Journal of Geomatics Science and Technology, vol. 16, no. 1, pp. 81-90, September 2026- Research article

Zero-Shot Tree Crown Delineation in Semi-Arid Urban Forests Using UAV Imagery and the Segment Anything Model

Mahnaz Babaie [1], Zahra Azizi [2,*], Siavash Bakhtiarvand Bakhtiari [3]

[1] Department of Remote Sensing and GIS, SR.C., Islamic Azad University, Tehran, Iran
Mahnaz.babaei@srbiau.ac.ir

[2] Department of Geoinformation and Geomatics Engineering, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran
zah_azizi@sbu.ac.ir

[3] Head of Department at Mobarakeh Steel Company, Isfahan, Iran
siavashbakhtiar369@gmail.com

(*Received: May 2025, Accepted: August 2026*)

Abstract

Urban forests play a critical role in mitigating environmental stresses in semi-arid cities by improving air quality, reducing urban heat island effects, enhancing carbon sequestration, and supporting ecosystem resilience. Accurate delineation of individual tree crowns (ITCs) is fundamental for urban forest inventory, biomass estimation, ecological monitoring, and sustainable green-space management. Recent advances in foundation models have created new opportunities for automated image segmentation; however, their applicability for tree crown delineation in semi-arid urban forests remains insufficiently investigated. In particular, the performance of zero-shot segmentation models under varying crown architectures and canopy densities is still poorly understood. This study evaluated the zero-shot capability of the Segment Anything Model (SAM) for ITC delineation using high-resolution UAV-derived RGB imagery acquired over a 1,600-ha urban afforestation area in southwestern Isfahan, Iran. A total of 5,974 reference trees representing three dominant species with contrasting crown structures, including Cercis siliquastrum, Robinia pseudoacacia, and Pinus eldarica, were used for quantitative evaluation. Model performance was assessed using Precision, Recall, F1-score, and Intersection over Union (IoU) metrics. The results revealed substantial differences among species. SAM achieved its highest performance for Cercis siliquastrum, characterized by well-separated and relatively uniform crowns, with an F1-score of 0.78 and an IoU of 0.52. In contrast, performance declined considerably for Robinia pseudoacacia (F1 = 0.39, IoU = 0.23) and Pinus eldarica (F1 = 0.35, IoU = 0.18), where crown overlap, irregular canopy geometry, and dense vegetation structure increased segmentation uncertainty. Overall, 3,487 trees were correctly delineated by the model. The findings demonstrate that SAM provides a promising training-free framework for large-scale urban tree crown mapping. Nevertheless, segmentation accuracy remains strongly dependent on canopy structure, crown morphology, and canopy density. This study provides one of the first comprehensive evaluations of the zero-shot Segment Anything Model in semi-arid urban afforestation environments and highlights both the opportunities and limitations of foundation models for operational urban forest monitoring, inventorying, and management. The results further emphasize the need for integrating structural information and advanced prompting strategies to improve segmentation performance in complex urban canopy conditions.

Keywords: Tree crown delineation; Zero-shot learning; UAV imagery; Semi-arid urban forests; Deep learning.

* Corresponding author

1. Introduction

Urban forests play a pivotal role in enhancing the sustainability of cities, particularly within semi-arid environments where water scarcity, elevated temperatures, and ecosystem vulnerability are prevalent. These forests contribute significantly to microclimate regulation, air purification, carbon sequestration, soil erosion reduction, and biodiversity preservation [1-2]. An essential prerequisite for effective urban forestry management and ecological monitoring is the accurate delineation of individual tree crowns (ITCs). This enables precise forest inventorying, above-ground biomass estimation, and assessments of ecosystem health.

Conventional ground-based surveys, while accurate, are often impractical and prohibitively expensive when applied to extensive afforestation areas [3-4]. Consequently, remote sensing technologies, particularly unmanned aerial vehicles (UAVs) paired with high-resolution RGB sensors, have gained prominence. Through structure-from-motion (SfM) techniques, these systems produce orthomosaics and canopy height models (CHMs), facilitating extraction of tree-level metrics like crown diameter and height across diverse urban settings [5].

Traditional remote sensing analysis methods—including local maxima filtering, watershed segmentation, and object-based image analysis—have proven useful for delineating tree crowns in simpler forest structures. However, they encounter significant limitations in structurally complex urban plantations, especially those featuring overlapping canopies and irregular crown geometries, resulting in elevated omission and commission errors [6-8].

