Efficient Large-scale Mapping of *Acacia Tortilis* Trees Using UAV-based Images and Transformer-based Semantic Segmentation Architectures

Mohamed Barakat A. Gibril[1], Abdallah Shanableh[1, 2, 3], Rami Al-Ruzouq[1, 2], Nezar Hammouri[1], Fouad Lamghari [4], Safa M. Ahmed[1], Ahmed Mansour[1], Ratiranjan Jena[1], Hani Shanableh[1], Mohammed Abdulraheem Almarzouqi[5], Nedal Salem Alafayfeh[5], Simon Zerisenay Ghebremeskel[4]

[1] GIS and Remote Sensing Center, Research Institute of Sciences and Engineering, University of Sharjah, Sharjah 27272, United Arab Emirates;
[2] Civil and Environmental Engineering Department, University of Sharjah, Sharjah, Sharjah 27272, United Arab Emirates;
[3] Scientific Research Center, Australian University, Kuwait;
[4] Fujairah Research Centre, Al-Hilal Tower, 3003, P.O. Box 666 Fujairah, United Arab Emirates;
[5] UAE Ministry of Climate Change and Environment, United Arab Emirates.

Keywords: Acacia trees, UAV, individual tree crown delineation, semantic segmentation, vision transformers.

Abstract

The *Acacia tortilis* tree, locally known as Al Samr, is one of the native trees in arid and semi-arid ecosystems. This type of tree thrives in challenging climate conditions and considerably contributes to desert ecosystems. However, Acacia trees are increasingly vulnerable to land degradation, degradation, grazing, urbanization, and the demand for wood as a fuel source. Given the ecological significance of Acacia trees and their vulnerability to various environmental threats, current information on their distribution and population is essential for effectively conserving and managing this native species. This study aims to map Acacia trees from unmanned aerial vehicle (UAV)-based images using deep learning techniques. First, a comprehensive field campaign was conducted to record the locations of Acacia trees within the study area. Thereafter, the Segment Anything model was fine-tuned to delineate tree boundaries from the UAV data, facilitating the preparation of ground-truth labels. Subsequently, Mask2Former, a semantic segmentation architecture utilizing a dual-attention vision transformer backbone, was implemented to segment the Acacia trees. The performance of the proposed architecture was compared against those of Mask2Former models based on alternative architectures, including Swin Transformer, Grounded Language-Image Pre-training, and EVA02, a transformer-based visual representation pre-trained model. Results demonstrated that the proposed approach outperformed the evaluated models, efficiently delineating Acacia trees and achieving a mean intersection-over-union of 83.43% and a mean F-score of 90.27.%. The proposed approach offers valuable means for building tree inventories, updating geospatial databases, and promoting sustainable management of native Acacia trees.

1. Introduction

Acacia tortilis, known as Umbrella thorn and Al Samr (local name), features a small to medium-sized slow growing tree predominantly found in arid and semi-arid ecosystems of eastern and northern Africa and the Middle East (Doran et al., 1983). Given the longevity of Acacia trees, an estimated average lifespan of about 200 years, and the ability to tolerate harsh environmental conditions, they play a significant role in desert ecosystems (Isaacson et al., 2017; Ross and Burt, 2015). These trees are highly drought-resistant and can tolerate extreme conditions, including high salinity, temperature, and seasonal waterlogging (AbdElRahman and Krzywinski, 2008).

In the United Arab Emirates (UAE), Acacia trees are primarily concentrated in the northeastern regions, where they are among the most visually prominent tree species (Brown and Feulner, 2023). In Abu Dhabi, located in the southwestern portion of the UAE, Acacia trees are primarily found on the gravel plains of the eastern regions, such as Al Ain and Jabal Hafit, with a smaller population in Sila in the west. The Acacia tree is an essential resource, providing essential livestock feed, protecting soil from erosion, and supporting beekeeping and honey production.

Acacia trees are under continuous threat due to land degradation, overgrazing, urban expansion, and the collection of wood for fuel (Environment Agency - Abu Dhabi, 2020). Numerous government initiatives aim to boost the population of Acacia trees. For example, in 2018, the UAE Ministry of Climate Change and Environment launched a program to distribute 54 million Acacia seeds free of charge to the public, encouraging the growth and expansion of these trees.

