Comparative Study Between Object Detection Models, for Olive Fruit Fly Identification

Margarida Victoriano[1][a], Lino Oliveira[1][b] and Hélder P. Oliveira[1,2][c]
[1]INESC TEC, FEUP Campus, Porto, Portugal
[2]Faculty of Sciences, University of Porto, Porto, Portugal

Keywords: Image Processing, Object Detection and Classification, Insect Management.

Abstract: Climate change is causing the emergence of new pest species and diseases, threatening economies, public health, and food security. In Europe, olive groves are crucial for producing olive oil and table olives; however, the presence of the olive fruit fly (*Bactrocera Oleae*) poses a significant threat, causing crop losses and financial hardship. Early disease and pest detection methods are crucial for addressing this issue. This work presents a pioneering comparative performance study between two state-of-the-art object detection models, YOLOv5 and YOLOv8, for the detection of the olive fruit fly from trap images, marking the first-ever application of these models in this context. The dataset was obtained by merging two existing datasets: the DIRT dataset, collected in Greece, and the CIMO-IPB dataset, collected in Portugal. To increase its diversity and size, the dataset was augmented, and then both models were fine-tuned. A set of metrics were calculated, to assess both models performance. Early detection techniques like these can be incorporated in electronic traps, to effectively safeguard crops from the adverse impacts caused by climate change, ultimately ensuring food security and sustainable agriculture.

1 INTRODUCTION

Climate change has considerable effects on the agricultural industry, especially in terms of pests and diseases. It is possible for pests and illnesses to thrive and spread as temperature and precipitation patterns change, increasing agricultural damage. Temperature variations may also affect the behavior and life cycle of pests, enabling them to spread more quickly and colonize new areas, increasing the risk of agricultural damage and making pest management more difficult. The timing of crop growth and the presence of natural pest predators can also be impacted by changes in precipitation and temperature, and this can lead to a mismatch between crop phenology and insect populations, which can exacerbate pest damage (Nazir et al., 2018).

The expected effects of climate change on the dynamics of pest and diseases in agriculture underline the significance of effective pest control strategies that take these factors into account. In order to reduce crop losses and use safer control management strategies that safeguard the crop, human health, and the environment, early pest detection is essential.

Our research focuses on the olive fruit fly (*Bactrocera oleae*), which is a significant threat to olive trees in the European Union, particularly in Spain and Italy, which have the largest cultivation areas (Eurostat, 2019).

Monitoring the quantity of pests that have been caught in a trap within a given time frame is one of the crucial aspects of confirming the occurrence of an outbreak. Experts typically have to go to the fields to visually inspect each trap manually to assess infestations, which can be a cumbersome and expensive task, especially in large and dispersed orchards (Shaked et al., 2017; Remboski et al., 2018; Hong et al., 2020). To address this, new approaches have been used to more effectively monitor this as well as other insects, including the installation of electronic traps supplied with cameras to record images of the traps that can then be evaluated using classification and detection models (Uzun et al., 2022; Le et al., 2021; Diller et al., 2022).

Due to the high percentage of EU cropland devoted to olive trees and the recurring outbreaks of olive fruit flies, monitoring and controlling pests in

[a] https://orcid.org/0009-0003-1840-4636
[b] https://orcid.org/0000-0003-1036-1072
[c] https://orcid.org/0000-0002-6193-8540

Victoriano, M., Oliveira, L. and Oliveira, H.
Comparative Study Between Object Detection Models, for Olive Fruit Fly Identification.
DOI: 10.5220/0012361800003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 2: VISAPP*, pages 458-465
ISBN: 978-989-758-679-8; ISSN: 2184-4321

olive orchards is crucial to preserving the health and productivity of olive trees as well as the quality of by-products like table olives and olive oil. Early detection of infestations and proactive approach through control methods, like precise application of pesticides, can reduce the need for more extensive interventions, assisting in the sustainable development of olive production while conserving the environment and human health (Caselli and Petacchi, 2021).

An approach that has been increasingly followed is to use images collected in real time, from traps placed on the ground, and computer vision algorithms to automatically detect the presence of pests.

