Transformer-Based Sunflower (*Helianthus annuus* L.) Recognition from Multi-Temporal UAV Orthomosaics

Esra Yildirim [1], Ismail Colkesen [1], Umut Gunes Sefercik [1]

[1] Gebze Technical University, Engineering Faculty, Department of Geomatics Engineering, 41400 Kocaeli, Turkey
(esrayildirim, icolkesen, sefercik)@gtu.edu.tr

Keywords: Smart Agriculture, Sunflower Detection, UAV Remote Sensing, Deep Learning, DETection TRansformer (DETR).

Abstract

The first appearance of inflorescence in sunflowers (*Helianthus annuus* L.) signifies the transition of the sunflower from the vegetative stage to the reproductive (R). At this growth period, accurate and automated detection of sunflower inflorescences is of utmost significance for sunflower yield estimation. Unmanned aerial vehicles (UAVs) have become essential in agricultural product detection due to their high spatial and temporal resolution data collection ability. With the rapid enhancements in deep learning, transformer architectures have emerged as a revolutionary paradigm, showing remarkable success in precision agriculture applications, including crop recognition and mapping. The main goal of this study is to investigate the potential of the DETection TRansformer (DETR) model in identifying sunflowers at the reproductive stage using multi-temporal UAV orthomosaics. To this end, orthomosaics were produced using high-resolution aerial photos collected with a DJI Phantom 4 Pro V2 UAV in a sunflower field located in Akyazı district of Sakarya province, during two reproductive periods of sunflower (R5.1 and R5.9). Utilizing the orthomosaics, two sunflower detection datasets were constructed to train and evaluate the model. The results revealed that the DETR performed better on the R5.9 growth stage ($AP_{0.50}$ = 92.40%, AR_{100} = 68.00%) than the R5.1 ($AP_{0.50}$ = 83.70%, AR_{100} = 53.90%). Furthermore, given increasing IoU thresholds, DETR demonstrated 16.4% and 29.8% improvements in AP and $AP_{0.75}$, respectively, at the R5.9 stage. The results highlighted that DETR could be a powerful tool for identifying sunflowers, especially at advanced growth stages, likely due to more distinct and developed features of inflorescences.

1. Introduction

Sunflower (*Helianthus annuus* L.) ranks among the world's top oilseed crops, value for its nutritional, medicinal, and economic benefits. It serves as a key raw material for various industries, including plastic, chemical, paint, soap, cosmetics, and biodiesel. Additionally, its pulp is utilized for animal feed production. Sunflowers are easily recognized by their single stem with a distinctly large inflorescence, and the size and number of these inflorescence play a crucial role in determining both yield and seed quality (Seiler, 1997). Global demand for sunflowers continues to rise. Production has grown significantly, increasing from 6.8 million tons harvested on 6.7 million hectares in 1961 to 54.3 million tons on 29.3 million hectares in 2022 (FAO, 2023). In Türkiye, this trend is particularly evident. According to the Turkish Statistical Institute (TUIK), sunflower cultivation expanded from 5.5 million decares in 2004 to 9.5 million in 2023, marking a 73% increase (TUIK, 2023). Production more than doubled during the same period, rising from 0.9 million tons to 2.2 million tons, marking a 144% growth. Yield per unit area also improved significantly, increased by 137%, and reached 498 kg per decare. As the demand for sunflowers continues to grow, effective monitoring and identification are essential for maximizing yields and maintaining sustainable production. However, traditional field observation presents challenges due to complex environmental conditions and the diverse morphology of sunflower inflorescences (Jing et al., 2024).

Remote sensing has proven to be an invaluable tool for monitoring large areas with high resolution, significantly improving the accuracy and efficiency of agricultural practices. The introduction of Unmanned Aerial Vehicles (UAVs) has further revolutionized this field, providing a cost-effective means to capture extremely detailed images with flexible flight schedules at low altitudes. These capabilities make UAVs particularly well-suited for detecting and monitoring even smallscale agricultural features, such as sunflower inflorescences, maize tassels, cotton seedlings, rapeseed flowers, and wheat spikes (Alzadjali et al., 2021; Zhao et al., 2021; Feng et al., 2023; Li et al., 2023; Chen et al., 2024; Jing et al., 2024).

In recent years, UAV imagery has become a key resource in sunflower crop monitoring, sparking numerous studies. For example, Pérez-Ortiz et al. (2015) introduced a weed mapping system for sunflowers, leveraging UAVs and machine learning to support site-specific weed control management. Similarly, Li et al. (2021) investigated sunflower lodging by applying various band combinations of multispectral UAV images and comparing the performance of Random Forest, SegNet, and U-Net algorithms. Bai et al. (2022) used RGB UAV imagery and a peak detection algorithm to count sunflowers at the seedling stage. Recently, Song et al. (2023) used multispectral UAV imagery and deep semantic segmentation models like SegNet, DeepLab, PSPNet to identify sunflower growth stages. These studies highlight the versatility and potential of UAV-based imagery, especially when integrated with deep learning (DL) algorithms, for advancing precision agriculture, particularly in monitoring and managing sunflower crops.

