

A Study on a Single-Tree Segmentation Method Based on UAV Imagery and a YOLOX Threshold Cascade

Tianliang Zhang^{1,2,a}, Zerun Liu^{1,2,b}, Jia Dong^{1,2,c}, Tang Song^{1,2,d}, Huan Mei^{1,2,e}, Zhiqiang Wang^{1,2,f,*}

¹*Institute of Applied Mathematics, Hebei Academy of Sciences, No. 46 South Youyi Street, Shijiazhuang, China*

²*Hebei Information Security Certification Technology Innovation Center, No. 46 South Youyi Street, Shijiazhuang, China*

^a*tianliangzn@126.com*, ^b*k659006002@qq.com*, ^c*dongjia851116@126.com*, ^d*867965226@qq.com*,
^e*574465982@qq.com*, ^f*wzhiq77@dingtalk.com*

**Corresponding author*

Keywords: UAV Remote Sensing; Single-Tree Segmentation; YOLOX; Watershed; Cascaded Framework

Abstract: Accurate individual tree segmentation is a core technical prerequisite for forest resource surveys, stand structure analysis, carbon stock estimation, and smart forestry monitoring. Traditional individual tree segmentation methods rely on manually designed features and global image processing, which are prone to over-segmentation, under-segmentation, and missed detections in scenarios with overlapping tree canopies, uneven lighting, and complex backgrounds. In contrast, pure deep learning instance segmentation models suffer from limitations such as high annotation costs, slow inference speeds, and insufficient edge accuracy. To address these challenges, this study uses planted forests in the Saihanba region of Hebei Province, China, as the research area. By constructing a dataset of high-resolution UAV imagery, we propose a single-tree segmentation framework that cascades the YOLOX object detection algorithm with a multi-stage optimized labeled watershed algorithm: First, a lightweight YOLOX model is used to perform precise single-tree detection and bounding box localization. Subsequently, within the detected boxes, grayscale enhancement, median filtering, morphological optimization, and label-guided watershed segmentation are executed to achieve fine-grained extraction of individual tree crowns. This method balances detection speed, segmentation accuracy, and practical applicability, providing lightweight, high-precision technical support for efficient surveys of planted forests, single-tree parameter inversion, and ecological monitoring. Comparisons of different IoU and confidence thresholds show that the optimal overall segmentation results are achieved when the confidence threshold is 0.1 and IoU is 0.5.

1. Introduction

Forests are the backbone of terrestrial ecosystems, performing critical functions such as carbon sequestration, oxygen release, biodiversity conservation, and timber supply. China has the world's

largest area of planted forests, making accurate, efficient, and large-scale tree-level resource monitoring an essential requirement for modern forestry management. Information on individual tree location, crown extent, and outline serves as foundational data for stand density estimation, growth monitoring, pest and disease early warning, and carbon sink quantification. Traditional manual surveys are inefficient, costly, time-consuming, and highly subjective, making it difficult to meet the demands of refined, dynamic management. Drone-based remote sensing, with its advantages of high spatiotemporal resolution, low cost, mobility, flexibility, and ease of operation, has become the mainstream technical approach for forest monitoring at small to medium scales. Individual tree segmentation based on high-resolution UAV imagery enables rapid acquisition of individual tree crown boundaries and spatial distribution, significantly improving the efficiency and accuracy of forest surveys. However, actual forest stands commonly present challenges such as intertwined and overlapping tree crowns, interference from light and shadows, wide variations in crown size, and complex understory backgrounds. These issues make it difficult for traditional image processing and single deep learning methods to simultaneously achieve robustness, accuracy, and efficiency.