Recent UAV-based studies in semi-arid urban zones within Iran have begun exploring segmentation techniques that integrate UAV imagery with manual or automatic delineation workflows [9-10]. For instance, Miraki and Sohrabi (via SSRN) assessed ITC delineation using the Segment Anything Model (SAM) across various stand types, reporting over 90% accuracy for well-defined crowns and demonstrating the potential for estimating green-space water requirements with high precision in semi-arid climates [11-12].This underscores SAM's capability in clearly structured contexts.

Beyond RGB data, the integration of auxiliary information such as LiDAR and hyperspectral imagery has shown promise in increasing segmentation accuracy [13]. Fusion strategies combining optical and LiDAR-derived features have yielded F1-score improvements of up to 12.6 percentage points (e.g., surface normals, intensity, height, LAI) and significantly enhanced species mapping in tropical settings [14]. Meanwhile, recent research by Ravindran et al. [15] examined the use of hyperspectral imagery and dimensionality reduction (PCA, UMAP, factor analysis) prior to conducting ITC segmentation with SAM. Although overall segmentation performance remained modest (average F1 $\approx$ 0.26), approaches utilizing UMAP and factor analysis achieved improvements at certain sites, highlighting the challenge and potential of hyperspectral integration.

A critical advancement in recent years has been the development of foundation models for computer vision—large-scale, pre-trained networks that excel in generalized, zero-shot tasks. The Segment Anything Model (SAM), introduced by Kirillov et al. [16], embodies this shift: trained on more than 11 million images with over a billion masks, SAM is designed to perform promptable segmentation across diverse domains without task-specific fine-tuning.

In the realm of remote sensing, Osco et al. [13] demonstrated that SAM can offer remarkable generalization across aerial and orbital imagery when guided through various prompts, including bounding boxes, points, or even textual labels. They further improved performance by combining prompt engineering with one-shot training techniques—strengthening SAM's adaptability while mitigating the need for extensive manual annotation.

Subsequent work has explored SAM's use in ecological remote sensing. Teng et al. [17] evaluated SAM against customized Mask R-CNN models for ITC instance segmentation in young tree plantations. While SAM out-of-the-box did not outperform tailored models, integrating digital surface model (DSM) data and fine-tuning prompts via parameter-efficient methods significantly enhanced outcomes. Likewise, BalSAM—introduced by another 2025 study—combined SAM with DSM integration across boreal, temperate, and tropical forests. These results highlight that, although vanilla SAM may underperform, structured enhancements such as prompt tuning and auxiliary data (e.g., DSM) can elevate segmentation performance, making SAM a promising candidate for scalable, forest-type-agnostic ITC mapping. These advances collectively set the stage for exploring SAM in urban semi-arid afforestation contexts—a domain that remains underexplored.

While foundation models like SAM offer data-efficient segmentation, their real-world applicability depends heavily on forest structure complexity, species crown morphology, and canopy density. Moreover, integrating UAV-derived DSM or spectral features could significantly mitigate segmentation errors in dense or overlapping canopies. However, most of these improvements relied on additional structural data, prompt optimization, or model adaptation, making it difficult to isolate the intrinsic zero-shot capability of SAM under operational conditions.

Collectively, previous studies indicate that SAM has considerable potential for tree crown delineation; however, its performance remains highly dependent on canopy structure, data characteristics, and the availability of supplementary information. Despite the growing interest in foundation models for remote sensing applications, evidence regarding their performance for individual tree crown delineation in semi-arid urban forests remains extremely limited. Most previous studies have focused on natural forests, agricultural systems, or temperate ecosystems and often relied on additional data sources such as DSMs, LiDAR, or model fine-tuning. Consequently, the transferability of zero-shot foundation models to urban afforestation environments characterized by heterogeneous canopy structures, varying crown morphologies, and water-limited conditions remains uncertain.

To address this knowledge gap, a comprehensive evaluation framework was developed to assess the zero-shot performance of the Segment Anything Model (SAM) using high-resolution UAV RGB imagery acquired over a large urban afforestation site in southwestern Isfahan, Iran. Specifically, we investigate SAM performance across three dominant tree species exhibiting contrasting crown architectures and canopy densities. The novelty of this study lies in providing one of the first comprehensive evaluations of the zero-shot Segment Anything Model (SAM) for individual tree crown delineation in semi-arid urban afforestation environments. Specifically, the study (1) quantifies the influence of crown morphology and canopy density on segmentation performance, (2) evaluates SAM across contrasting broadleaf and coniferous species, and (3) identifies key limitations and opportunities for applying foundation models in operational urban forest monitoring.