Given that Acacia trees are spread across vast, hard-to-reach areas (such as isolated or rugged terrain), field-based estimations of their extent and population are not only time-consuming but also costly and challenging. Remote sensing has been an indispensable means for mapping and monitoring vegetation and tree species. In recent years, unmanned aerial vehicle (UAV)-based remote sensing platforms and sensors have witnessed significant advancements, resulting in a notable growth in the availability and use of very-high spatial resolution remotely sensed data for mapping and monitoring vegetation and tree species.

Deep learning (DL) techniques, particularly convolutional neural networks (CNN) and vision transformers, have been extensively used in identifying, detecting, and mapping individual tree crowns from UAV-based images using different vision tasks. Some of the widely used DL tasks include object detection (Dakov and Petrova-Antonova, 2024), semantic (Luo et al., 2024), and instance segmentation (Gibril et al., 2024, 2022; Xie et al., 2024).

Over the past few years, various semantic segmentation architectures, such as U-Net (Ronneberger et al., 2015: Gazzea et al., 2022), leveraging various backbones (CNN and transformer-based architectures), have been widely adopted for delineating

https://doi.org/10.5194/isprs-annals-X-G-2025-285-2025 |

tree crowns from remotely sensed data. Given the strong ability of deep vision transformers in capturing global and contextual information from remotely sensed data, different transformer-based semantic segmentation architectures have shown superior performance in mapping tree crowns and improved accuracy and efficiency (Gibril et al., 2023). Lin et al. (2024) assessed various CNN and transformer-based models for mapping olive trees using high-resolution satellite data at a sub-national scale. Their findings revealed that transformer-based models surpassed CNN-based models in accurately identifying olive trees at the pixel level. Al-Ruzouq et al. (2024) underscored the effectiveness and feasibility of deep vision transformers for large-scale segmentation of individual date palm trees using multi-city WorldView-3 satellite datasets.

To the best of the authors' knowledge, accurate information regarding the distribution and population of Acacia trees in the UAE is either scarce or unavailable. Large-scale mapping of Acacia trees in diverse urban and agricultural landscapes presents challenges due to the coarse spatial resolution of satellite images. Likewise, the limited spectral resolution of UAV-based RGB images complicates precise tree identification. This study leverages transformer-based models to integrate global and contextual information, enabling the accurate mapping of Acacia trees across the Fujairah Emirate using extensive UAV-based datasets.

2. Materials and Methods

2.1 Study Area and Dataset

The study area covers multiple urban and farmlands across the Fujairah and Sharjah emirates, including Fujairah City and Kalba, with a total area of 25 km². The experimental site encompasses a wide range of tree species, such as *Phoenix dactylifera L*. (date palm), *Prosopis cineraria* (Ghaf), Christ's thorn jujube (Sidr), *Prosopis juliflora* (Mesquite), *Azadirachta indica* (Neem), and *A. tortilis*.

The dataset for this study was collected using the senseFly eBee X, a fixed-wing, survey-grade UAV system. The UAV was equipped with a 20 MP S.O.D.A. (Sensor Optimized for Drone Applications) ultra-compact digital camera. The flight altitude was set at 122 m following permissions granted by civil aviation authorities. The data acquisition was conducted with 70% horizontal and 40% vertical overlaps. The images were captured on clear, cloud-free days between 9:00 a.m. and 1:00 p.m., with a ground sampling distance of 2.5 cm per pixel.

2.2 Field Campaign

In the initial phase of the analysis, an extensive field campaign was undertaken to collect the coordinates of representative ground-truth data for Acacia trees. These data serve as a critical reference for developing and evaluating DL models (Piragnolo et al., 2021). The coordinates and photos of the various Acacia trees were recorded using the ArcGIS Field Map mobile application. The Acacia trees exhibited significant variation in crown size, height, degree of greenness, and surrounding landscape. A total of 9100 Acacia trees were collected and widely distributed across the study area.