The purpose of this research is to provide a thorough evaluation of the performance of two state-of-the-art, fast, and accurate object identification algorithms, YOLOv5 and YOLOv8, in a novel area. To accomplish this, we fine-tuned the YOLOv5 and YOLOv8 models using trap images collected from two locations: Portugal and Greece. These computer vision models combined with the prediction monitoring process can be used to create a reliable and effective method for controlling olive fruit fly populations and protecting olive harvests.

2 RELATED WORK

Object detection is a crucial field in computer vision that deals with identifying and locating objects in images or videos. With the advent of deep learning techniques, there has been substantial progress in object detection in recent years. Insect detection and classification is a particular domain of object detection that has significant applications in agriculture, pest control, and disease prevention. In this section, we will examine the current state-of-the-art in insect detection and classification, with a particular emphasis on deep learning methods.

Deep learning is becoming increasingly significant for insect classification and identification due to its capacity to automatically extract attributes from raw visual data and learn intricate correlations between them. For a variety of purposes, including agriculture, pest management, and disease prevention, insect classification and identification are essential. Accurate identification and classification of insects can be accomplished through the use of deep learning, which can contribute in the creation of efficient management and control techniques for insect populations (Bjerge et al., 2022; Diller et al., 2022; Mamdouh and Khattab, 2021; Tetila et al., 2020).

Diller et al. (Diller et al., 2022) conducted a study to train deep learning algorithms for the identification of three major fruit fly species - *Ceratitis capitata*, *Bactrocera dorsalis*, and *B. zonata*. The algorithm was trained on images collected from laboratory colonies of two species and tested in the field in five different countries, and on images captured by an electronic trap built by the authors. The electronic trap mostly contained only one specimen, which made it easier to label each image and apply that label to each insect found in it. To address the issue of multiple insect species in field images, the authors manually labelled each occurrence by drawing bounding boxes around them using specific software. Several data augmentation methods were applied, such as horizontal flipping, vertical flipping, rotations and random brightness and contrast changes, to enhance the dataset and prevent overfitting. The model was trained using the Faster R-CNN ResNet50 algorithm and achieved an average precision of 87.41% to 95.80% for all three target classes. The model also showed robustness to distinct lighting conditions and accuracy on clustered samples. The classification accuracy for all classes in the dataset collected on the field was 98.29%, and the electronic trap obtained high precision and accuracy for the three target species, ranging from 86% to 97%. The proposed model showed good ability to distinguish between the target classes and non-target classes.

Mamdouh and Khattab (Mamdouh and Khattab, 2021) proposed a framework to detect and count olive fruit flies in images captured by smart traps in olive groves. The framework includes preprocessing of the input dataset by normalizing it with yellow mean color to unify the background color and make it illumination invariant. The dataset images have varying lighting settings caused by varying light intensities and shadows cast by objects, so data augmentation techniques were applied (flipping, shifting and rotations) to increase the variability of images, thus increasing accuracy. To reduce overfitting, the researchers added a Leeds butterfly dataset to the training data. The researchers used YOLOv4, a single-stage object detection paradigm, which requires only one network pass to identify objects. Bounding box detection is modelled as a regression task, and classes are modelled as conditional probabilities. With a pre-trained network, the final layer is fine-tuned based on the input dataset. The network size was changed to detect small objects, and the anchor boxes were clustered based on the dataset bounding boxes dimensions by running k-means with nine clusters. The performance of the proposed framework was evaluated using several metrics, including mean average precision, precision, recall, and F1-score. The performance metrics of the proposed framework were better, comparing to the baseline model, although the training time has greatly increased. The proposed framework obtained 84% precision, 97% recall, 90% F1-Score, and 96.68% mAP, while the baseline model achieved 73% precision, 90% recall, 81% F1-Score, and 87.59% mAP.