2. Materials and Methods

2.1. Study Area

The study was conducted in a semi-arid urban afforestation site located southwest of Isfahan, Iran, within the reforested lands of the Mobarakeh Steel Company. This area was established as part of a long-term environmental sustainability initiative aimed at mitigating industrial ecological impacts. The afforestation site spans approximately 1,600 hectares, situated at elevations ranging from 1,500 to 1,600 meters above sea level, with geographical coordinates of approximately 32°15′53″N, 51°25′29″E .Figure 1 illustrates the geographical location of the study area in southwestern Isfahan, Iran. The urban afforestation site established by Mobarakeh Steel Company is highlighted with a red circle.

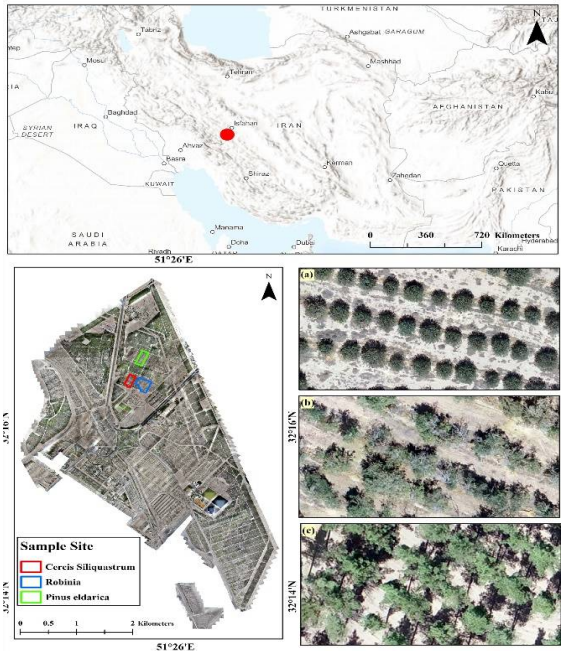


Figure 1. Location of the study area in southwestern Isfahan, Iran (highlighted in red).

The climate is characterized as arid to semi-arid, with hot, dry summers where daytime temperatures exceed 35°C and minimal annual precipitation, primarily occurring as rainfall or, less frequently, snow in higher elevations. Prevailing winds are predominantly from the northwest and northeast, and relative humidity remains low throughout most of the year. Due to proximity to Iran's central deserts, the ecosystem exhibits sparse native vegetation and limited natural regeneration capacity.

Seventeen different tree species have been introduced to this area as part of the reforestation

effort. Among these, the most dominant species include *Cupressus* arizonica (Arizona cypress), *Cercis siliquastrum* (Judas tree), *Pinus eldarica* (Afghan pine), *Robinia pseudoacacia* (Black locust), *Juniperus communis* (common juniper), *Ailanthus altissima* (tree-of-heaven), and *Amygdalus scoparia* (wild almond).

The experimental design focused on three homogeneous test sites, each dominated by one of the following species: *Cercis siliquastrum*, *Robinia pseudoacacia*, and *Pinus eldarica*. These were selected based on their crown structure, spatial arrangement, and ecological relevance.

2.2. UAV Data Acquisition

High-resolution RGB imagery was acquired using a DJI Phantom 4 Pro UAV equipped with a 1/2.3″ CMOS sensor and a 20-megapixel camera. Flights were conducted in June and July 2022 under clear sky conditions to minimize shadow effects. The UAV was flown at an altitude of 127 m with 80% front and side overlap, capturing images in JPEG format (3,000 × 3,000 pixels, 8.8 mm focal length). The ground sampling distance (GSD) was approximately 2–3 cm.

The collected imagery was processed in Agisoft Metashape Professional to generate orthomosaics, digital surface models (DSMs), and 3D point clouds, which served as the base for tree crown delineation.

2.3. Ground Reference Data

A comprehensive ground-truth dataset was provided by the Mobarakeh Steel Company's Environmental Monitoring Division, including 5,974 individual tree records. Each record contained the tree species, spatial coordinates, and estimated crown diameter.

The trees were distributed across the three selected sites as follows (Figure 2):

- Site a: *Cercis siliquastrum* (1,721 individuals),
- Site b: *Robinia pseudoacacia* (2,213 individuals),
- Site c: *Pinus eldarica* (2,040 individuals).



Figure 2. 3D visualization of individual tree crowns in the selected test sites.