Currently, methods for individual tree segmentation based on UAV imagery are primarily divided into two categories: traditional image processing methods and deep learning methods. Traditional methods are represented by the local maximum method, clustering algorithms, and the labeled watershed algorithm^[1,2]. The local maximum method locates individual trees by extracting crown vertices, but it is prone to over-segmentation, where a single crown may be represented by multiple vertices; clustering algorithms partition regions based on pixel similarity, but they are sensitive to noise and rely on empirical parameter settings; the labeled watershed algorithm can extract continuous crown boundaries, but it is susceptible to image noise and background interference, leading to prominent over-segmentation and under-segmentation issues^[3-5]. Deep learning methods, leveraging their powerful feature extraction capabilities, have demonstrated significant advantages in single-tree detection and segmentation tasks^[6]. In single-tree detection, object detection models such as the YOLO series and Faster R-CNN have been widely adopted, enabling rapid extraction of single-tree locations and bounding boxes^[7,8]. Among these, YOLOX, as an improved model in the YOLO series, adopts an anchor-free design, decouples the detection head from the SimOTA label matching strategy, and performs excellently in small-object detection and resistance to complex background interference; however, its application in forestry single-tree detection remains relatively limited^[9]. Regarding individual tree segmentation, semantic segmentation models such as U-Net and Mask R-CNN can directly extract pixel-level boundaries of tree crowns; however, they suffer from high annotation costs, complex model training, and demanding hardware requirements^[10].

To address the aforementioned issues, this study proposes a method for segmenting individual trees in UAV imagery based on YOLOX and the Multi-Stage Optimized Marker-Controlled Watershed Algorithm (MOMW). The main research activities include: constructing a dataset of high-resolution UAV imagery of planted forests, and completing data annotation, segmentation, and enhancement; developing a single-tree detection model based on the YOLOX framework and optimizing training strategies to improve the detection accuracy of trees with small crowns; proposing the Multi-Stage Optimized Marker-Controlled Watershed Algorithm (MOMW), which addresses the noise interference and over-segmentation issues of traditional watershed algorithms through global grayscale enhancement, median filtering, and optimized morphological opening operations; and systematically evaluating segmentation performance under different confidence thresholds and IoU thresholds to validate the method's adaptability across various forest stand conditions.

2. Materials and Methods

2.1. Overview of the Study Area

The study area is located in the vicinity of the Saihaba Mechanical Forest Farm within Weichang Manchu and Mongol Autonomous County, Chengde City, Hebei Province, China. It lies in the transitional zone between the Inner Mongolia Plateau and the Northern Hebei Mountains, at an elevation of 1,010–1,940 m. The study area is dominated by planted forests, with dominant tree species including Mongolian pine and larch. The trees are 5–8 years old, with relatively high stand density and significant variation in crown size. There is varying degrees of canopy overlap, and large areas of grassland and shrubbery are distributed throughout the forest. This represents a typical research setting for coniferous plantations and provides an ideal experimental site for this study.

2.2. Data Collection and Preprocessing

2.2.1. Collection Equipment

This experiment utilized a DJI Mavic 3 drone as the data acquisition platform, equipped with a visible-light imaging sensor featuring 20 million effective pixels and a 24mm equivalent focal length, capable of saving images in both JPG and RAW formats. The device employs a DJI remote controller for flight path control and real-time image transmission. It is lightweight, offers long battery life, and delivers consistent image quality, making it well-suited for large-scale field operations in forestry.

2.2.2. Collection Plan

Data collection took place in August 2025, a period when vegetation in the study area was thriving. To ensure uniform illumination of the imagery and avoid interference from backlighting and shadows, data collection was scheduled for sunny days between 10:00 a.m. and 2:00 p.m., when the sun's altitude angle exceeded 45°. For this experiment, the flight altitude was set at 30 m and the flight speed at 5 m/s, corresponding to a ground resolution of approximately 1.0 cm per pixel. The actual acquired images and their ground resolution display are shown in Figure 1.