Traditionally, agricultural crop detection and mapping have relied on image processing techniques and machine learning methods. However, these approaches often face challenges due to their reliance on manually crafted features, which can limit the algorithm's robustness in complex environments with uneven illumination environments (Wei et al., 2022). The rise of DL, particularly convolutional neural networks, has transformed remote sensing applications, offering significant improvements in precision agriculture. The DL has been applied to various tasks, including crop detection, disease recognition, and yield estimation, with notable success. For example, Chen et al. (2024)

developed an improved DL-based object detection model, YOLOv8n, to detect maize tassels using UAV imagery. In another study, Jing et al. (2024) introduced a Sunflower-YOLO model, an improved version of the YOLOv7-tiny network, for identifying sunflower capitula from RGB UAV imagery. Li et al. (2023) focused on rapeseed inflorescences, proposing an upgraded YOLOv5 algorithm to count them using RGB UAV imagery. For the detection of sunflowers at different phenological stages from multi-temporal RGB UAV orthomosaics, Meanwhile, Yildirim et al. (2024) evaluated U-Net and DeepLab V3 deep semantic segmentation models to detect sunflowers at various phenological stages using multi-temporal RGB UAV orthomosaics. In cotton seedling detection, Feng et al. (2023) employed YOLOv5, YOLOv7, and CenterNet on multispectral UAV images collected at six different dates. They concluded that YOLOv7 outperformed other models in identifying and counting cotton seedlings, and seedlings were better recognized at 23 days after sowing. These studies highlight the growing role of DL in overcoming the limitations of traditional methods, enabling more accurate and efficient crop monitoring and mapping in precision agriculture.

Recently, transformer-based architectures have attracted significant attention for their superior performance in natural language processing and thus applied to image vision tasks, performing better than popular DL-based convolutional neural networks (CNNs) due to their special decoder-encoder structure (Zhou et al., 2022a). Transformers differ from traditional CNN backbones by producing global receptive fields instead of local ones, enhancing their effectiveness in detecting targets within complex scenes (Zhou et al., 2022b). Transformer architectures can leverage attention mechanisms to map the global dependencies between inputs and outputs (Choi et al., 2023). In this regard, transformers have gained popularity as a cutting-edge and innovative technology in the agricultural field. They have been widely adopted in the literature to address diverse precision agriculture applications with high performance (Xie et al., 2024). For example, Li et al. (2024) developed SoybeanNet harnessing the transformer backbone to count soybean pods using UAV images and achieved a counting accuracy of 84.51%. To detect wheat heads from UAV images, Zhu et al. (2022) proposed three transformer architectures (i.e., FR-Transformer, R-Transformer, and FR-Transformer) by combining Faster R-CNN, RetinaNet, and YOLOv3 with transformer backbones. Results revealed that the FR-Transformer outperformed state-of-the-art models by AP_{50} of 88.3%. For the rice weed segmentation using UAV imagery, Guo et al. (2025) introduced CTFFNet by combining CNN and transformer architectures. The results indicated that the proposed method outperformed single CNN or transformer algorithms, achieving the highest segmentation accuracy for complex and variable weed shapes, with an MIoU of 72.8%.

Although many studies show the potential of combining transformer-based architectures and UAV data for detecting and mapping crops, it is seen that limited studies have been focused on detecting and mapping sunflower inflorescences. Motivated by this research gap, this study aimed to explore the potential use of DETection TRansformer (DETR) for identifying sunflowers at different phenological stages using multi-temporal UAV data. To the best of our knowledge, this is the first study using the transformer in the sunflower detection domain. To this end, two different dated UAV flights were conducted in the study area, Sakarya, Turkey, and two high-resolution RGB orthomosaics representing two reproductive stages of sunflower growth (i.e., R5.1 and R5.9) were generated. Using orthomosaics, two sunflower detection datasets were constructed, and the performance of the DETR model was investigated using COCO metrics. This work also contributed to the literature by building sunflower datasets at different plant growth stages, which can be benchmark datasets for sunflower detection tasks.

2. Study Area

The study was carried out in a sunflower parcel within the fields managed by Sakarya Maize Research Institute in the Arifiye district of Sakarya province. Sakarya province is located in the northwestern part of Türkiye between 29°57' and 30°53' east longitudes and 40°17' and 41°13' north latitudes. The province is a transitional climate area where the Black Sea and Mediterranean climates are effective, with hot and humid summers and rainy, warm winters. The province has a total area of 5,015 km^2, 44% of which is plateaus, 34% is mountains and 22% is plains. The agricultural areas where cultivated field crops, vegetables, and fruits cover approximately 48% of the province's territory. Its rich vegetation is mainly forested, and important agricultural products such as wheat, potatoes, beet, maize, sunflower, and hazelnut are grown in the province. With 10,970 tons of sunflower production, Sakarya province meets 0.57% of Türkiye's sunflower production (TUIK, 2023).

Sunflower has strong adaptation thanks to its ability to be grown in different climate conditions and soil types. According to plant development, sunflower growth consists of two main stages: the vegetative and the reproductive (Schneiter and Miller, 1981). The reproductive stage (R) consists of nine sub-stages, starting with the initial formation of the plant inflorescences and ending with the completion of plant physiological maturity. In this study, field studies were conducted to determine the inflorescences of sunflower plants in two different reproductive stages (i.e., R5.1 and R5.9) in Kirazca Agricultural Research Area managed by Sakarya Maize Research Institute, and training and test sites were established in the sunflower field designed to be used in experimental research.