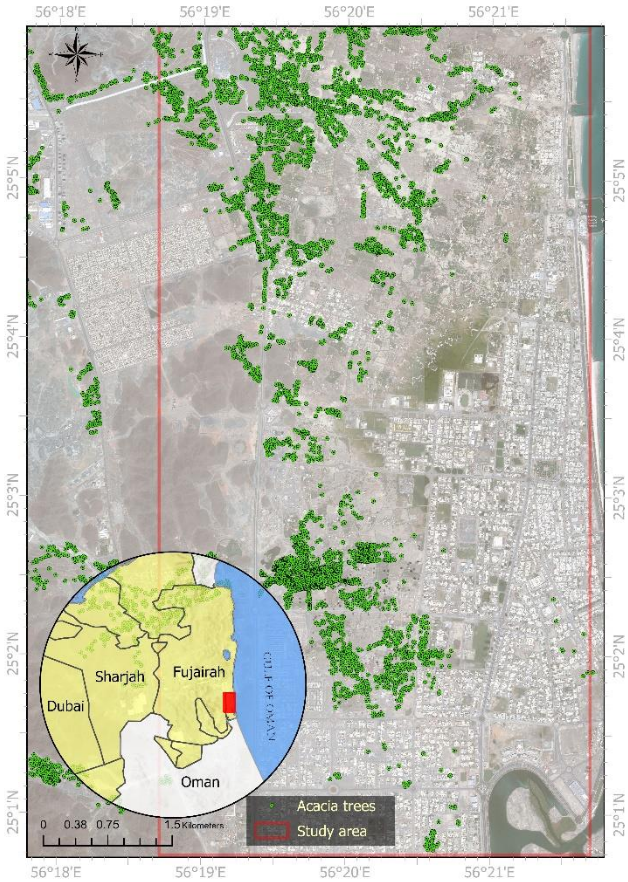

Figure 1. Geographical location of the study area.

2.3 Data Preparation

Manual annotation of Acacia trees in large-scale UAV data is a time-consuming task. This study utilizes Meta's Segment Anything model (Kirillov et al., 2023: Pirotti et al., 2017) along with a low-rank-based fine-tuning strategy to accelerate this process (Zhang and Liu, 2023). The approach involves fine-tuning the SAM on a small dataset to delineate tree boundaries. Subsequently, the automatically generated boundaries from the fine-tuned model are refined and improved. The final Acacia tree boundaries were selected based on ground-truth data and expert image interpretation. The preparation of data for semantic segmentation involves organizing it into image-mask pairs, with each mask accurately representing the delineated Acacia trees in the corresponding image. The study area was divided into three zones for training, validation, and testing. Each zone was further subdivided into 1024×1024 image-mask pairs. A total of 11,067 pairs were allocated for training, 1010 for validation, and 800 for testing the models. Figure 2 illustrates an example of image-mask pairs selected from the training dataset.

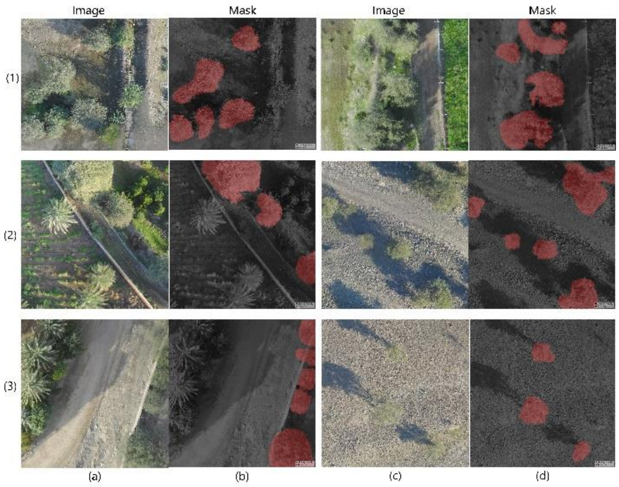

Figure 2. Examples of image-mask pairs representing Acacia trees.

2.4 Semantic Segmentation

Semantic segmentation models are widely utilized in DL for pixel-level classification across a broad range of remote sensing applications. In this study, we utilized the Masked-attention Mask Transformer (Mask2Former) architecture (Cheng et al., 2022), with Dual Attention Vision Transformers (DaViT) (Ding et al., 2022) as the backbone, to map Acacia trees from UAV-based images.