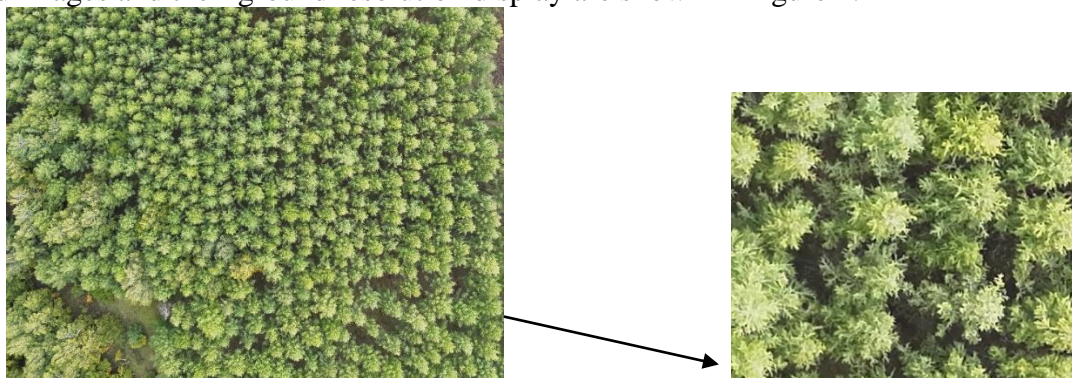


Figure 1: Single image and ground resolution display.

2.3. YOLOX-Based Single-Tree Detection Model

2.3.1. YOLOX Model Architecture

YOLOX is an anchor-free, single-stage object detection model based on YOLOv3. It achieves an excellent balance between detection accuracy and speed, making it particularly suitable for small-

object detection tasks. The model primarily consists of four components: the input layer, the backbone network, the neck feature fusion network, and the detection head. The model architecture is shown in Figure 2. The input stage employs adaptive image resizing to standardize inputs of varying sizes to 640×640 pixels; it also incorporates Mosaic data augmentation to enrich target scale and background information. The backbone network consists of a Focus module, a CBS (Conv_BN_SiLU) module, residual blocks, and an SPP (Spatial Pyramid Pooling) module. The Focus module performs downsampling via skip-pixel sampling to prevent loss of feature information; the SPP module fuses features from different receptive fields through multi-scale pooling. The neck network adopts a PANet path aggregation network, which effectively combines shallow-layer texture features with deep-layer semantic features through bidirectional feature fusion (top-down and bottom-up), thereby enhancing small-object detection capabilities. The detection head adopts a decoupled head architecture, separating the classification and regression tasks into two independent branches, thereby resolving the task conflict issues inherent in traditional coupled heads; it employs an anchor-free design, eliminating the need for pre-defined anchor boxes and reducing model complexity.

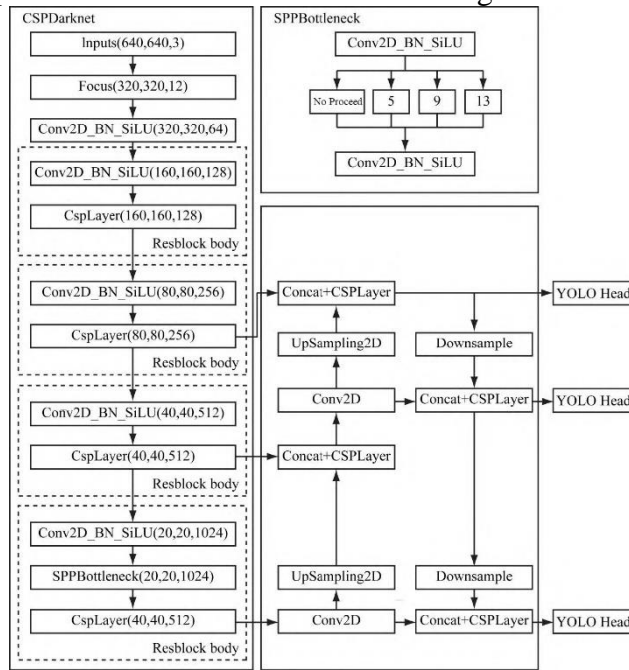


Figure 2: YOLOX Network Architecture

2.3.2. Model Training Strategy

This paper employs a transfer learning approach, using weights pre-trained on the COCO dataset to initialize the model, followed by global fine-tuning using the collected data. Key training parameters are set as follows: The batch size is set to 16, with a total of 100 training epochs. The Stochastic Gradient Descent with Momentum (SGDM) algorithm is used for optimization, with a momentum of 0.937, an initial learning rate of $5e-4$, a learning rate decay factor of 0.99, and a weight decay coefficient of $5e-4$. Model performance is evaluated on the validation set after each epoch, and the weights yielding the lowest validation set loss are saved as the final detection model.

