

Assessing Pretrained GIS Deep Learning Models for Urban Tree Inventory Mapping

Moyo Velile, Kačmarík Michal

Department of Geoinformatics, Faculty of Mining and Geology, VŠB-Technical University of Ostrava, Ostrava-Poruba, 70 800, Czech Republic, velile.moyo.st@vsb.cz, michal.kacmarik@vsb.cz

Abstract

Urban tree inventories are essential for sustainable city planning, environmental monitoring, and climate adaptation, yet traditional field-based surveys are time-consuming and costly. The increasing availability of high-resolution RGB aerial imagery, combined with deep learning, enables automated individual tree detection (ITD) in complex urban environments. This study evaluates and compares two deep learning-based object detection approaches for urban tree detection using RGB aerial imagery: a pretrained Tree Detection model implemented in ESRI ArcGIS Pro based on the DeepForest framework and a YOLOv9-based detection model implemented via the Deepness plugin in QGIS. Both approaches were applied to a 0.78 km² urban area in the Poruba district of Ostrava, Czech Republic, and validated using municipal tree inventory data as ground truth. Detection performance was assessed through spatial matching between detected trees and reference points using a fixed distance threshold and quantified using precision, recall, and F1-score. The results reveal a clear trade-off between detection completeness and detection reliability. The ArcGIS Pro approach achieved higher recall (0.63), detecting a larger proportion of reference trees, but showed lower precision (0.32) due to a high number of false positives. In contrast, the QGIS Deepness approach produced more conservative detections with higher precision (0.64) but substantially lower recall (0.26). These findings highlight the importance of selecting detection methods based on application-specific requirements in urban tree inventory mapping.

Keywords: individual tree detection, urban forestry, deep learning, object detection, RGB aerial imagery

Introduction

Urban trees play an important role in supporting healthier and more livable cities by improving air quality, reducing heat, and contributing to overall environmental comfort. Accurately detecting and quantifying individual trees is important because it allows cities to monitor their urban forests, maintain updated inventories and understand how tree distributions change over. With the growing availability of high-resolution RGB imagery from aerial surveys and drones, it is possible to detect and map individual trees more efficiently. This process, known as individual tree detection (ITD), focuses on automatically locating and delineating tree crowns from imagery (Hu et al., 2017; Michez et al., 2016). Traditional field surveys are often slow and expensive, while automated ITD provides a faster, repeatable way to track trees across large urban areas. These maps support better decisions related to urban heat mitigation, biodiversity planning, and long-term green-infrastructure management.

Early ITD studies relied on classical image-processing techniques such as watershed segmentation, local maxima filtering, and mean-shift clustering to separate tree crowns based on colour, brightness, and texture patterns (Yu et al., 2022; Wu et al., 2023). Although these techniques can perform adequately in simpler environments, they often face difficulties in dense or irregular canopy structures, where overlapping crowns, shadows, and nearby buildings introduce substantial complexity. As the field progressed, machine-learning and deep-learning approaches began to offer more powerful solutions by learning different features directly from labelled examples. Classical machine-learning algorithms such as Support Vector Machines

and Random Forests were applied to classify tree objects using a mix of spectral, geometric, and contextual information (Kouvaras & Petropoulos, 2024; Brodić et al., 2022). More recently, deep learning has become the leading methodology for ITD, with models such as Mask R-CNN, Faster R-CNN, YOLO, and DeepForest widely used for detecting and delineating tree crowns from RGB imagery (Zhang et al., 2022; Stark et al., 2025). Although these models typically achieve higher accuracy than earlier approaches, there are still open questions about their generalization across different environments, the amount of training data required, and which architectures perform best in complex urban settings. These challenges highlight the need for more comparative studies to evaluate the strengths and limitations of state-of-the-art ITD methods based solely on RGB imagery.

The aim of this research is to systematically evaluate deep learning architectures implemented in vastly used GIS software for object detection of individual trees from RGB aerial imagery. By integrating predictions with ground-truth tree data, the study investigates each model's accuracy, generalization ability, and sensitivity to training data conditions in urban landscapes.

Data and Method

Area of Interest

The area of interest is located in the Poruba district of Ostrava, Czech Republic. It covers approximately 0.78 km² and is bounded by Opavská Street, Hlavní třída, 17. listopadu Street, and Porubská Street. The study area represents a heterogeneous urban environment with residential buildings, transportation infrastructure, and managed green spaces.