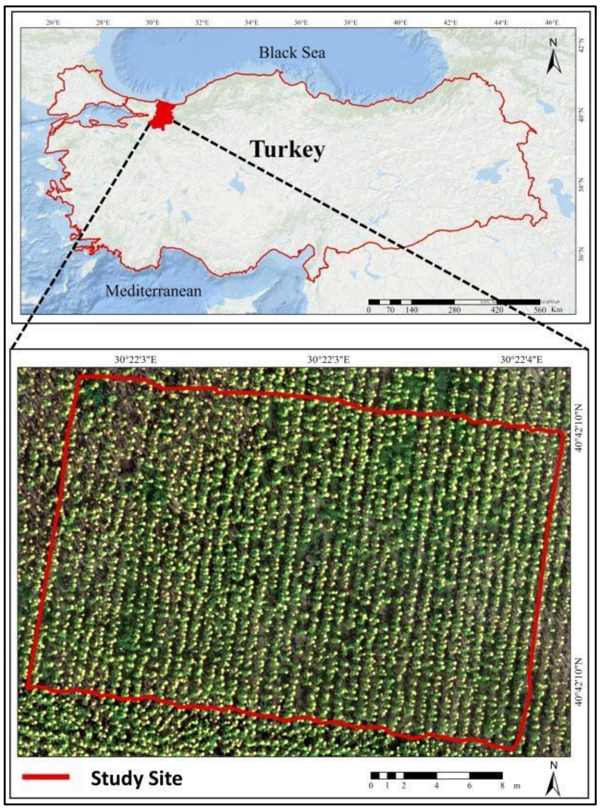

Figure 1. Location map of the study area and experimental site.

3. Methodology

In this current work, which aimed to detect sunflower inflorescences with the integration of transformer-based architectures and UAV technologies, DETR was evaluated as the transformer model and RGB UAV orthomosaic data generated at two different growth stages of sunflower were utilized as the main data source. For this purpose, the sunflower detection performance of the transformer model at different phenological stages is specifically investigated in detail. The methodology followed by the purpose of the study comprises three primary steps: (i) Collection high-resolution UAV aerial photographs, photogrammetric processing, and orthomosaic generation, (ii) creation of ground-truth data by manual labeling on orthomosaics, cropping of orthomosaics, data augmentation and division of the dataset into training, validation, and test datasets, (iii) construction of the DETR model, hyperparameter optimization, training and validation of the model with transfer learning approach, performance analysis and testing of the model.

3.1 UAV Data Acquisition and Orthomosaic Production

In UAV imaging, according to the targeted quality of final products, different flight parameters are applied. In this study, to achieve high geometric and spectral description quality in produced orthomosaics, bundle-grid (double-grid) flight missions were applied from 50 m altitude with 80% and 60% front and side overlap ratios and 70° viewing angle. With the advantage of bundle-grid flights in both North-South and EastWest directions and oblique viewing, the façades of the sunflowers were also imaged. As mentioned above, flights were completed by using DJI Phantom IV V2 UAV which has 20 MP resolution Sony Exmor RGB camera with 8.8 mm focal length. In line with the applied flight parameters, the aerial photos were obtained with an average ground sampling distance (GSD) of ~1.47 cm and each region of the stereo models to be generated was represented with at least nine aerial photos.

The photogrammetric processing steps for orthomosaic production were implemented in Agisoft Metashape software by the workflow given in Figure 2. The aerial photos were matched by utilizing Structure from Motion (SfM) technique which has three major steps: (i) feature extraction for an individual UAV image, (ii) feature matching for each UAV image pair, and (iii) parameter solving based on iterative bundle adjustment (Jiang et al., 2020). Following image matching, absolute orientation was performed by using 14 polycarbonate mobile ground control points (GCPs) established in the field before UAV flights. The root mean square error (RMSE) of the GCPs used in absolute orientations was calculated as ±1.5 cm.

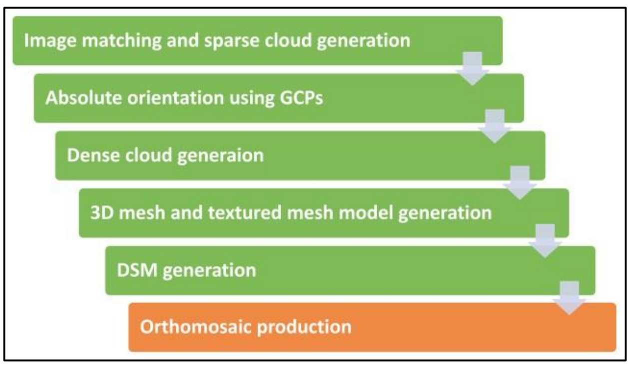

Figure 2. Photogrammetric processing steps implemented for orthomosaic production in Agisoft Metashape.

The produced orthomosaics in the same study area for two different dates are shown in Figure 3.