Tetila et al. (Tetila et al., 2020) proposed a framework for classifying and counting insect pests in soybean fields using images captured in the field. The authors evaluated three deep neural networks trained with three different strategies: fine-tuning of the network layers with the weights obtained from ImageNet, complete network with the weights initialized randomly, and transfer learning with the weights obtained from ImageNet. The proposed solution follows the SLIC Superpixels method to segment the insects by employing the k-means algorithm to generate the superpixels that represent the similar regions. Image annotation is performed by specialists to compose a dataset for training and testing purposes. The obtained dataset comprises seven insect classes. A CNN is then trained to extract visual characteristics from superpixels and produce the classification model to categorize the insect images. In the post-processing stage, a plantation image is segmented using the SLIC method, and each segment's superpixels are assigned to a certain class. To identify each superpixel using the CNN-trained classification model, the system scans the image from left to right and top to bottom while parallelly displaying the class's color. Thus, by supplying one class per segment, a colorful map is produced. The number of insects is determined by adding up the insects from each superpixel class that the system has identified. The dataset was augmented using rotation, scaling, scrolling, and zooming. For image classification, the authors trained with Inception-Resnet-v2, ResNet50, and DenseNet-201. The DenseNet-201 obtained the best accuracy, 94.89%, followed by Resnet-50 with 93.78% and then Inception-Resnet-v2 with 93.40%.

Bjerge et al. (Bjerge et al., 2022) developed a real-time insect detection and classification algorithm using YOLOv3 deep learning model. The study aimed to identify eight different insect species by exploring various combinations of network parameters, such as image size, kernel size, and number of training iterations, to obtain the best model performance. The authors acquired a dataset of 5757 images containing insects and 2121 background images without insects from *Sedum* plants and *Calluna vulgaris* plants in Denmark, which were manually labelled with bounding boxes surrounding the insects. The dataset was augmented using default settings of the algorithm to increase the size of the training dataset. The authors evaluated the performance of the model using several metrics, including precision and recall, during training and testing phases. The average training precision was 95%, and the average training recall was 91%. For testing, the average precision and recall were 72% and 73%, respectively, and the mean average precision (mAP) was 87%. Overall, the study demonstrated the ability of YOLOv3 to accurately detect and classify different insect species in real-time.

In summary, while the reviewed studies demonstrate the potential of deep learning algorithms for insect classification and detection, there are still opportunities for further improvement in terms of data augmentation, labelling, pre-training techniques and the use of different networks, namely YOLOv5 and YOLOv8, especially for the detection of the olive fruit fly, due to the lack of comprehensive studies on this particular insect.

3 PROPOSED METHODOLOGY

Our study introduces a comparison between two models that can identify the olive fruit fly, also known as *Bactrocera oleae*, a pest that has a significant negative impact on the olive oil industry. We used YOLOv5 (Jocher, 2020) and YOLOv8 (Jocher et al., 2023) architectures for this purpose, due to their high accuracy shown on several object detection benchmarks, fast inference and flexibility. Our framework involves multiple steps, including dataset preparation, application of data augmentation techniques to mimic real-world scenarios, and training with both models on a cross-cohort dataset of olive fruit fly trap images. To assess the performance of both models, we used three-fold cross-validation with hyperparameter tuning.

3.1 Dataset Description

The CIMO-IPB dataset (Pereira, 2023) contains 321 pictures, collected in Portuguese olive groves, mostly exhibiting olive fruit flies, trapped in yellow sticky traps. Experts annotated each image in the dataset with spatial identification of the olive fruit fly using the labelImg tool (Tzutalin, 2015). The insects in the images were manually outlined using bounding boxes and labelled, as shown in Figure 1.

The DIRT dataset (Kalamatianos et al., 2018) comprises 848 pictures, predominantly exhibiting olive fruit flies trapped in McPhail traps from 2015 to 2017 at different sites in Corfu, Greece. While the majority of the images were procured via the e-Olive mobile application, allowing users to upload captures and photographs, they were obtained using a range of

Figure 1: CIMO-IPB Dataset Image Sample with Annotations (Pereira, 2023).

gadgets like smartphones, tablets, and cameras, which makes the dataset non-standardized. Each image in the dataset contains spatial identification of the olive fruit fly, annotated by experts, using the labelImg tool, where the insects in the images were outlined manually using bounding boxes and tagged. Figure 2 shows an annotated image retrieved from the DIRT dataset.

Figure 2: DIRT Dataset Image Sample with Annotations (Kalamatianos et al., 2018).

Both datasets contain other insects, but those instances will be considered as background.