To validate the segmentation results, circular buffers with a radius of 2 m were generated around each ground-referenced tree location. The buffer radius was selected based on field measurements indicating that the average crown diameter of the dominant tree species within the study area was approximately 4 m. Consequently, a 2-m radius buffer provided a realistic representation of the average crown extent and served as a spatial reference for matching predicted crowns with field observations during accuracy assessment.

2.4. Segment Anything Model (SAM) for Zero-Shot Crown Delineation

The Segment Anything Model (SAM), developed by Meta AI, was employed for individual tree crown delineation from UAV-derived orthomosaic imagery. SAM is a foundation model designed for promptable image segmentation and is based on a transformer architecture consisting of three main components: (i) an image encoder, (ii) a prompt encoder, and (iii) a lightweight mask decoder. Unlike conventional deep learning approaches that require large task-specific training datasets and model retraining, SAM utilizes knowledge learned from billions of annotated masks, enabling robust zero-shot segmentation across diverse image types and environmental conditions.

In this study, the ViT-H (Vision Transformer-Huge) version of SAM was implemented using its Automatic Mask Generation framework. The model was applied in a fully zero-shot manner, without any manual prompting, parameter adjustment, fine-tuning, retraining, or incorporation of field reference data during the segmentation process. Consequently, all crown delineations were generated solely from the visual information contained in the UAV imagery, allowing the intrinsic segmentation capability of SAM to be evaluated under operational forestry conditions.

To facilitate efficient processing and reduce memory requirements, the orthomosaic images were divided into 800 × 800 pixel tiles with a 100-pixel overlap between adjacent tiles. The overlap was introduced to minimize edge effects and preserve crown continuity across tile boundaries. Each image tile was processed independently using the default SAM configuration to generate candidate segmentation masks representing potential tree crowns.

Following segmentation, masks generated from adjacent tiles were merged and reprojected into the original spatial reference system. Duplicate or

overlapping crown segments occurring within tile-overlap regions were removed during the mosaicking procedure to produce a seamless crown delineation layer. The resulting crown polygons were then exported as vector shapefiles and used for subsequent spatial accuracy assessment against the ground-reference dataset.

The overall workflow consisted of four sequential steps: (1) UAV image acquisition and orthomosaic generation, (2) image tiling and automatic segmentation using SAM, (3) merging and post-processing of segmentation masks, and (4) quantitative accuracy assessment using field-referenced tree locations and crown buffers.

2.5. Accuracy Assessment

Segmentation performance was evaluated by comparing the predicted crown polygons generated by the Segment Anything Model (SAM) with ground-referenced tree locations. A buffer-based matching procedure was adopted to account for positional uncertainties in both field observations and image-derived crown boundaries. Circular buffers with a radius of 2 m were generated around each reference tree location. The buffer radius was selected based on field measurements indicating that the average crown diameter of trees in the study area was approximately 4 m. Consequently, the buffer represented the average crown extent surrounding each tree and provided a practical spatial tolerance for matching predicted crowns with field observations.

A one-to-one matching strategy was adopted, whereby each reference tree could be matched to only a single predicted crown polygon. A predicted crown polygon was classified as a true positive (TP) when it intersected a reference buffer. Predicted crowns that did not intersect any reference buffer were considered false positives (FP), whereas reference trees without an intersecting predicted crown were classified as false negatives (FN). These matching results formed the basis for the quantitative accuracy assessment.

2.5.1. Precision, Recall, and F1-Score

Detection accuracy was quantified using Precision, Recall, and F1-score, calculated from the numbers of true positives (TP), false positives (FP), and false negatives (FN) obtained through the buffer-based matching procedure:

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

where TP represents correctly detected tree crowns, FP denotes erroneously segmented objects, and FN indicates omitted tree crowns. The F1-score provides a balanced measure of detection performance by simultaneously accounting for commission and omission errors.

2.5.2. Intersection over Union (IoU)

In addition to the detection-based metrics, the spatial agreement between predicted crown polygons and reference buffers was quantified using the Intersection over Union (IoU) metric:

$$IoU = |P \cap R| / |P \cup R|$$

where P represents the predicted crown polygon and R represents the corresponding reference buffer. IoU values range from 0 to 1, with higher values indicating greater spatial overlap and better delineation quality. In this study, IoU was used as a descriptive measure of crown delineation accuracy rather than as the primary criterion for assigning true positives.

The combined use of detection-based metrics (Precision, Recall, and F1-score) and a spatial-overlap metric (IoU) provides a comprehensive evaluation of both tree detection success and crown delineation quality.