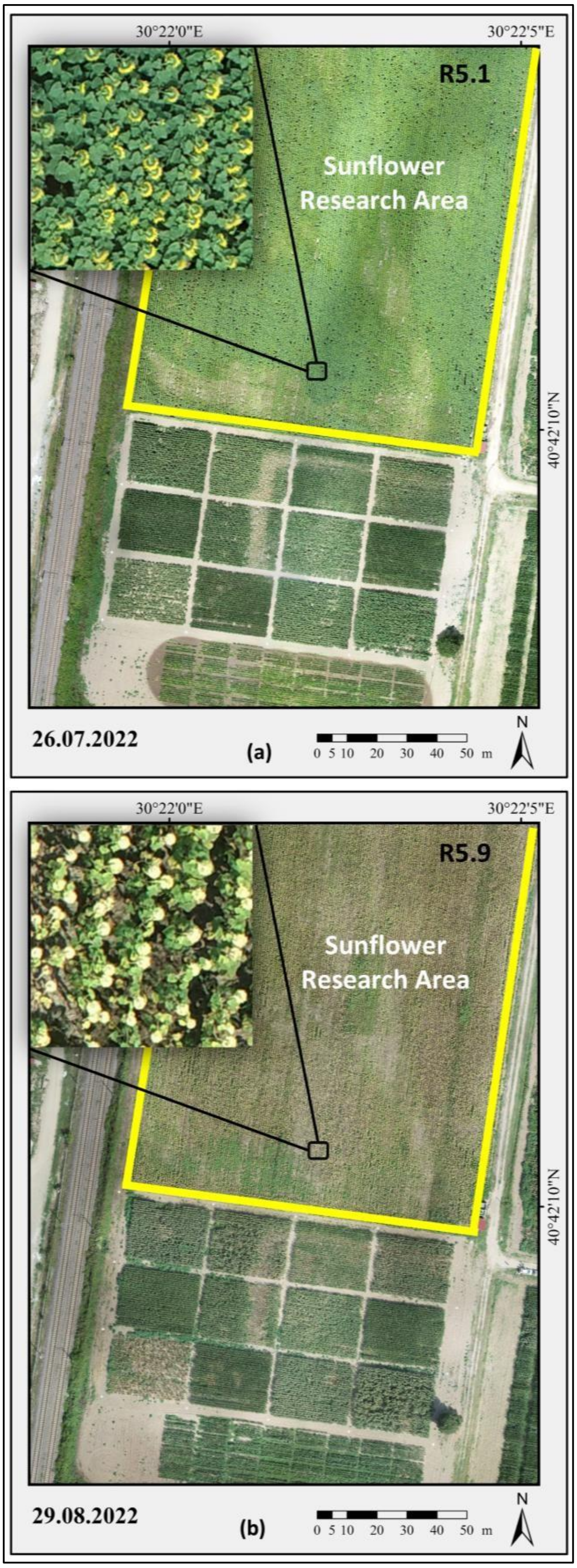

Figure 3. UAV orthomosaics produced at phenological growth stages of (a) R5.1, and (b) R5.9 for the study area.

3.2 Sunflower Data Set

To create a labeled dataset to be used in the model construction, sunflower inflorescences on generated orthomosaics were manually labeled with visual interpretation in ArcGIS Pro 3.0.3 software. In this manner, approximately 2,000 sunflower instances were labeled separately on each orthomosaic within the study area, as shown in Figure 4. Since processing very highresolution UAV data is labor-intensive and time-consuming, orthomosaics were divided into 256×256 pixel image patches with 50% overlap (i.e., 128×128 pixels) to avoid GPU limitations. Image overlap was used to increase the number of training samples and to prevent the edge problem for sunflowers located in the image borders, ensuring that each sample falls within at least one image patch. In addition, synthetic data augmentation was applied by flipping, cropping, and rotating the training images by 90º, 180º, and 270º to increase the number of training samples and improve model performance. As a result, the R5.1 training dataset contains a total of 982 images, and the R5.9 training dataset contains 1186 images. Afterwards, the created training datasets were randomly divided into training, validation, and test datasets at 70%, 20%, and 10% ratios, respectively, for training, validation, and testing of the model.

