Segmentation and Identification of Mediterranean Plant Species

Parminder Kaur[1], Daniela Gigante[2], Marco Caccianiga[3], Simonetta Bagella[4], Claudia Angiolini[5,6], Manolo Garabini[7], Franco Angelini[7], and Paolo Remagnino[1]

[1] Department of Computer Science, Durham University, Stockton Road, Durham DH1 3LE, UK
{parminder.kaur,paolo.remagnino}@durham.ac.uk
[2] Department of Agricultural, Food and Environmental Sciences, University of Perugia, Borgo XX giugno 74, 06121 Perugia, Italy
daniela.gigante@unipg.it
[3] Department of Biosciences, Università Sdegli Studi di Milano, Via Celoria 26, 20133 Milano, Italy
marco.caccianiga@unimi.it
[4] Department of Chemical, Physical, Mathematical and Natural Science, Università di Sassari, Via Piandanna 4, 07100 Sassari, Italy
bage@uniss.it
[5] Department of Life Sciences, Università degli studi di Siena, Via PA. Mattioli 4, 53100 Siena, Italy
[6] NBFC, National Biodiversity Future Center, Palermo, Italy
claudia.angiolini@unisi.it
[7] Centro di Ricerca "Enrico Piaggio" and Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Largo Lucio Lazzarino 1, 56126 Pisa, Italy
{manolo.garabini,frncangelini}@gmail.com

Abstract. Recently, object recognition and image segmentation have gained much attention in the computer vision field and image processing for effective object localisation and identification. Researchers have applied semantic segmentation and instance segmentation in diverse application areas. However, the least research has been performed in natural habitat monitoring or plant species identification in natural environments/surroundings. For this study, we composed a real image dataset from four habitats: forests, dunes, grasslands, and screes from various locations in Italy. Habitat expert botanists annotated the data using bounding box annotations which have been further utilised to generate the plant species masks using the recently proposed Segment Anything Model (SAM) for segmentation, localisation, and identification tasks. Extensive experimentation has been performed on habitat data with bounding boxes and masks using YOLOv8 detection and segmentation models. Comparative analysis of models, model training with different train data percentages, and the importance of masks over bounding boxes have been studied and discussed.

Keywords: Deep learning · Object identification · Plant species recognition · Instance segmentation · Habitat monitoring.

1 Introduction

Natural habitats play a vital role in the survival of humans, flora and fauna. So, they are required to be monitored and preserved for the existence of life on Earth. Although artificial intelligence is being utilised in several applications nowadays, monitoring of natural habitats is still performed chiefly through field observations by human experts, especially terrestrial habitats [1]. There are several existing challenges for monitoring the conservation status of natural habitats defined according to the Habitats Directive (92/43/EEC), such as:(1) human operators are the only practical options for monitoring; (2) human involvement provides a significant amount of subjectivity in this process, and this lessens the consistency and comparability of relevés; and (3) each habitat monitoring can be done during a limited period in a year, so with an increase in number of habitats, more number of professional surveyors are required. So, to overcome these challenges, the latest deep learning techniques can be utilised to automate the whole habitat monitoring process or to assist human experts in the field.

Lately, deep classification, detection and segmentation architectures have been widely utilised in various fields, including plant image classification, plant or leaf disease detection, plant phenotyping, and habitat monitoring through satellite images. Recently, automatic habitat or environment monitoring has been carried out using remote sensing images [25]. For instance, a NaturaSat software tool has been introduced to monitor habitats using satellite images in [13]. There are multiple public datasets available for plant/leaf recognition and plant/leaf disease detection. As an exemplar, PlantCLEF [5] organises a challenge yearly for plant specie identification by providing massive image data comprising several plant types. However, there is a scarcity of plant image data taken in the field in its habitat and annotated by human experts. Motivated by the above issues, we composed real image data of target species such as typical species (TS), characteristic species (CS), alien species (AS) and early warning species (EWS) from four habitats: forests, dunes, grasslands, and screes. The presence of TS and CS indicates favourable habitat conservation status, while the existence of EWS and AS is not a good sign for habitat health [1].