Data Sets

The primary dataset used in this study was high-resolution RGB aerial orthophotos obtained from the Orthophoto of the Czech Republic (Ortofoto ČR) provided by the Czech Office for Surveying, Mapping and Cadastre (ČÚZK) through a public Esri ArcGIS Server view service. The orthophotos were accessed via WMTS/WMS services and are provided as pre-generated map tiles in the S-JTSK / Krovak East North coordinate reference system (EPSG:5514).

The orthophoto data have a spatial resolution of 12,5 cm per pixel, enabling detailed visual discrimination of individual tree crowns. The imagery is stored as tiled raster data with a tile size of 256 × 256 pixels, supporting efficient visualization and analysis at large scales.

Additionally, ground reference data were provided by the municipal tree inventory, consisting of a point-based dataset representing individual trees within the study area. Each record includes a unique tree identifier, species information, and measured tree characteristics such as total height, stem height, crown spread, and diameter at breast height (DBH), along with inspection and condition attributes. The spatial locations of these trees were used as ground truth for validating tree detection results.

Instrumentation

The study used ArcGIS Pro with pretrained deep learning packages (DLPKs) provided by Esri (<https://www.esri.com/en-us/arcgis/deep-learning-models>) for tree detection in high-resolution imagery. The Tree Detection model is based on the DeepForest framework and employs a RetinaNet object detection architecture implemented in the PyTorch torchvision library.

The model was designed for use with 8-bit, three-band RGB imagery at spatial resolutions between 10 and 25 cm, matching the characteristics of the orthophoto data used in this study. The model was pretrained using the NEON Tree Crowns and NEON Tree Benchmark datasets, provided by the Weecology Lab at the University of Florida. Model fine-tuning was not

performed, as the pretrained models cannot be further trained using ArcGIS Pro tools. (<https://www.arcgis.com/home/item.html?id=4af356858b1044908d9204f8b79ced99>)

In addition to ESRI ArcGIS Pro, open-source geographic information system software QGIS was used with the Deepness plugin, which enables inference of deep neural network models in ONNX format for raster-based object detection. The Deepness plugin supports deep learning–based analysis of orthophotos, allowing processing to be limited to user-defined spatial extents and integration of model outputs as GIS layers.

Tree-top detection was performed using a pretrained YOLOv9-based object detection model from the Deepness Model Zoo

(https://qgis-plugin-deepness.readthedocs.io/en/latest/main/main_model_zoo.html).

The model was designed for high-resolution aerial imagery with an input resolution of 10 cm per pixel and an input tile size of 640 × 640 pixels. The model was trained on a combination of publicly available tree datasets and outputs point-based detections representing individual treetops.

Methodology

Individual tree detection was performed using two deep learning–based object detection approaches applied to high-resolution RGB aerial imagery. The first approach used the pretrained Tree Detection deep learning package in ArcGIS Pro to generate bounding boxes representing individual tree detections. The second approach used the Deepness plugin in QGIS to produce bounding-box detections of treetops.

For both methods, bounding box centroids were extracted and used as point representations of detected tree locations. Municipal tree inventory data, provided as point locations of individual trees, were used as ground reference data. All datasets were projected to a common coordinate reference system and clipped to the same area of interest.

Detection performance was evaluated using nearest-neighbor spatial matching between detected tree centroids and ground truth tree points with a fixed distance threshold of 3 m. Two complementary spatial joins were performed: detected trees were matched to ground truth points to identify true positives and false positives, and ground truth points were matched to detected trees to identify true positives and false negatives. Based on these classifications, standard object detection metrics (precision, recall, and F1-score) were computed to quantify detection reliability and completeness.

Evaluation Metrics

Detection performance was evaluated using three standard object detection metrics: precision, recall, and F1-score. Precision measures the proportion of detected trees that correctly correspond to reference trees and reflects the reliability of the detection results. Recall measures the proportion of reference trees that were successfully detected and reflects the completeness of the detection results. The F1-score combines precision and recall into a single metric, providing a balanced assessment of overall detection performance.

Precision, recall, and F1-score were calculated from the numbers of true positives (TP), false positives (FP), and false negatives (FN) using the following equations:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

These metrics were used to compare the detection performance of the ArcGIS Pro and QGIS Deepness object detection approaches.