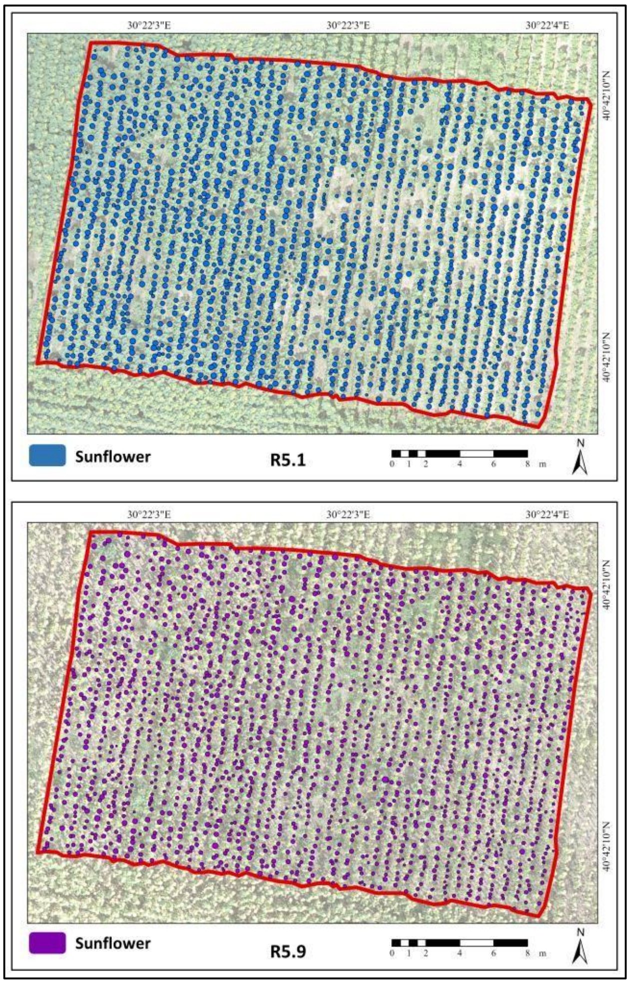

Figure 4. Ground truth sunflower labels generated for the study area on two orthomosaics.

3.3 DETection Transformer (DETR)

Transformers have emerged as popular and are revolutionizing architectures in DL, achieving state-of-the-art results across various NLP and computer vision benchmarks. Numerous adaptations of the original transformer model, first introduced in 2017 (Vaswani et al.), have been developed to deal with problems, including computational demands and data efficiency.

The transformer-based object detection model used in the study, DEtection Transformer (DETR) treats object detection as a direct set prediction task within a transformer-based encoder-decoder structure. This paradigm eliminates the need for anchor generation or the non-maximum suppression processes. The DETR framework consists of three main components: a CNN backbone, an encoder-decoder transformer, and a feed-forward network (Figure 5). The CNN backbone captures intricate features from input data, producing a feature map. The feature representation is then flattened and supplemented with positional encoding before being sent to a transformer encoder. The transformer decoder, utilizing a fixed set of learned positional embeddings known as object queries, attends to the encoder's output. Each decoder output embedding is passed through a shared feed-forward network (FFN) to predict either a detection (class and bounding box) or a "no object" classification. (Carion et al., 2020).

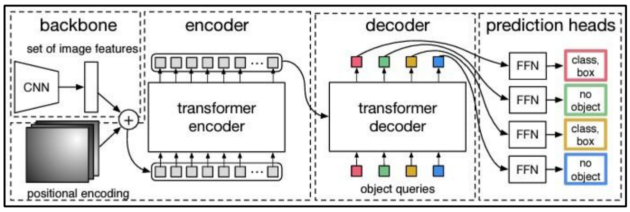

Figure 5. Architecture of the DETR (Carion et al., 2020).

3.4 Accuracy Assessment Metrics

To evaluate the performance of the DETR model, Microsoft Common Objects in Context (MS COCO) evaluation metrics, which are frequently used in deep learning-based object detection studies, were utilized. In the COCO metrics, AP and AR are computed by averaging across 10 IoU thresholds ranging from 0.50 to 0.95 in increments of 0.05. Additionally, $AP_{0.5}$ and $AP_{0.75}$ represent AP values calculated at fixed IoU thresholds of 0.50 and 0.75, respectively. AR_1, AR_{10}, and AR_{100} correspond to the maximum recalls averaged over all categories and IoU thresholds, considering 1, 10, and 100 detections per image, respectively (Lin et al., 2014).

The IoU quantifies the overlap between the model's predicted bounding box (B_p) and the ground truth bounding box (B_{gt}). At an IoU threshold of 0.5, predictions with IoU values below 0.5 for a specific class are categorized as false positives (FP), while predictions with IoU values above 0.5 are referred to as true positives (TP). Ground truth boxes not matched with predictions are labeled false negatives (FN). Precision and recall metrics are derived using TP, FP, and FN. The AP reflects the shape of the precision/recall curve, calculated as the averaged precision at a set of eleven equally spaced recall levels [0, 0.1, ..., 1]. Also, AR is the recall averaged over all IoU values between 0.5 and 1.0. The formulas for IoU, precision, and recall are given below:

https://doi.org/10.5194/isprs-archives-XLVIII-M-6-2025-309-2025 |

$$IoU = \frac{area(B_p \cap B_p)}{area(B_p \cap B_{gt})} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4. Results and Discussions

This study conducted all experiments using Python programming language with the PyTorch 2.4.0 deep learning framework in the cloud-based Google Colab Pro environment, which offers access to the NVIDIA A100 GPU. The DETR is trained separately using two sunflower datasets with the AdamW optimizer with the initial transformer's learning rate to 10^{-4}, the backbone's to 10^{-5}, and weight decay to 10^{-4}. The training of the models was initialized with ResNet-50 backbone weights pre-trained on the COCO dataset, and both models were trained for 100 epochs. After training the DETR, the training and validation loss graphs were monitored on TensorBoard, and it was observed that both loss curves gradually decreased as the number of iterations increased (Figure 6). The fact that the training loss and validation loss tend to decrease indicates that overfitting was not observed in the models and that the models were able to generalize what they learned in the training dataset to the validation dataset. On the other hand, it was clear that the loss values of the model trained with the R5.9 dataset converged more than the R5.1.