This article mainly focuses on monitoring the conservation status of natural habitats by identifying the target species in the four habitats to assist humans in assessing the habitat conservation status. After collecting images from the four habitats, they are annotated by the domain experts in their respective habitats for plant localisation and identification. We utilise the Segment Anything Model (SAM) [11] for converting bounding boxes to segmentation masks as they provide a precise understanding of the plant's shape, size, and position. Bounding box annotation in itself is an expensive task, and polygon annotation is even more complex and time-consuming if done manually, especially in the case of plants due to their complex shapes, growth in groups, and mixing up with other plants and backgrounds. However, SAM made this job more accessible by introducing a framework which can be used to retrieve appropriate segment masks as per the given bounding boxes. Then the YOLOv8 detection and segmentation models have been utilised to analyse their performance on this novel plant species data and to answer an essential question of whether the expensive polygon annotation is required or the results are almost the same even with bounding box annotation.

Contributions

The major contributions of this study are as follows:

1. To overcome the annotated data scarcity of plant species, we collected data from four habitats: forests, dunes, grasslands, and screes and ground truth generation using bounding box annotations by botanists.
2. Proposed a pipeline for plant species identification and monitoring the health of natural habitats using state-of-the-art YOLOv8 segmentation model and SAM [11].
3. To analyse whether the polygon annotation (segment mask) performs better than bounding box annotation for plant species localisation and recognition.

The rest of the article is organised as follows: Section 2 presents the related literature in plant-related tasks, YOLOv8, and SAM, Section 3 provides a brief introduction to SAM, YOLOv8, and the proposed framework, Section 4 demonstrates all the experimentation and the obtained results, and Section 5 concludes the study.

2 Related work

Researchers are mostly utilising miscellaneous deep learning approaches for plant identification due to their better performance over traditional approaches and improved accuracy in other application areas [2]. Authors in [22] have fine-tuned the pre-trained self-supervised vision transformer (ViT) on ImageNet data for plant image identification and secured a first place in PlantCLEF2022 [5] challenge. After experimentation, it has been shown that the data can be utilised to pre-train a model for plant disease recognition or other plant-related tasks. In [17], authors have introduced a new dataset consisting of 100 ornamental plant species collected from the Beijing Forestry University campus. They implemented a 26-layer ResNet model for plant identification. A lightweight deep convolutional neural network, Ayur-PlantNet, is introduced in [14] for 40 Ayurvedic plant species classification. First, plant segments are retrieved from the images and then further classified for identification.

Tea bud and leaf target detection have been performed by YOLOv8 in [21] to improve the accuracy of tea-picking robots in locating tea bud picking points in complex environment. Authors in [26] have utilised the YOLOv8 segmentation model for jujube fruit instance segmentation. Different YOLO models are compared to find the best segmentation model for the required task. The proposed YOLOseg-Jujube is robust, fast, has less computation cost, can identify jujube fruit ripeness stages, and is even accessible for real-time low-power device applications. An optimised version of YOLOv8 has been proposed in [10] by incorporating Simulated Annealing (SA) for finding the optimal solution of the loss function in the last layer of CNN to apply in crop prevention from diseases and insect pests. After experimentation, it has been found that the modified version of YOLOv8 outperforms YOLOv7 in disease and pest identification.

The Segment Anything Model (SAM) has been exploited in various application areas for multiple purposes. The SAM's performance has been tested on a substantial medical dataset in [7] and found that it shows better results in prompt mode (prompt points and bounding boxes for mask generation) rather than everything mode (mask generation for all objects). In digital pathology, SAM showed excellent segmentation performance for large connected objects; however, unsatisfying performance for dense instance segmentation even after providing several prompts [3]. Authors in [20] have proposed Leaf Only SAM by merging post-processing steps with SAM to segment potato leaves. Leaf Only SAM is compared with Mask RCNN (finetuned on a small potato leaf dataset), and Mask RCNN outperformed Leaf Only SAM. The advantage of Leaf Only SAM is that no annotated data and extra training are required, so it can be utilised in that scenario. Authors in [24] have proposed a novel Inpaint Anything (IA) framework by integrating SAM and image inpainting.

3 Proposed methodology

This section briefly introduces SAM and YOLOv8, and the proposed pipeline for localisation and identification of plant species. The bounding box annotations are transformed to segment masks using SAM, and then species are recognised using YOLOv8.