2.4. MOMW Algorithm

Within the individual tree bounding boxes detected by YOLOX, a multi-stage optimized thresholding algorithm is employed for detailed canopy segmentation. Building upon the traditional thresholding algorithm, this method incorporates three preprocessing steps: global grayscale

enhancement, median filtering for noise reduction, and optimized morphological opening operations. These steps effectively address the issue of over-segmentation caused by background interference and noise. The algorithm workflow is as follows:

Global Gray-Scale Enhancement: The RGB image within the detection bounding box is converted to a grayscale image, and global histogram equalization is applied to enhance the contrast between the tree canopy and the background. The original grayscale levels are mapped to a uniform distribution via a cumulative distribution function, expanding the grayscale dynamic range and making the tree canopy edges clearer.

Median Filtering for Noise Reduction: A 3×3 sliding window is used to apply median filtering to the grayscale image, removing salt-and-pepper noise and fine noise on the tree crown surface. As a nonlinear filtering method, median filtering preserves the edge features of the tree crown while reducing noise, thereby avoiding contour blurring caused by excessive smoothing.

Optimized Morphological Opening: The filtered grayscale image is binarized, and two rounds of morphological opening operations are performed using a 3×3 structural element. First, an erosion operation is applied to sever fine connections between the tree canopy and the background, as well as between tree canopies. Then, a dilation operation is performed to restore the canopy shape and eliminate internal voids within the canopy.

Labeled watershed segmentation: Based on the binary image, foreground labels (crown regions) and background labels are extracted. These labels are used as seed points to guide the watershed algorithm for segmentation, ultimately yielding pixel-level boundaries of individual tree crowns.

3. Results and Analysis

3.1. Analysis of Model Detection Performance

This study utilized orthophotos captured by drones and employed the LabelImg annotation tool to perform bounding box annotations on individual trees. To accommodate the model's input dimensions, the images were cropped into 640×640 -pixel tiles, thereby constructing a dedicated dataset for single-tree detection. The dataset comprises a total of 102 valid images, covering individual tree samples from diverse habitats and of varying sizes, ensuring the dataset's diversity and representativeness. The dataset was randomly split into training, validation, and test sets at a ratio of 7:1.5:1.5. The performance of the trained YOLOX model was evaluated using the test set data. The trained YOLOX model demonstrated excellent detection performance for individual trees in UAV imagery, achieving an accuracy rate of 85.8% and effectively distinguishing individual trees from similar background objects such as understory shrubs. Subsequently, the segmentation performance of the MOMW algorithm was evaluated using the manually interpreted crown boundaries of individual trees as ground truth. As shown in Figure 3, the MOMW algorithm achieves significant segmentation accuracy, primarily due to three factors: first, segmentation within the detection bounding box reduces interference from background pixels; second, global grayscale enhancement improves the contrast between the tree canopy and the background; and third, median filtering and morphological opening operations effectively remove noise, disconnect adjacent tree canopies, and reduce over-segmentation and under-segmentation.