We merged both image datasets, in order to add more variability to the data. By combining these two datasets, we were able to create a more comprehensive and more diverse dataset that included images of the same insect species but with different backgrounds, lighting conditions, and poses, allowing the model to generalize better to new, unseen data.

3.2 Data Preprocessing

In order to use the labels with the YOLOv5 and YOLOv8 object detection model, the labels first had to be transformed from the PASCAL VOC XML format to the YOLO format. The annotations in the YOLO format are described by a text file, for each image, where each line represents an object. Each object is described with an identifier, followed by the x-coordinate and the y-coordinate of the center of the bounding box, and the width and height of the bounding box. We followed a common practice of dividing the dataset into training, validation, and testing sets, where 80% was assigned for training, 10% for validation, and 10% for testing. This enabled us to employ most of the data for model training while holding back a section to evaluate its performance. Furthermore, to enhance the model's precision and robustness, we implemented three-fold cross-validation, on the training set, in addition to the split.

3.3 Data Augmentation Techniques

Performing object detection on trap images captured outdoors may pose some challenges that may undermine the model's accuracy and robustness. Some of these difficulties, for instance, include variations in lighting, weather conditions, and the presence of blurry or low-quality images. To overcome this, data augmentation is a crucial stage in the process of training models. The goal is to increase the variety and volume of data that is available, which will boost the model's precision, resilience, and generalizability. Figure 3 shows the data augmentation techniques that were applied to the dataset.

We applied horizontal (Figure 3b) and vertical flips (Figure 3c), which is an efficient way to create mirrored versions of the original images. This increases the quantity of training data that is available and enables the model to learn to detect objects from a variety of orientations.

We also applied blurring (Figure 3d), which replicates fluctuations in focus and camera movement that may come from external factors. Similarly, random brightness modification (Figure 3g) can aid in simulating changes in lighting circumstances, which can be influenced by weather or sun position. We also applied multiplicative noise (Figure 3e and Figure 3f), which adds noise to the images to help the model learn to recognize insects under noisy conditions.

By using JPEG compression (Figure 3h), we compressed the images to enable the model to learn to recognize insects under low-resolution conditions. This is particularly useful in the context of capturing pictures in the field, where images need to be compressed to reduce their size for transmission or storage, leading to poorer image quality.

3.4 YOLOv5 Architecture

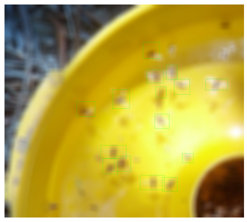
YOLO (You Only Look once) (Redmon et al., 2016) is a real-time object detection algorithm. YOLOv5 (Jocher, 2020) is the fifth version of the YOLO algorithm and is designed to be faster, more accurate, and easier to use than its predecessors.

YOLOv5 is a single-stage object detector and is composed of three main components: the backbone,


(a) Original image.


(b) Horizontal Flip.


(c) Vertical Flip.


(d) Blurr.


(e) Multiplicative Noise.


(f) Multiplicative Noise with varying factor.


(g) Random Brightness Contrast.

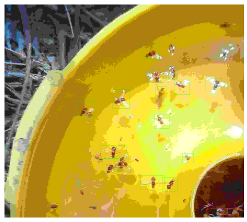

(h) JPEG Compression.

Figure 3: Data augmentation techniques applied to the olive fruit fly image dataset. This helped to improve the performance of the models to detect olive fruit flies in different orientations and in different conditions.

the neck and the head.

Yolov5 combined the cross stage partial network (CSPNet) (Wang et al., 2019) with Darknet to create CSPDarknet, which operated as the network's backbone. By including gradient changes into the feature map, this integration decreases the parameters and FLOPS of the model and addresses the issue of redundant gradient information in large-scale backbones. This not only guarantees quicker and more precise inference, but also minimizes the dimension of the whole model.

In order to improve information flow, Yolov5 additionally uses the path aggregation network (PANet) (Wang et al., 2020) as its neck. The dissemination of low-level features is enhanced by PANet's novel feature pyramid network (FPN) structure, which strengthens the bottom-up path. Moreover, adaptive feature pooling is used to link feature grids and all feature levels, enabling helpful information at each level to spread immediately to the next subnetwork.