3. Results

Following the methodology described above, the performance of the Segment Anything Model (SAM) was evaluated across the three study sites using both detection-based and spatial-overlap metrics. A total of 5,974 reference trees were assessed, of which 3,487 were correctly delineated by the model, corresponding to an overall detection rate of approximately 58.4%. Segmentation performance varied considerably among species and was strongly influenced by crown morphology, canopy density, and crown overlap. The results are presented in terms of overall segmentation accuracy, species-specific

performance, crown delineation quality, and common segmentation error patterns.

3.1. Overall Segmentation Performance

The Segment Anything Model (SAM) successfully delineated 3,487 out of 5,974 reference trees across the three study sites, corresponding to an overall detection rate of approximately 58.4%.

Segmentation performance varied substantially among the studied species, indicating that SAM responded differently to contrasting crown structures and canopy configurations. Among the evaluated species, Cercis siliquastrum exhibited the highest segmentation accuracy, achieving the greatest values for Precision, Recall, F1-score, and IoU. In contrast, considerably lower performance was observed for Robinia pseudoacacia and Pinus eldarica. The F1-score obtained for Cercis was approximately twice that achieved for Pinus and substantially higher than that obtained for Robinia, highlighting pronounced species-dependent variability in segmentation performance.

Table 1. Detection statistics and segmentation accuracy metrics for the three evaluated tree species.

Species	Trees	TP	FP	FN	Precision	Recall	F1-score	IoU
Cercis siliquastrum	1721	1388	424	333	0.76	0.80	0.78	0.52
Robinia pseudoacacia	2213	1237	2894	976	0.29	0.55	0.39	0.23
Pinus eldarica	2040	862	2005	1174	0.30	0.42	0.35	0.18

Differences among species were also reflected in the detection statistics. While the majority of Cercis trees were successfully identified, omission and commission errors increased markedly in Robinia and Pinus stands. These results indicate that segmentation quality was not uniform across the study area and was strongly influenced by species-specific canopy characteristics. Detailed accuracy metrics for each species are presented in Table 1, while Figure 3 summarizes the corresponding performance patterns.

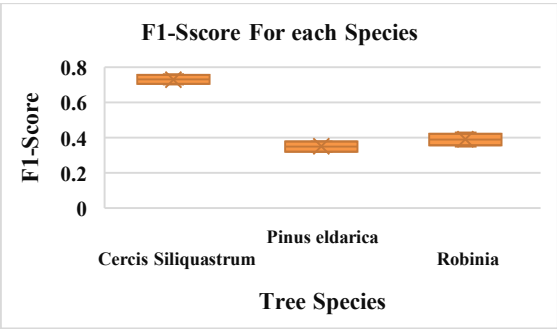


Figure 3. F1-score values of the SAM model for each tree species.

3.2. Intersection over Union (IoU) and Tree Detection Counts

IoU analysis confirmed these trends. SAM achieved an IoU of 0.52 for *Cercis*, reflecting moderate agreement with reference boundaries, while IoU was considerably lower for *Robinia* (0.23) and *Pinus* (0.18). Out of 5,974 reference trees, 3,487 were correctly delineated by SAM. For *Cercis*, 1,388 of 1,721 trees were correctly detected (≈80%), compared to lower rates for *Robinia* and *Pinus*. Table 1 show Confusion matrix–based results of the SAM model for crown delineation across the three studied species. The table reports true positives (TP), false positives (FP), false negatives (FN), total reference trees, and IoU.

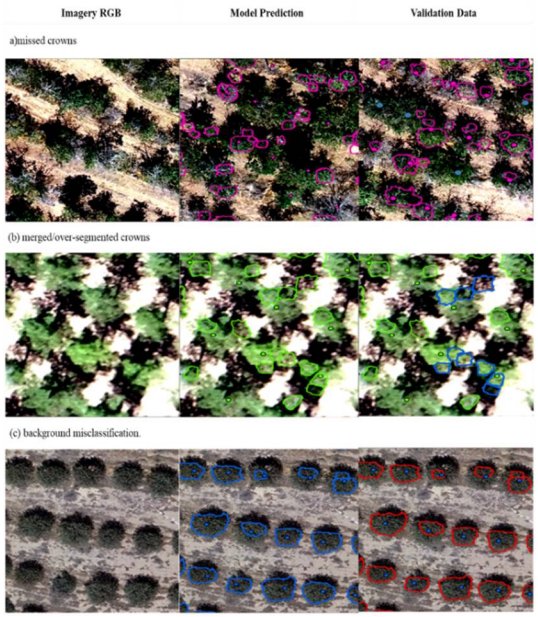


Figure 4. Examples of segmentation errors in the SAM model across the study sites. The panels show: (a) missed crowns (validation data), (b) over-segmented or merged crowns (model prediction), and (c) background misclassification (RGB image).