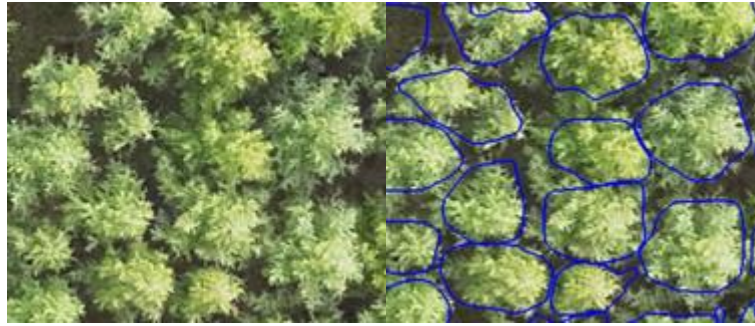


Figure 3: Segmentation results of the MOMW algorithm

3.2. Impact of Different Parameters on Segmentation Performance

The confidence threshold determines the filtering criteria for YOLOX detection boxes and directly affects the quality of samples for subsequent segmentation, while the IoU threshold influences the determination of individual tree segmentation results. To analyze the impact of the detection model's confidence threshold and IoU threshold on segmentation accuracy, this study set four thresholds—0.1, 0.3, 0.5, and 0.7—to examine their effects on segmentation performance. As the confidence threshold increases, segmentation accuracy gradually decreases; when the confidence threshold exceeds 0.1, some true individual trees are missed. As the IoU threshold increases, segmentation accuracy gradually decreases; the segmentation performance is optimal when the IoU threshold is set to 0.5. This is because the segmentation boundaries in areas with overlapping tree canopies and blurred edges are difficult to align perfectly with the ground truth, and a high IoU threshold may classify these reasonable segmentation results as errors. In this study, the best segmentation results were achieved with a confidence threshold of 0.1 and an IoU threshold of 0.5, as shown in Figure 4. In practical applications, appropriate confidence and IoU thresholds can be selected based on specific requirements.

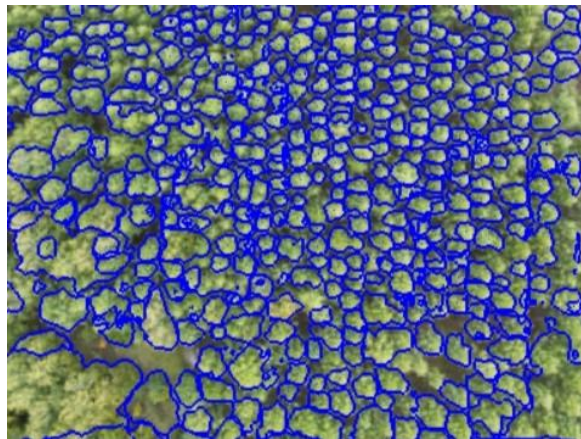


Figure 4: Single-tree segmentation results

4. Discussion

4.1. Analysis of Method Advantages

The two-stage framework proposed in this study, consisting of “YOLOX single-tree detection combined with multi-stage optimized watershed segmentation,” effectively addresses the numerous limitations of traditional single-tree segmentation methods and offers the following significant

advantages: YOLOX's anchor-free design and decoupled detection head effectively capture the features of trees with small crowns, significantly reducing the false negative rate and providing reliable bounding boxes for subsequent segmentation; performing segmentation within the detected bounding boxes excludes background pixels, greatly reducing the interference of the background on the watershed algorithm; The multi-stage optimized preprocessing steps effectively address the noise sensitivity and over-segmentation issues of traditional watershed algorithms. Global grayscale enhancement improves the contrast between the tree canopy and the background, median filtering removes noise, and morphological opening operations disconnect tree canopy connections, resulting in more accurate segmentation boundaries; The method described in this paper achieves high-precision single-tree segmentation using only visible-light imagery from consumer-grade drones, resulting in low data acquisition costs and simple operation; simultaneously, the YOLOX and MOMW algorithms have low computational requirements and can be deployed on drone edge computing platforms to enable real-time single-tree segmentation, making them highly practical.

4.2. Limitations and Future Directions

This study still has certain limitations that require further refinement in future research:

Generalizability across tree species: This study was conducted exclusively on artificial forests of hybrid Korean pine and larch; the model's generalizability to other tree species (such as broadleaf trees and mixed forests) remains to be verified. Future work will involve constructing multi-species, multi-region single-tree segmentation datasets to enhance the model's adaptability to various scenarios.

Segmentation of Occluded Canopies: For trees severely obscured by adjacent canopies, there is still room for improvement in detection and segmentation accuracy. Future work will incorporate multispectral imagery and LiDAR data to fuse spectral, textural, and 3D structural information, thereby enhancing the segmentation of occluded canopies.

Model Lightweighting: Although YOLOX has a relatively small number of parameters, its real-time performance on resource-constrained edge devices still requires optimization. In future work, we will adopt lightweighting techniques such as model pruning and quantization to further reduce the model size and improve inference speed.

Parameter Automation: Currently, parameters such as confidence thresholds and intersection-over-union ratios still require manual configuration. In future work, we will investigate adaptive parameter optimization methods to achieve fully automated operation of the algorithm.