The Mask2Former architecture entails a backbone architecture, a pixel decoder, and a transformer decoder. Mask2Former incorporates low- and high-resolution features and limits computational growth by using a multiscale deformable attention transformer (Zhu et al., 2020) as a pixel decoder. The transformer decoder, which contains a masked attention operator, receives multiscale feature maps and focuses on features within the foreground of the predicted mask for each query instead of considering the entire feature map.

This study adopted DAViT architecture as a backbone network of Mask2Former. DAViT captures local and global features using two complementing attention mechanisms: spatial window and channel group attention. Although the spatial attention mechanism focuses on local details by considering interactions across spatial locations within the image, the channel attention mechanism captures global and contextual information through attention across different feature channels. The performance of Mask2Former with a DaViT backbone was compared with that of Mask2Former with other backbones, including Swin transformer (Liu et al., 2021), Transformer-based visual representation pre-trained (EVA02) (Fang et al., 2024), and Grounded Language-Image Pre-training (GLIP) (Li et al., 2022).

2.5 Accuracy Metrics

The segmentation quality of Acacia trees was evaluated using two standard semantic segmentation metrics, namely, mean intersection over union (mIoU) and mean F-score (mF-score). The formulas for calculating mIoU and mF-score are expressed in Equations 1–4. Mask2Former models with different backbones were evaluated on the validation dataset every 5000 iterations, with the best-performing weights chosen to assess their performance on the testing dataset.

$$\mathrm{IoU} = \frac{\mathrm{TP}}{(\mathrm{FP} + \mathrm{TP} + \mathrm{FN})}, \quad (1)$$

$$\mathrm{mIoU} = \frac{1}{2}(\mathrm{IoU}_{\mathrm{backgrounds}} + \mathrm{IoU}_{\mathrm{Acacia\ tree}}), \quad (2)$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \quad (3)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad (4)$$

$$\text{F-score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (5)$$

$$\mathrm{mF-score} = \frac{1}{2}(\text{F-score}_{\mathrm{backgrounds}} + \text{F-score}_{\mathrm{Acacia\ tree}}), \quad (6)$$

where TP = true positive;
FP = false positive;
FN = false negative.

3. Results

This study aimed to harness vision transformers' capabilities of capturing global and contextual features, enabling accurate mapping of Acacia trees across diverse urban and heterogeneous agricultural landscapes. The Mask2Former architecture, integrated with the DAViT backbone network, was utilized to efficiently leverage spectral and spatial information. Additionally, the performance of Mask2Former was evaluated with various backbone architectures, including Swin Transformer, EVA02, and GLIP. The tiny version of these models was selected in this investigation. The experiments were conducted using the PyTorch, MMsegmentation, and MMPretrain frameworks. The models were trained and evaluated on a 64 GB of RAM workstation powered by an NVIDIA Titan RTX graphics card, which provides 24 GB of dedicated memory. The proposed model was trained using the AdamW optimizer, with a batch size of 2 and a learning rate of 0.0001.

Model complexity, often assessed based on memory consumption and computational demands, plays a crucial role in comparing different deep learning (DL) models. More intricate and deeper DL architectures typically involve a higher number of parameters and sophisticated design structures. The total number of parameters for each backbone is as follows: DAViT has 28.36 M parameters, Swin Transformer has 28.29 M, GLIP has 27.521 M, and EVA02 has 5.759 M. The training time for Mask2Former with the DAViT backbone was 40.89 h, 36.3 h with EVA02, 24.1 h with GLIP, and 19.4 h with Swin Transformer. Figure 3 illustrates the computed mF-scores at every 5000 training iterations for each evaluated Mask2Former model with different backbones.