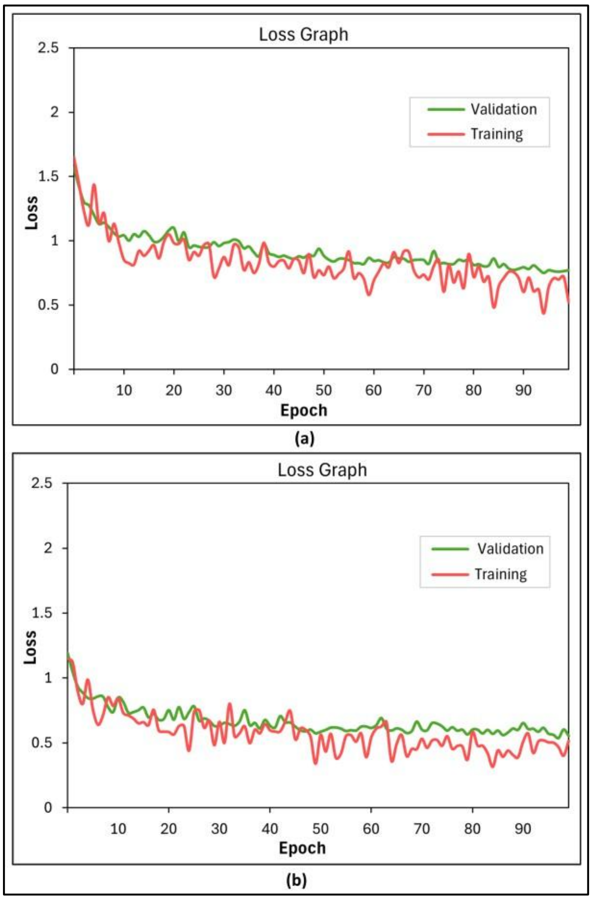

Figure 6. Loss curve of the DETR model for (a) R5.1 and (b) R5.9 sunflower growth stages during the training process.

After the training and validation, the performance of the DETR model on the test dataset was investigated using COCO metrics (Table 1). The results showed that DETR consistently performed better on the R5.9 growth stage of the sunflower across all metrics. For the $AP_{0.5}$ metric, DETR achieved a score of 92.4% on the R5.9 stage, compared to 83.7% on the R5.1 stage. As the IoU thresholds increased, the model maintained superior performance on the R5.9 dataset. Notably, it achieved an $AP_{0.75}$ score of 77.3%, significantly higher than the 47.5% observed for the R5.1 dataset. When considering overall AP, the DETR model reached 62.8% on the R5.9 stage, outperforming the 46.4% score on the R5.1 stage. This suggests that the more developed and distinct inflorescence features in the R5.9 stage contributed to improved localization accuracy. On the other hand, the DETR model showed superior performance on the R5.9 stage in terms of AR metrics (AR_1, AR_{10}, AR_{100}). In particular, it performed better on the R5.9 dataset than the R5.1 dataset by a large margin in terms of the AR_{100} metric, which considers at most 100 detections per image. As a whole, the results revealed that the DETR model could identify sunflower inflorescences more accurately in higher growth stages, namely R5.9. Overall, these results indicated that the DETR model was better equipped to identify sunflower inflorescences at more advanced growth stages, such as R5.9, where the inflorescence features were more distinct and well-defined.

Metric	Dataset	
	26 July (R5.1)	**29 August (R5.9)**
AP	0.464	<u>0.628</u>
AP0.5	0.837	<u>0.924</u>
AP0.75	0.475	<u>0.773</u>
AR_1	0.011	<u>0.016</u>
AR_{10}	0.112	<u>0.156</u>
AR100	0.539	<u>0.680</u>

Table 1. Performance of DETR on two sunflower growth stages.

To further investigate the model's performance, the sunflower prediction results on the test dataset were quantitatively analyzed. The results clearly showed that the DETR model tends to produce more false alarms, missed detections, and inaccurate bounding boxes on the R5.1 growth stage of sunflowers (Figure 7). At this stage, it was observed that the model tended to produce more FN and FP when multiple sunflowers were adjacent or overlapping. There were also problems with the precise positioning of the predicted TP boxes.

In particular, adjacent sunflower inflorescences could not be detected separately and could be estimated with a single bounding box. Furthermore, various detection errors were observed due to objects being truncated at the image edges.

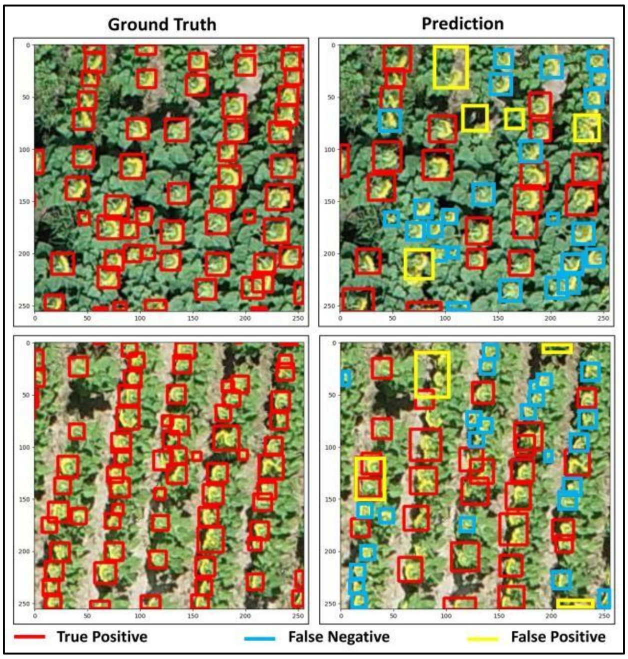

Figure 7. Sunflower prediction results on R5.1 growth stage.