3.1 Segment Anything Model (SAM)

Segment Anything Model (SAM) is a foundation model for image segmentation which has been recently proposed by Meta [11]. It has been trained with 1 billion masks on 11 million images and has a splendid zero shot segmentation capability. SAM utilises an MAE [6] pre-trained Vision Transformer (ViT) [4] with minimal adaptations to process high resolution inputs. It was tested on 23 datasets to evaluate its zero shot transfer potential and it showed excellent results. Given an image as input to SAM, it segments the whole image and automatically generates masks. It also provides an option of providing a prompt for mask generation as per requirement. These prompts include prompt points, bounding box coordinates, or a combination of both. We can also specify prompts for adding or removing the mask area to generate the exact mask shape.

3.2 You Only Look Once (YOLO)

YOLO (You Only Look Once) was initially proposed by Joseph Redmon and it was the first model to detect the bounding boxes and predict the class probabilities by processing the image in one go [15]. The object detection models before YOLO were remodelling the classification models to perform detection. However, in YOLO, they framed the object detection problem as a regression problem to spatially separated bounding boxes and associated class probabilities. Afterwards, multiple improved YOLO versions were introduced by various researchers, and the YOLO models transformed the computer vision field. They have been applied in miscellaneous research areas and found to be quite efficient. YOLO models are pre-trained on massive datasets such as ImageNet and COCO. So, the pre-trained model weights can be easily used to further train the model on custom datasets with less number of instances and obtain good results. These models can produce high accuracy with small model sizes, and they are faster to train as well [16].

The latest version of the YOLO series is YOLOv8, which was proposed by the Ultralytics team in the early 2023 [18]. It is an open-source state-of-the-art model distributed under the General Public License [8]. Glenn Jocher introduced YOLOv5 after minor changes in the YOLOv3 model [9]. YOLOv8 is the further improved version of YOLOv5. The significant changes incorporated in YOLOv8 are anchor-free detection, mosaic augmentation, and updates in the convolution blocks used in the model, such as replacing the C3 module with the C2f module. The YOLOv8 architecture and major changes are shown in [19].

3.3 Framework

Figure 1 illustrates the step-by-step procedure followed for plant species identification and localisation. Firstly, the plant species images are collected from four habitats. Habitat expert botanists annotate the images with bounding box annotations using the Labelbox tool [12], which are then exported as a single JSON file per habitat. Each JSON file is converted to the YOLO object detection annotation format text files corresponding to each image in the dataset. YOLOv8 object detection model is trained using this transformed dataset which is then capable of predicting bounding boxes along with species class and confidence score on test images. The bounding boxes from the YOLO detection data format are utilised to generate masks using the Segment Anything Model (SAM), which are then saved in the TXT file (along with class information) corresponding to each image which can be directly provided to YOLOv8 segmentation model for training. The trained segmentation model can predict the bounding boxes and masks along with class and confidence scores on given query images.

4 Experimental analysis

4.1 Dataset

The dataset consists of images comprising the selected target species from four habitats: forests, dunes, grasslands, and screes. Figure 2 shows the sample images of one species from each habitat. Each species name is mentioned under the corresponding image. Human experts have composed the data by visiting the

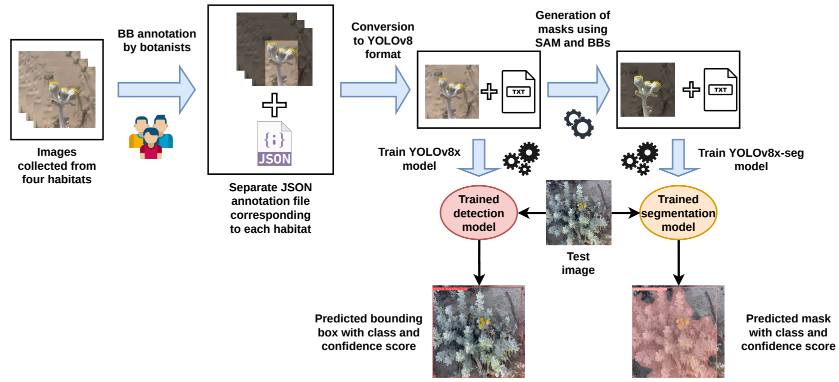

Fig. 1. Proposed pipeline for plant species detection, segmentation, and localisation

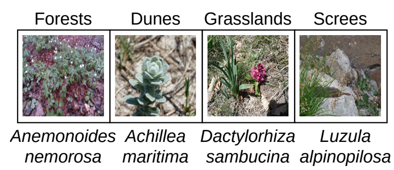

Fig. 2. Sample data from all habitats

habitat locations and doing field operations at the exact time of the year as per plant blooming. The collected images are annotated by botanists using bounding box annotations around the appropriate plant species in the image using the LabelBox tool [12]. Table 1 shows the target plant species considered for the experimentation from respective habitats. The number of training, validation and testing data instances are displayed in Table 2. The train, validation, and test splits are 70, 20, and 10.