Table 1 presents the highest mIoU and mF-score achieved on the validation and testing datasets. The proposed approach, Mask2Former with the DAViT backbone, outperformed the other models on the validation and testing datasets. This approach achieved mIoUs of 82.3% and 84.06% and mF-scores of 89.5% and 90.7% on the validation and testing data, respectively. The Mas2Former models based on GLIP and Swin Transformer also demonstrated strong performance on the testing dataset. Specifically, the GLIP-based model achieved an mF-score of 90.42% and a mIoU of 83.67%, while the Swin Transformer-based model attained an mF-score of 90.27% and an mIoU of 83.43%.

https://doi.org/10.5194/isprs-annals-X-G-2025-285-2025

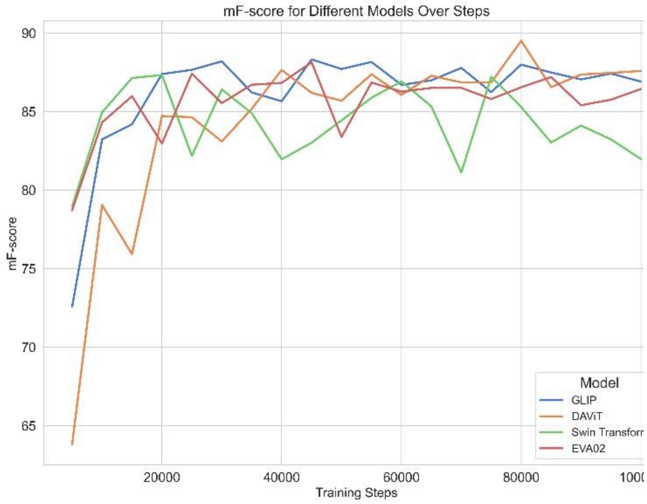

Figure 3. mF-score values computed from the validation dataset over the training steps for each evaluated Mask2Former model with different backbones.

Backbone	Validation		Testing	
	mIoU	mF-score	mIoU	mF-score
Swin transformer	79.28	87.33	83.43	90.27
GLIP	80.63	88.32	83.67	90.42
EVA-2	81	88.60	81.7	89.0
DAViT	82.3	89.50	84.06	90.7

Table 1. Experimental results of the evaluated segmentation architectures.

Figure 4 presents a set of images selected from the testing dataset (first column), along with their ground-truth data (second column) and the segmentation results of the proposed approach (third column). The proposed approach successfully segmented Acacia trees of varying sizes and in diverse surrounding environments (Figure 2). Moreover, the model effectively recognized and delineated Acacia trees in images collected on different dates and times despite variations in shadow locations due to changing lighting conditions.

The field campaign in this study involved capturing the locations of only 9100 Acacia trees distributed over a large area for model development. Given the vast coverage of the study area and the scattered nature of these samples, the number of labeled trees must be augmented using image interpretation of UAV data and Google Street View images. This approach leveraged Acacia trees' distinct appearance and shadows to identify additional samples. However, the process faced challenges due to significant variations in the color of Acacia trees—depending on their water content—and the potential inclusion of similar species, such as *Prosopis juliflora*, and the presence of artifacts in some parts of the images. Although the image interpretation helped in expanding the dataset, it may have introduced minor inaccuracies due to misclassification.

The proposed model successfully delineated Acacia trees in heterogeneous scenes (Figure 5a). The model accurately identifies the visible portion of an Acacia tree, even though another tree obscures part of it (Figure 5b). However, in some instances, such as in Figure 3c, the model struggles to detect Acacia trees in mountainous areas where the trees blend with the background (indicated by the yellow rectangle). Additionally, minor misclassifications were observed, with some parts of *P. juliflora* being mistaken for Acacia trees (Figure 3d).