Results

A total of 2,810 reference trees were present within the study area based on the municipal tree inventory. The QGIS Deepness approach detected 889 tree objects, while the ArcGIS Pro approach detected 5,812 tree objects. Figure 1 below presents the study area in Poruba, and shows a visual comparison between municipal reference trees and tree detections produced by the ArcGIS Pro and QGIS models.



Fig. 1. Tree detections shown over the RGB orthophoto within the study area.

Table 1 summarizes the object detection performance of the ArcGIS Pro and QGIS Deepness approaches based on precision, recall, and F1-score.

Table 1. Object detection performance metrics.

Method	Precision	Recall	F1-score
QGIS Deepness (YOLOv9)	0.64	0.26	0.37
ArcGIS Pro (DeepForest / RetinaNet)	0.32	0.63	0.42

Using spatial matching with a 3 m distance threshold, the ArcGIS Pro model achieved a higher recall (0.63), detecting a larger proportion of the reference trees, but produced a large number

of false positives, resulting in lower precision (0.32). In contrast, the QGIS Deepness model detected fewer trees overall and achieved higher precision (0.64), but exhibited low recall (0.26), missing a large proportion of reference trees. The resulting F1-scores were 0.42 for ArcGIS Pro and 0.37 for QGIS Deepness, reflecting a trade-off between detection completeness and detection reliability.

Conclusion

This study shows that pretrained deep learning models for urban tree detection exhibit contrasting performance characteristics when applied to high-resolution RGB aerial imagery. The ArcGIS Pro approach favors detection completeness and is therefore more suitable for applications such as inventory updating, whereas the QGIS Deepness approach produces more conservative detections with fewer false positives. These findings indicate that model selection should be guided by the specific objectives of the application, and that combining detection strategies or refining model parameters may improve overall performance in urban environments.

References

- Brodić, N., Cvijetinović, Ž., Milenković, M., Kovačević, J., Stančić, N., Mitrović, M., & Mihajlović, D. (2022). Refinement of Individual Tree Detection Results Obtained from Airborne Laser Scanning Data for a Mixed Natural Forest. *Remote Sensing*, 14(21). <https://doi.org/10.3390/rs14215345>
- Hu, X., Chen, W., & Xu, W. (2017). Adaptive mean shift-based identification of individual trees using airborne LiDAR data. *Remote Sensing*, 9(2). <https://doi.org/10.3390/rs9020148>
- Kouvaras, L., & Petropoulos, G. P. (2024). A Novel Technique Based on Machine Learning for Detecting and Segmenting Trees in Very High Resolution Digital Images from Unmanned Aerial Vehicles. *Drones*, 8(2). <https://doi.org/10.3390/drones8020043>
- Michez, A., Piégay, H., Lisein, J., Claessens, H., & Lejeune, P. (2016). Mapping of riparian vegetation using unmanned aerial system imagery: A supervised classification approach. *International Journal of Applied Earth Observation and Geoinformation*, 44, 88–94. <https://doi.org/10.1016/j.jag.2015.06.014>
- Stark, T., Pardo, J., Wurm, M., Leichtle, T., Martin, K., & Taubenbock, H. (2025). Urban Tree Detection: Comparing YOLOv8 and DeepForest for Accurate Single-Tree Identification from Aerial Imagery. 2025 Joint Urban Remote Sensing Event, JURSE 2025. <https://doi.org/10.1109/JURSE60372.2025.11076000>
- Wu, H., Zhuang, M., Chen, Y., Meng, C., Wu, C., Ouyang, L., Liu, Y., Shu, Y., Tao, Y., Qiu, T., & Li, J. (2023). Urban Treetop Detection and Tree-Height Estimation from Unmanned-Aerial-Vehicle Images. *Remote Sensing*, 15(15). <https://doi.org/10.3390/rs15153779>
- Yu, K., Hao, Z., Post, C. J., Mikhailova, E. A., Lin, L., Zhao, G., Tian, S., & Liu, J. (2022). Comparison of Classical Methods and Mask R-CNN for Automatic Tree Detection and Mapping Using UAV Imagery. *Remote Sensing*, 14(2). <https://doi.org/10.3390/rs14020295>
- Zhang, L., Lin, H., & Wang, F. (2022). Individual Tree Detection Based on High-Resolution RGB Images for Urban Forestry Applications. *IEEE Access*, 10, 46589–46598. <https://doi.org/10.1109/ACCESS.2022.3171585>