Table 1. Various plant species from four habitats

Dunes	**Forests**	**Grasslands**	**Screes**
Carpobrotus acinaciformis	*Anemonoides nemorosa*	*Asphodelus macrocarpus*	*Luzula alpinopilosa*
Achillea maritima	*Corydalis cava*	*Dactylorhiza sambucina*	*Geum reptans*
Calamagrostis arenaria	*Doronicum columnae*	*Orchis morio*	*Oxyria digyna*
Eryngium maritimum			*Cerastium pedunculatum*
Pancratium maritimum			*Cerastium uniflorum*
Thinopyrum junceum			*Leucanthemopsis alpina*
			Ranunculus glacialis
			Saxifraga bryoides

Table 2. Training, testing and validation data splits considered for the experiments

Habitat/Split	**Train**	**Val**	**Test**	**Total**
Grasslands	645	185	92	922
Forests	765	217	110	1092
Screes	428	123	61	612
Dunes	424	121	60	605
All habitats	2262	646	323	3231

4.2 Evaluation metrics

The typical evaluation metrics for object detection and segmentation have been chosen for evaluating the model performance, such as precision, recall, F1 score, confusion matrix, mAP50, and mAP50-95 [23]. mAP50 depicts the mean Average Precision at an IoU (Intersection over Union) threshold of 0.5 and mAP50-95 over IoU thresholds of $0.5 - 0.95$ in steps of 0.05. The results section also displays bar charts, PR, and F1 score curves for performance visualisation.

4.3 Model training

YOLOv8 detection and segmentation models have been trained for 200 epochs with batch size 8 and image size 640x640. Model training is performed on default YOLOv8 hyper-parameter values. All the experimentation has been performed on Intel Core i9 24-Core Processor (Up to 5.8GHz) and 12GB NVIDIA GEFORCE RTX 3060 GPU.

4.4 Results

All the results are obtained on test data only, and test data splits (number of instances) are mentioned in Subsection 4.1. The highest values are highlighted in bold wherever required. YOLOv8x segmentation model has been exploited for its better performance than other versions for all the individual segmentation experimentation. Table 3 compares YOLOv8x object detection and segmentation models on all habitats and considers all classes of each habitat. *Inst.* represents that particular class's total number of instances. The segmentation (box or mask) results are better in the case of forests and grasslands; however, the detection results (only bounding box) are better for dunes and screes. There could be multiple reasons behind these results (dunes and screes): (1) Dunes and screes data have less number of instances and more classes (plant species) than forests and grasslands; (2) Most of the images contain a single large focused plant, so easier to represent with a bounding box as it almost covers the whole image; (3) SAM could not properly generate masks for dunes and screes. Table 4 presents the performance metric values for bounding boxes and masks obtained using the yolov8x segmentation model trained on all habitats data.

Table 3. Comparison of object detection and segmentation model results on all habitats

Classes	Inst.	Precision			Recall			mAP50			mAP50-95		
		Detect	Segment		Detect	Segment		Detect	Segment		Detect	Segment	
		Box	Box	Mask	Box	Box	Mask	Box	Box	Mask	Box	Box	Mask
Forests	552	0.741	**0.812**	0.794	0.708	**0.741**	0.718	0.755	**0.778**	0.751	0.431	**0.605**	0.495
Dunes	78	**0.871**	0.809	0.676	**0.753**	0.52	0.438	**0.799**	0.604	0.488	**0.628**	0.454	0.232
Grasslands	740	0.768	**0.791**	0.777	0.752	**0.764**	0.739	0.787	**0.813**	0.773	0.366	**0.549**	0.439
Screes	100	**0.825**	0.438	0.421	**0.504**	0.496	0.48	**0.571**	0.417	0.395	**0.379**	0.259	0.227