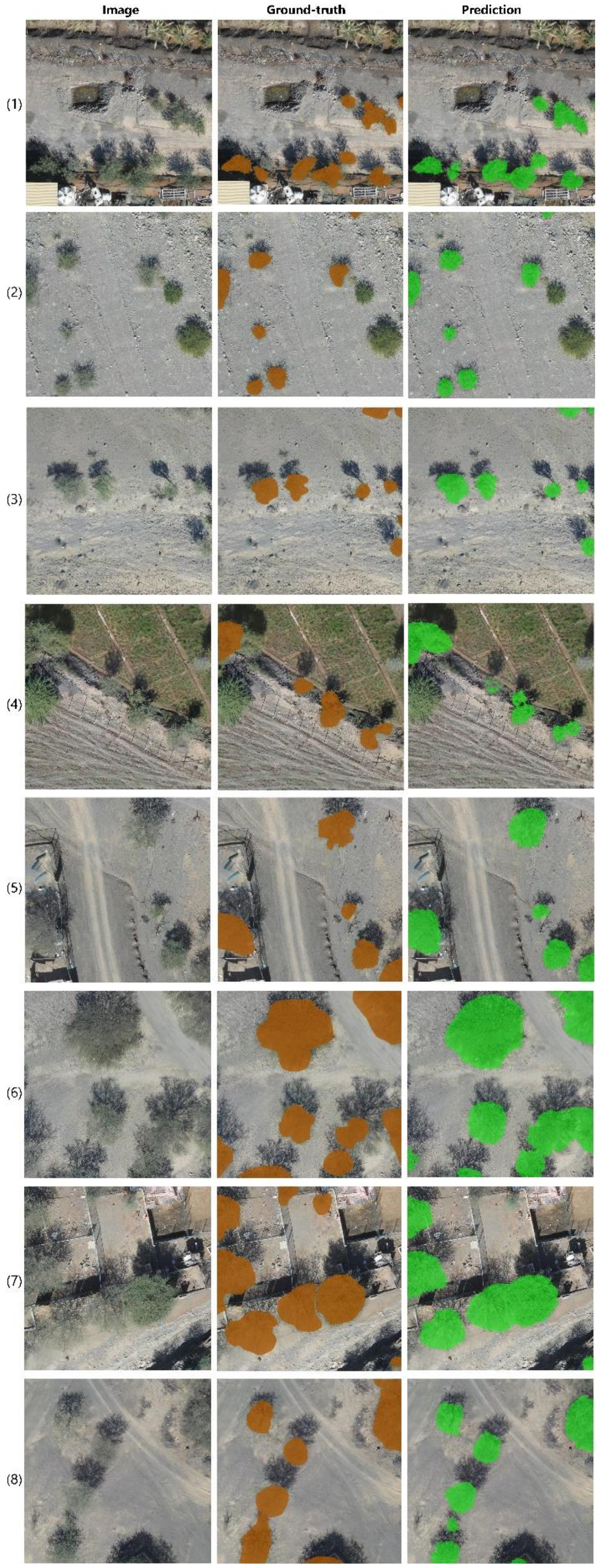

Figure 4. Results of the proposed approach based on selected images from the testing dataset.

https://doi.org/10.5194/isprs-annals-X-G-2025-285-2025 | © Author(s) 2025. CC BY 4.0 License.

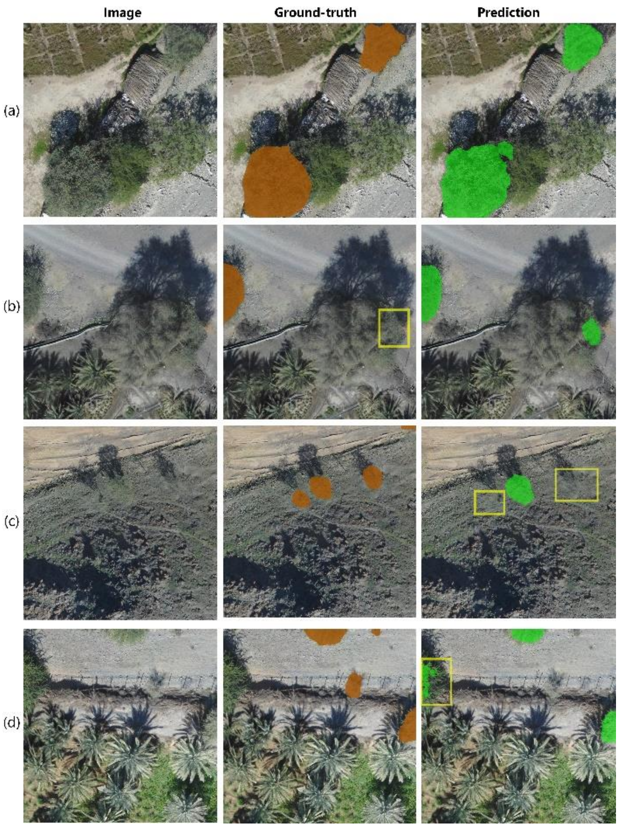

Figure 5. Examples of the proposed model's performance in detecting Acacia trees across various challenging scenes.