On the other hand, much fewer false alarms and missed detections were observed in the prediction results at the R5.9 growth stage compared to R5.1 (Figure 8). Similar to R5.1, there were also problems at this stage where multiple sunflowers were located adjacently and at the image edges. Overall, sunflower detection at the R5.9 stage produced more accurate and precise results due to the more mature sunflower inflorescences and their obvious boundaries.

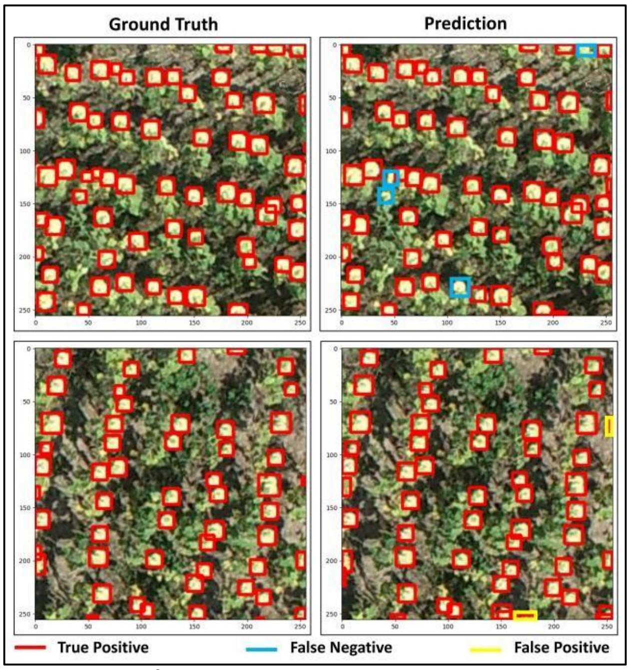

Figure 8. Sunflower prediction results on R5.9 growth stage.

5. Conclusions

Identifying sunflowers from high-resolution UAV data is vital in achieving high yield and ensuring sustainable production. In this study, the transformer-based object detection method, DETection TRansformer (DETR), was employed to detect sunflower heads at different phenological stages using multi-temporal UAV orthomosaics. The results revealed that the DETR showed superior performance in identifying sunflower heads at the later reproductive stage (R5.9) compared to the earlier stage (R5.1), likely due to more distinct and developed features at this advanced growth stage. These results may stem from the mature inflorescences of the sunflower plant, which has more distinct and developed features and borders in the later reproductive stages. Overall, results highlight that DETR could be a powerful tool for identifying sunflowers at different phenological stages.

Although this study highlights that the combination of UAV technology and transformer architectures offers effective solutions for the detection of agricultural products, some issues still need to be developed and planned to be realized in future studies. In this manner, it is necessary to utilize UAV elevation data (Digital Elevation Model, Canopy Height Model, etc.), multispectral data, and vegetation indices such as NDVI as datasets in the training of the DL model and to examine their effects on model performance. Another is to examine the model performance on sunflower data taken from a different study site to test the model's generalization capability. Explainable artificial intelligence techniques such as Grad-CAM should be used to interpret model predictions and improve model performance accordingly.

References

Alzadjali, A., Alali, M.H., Veeranampalayam Sivakumar, A.N., Deogun, J.S., Scott, S., Schnable, J.C., Shi, Y., 2021. Maize tassel detection from UAV imagery using deep learning. *Frontiers in Robotics and AI*, 8, 600410.

Badrinarayanan V., Kendall A., Cipolla R., 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481-2495.

Bai, Y., Nie, C., Wang, H., Cheng, M., Liu, S., Yu, X., Shao, M., Wang, Z., Wang, S., Tuohuti, N., Shi, L., Ming, B., Jin, X., 2022. A fast and robust method for plant count in sunflower and maize at different seedling stages using high-resolution UAV RGB imagery. *Precision Agriculture*, 23(5), 1720-1742.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds.), *Computer Vision – ECCV 2020, Lecture Notes in Computer Science*, Springer, Cham, 213-229. doi.org/10.1007/978-3-030-58452-8_13

Chen, J., Fu, Y., Guo, Y., Xu, Y., Zhang, X., Hao, F., 2024. An improved deep learning approach for detection of maize tassels using UAV-based RGB images. *International Journal of Applied Earth Observation and Geoinformation*, 130, 103922.

Choi, S.R., Lee, M., 2023. Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. *Biology*, 12(7), 1033.

FAO, 2023. FAOSTAT, Crops and Livestock Products. https://www.fao.org/faostat/en/#data/QCL, (Accessed in July 2024).