Figure 3 shows the segmentation model's performance when trained on different train percentages of training data on forests, keeping the validation and test

Table 4. Yolov8x segmentation model performance metrics for all habitats

Classes	Inst.	Bounding box				Mask			
		Prec	**Rec**	**mAP50**	**mAP50-95**	**Prec**	**Rec**	**mAP50**	**mAP50-95**
All	1470	0.735	0.604	0.645	0.469	0.616	0.581	0.576	0.35
Asphodelus macrocarpus	463	0.806	0.692	0.782	0.502	0.723	0.693	0.696	0.364
Dactylorhiza sambucina	174	0.874	0.799	0.854	0.577	0.785	0.787	0.829	0.528
Orchis morio	103	0.764	0.835	0.847	0.613	0.695	0.845	0.837	0.484
Anemonoides nemorosa	227	0.866	0.858	0.909	0.652	0.775	0.817	0.15	0.444
Corydalis cava	99	0.611	0.333	0.426	0.277	0.568	0.394	0.436	0.245
Doronicum columnae	226	0.92	0.956	0.973	0.888	0.897	0.96	0.977	0.811
Carpobrotus acinaciformis	15	0.905	0.639	0.873	0.698	0.788	0.667	0.783	0.349
Achillea maritima	17	0.73	0.765	0.847	0.631	0.551	0.765	0.658	0.406
Calamagrostis arenaria	8	0.605	0.75	0.669	0.537	0.466	0.656	0.521	0.246
Eryngium maritimum	13	0.767	0.692	0.747	0.559	0.656	0.769	0.655	0.445
Pancratium maritimum	11	0.8	0.727	0.658	0.38	0.317	0.364	0.249	0.124
Thinopyrum junceum	14	0.6	0.357	0.463	0.384	0.563	0.37	0.478	0.242
Luzula alpinopilosa	6	0.832	0.825	0.755	0.548	0.784	0.833	0.755	0.47
Geum reptans	20	0.595	0.35	0.426	0.253	0.53	0.35	0.406	0.233
Oxyria digyna	5	0.606	0.327	0.323	0.217	0.636	0.363	0.268	0.208
Cerastium pedunculatum	12	0.592	0.486	0.37	0.205	0.325	0.333	0.281	0.109
Cerastium uniflorum	7	0.943	0.571	0.582	0.416	0.884	0.571	0.582	0.358
Leucanthemopsis alpina	13	0.586	0.231	0.248	0.146	0.311	0.154	0.208	0.141
Ranunculus glacialis	17	0.791	0.588	0.807	0.664	0.695	0.672	0.806	0.59
Saxifraga bryoides	20	0.511	0.3	0.338	0.237	0.365	0.25	0.286	0.211

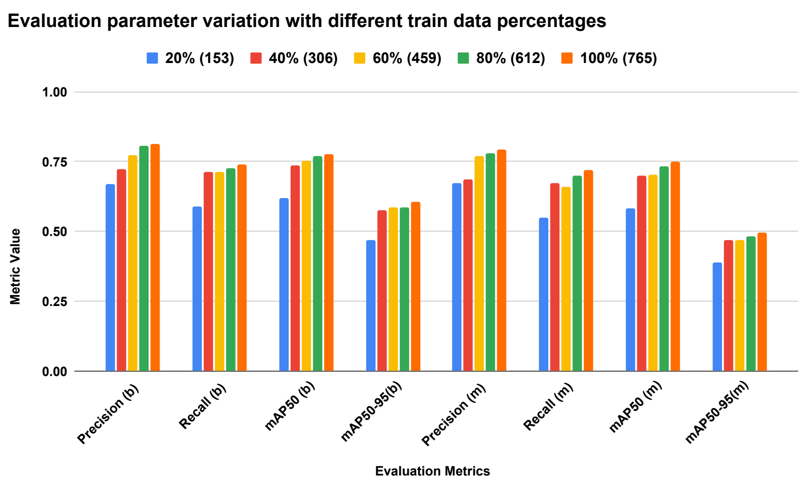

Fig. 3. Comparison of evaluation metrics at different train data percentages on forests

data the same. The train samples are picked randomly manually from the whole train data portion for each training. All metrics have the highest values when the model is trained on 100% of train data. Figure 4 demonstrates the comparison of different YOLOv8 segmentation model versions on all habitat data for all classes in the form of a bar chart. Figure 5 shows the mask F1 plot for all the habitats data obtained using YOLOv8x segmentation model, and Figure 6 demonstrates the mask Precision-Recall (PR) curve. The confusion matrix obtained from all habitat test data is displayed in Figure 7. It can be visualised in the confusion matrix that the high number of test instances are predicted as "background", and the actual "background" is indicated as different plant species. So, the model is getting confused between the background and the actual plant, which is evident in this complex application (plants resemble the background).