3.3. Qualitative Analysis and Error Modes

Visual assessment of segmentation results revealed three common error patterns:

1. Missed crowns – individual trees in dense clusters were frequently omitted (Figure 4a).
2. Over-segmentation – irregular crowns were often split into multiple masks (Figure 4b).

3. Background confusion – non-tree objects (soil, built structures) were occasionally misclassified as crowns. But this does not apply to trees that are placed next to each other at a certain distance. (Figure 4c).

3.4. Influence of Crown Structure and Spatial Distribution

Species-specific results are closely tied to canopy architecture. *Cercis siliquastrum* trees, planted in regular spacing with minimal crown overlap, were delineated with high accuracy. Conversely, *Robinia* crowns (umbrella-shaped, overlapping) and *Pinus* crowns (dense coniferous canopies) exhibited frequent boundary merging, reducing segmentation precision. These structural effects are visualized in Figure 5, which compares SAM predictions with ground-truth boundaries across the three sites.



Figure 5. Crown segmentation performance of SAM for the three test species. Left to right: RGB imagery, ground-truth buffers, and SAM model predictions.

3.5. Factors Affecting Model Accuracy

Three main factors influenced segmentation performance:

- Crown visual complexity: species with distinct, uniform crowns achieved higher accuracy.
- Canopy density and heterogeneity: dense stands with high crown contact reduced boundary clarity.
- Image acquisition conditions: illumination, spatial resolution, and shadowing affected detection precision. UAV flights under clear skies improved results, though higher resolution increased computational costs.

4. Discussion

The results of this study demonstrate that the performance of the Segment Anything Model (SAM) for individual tree crown (ITC) delineation in a semi-arid urban afforestation context is highly dependent on species-specific crown architecture and canopy density. SAM achieved its highest accuracy for *Cercis siliquastrum* (F1 = 0.78; IoU = 0.52), a species with spherical and well-separated crowns that facilitated mask generation and reduced boundary ambiguity. In contrast, the accuracy was considerably lower for *Robinia pseudoacacia* (F1 = 0.39; IoU = 0.23) and *Pinus eldarica* (F1 = 0.35; IoU = 0.18), where overlapping crowns, umbrella-like shapes, and vertically layered coniferous canopies obscured crown boundaries and reduced both detection and boundary adherence. These findings clearly indicate that SAM's performance is more robust in homogeneous, regularly spaced plantations than in dense or heterogeneous stands, supporting previous observations regarding the influence of canopy structure on automated crown delineation [17-18].

The identification of error modes further supports this interpretation. The most common errors included missed crowns in tightly clustered canopies, over-segmentation of irregular crowns into multiple masks, and background misclassification where non-tree objects such as bare soil or built surfaces were incorrectly identified as crowns. Such errors were particularly pronounced in the *Robinia* and *Pinus* plots, while *Cercis* trees, being more isolated, were delineated with fewer mistakes. These patterns are in line with previous UAV-based segmentation studies in semi-arid and urban contexts, which have consistently shown reduced accuracy under conditions of crown overlap, crown aggregation, or structural complexity [7-8]. The results therefore confirm that crown separation remains one of the principal challenges for zero-shot segmentation models operating in complex forest environments.

Comparison with conventional ITC delineation methods highlights both the strengths and limitations of SAM. Traditional approaches such as local maxima filtering, watershed segmentation, or object-based image analysis are often constrained by their sensitivity to canopy overlap and irregular shapes, resulting in elevated omission and commission errors [12,19]. While SAM does not eliminate these challenges, its zero-shot capability provides a

significant advantage by avoiding the need for manually annotated training datasets and site-specific model calibration. This characteristic substantially reduces implementation effort and facilitates transferability across locations. The observed trade-off between generalizability and segmentation accuracy reflects broader trends in the adoption of foundation models for remote sensing applications, where scalability and adaptability are often prioritized over task-specific optimization [14, 16].