Feng, Y., Chen, W., Ma, Y., Zhang, Z., Gao, P., Lv, X., 2023. Cotton seedling detection and counting based on UAV multispectral images and deep learning methods. *Remote Sensing*, 15(10), 2680.

Guo, Z., Cai, D., Jin, Z., Xu, T., Yu, F., 2025. Research on unmanned aerial vehicle (UAV) rice field weed sensing image segmentation method based on CNN-transformer. *Computers and Electronics in Agriculture*, 229, 109719.

Jiang, S., Jiang, C., Jiang, W., 2020. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 230-251.

Jing, R., Niu, Q., Tian, Y., Zhang, H., Zhao, Q., Li, Z., Zhou, X., Li, D., 2024. Sunflower-YOLO: Detection of sunflower capitula in UAV remote sensing images. *European Journal of Agronomy*, 160, 127332.

Li, G., Han, W., Huang, S., Ma, W., Ma, Q., Cui, X., 2021. Extraction of sunflower lodging information based on UAV multi-spectral remote sensing and deep learning. *Remote sensing*, 13(14), 2721.

Li, J., Li, Y., Qiao, J., Li, L., Wang, X., Yao, J., Liao, G., 2023. Automatic counting of rapeseed inflorescences using deep learning method and UAV RGB imagery. *Frontiers in Plant Science*, 14, 1101143.

Li, J., Magar, R.T., Chen, D., Lin, F., Wang, D., Yin, X., Zhuang, W., Li, Z., 2024. SoybeanNet: Transformer-based convolutional neural network for soybean pod counting from Unmanned Aerial Vehicle (UAV) images. *Computers and Electronics in Agriculture*, 220, 108861.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.), *Computer Vision – ECCV 2014*, *Lecture Notes in Computer Science*, vol 8693, Springer, Cham, 740-755. doi.org/10.1007/978-3-319-10602-1_48

Pérez-Ortiz, M., Peña, J.M., Gutiérrez, P.A., Torres-Sánchez, J., Hervás-Martínez, C., López-Granados, F., 2015. A semisupervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing*, 37, 533-544.

Redmon J., Divvala S., Girshick R., Farhadi A., 2016. You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 779-788.

Schneiter, A.A., Miller, J.F., 1981. Description of sunflower growth stages. *Crop Science*, 21(6), 901-903.

Seiler, G.J., 1997. Anatomy and morphology of sunflower. In: A. A. Schneiter (ed.), *Sunflower Technology and Production*, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 67-111. doi.org/10.2134/agronmonogr35.c3.

Song, Z., Wang, P., Zhang, Z., Yang, S., Ning, J., 2023. Recognition of sunflower growth period based on deep learning from UAV remote sensing images. *Precision Agriculture*, 24(4), 1417-1438.

TUIK, 2023. Central Dissemination System. https://biruni.tuik.gov.tr/medas/?locale=tr, (Accessed in July 2024).

Xie, W., Zhao, M., Liu, Y., Yang, D., Huang, K., Fan, C., Wang, Z., 2024. Recent advances in Transformer technology for agriculture: A comprehensive survey. *Engineering Applications of Artificial Intelligence*, 138, 109412.

Wei, L., Luo, Y., Xu, L., Zhang, Q., Cai, Q., Shen, M., 2021. Deep convolutional neural network for rice density prescription map at ripening stage using unmanned aerial vehicle-based remotely sensed images. *Remote Sensing*, 14(1), 46.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Yang, X., Ye, Y., Li, X., Lau, R.Y., Zhang, X., Huang X., 2018. Hyperspectral image classification with deep learning models. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9), 5408-5423.

Yildirim, E., Colkesen, I., Sefercik, U.G., 2024. Identification of sunflowers (Helianthus annuus L.) from multi-temporal UAV orthomosaics using deep learning models. *9th Advanced Engineering Days*, Tabriz, Iran, 782-785.

Yildirim, E., Nazar, M., Sefercik, U.G., Kavzoglu, T., 2022. Stone Pine (Pinus Pinea L.) Detection from High-Resolution UAV Imagery Using Deep Learning Model. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, 441-444.

Zhao, J., Zhang, X., Yan, J., Qiu, X., Yao, X., Tian, Y., Zhu, Y., Cao, W., 2021: A wheat spike detection method in UAV images based on improved YOLOv5. *Remote Sensing*, 13(16), 3095.

Zhou, C., Ye, H., Sun, D., Yue, J., Yang, G., Hu, J., 2022a. An automated, high-performance approach for detecting and characterizing broccoli based on UAV remote-sensing and transformers: A case study from Haining, China. *International Journal of Applied Earth Observation and Geoinformation*, 114, 103055.

Zhou, Q., Huang, Z., Zheng, S., Jiao, L., Wang, L., Wang, R., 2022b. A wheat spike detection method based on Transformer. *Frontiers in Plant Science*, 13, 1023924.

Zhu, J., Yang, G., Feng, X., Li, X., Fang, H., Zhang, J., Bai, X., Tao, M., He, Y., 2022. Detecting wheat heads from UAV low-altitude remote sensing images using deep learning based on transformer. *Remote Sensing*, 14(20), 5141.