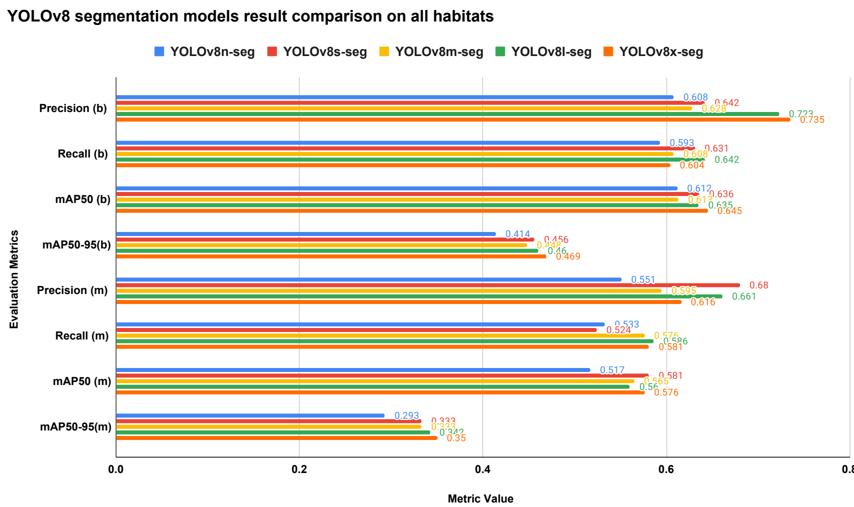

Fig. 4. Comparison of YOLOv8 segmentation models on all habitat test data

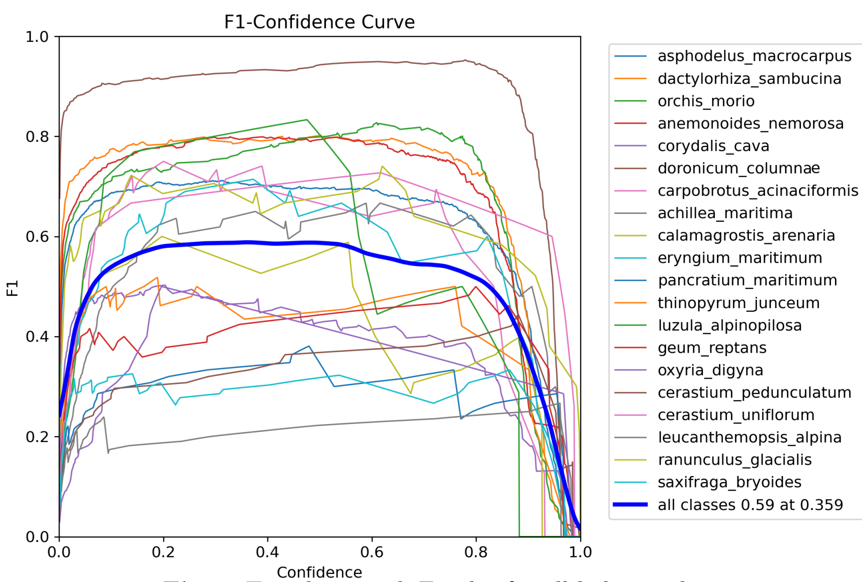

Fig. 5. Test data mask F1 plot for all habitats data

5 Conclusion

In this paper, we have experimented on real plant species image data collected from four habitats: forests, dunes, grasslands, and screens. Botanists annotated the images using bounding box annotations. The bounding boxes are used as a prompt to SAM to retrieve corresponding masks. The YOLOv8 object detection