To the best of the author's knowledge, this study represents one of the first efforts to map Acacia trees from large-scale UAV-based imagery. The proposed method provides an automated approach for building a comprehensive dataset of this native species and can be adapted for mapping other native trees across the UAE. Although the proposed approach outperformed the evaluated architectures on the validation and testing datasets, it required a longer processing time. Future improvements could include refining the training and testing datasets and incorporating additional samples from diverse UAE regions to enhance the model's generalizability. The effects of integrating CNN and transformer-based features might need to be further investigated in future studies. Considering the high cost and time investment of extensive field campaigns, future studies should also explore semi-supervised DL models to effectively map Acacia trees from a limited amount of labelled data.

4. Conclusion

The Acacia tree, one of the UAE's native species, is a vital resource, providing essential livestock feed, preventing soil erosion, and supporting beekeeping and honey production, thereby substantially contributing to the UAE ecosystem. This study aimed to harness the capabilities of deep vision transformers in capturing global and contextual information from the data to map Acacia trees using large-scale UAV imagery. In this study, the Mask2Former architecture, a semantic segmentation architecture utilizing a dual attention vision transformer (DAViT) backbone, was utilized. DAViT efficiently captures and integrates the spatial and spectral features of the data. The performance of the proposed approach was compared with those of Mask2Former based on Swin transformer, GLIP, and EVA02. The proposed architecture demonstrated strong performance, outperforming the evaluated models and achieving an mIoU of 84.06% and an mF-score of 90.7%. This study underscores the effectiveness of vision transformers for large-scale mapping of Acacia trees using UAV imagery. The proposed architecture shows potential for developing tree inventories, updating geospatial databases, and facilitating the sustainable management of native Acacia trees. Furthermore, this approach can be adapted to map other tree species.

Acknowledgements

The authors extend their gratitude to the UAE Ministry of Climate Change and Environment for providing the UAV data used in this research. They also sincerely appreciate the financial support from the Fujairah Research Centre (Funded Project Number: 133049) and the research facilities provided by the University of Sharjah.

References

AbdElRahman, H.F., Krzywinski, K., 2008. Environmental effects on morphology of Acacia tortilis group in the Red Sea Hills, Northeastern Sudan and South-Eastern Egypt. For. Ecol. Manage. 255, 254–263. https://doi.org/10.1016/j.foreco.2007.09.021

Al-Ruzouq, R., Gibril, M.B.A., Shanableh, A., Bolcek, J., Lamghari, F., Hammour, N.A., El-Keblawy, A., Jena, R., 2024. Spectral–Spatial transformer-based semantic segmentation for large-scale mapping of individual date palm trees using very high-resolution satellite data. Ecol. Indic. 163, 112110. https://doi.org/10.1016/j.ecolind.2024.112110

Brown, G., Feulner, G.R., 2023. The Vegetation of the United Arab Emirates and Ecosystem Management Issues, in: A Natural History of the Emirates. Springer, pp. 121–159.

Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R., 2022. Masked-attention Mask Transformer for Universal Image Segmentation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2022-June, 1280–1289. https://doi.org/10.1109/CVPR52688.2022.00135

Dakov, V., Petrova-Antonova, D., 2024. Urban Tree Detection from Remote Sensing Data based on DeepForest Model, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 35–41. https://doi.org/10.5194/isprs-annals-X-4-W4-2024-35-2024

Ding, M., Xiao, B., Codella, N., Luo, P., Wang, J., Yuan, L., 2022. DaViT: Dual Attention Vision Transformers, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-031-20053-3_5

Doran, J.C., Turnbull, J.W., Boland, D.J., Gunn, B. V, 1983. Handbook on seeds of dry-zone acacias. A guide for collecting, extracting, cleaning, and storing the seed and for treatment to promote germination of dry-zone acacias.

Environment Agency - Abu Dhabi, 2020. The Abu Dhabi Red List of Species.