Recent studies provide further context for interpreting the results of this research. Teng et al. [17] demonstrated that incorporating structural information such as digital surface models (DSM) significantly improved SAM's performance for crown delineation, suggesting that integrating UAV-derived DSM data could enhance results in dense stands such as those of *Robinia* and *Pinus*. Similarly, [20] showed that automatic prompt generation could reduce segmentation errors in remote sensing applications, while Vaschetti et al. [21] introduced a SAM-based pipeline (TreePseCo) that combined center-point detection and guided prompts to achieve higher accuracy across different forest types. These studies suggest that, although SAM alone may not consistently outperform specialized models such as Mask R-CNN or U-Net [19], its adaptability through prompt engineering, guided segmentation, and multi-source data fusion makes it a promising candidate for operational forest monitoring and urban green-space management.

An important contribution of this study is that it provides one of the first evaluations of a zero-shot foundation model for tree crown delineation in a semi-arid urban afforestation environment. Unlike most previous studies that focused on natural forests or relied on supervised deep-learning workflows, the present research assessed SAM without any retraining or fine-tuning, thereby providing a realistic evaluation of its out-of-the-box performance under operational conditions. The ability of SAM to achieve reliable delineation in *Cercis* plantations demonstrates its potential for rapid urban forest inventories and large-scale monitoring programs where annotated datasets are unavailable or prohibitively expensive to produce.

The results also have practical implications for urban forest management. Accurate delineation of individual tree crowns is essential for estimating tree density, monitoring plantation health, assessing survival rates, and supporting ecosystem-service evaluations. The strong performance observed in regularly spaced plantations suggests that SAM could substantially reduce the time and labor required for crown mapping in urban afforestation projects. However, the reduced accuracy in *Robinia* and *Pinus* stands indicates that supplementary methods—whether through structural data integration, topology-aware post-processing, or hybrid deep-learning pipelines—are necessary for achieving reliable results in more complex canopy environments.

Overall, the findings demonstrate that SAM represents a significant advancement toward scalable and transferable ITC delineation. Nevertheless, crown overlap, irregular crown geometry, and canopy heterogeneity remain important constraints that limit segmentation accuracy. Future research should therefore focus on integrating foundation models with height information, automated prompting strategies, and species-aware post-processing techniques to improve performance across diverse urban and forest ecosystems.

5. Conclusion

This study evaluated, for the first time, the applicability of the Segment Anything Model (SAM) as a zero-shot foundation model for individual tree crown delineation in a semi-arid urban afforestation environment. The results demonstrated that SAM performs reliably in plantations characterized by regular spacing and clearly separated crowns, whereas its accuracy declines in stands with dense, overlapping, or structurally complex canopies. These findings highlight crown morphology and canopy organization as key factors governing segmentation success.

SAM demonstrates strong potential as a scalable and cost-effective approach because it does not require task-specific training data or model retraining. At the same time, its limitations in complex canopy conditions suggest that supplementary strategies, such as incorporating structural information, automated prompting, or topology-aware post-processing techniques, are necessary to improve performance.

Overall, this research highlights both the opportunities and the current constraints of applying foundation models in ecological monitoring. SAM can serve as a valuable baseline tool for operational tree crown mapping in urban afforestation projects, while future developments should focus on hybrid workflows that combine the flexibility of foundation models with complementary data sources to achieve more robust and accurate crown delineation across diverse forest conditions.

References

[1] E. G. McPherson, N. van Doorn, and J. de Goede, "Structure, function and value of street trees in California, USA," *Urban Forestry & Urban Greening*, vol. 17, pp. 104–115, 2016.

[2] A. Harikumar, F. Bovolo, and L. Bruzzone, "A local projection-based approach to individual tree detection and 3-D crown delineation in multistoried coniferous forests using high-density airborne LiDAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1168–1182, Feb. 2019, doi: 10.1109/TGRS.2018.2865014.

[3] F. E. Fassnacht, E. Schmidt-Riese, T. Kattenborn, and J. Hernández, "Explaining Sentinel-2-based dNBR and RdNBR variability with reference data from the bird's eye (UAS) perspective," *International Journal of Applied Earth Observation and Geoinformation*, vol. 95, Art. no. 102262, 2021, doi: 10.1016/j.jag.2020.102262.

[4] S. A. U. V. Natesan, "Individual tree species identification using Dense Convolutional Network (DenseNet) on multitemporal RGB images from UAV," *Journal of Unmanned Vehicle Systems*, vol. 8, no. 4, pp. 310–333, Dec. 2020, doi: 10.1139/juvs-2020-0007.

[5] J. Talebi and Z. Azizi, "Enhancing land feature classification with the BTR Extractor: A novel software package for high-accuracy analysis of aerial laser scan data," *MethodsX*, vol. 14, Art. no. 103090, 2025, doi: 10.1016/j.mex.2025.103090.