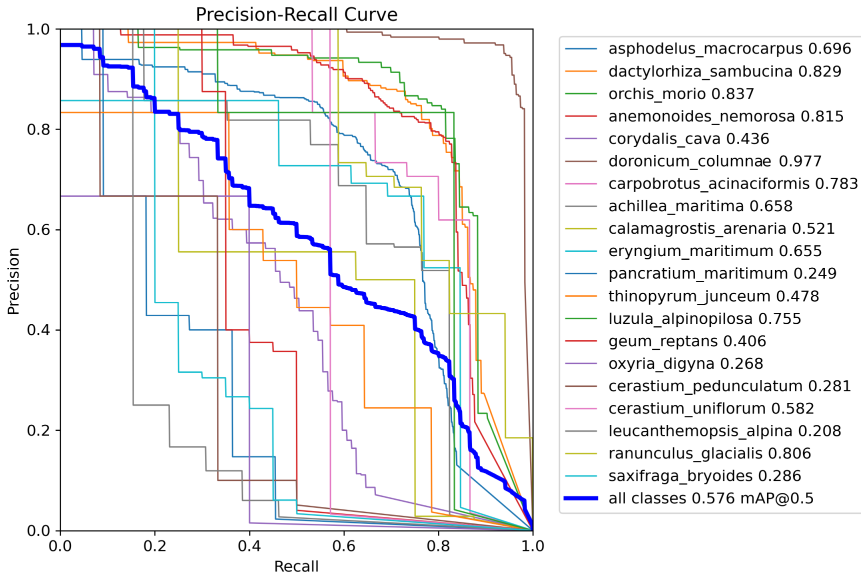

Fig. 6. Test data mask PR curve for all habitats data

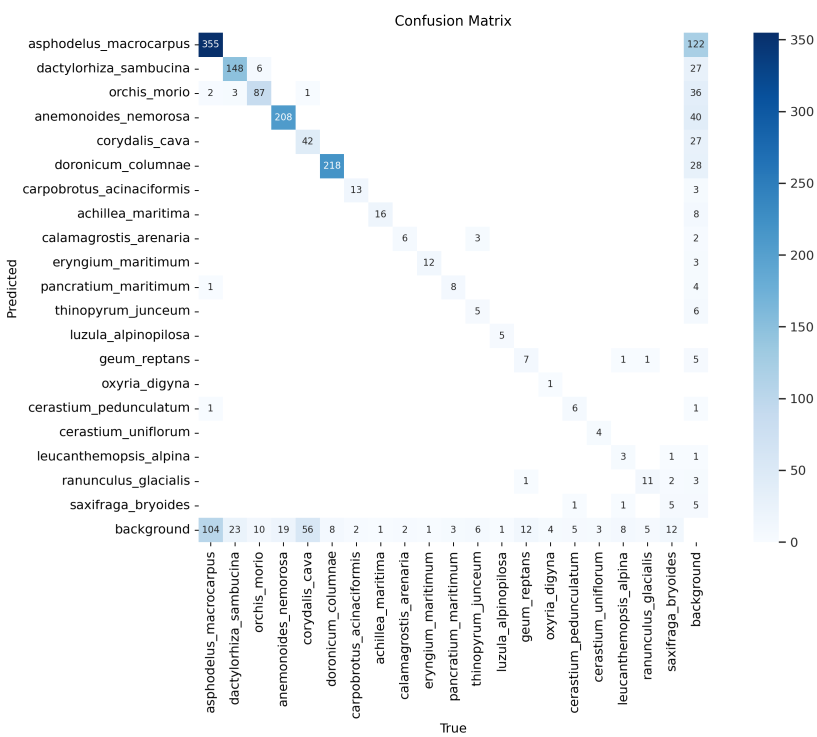

Fig. 7. Confusion matrix for all habitats data

and segmentation model has been compared to analyse whether the segment masks or polygon annotation provide better results than the bounding boxes. After extensive experimentation, we got mixed results. For two habitats (dunes and screes), detection metric values are higher; for the other two, segmentation metric values are higher (forests and grasslands). Even for forest and grassland segmentation, mostly the box evaluation metrics have higher values than the mask metrics. It can be inferred that SAM can be utilised for generating masks if required; however, it will not certainly provide an appropriate localisation with an exact plant species shape due to their complex nature and similar background. More experiments are also performed to compare different versions of the YOLOv8 segmentation model and train the model with diverse percentages of train data. Presently, we are collecting more data (images and videos) via field operations and using a robot which we will utilise further for future experimentation, and also try to automate the monitoring of habitats using the robot entirely.