Fang, Y., Sun, Q., Wang, Xinggang, Huang, T., Wang, Xinlong, Cao, Y., 2024. EVA-02: A visual representation for neon genesis. Image Vis. Comput. 149.

https://doi.org/10.1016/j.imavis.2024.105171

Gazzea, M., … Arghandeh, R., 2022. Tree Species Classification Using High-Resolution Satellite Imagery and Weakly Supervised Learning. *IEEE Trans. Geosci. Remote Sensing* 60, 1–11. https://doi.org/10.1109/TGRS.2022.3210275

Gibril, M.B.A., Shafri, H.Z.M., Al-Ruzouq, R., Shanableh, A., Nahas, F., Al Mansoori, S., 2023. Large-Scale Date Palm Tree Segmentation from Multiscale UAV-Based and Aerial Images Using Deep Vision Transformers. Drones 7. https://doi.org/10.3390/drones7020093

Gibril, M.B.A., Shafri, H.Z.M., Shanableh, A., Al-Ruzouq, R., Hashim, S.J. bin, Wayayok, A., Sachit, M.S., 2024. Large-scale assessment of date palm plantations based on UAV remote sensing and multiscale vision transformer. Remote Sens. Appl. Soc. Environ. 34, 101195. https://doi.org/10.1016/j.rsase.2024.101195

Gibril, M.B.A., Shafri, H.Z.M., Shanableh, A., Al-Ruzouq, R., Wayayok, A., Hashim, S.J. bin, Sachit, M.S., 2022. Deep convolutional neural networks and Swin transformer-based frameworks for individual date palm tree detection and mapping from large-scale UAV images. Geocarto Int. 37, 18569–18599. https://doi.org/10.1080/10106049.2022.2142966

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, Las Vegas, NV, USA, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

Isaacson, S., Ephrath, J.E., Rachmilevitch, S., Maman, S., Ginat, H., Blumberg, D.G., 2017. Long and short term population dynamics of acacia trees via remote sensing and spatial analysis: Case study in the southern Negev Desert. Remote Sens. Environ. 198, 95–104. https://doi.org/10.1016/j.rse.2017.05.035

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R., 2023. Segment Anything, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371

Li, L.H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., Wang, L., Yuan, L., Zhang, L., Hwang, J.N., Chang, K.W., Gao, J., 2022. Grounded Language-Image Pre-training. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2022-June, 10955–10965. https://doi.org/10.1109/CVPR52688.2022.01069

Lin, C., Zhou, J., Yin, L., Bouabid, R., Mulla, D., Benami, E., Jin, Z., 2024. Sub-national scale mapping of individual olive trees integrating Earth observation and deep learning. ISPRS J. Photogramm. Remote Sens. 217, 18–31. https://doi.org/10.1016/j.isprsjprs.2024.08.003

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, pp. 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

Luo, T., Gao, W., Belotserkovsky, A., Nedzved, A., Deng, W., Ye, Q., Fu, L., Chen, Q., Ma, W., Xu, S., 2024. VrsNet - density map prediction network for individual tree detection and counting from UAV images. Int. J. Appl. Earth Obs. Geoinf. 131. https://doi.org/10.1016/j.jag.2024.103923

Pirotti, F., Kobal, M., Roussel, J.R., 2017. A Comparison of Tree Segmentation Methods Using Very High Density Airborne Laser Scanner Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W7, 285–290. https://doi.org/10.5194/isprs-archives-XLII-2-W7-285-2017

Piragnolo, M., ... Grigolato, S., 2021. Responding to Large-Scale Forest Damage in an Alpine Environment with Remote Sensing, Machine Learning, and Web-GIS. *Remote Sensing* 13, 1541. https://doi.org/10.3390/rs13081541

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Ross, Z., Burt, J., 2015. Unusual canopy architecture in the umbrella thorn acacia, Vachellia tortilis (= Acacia tortilis), in the United Arab Emirates. J. Arid Environ. 115, 62–65. https://doi.org/10.1016/j.jaridenv.2015.01.004

Xie, Y., Wang, Y., Sun, Z., Liang, R., Ding, Z., Wang, B., Huang, S., Sun, Y., 2024. Instance segmentation and stand-scale forest mapping based on UAV images derived RGB and CHM. Comput. Electron. Agric. 220, 108878. https://doi.org/10.1016/j.compag.2024.108878

Zhang, K., Liu, D., 2023. Customized Segment Anything Model for Medical Image Segmentation. pp. 1–16.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J., 2020. Deformable DETR: Deformable Transformers for End-to-End Object Detection 1–16.