5. Conclusions

This study addresses the challenges in single-tree segmentation of UAV imagery, including severe background interference, high rates of missed detection for small-canopy trees, and prominent over-segmentation and under-segmentation issues. We propose a single-tree segmentation method based on YOLOX and a multi-stage optimized thresholding approach. Using planted forests of *Pinus sylvestris* var. *mongolica* and *Larix* spp. in the Saihanba region of Hebei Province, China, as the study area, we experimentally validated the effectiveness and superiority of the method. The main conclusions are as follows:

- 1) A high-resolution UAV image dataset comprising 10 image samples of single trees in artificial forests of *Pinus sylvestris* var. *mongolica* and *Larix* spp. was constructed, providing a reliable data foundation for training deep learning models. The single-tree detection model based on YOLOX achieved excellent accuracy, with a precision of 85.8%, and was capable of effectively detecting individual trees with varying crown sizes.

- 2) The proposed method, based on YOLOX and a multi-stage optimized thresholding algorithm,

performs fine-grained crown segmentation within individual tree bounding boxes. Through global grayscale enhancement, median filtering, and morphological opening operations, the accuracy of individual tree segmentation was significantly improved.

This study validates the feasibility and superiority of the “detection + segmentation” two-stage framework for individual tree segmentation in UAV imagery. It provides an efficient and high-precision technical approach for resource monitoring at the individual tree level in planted forests, offering broad application prospects.

Acknowledgements

This research was supported by the Special Science and Technology Project for the Construction of the Chengde National Innovation Demonstration Zone for the Sustainable Development Agenda under Grant 202302F007.

References

- [1] Li D, Zhang J J, & Zhao M X.(2019). *Extraction of Stand Factors in UAV Image Based on FCM and Watershed Algorithm*. *SCIENTIA SILVAE SINICAE*, 55(05),180-187.
- [2] Gu Z X, & Pei F R.(2022). *Single Tree Recognition Algorithm Based on Multi-Layer K-means in Forest Point Cloud*. *FOREST RESOURCES MANAGEMENT*,(01),124-131.<https://doi.org/10.13466/j.cnki.lyzygl.2022.01.015>.
- [3] Wang X S & Yang G.(2018). *A multi-objective optimization segmentation method for tree image based on fusion clustering and classification algorithm*. *JOURNAL OF BEIJING FORESTRY UNIVERSITY*,40(12),124-131.<https://doi.org/10.13332/j.1000-1522.20180160>.
- [4] Huo L N & Zhang X L.(2021). *Individual Tree Information Extraction and Accuracy Evaluation Based on Airborne LiDAR Point Cloud by Multilayer Clustering Method*. *SCIENTIA SILVAE SINICAE*, 57(01),85-94.
- [5] Sun G, Xin Y & Yan C.(2019). *Estimating Single Tree Height Based on Improved K-means Algorithm*. *FOREST ENGINEERING*,35(06),31-35.<https://doi.org/10.16270/j.cnki.slgc.2019.06.022>.
- [6] Zeng J.(2025). *Research on deep learning-based individual tree segmentation methods for Liriodendron plantations using UAV high-resolution imagery*(Master's Thesis, Nanjing Forestry University). *Master's degree*. <https://doi.org/10.27242/d.cnki.gnjlu.2025.000459>.
- [7] Wu K, Lai T, Ye J M, et al.(2026). *Individual Tree Detection and Crown Width Estimation Using UAV Imagery based on SSD and Faster R-CNN Algorithms*. *REMOTE SENSING TECHNOLOGY AND APPLICATION*,41(01),166-177.
- [8] Wang K, Chen Y L & Xu H L.(2024). *Detection of individual trees of bayberry using improved YOLOv7*. *Transactions of the Chinese Society of Agricultural Engineering*, 40(11),159-167.
- [9] Ge Z, Liu S, Wang F, et al. *YOLOX: Exceeding YOLO Series in 2021*[J]. *arXiv preprint arXiv:2107.08430*, 2021.
- [10] Chen H, Quan Y, Ma X T, et al.(2026) *Individual-tree identification of Korean pine based on UAV imagery and Mask R-CNN*. *Chinese Journal of Applied Ecology*,1-13.<https://doi.org/10.13287/j.1001-9332.202605.007>