Acknowledgement

This research was supported by Grant agreement No. 101016970, European Union's Horizon 2020 Research and Innovation Programme - ICT-47-2020.

References

1. Angelini, F., Angelini, P., Angiolini, C., Bagella, S., Bonomo, F., Caccianiga, M., Della Santina, C., Gigante, D., Hutter, M., Nanayakkara, T., et al.: Robotic monitoring of habitats: the natural intelligence approach. IEEE Access (2023)
2. Arya, S., Sandhu, K.S., Singh, J., Kumar, S.: Deep learning: As the new frontier in high-throughput plant phenotyping. Euphytica **218**(4), 47 (2022)
3. Deng, R., Cui, C., Liu, Q., Yao, T., Remedios, L.W., Bao, S., Landman, B.A., Wheless, L.E., Coburn, L.A., Wilson, K.T., et al.: Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging. arXiv preprint arXiv:2304.04155 (2023)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
5. Goëau, H., Bonnet, P., Joly, A.: Overview of plantclef 2022: Image-based plant identification at global scale. In: CLEF 2022-Conference and Labs of the Evaluation Forum. vol. 3180, pp. 1916–1928 (2022)
6. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16000–16009 (2022)
7. Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., et al.: Segment anything model for medical images? arXiv preprint arXiv:2304.14660 (2023)
8. Jocher, G.: Ultralytics yolov8 github. https://github.com/ultralytics/ultralytics, 2023 [Online]

9. Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Yifu, Z., Wong, C., Montes, D., et al.: ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation. Zenodo (2022)
10. Kang, J., Zhao, L., Wang, K., Zhang, K., et al.: Research on an improved yolov8 image segmentation model for crop pests. Advances in Computer, Signals and Systems **7**(3), 1–8 (2023)
11. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
12. Labelbox: Labelbox. https://labelbox.com, 2023 [Online]
13. Mikula, K., Šibíková, M., Ambroz, M., Kollár, M., Ožvat, A.A., Urbán, J., Jarolímek, I., Šibík, J.: Naturasat—a software tool for identification, monitoring and evaluation of habitats by remote sensing techniques. Remote Sensing **13**(17), 3381 (2021)
14. Pushpa, B., Rani, N.: Ayur-plantnet: An unbiased light weight deep convolutional neural network for indian ayurvedic plant species classification. Journal of Applied Research on Medicinal and Aromatic Plants **34**, 100459 (2023)
15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
16. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
17. Sun, Y., Liu, Y., Wang, G., Zhang, H., et al.: Deep learning for plant identification in natural environment. Computational intelligence and neuroscience **2017** (2017)
18. team, U.: Ultralytics yolov8 docs. https://docs.ultralytics.com/, 2023 [Online]
19. Ultralytics: Ultralytics yolov8. https://github.com/ultralytics/ultralytics/issues/189, 2023 [Online]
20. Williams, D., MacFarlane, F., Britten, A.: Leaf only sam: A segment anything pipeline for zero-shot automated leaf segmentation. arXiv preprint arXiv:2305.09418 (2023)
21. Xu, F., Li, B., Xu, S.: Accurate and rapid localization of tea bud leaf picking point based on yolov8. In: China National Conference on Big Data and Social Computing. pp. 261–274. Springer (2023)
22. Xu, M., Yoon, S., Jeong, Y., Lee, J., Park, D.S.: Transfer learning with self-supervised vision transformer for large-scale plant identification. In: International conference of the cross-language evaluation forum for European languages (Springer;). pp. 2253–2261 (2022)
23. Yan, B., Fan, P., Lei, X., Liu, Z., Yang, F.: A real-time apple targets detection method for picking robot based on improved yolov5. Remote Sensing **13**(9), 1619 (2021)
24. Yu, T., Feng, R., Feng, R., Liu, J., Jin, X., Zeng, W., Chen, Z.: Inpaint anything: Segment anything meets image inpainting. arXiv preprint arXiv:2304.06790 (2023)
25. Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., et al.: Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment **241**, 111716 (2020)
26. Zhao, H., Xu, D., Lawal, O.M., Lu, X., Ren, R., Wang, X., Zhang, S.: Jujube fruit instance segmentation based on yolov8 method. Available at SSRN 4482151 (2023)