[6] Z. Azizi and M. Miraki, "Individual urban trees detection based on point clouds derived from UAV-RGB imagery and local maxima algorithm: A case study of Fateh Garden, Iran," *Environment, Development and Sustainability*, vol. 26, no. 1, pp. 2331–2344, Jan. 2024, doi: 10.1007/s10668-022-02853-3..

[7] Z. Azizi, A. Hosseini, and Y. Iranmanesh, "Estimating biomass of single oak trees using terrestrial photogrammetry," *Journal of Environmental Science and Technology*, vol. 19, no. 4, pp. 81–93, 2017.

[8] H. Huang, X. Li, and C. Chen, "Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 7, pp. 2256–2269, Jul. 2018, doi: 10.1109/JSTARS.2018.2830410.

[9] K. Korznikov, D. Kislov, T. Petrenko, V. Dzizyurova, J. Doležal, P. Krestov, and J. Altman, "Unveiling the potential of drone-borne optical imagery in forest ecology: A study on the recognition and mapping of two evergreen coniferous species," *Remote Sensing*, vol. 15, no. 18, Art. no. 4394, 2023, doi: 10.3390/rs15184394.

[10] M. Miraki and H. Sohrabi, "Integrating UAV imagery, manual, and automatic crown delineation for estimating green space water requirements in semi-arid climate," *SSRN Electronic Journal,* 2024, doi: 10.2139/ssrn.4953253.

[11] A. Daryaei, H. Sohrabi, C. Atzberger, and M. Immitzer, "Fine-scale detection of vegetation in semi-arid mountainous areas with focus on riparian landscapes using Sentinel-2 and UAV data," *Computers and Geosciences*, vol. 147, Art. no. 105686, 2021, doi: 10.1016/j.cageo.2020.105686.

[12] H. Zhong, Z. Zhang, H. Liu, J. Wu, and W. Lin, "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*, vol. 15, no. 2, Art. no. 293, 2024, doi: 10.3390/f15020293.

[13] L. P. Osco, Q. Wu, E. L. de Lemos, W. N. Gonçalves, A. P. M. Ramos, J. Li, and J. Marcato Junior, "The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot," *International Journal of Applied Earth Observation and Geoinformation*, vol. 124, Art. no. 103540, 2023, doi: 10.1016/j.jag.2023.103540.

[14] R. Ravindran, Y. Treitz, and D. Iwaszczuk, "Investigation on dimensionality reduction methods for tree-crown segmentation in hyperspectral imagery using Segment Anything Model," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-G-2025, pp. 705–712, 2025, doi: 10.5194/isprs-annals-X-G-2025-705-2025.

[15] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment Anything," *in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* Paris, France, 2023, pp. 4015–4026, doi: 10.1109/ICCV51070.2023.00371.

[16] M. Teng, A. Ouaknine, E. Laliberté, Y. Bengio, D. Rolnick, and H. Larochelle, "Assessing SAM for tree crown instance segmentation from drone imagery," Computer Vision and Pattern Recognition, 2025. https://arxiv.org/abs/2503.20199.

[17] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 173, pp. 24–49, Mar. 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[18] C. Hu, S. Dai, S. Wu, Q. Ye, and H. Yan, "Per-SAM-MCPA: A lightweight framework for individual tree crown segmentation from UAV imagery," *Remote Sensing*, vol. 18, no. 10, Art. no. 1559, 2026, doi: 10.3390/rs18101559.

[19] J. Karthigesu, T. Owari, S. Tsuyuki, and T. Hiroshima, "Improving individual tree crown detection and species classification in a complex mixed conifer–broadleaf forest using two machine learning models with different combinations of metrics derived from UAV imagery," *Geomatics*, vol. 5, no. 3, Art. no. 32, 2025, doi: 10.3390/geomatics5030032.

[20] J. L. Vaschetti, E. Arnaudo, and C. Rossi, "TreePseCo: Scaling individual tree crown segmentation using large vision models," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-M-7, pp. 275–282, 2025, doi: 10.5194/isprs-archives-XLVIII-M-7-2025-275-2025.

[21] A. Carraro, M. Sozzi, and F. Marinello, "The Segment Anything Model (SAM) for accelerating the smart farming revolution," *Smart Agricultural Technology*, vol. 6, Art. no. 100367, 2023, doi: 10.1016/j.atech.2023.100367.