The model uses SPPF, which is a variant of Spatial Pyramid Pooling (SPP) (He et al., 2014) that combines data from inputs and produces a fixed-length output, without slowing down the network. The most important context features are separated, and the receptive field is greatly expanded.

Yolov5's head (Redmon and Farhadi, 2018) also produces feature maps in three distinct sizes to provide multi-scale prediction and aid in the detection of small, medium, and large objects. It predicts the object bounding boxes, the scores and the object classes.

YOLOv5 architecture overview is described in Figure 4.

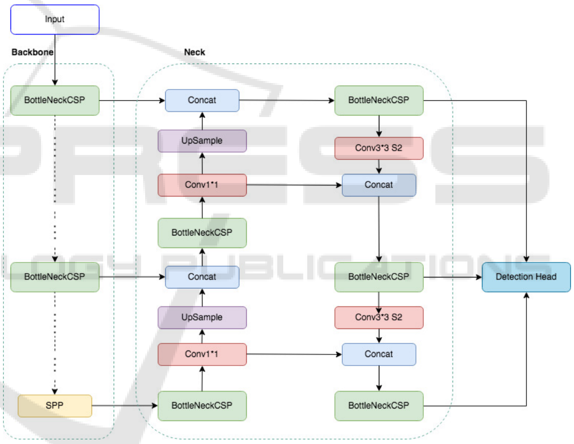


Figure 4: YOLOv5 Architecture.

3.5 YOLOv8 Architecture

YOLOv8 (Jocher et al., 2023) is the successor of the YOLOv5 algorithm and presents an integrated framework for training models for object detection, instance segmentation and image classification.

YOLOv8 introduces a new backbone, Darknet-53, that is faster and more accurate than previous versions. The effectiveness of this version is due to the use of a larger feature map and a more effective convolutional network, which leads to faster and more accurate object detection. The model's capacity to identify patterns and objects in the data is improved by the larger feature map's ability to support more complex feature interactions. It also helps to lessen overfitting and model training time.

Moreover, YOLOv8 makes use of feature pyramid networks, which boosts overall accuracy by better identifying objects of various sizes. Feature Pyramid networks use different feature map scales in conjunction with skip connections to more accurately predict larger and smaller objects, analogous to making predictions on different image sizes.

Unlike the YOLOv5, the YOLOv8 introduced an anchor-free detection head. It does not use anchor boxes to predict the offset from a predefined anchor box. Instead, it directly predicts the object's center in an image.

Figure 5 illustrates YOLOv8 architecture.

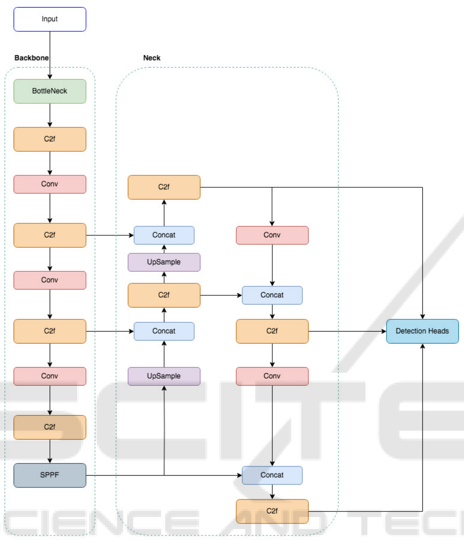


Figure 5: YOLOv8 Architecture.

3.6 Training and Testing

We conducted three-fold cross-validation with hyperparameter tuning on the YOLOv5 and YOLOv8 models using previously split training data. This allowed us to assess the performance of both models on different subsets of the data and optimize the hyperparameters.

The hyperparameters we experimented with are summarized on the Table 1.

Table 1: Range of hyperparameters varied during experimentation.

Hyperparameter	Range
Number of Epochs	50 - 200
Batch Size	8 - 20
Learning Rate	0.0001 - 0.1
Weight Decay	0.0005 - 0.001
Optimizers	SGD and Adam

4 RESULTS EVALUATION

Precision, recall, and mean average precision were used to evaluate the models' performance after cross-validation.

We retrained both models using the best performing hyperparameters, obtained with cross-validation, as shown in Table 2.

