer scientist between the Royal

Society Te Apārangi (RSNZ) and the Chinese Academy of Sciences (CAS).Dr. Yan is a guest professor at the Chinese Academy of Sciences and has been a visiting professor at both the University of Auckland in New Zealand and the National University of Singapore. In 2022, he was recognized as one of the world's top 2% most-cited scientists by Stanford University, as acknowledged by AUT. Currently, he holds the position of Chair of the ACM Multimedia Chapter of New Zealand and is a member of ACM, a senior member of IEEE, and a member of the IEEE Technical Committee. His email is weiqi.yan@aut.ac.nz.

Dr. Cong Xu is a lecturer and postdoctoral researcher at the School of Forestry, University of Canterbury. Dr. Xu is focused on utilizing remote sensing technology to evaluate forests and natural environments. Her work involves processing and analyzing LiDAR and satellite imagery to classify forest types, monitor forest dynamics and land cover changes, and estimate forest structure and biophysical variables using machine learning.Dr. Xu's research primarily concentrates on employing remote sensing methods to describe small-scale plantations in New Zealand, such as woodlots. This includes mapping, tree species classification, and estimating forest age and stand variables. She serves as the co-director of the Remote Sensing and Geospatial Analysis (RSGA) Laboratory at the School of Forestry. Her email is cong.xu@canterbury.ac.nz.

As the corresponding author, Yanjie Li is an Associate Researcher at the Research Institute of Subtropical Forestry, Chinese Academy of Forestry. His main research areas include the collection and evaluation of forest germplasm resources and forest phenomics. He currently serves as a Youth Editorial Board member for the high-starting-point new journal Plant Phenomics under the China Science and Technology Journal Excellence Action Plan, a guest editor for Frontiers in Forests and Global Changes, and a member of the Digital Forestry Professional Committee of the Zhejiang Forestry Society. In 2019, he received funding support from the National High-level Overseas Talent Program and the Overseas Returnees Serving the Country Action Plan by the Ministry of Human Resources and Social Security of China.As the first author or corresponding author, he has published over 20 papers in academic journals such as Plant Phenomics under the China Science and Technology Journal Excellence Action Plan, Industrial Crops and Products, and Forestry Research. Among these, 23 papers are indexed by SCI, with 8 in the first quartile of the Chinese Academy of Sciences (CAS). He has also published an electronic monograph titled *Basics of R Language and Spectral Modeling* and holds 2 authorized invention patents. His email is

aj7105@gmail.com.

Declaration of interests

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

■ **Data Availability**

☐ The research data is not available for sharing.

☑ Authors are required to provide the available means if the data can be shared.

If Yes, please provide your text here:The data mentioned in this paper are available on request from the corresponding author.

■ **Declaration of generative AI in scientific writing**

☑ The manuscript was not authored with the assistance of artificial intelligence (AI).

☐The utilization of artificial intelligence (AI)–generated text in an article must be disclosed.

■ **Ethics Statement**

☑Not applicable: This manuscript does not include human or animal research.

☐If this manuscript involves research on animals or humans, it is imperative to disclose all approval details.

If Yes, please provide your text here:

Author Contributions

☑It is imperative to provide clear specifications regarding the tasks undertaken by each author in the manuscript work.

Qingwei Meng conducted the experiment and wrote the manuscript. **Yanjie Li** designed the study, supported the data collection and field experiments and performed revisions of the manuscript. **Wei Qi Yan, Zhaoxu Zhang and Xia Hao** supervised experiments and performed revisions of the manuscript. **Cong Xu**, **Hui Chen** and **Wei Liu** supervised experiments and performed revisions of the manuscript, all authors read and approved the final manuscript.