Table 2: Best performing hyperparameters in YOLOv5 and YOLOv8.

Model	Number of Epochs	Batch Size	Learning Rate	Weight Decay	Optimizer
YOLOv5	150	8	0.01	0.0005	SGD
YOLOv8	90	8	0.01	0.0005	SGD

Finally, we tested the retrained models on the 10% previously split set.

The test results show that YOLOv5 and YOLOv8 perform similarly under identical training conditions. YOLOv5 achieved a precision of 0.8, recall of 0.75, and mAP of 0.82 at an IoU threshold of 0.5 during cross-validation. After retraining, its precision slightly dropped to 0.76, while recall and mAP remained constant.

On the other hand, YOLOv8, during cross-validation, attained a precision of 0.75, recall of 0.73, and mAP of 0.79. After retraining, its precision improved to 0.79, with recall staying at 0.74 and mAP increasing to 0.82.

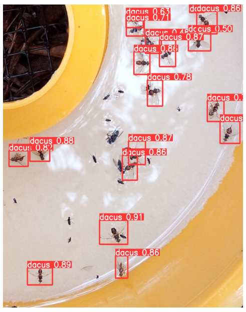
In summary, both models performed well in insect detection tasks. YOLOv5 demonstrated higher precision during cross-validation, whereas YOLOv8 showed significant improvement in precision and mAP after retraining. Notably, YOLOv8 trained faster than YOLOv5, which is crucial for efficient resource utilization in object detection tasks.

Table 3 shows the comparison between the obtained results.

Table 3: Results comparison between YOLOv5 and YOLOv8.

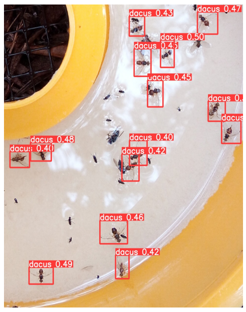
Metric	3-fold Cross Validation		Testing with Best Parameters	
	YOLOv5	YOLOv8	YOLOv5	YOLOv8
Precision	0.80	0.75	0.76	0.79
Recall	0.75	0.73	0.75	0.74
mAP@.5	0.82	0.79	0.82	0.82

Figure 6 shows the detection results obtained using YOLOv5 and YOLOv8 models, respectively, on samples from both datasets. On the DIRT dataset, both models detected correctly the olive fruit fly instances; however, YOLOv5 predictions had higher confidence values (Figure 6a), in comparison to YOLOv8 confidence values (Figure 6c). On the other hand, on the CIMO-IPB dataset, both models performed similarly in terms of the detection confidence values, as seen in Figure 6b and Figure 6d. The



(a) Detection results on DIRT dataset sample using YOLOv5.

(b) Detection results on CIMO-IPB dataset sample using YOLOv5.



(c) Detection results on DIRT dataset sample using YOLOv8.

(d) Detection results on CIMO-IPB dataset sample using YOLOv8.

Figure 6: Detection results using YOLOv5 and YOLOv8.

CIMO-IPB dataset posed a set of challenges. Many images were marred by shadows, cast by natural elements or artificial structures (like the phone shadow in Figure 6b), which can obscure crucial details. Additionally, intense sunlight created areas of overexposure, making it difficult to discern finer features. Furthermore, a lack of precise focus in some images compounded the complexity of the dataset. Despite the lower scores on these images, the models were still able to correctly predict the target class. This is a positive result, as it indicates that the models are robust to real-world challenges such as shadows, intense sun, and lack of focus. In real-world applications, it is important to have models that can still perform well even when faced with these challenges.

In the literature review, it is noted that a method proposed by (Mamdouh and Khattab, 2021) employing the DIRT dataset yielded superior results compared to our own approach. While both methodologies share the foundation of the DIRT dataset, a critical distinction arises from the integration of an additional dataset, the CIMO-IPB dataset, in our work. The CIMO-IPB dataset encompasses images of the same target object, the olive fruit fly, yet the conditions and settings of capture significantly diverge from those in the DIRT dataset. CIMO-IPB dataset comprises images of the same target object, albeit captured under vastly different conditions and settings. This strategic integration aims to enhance the robustness and generalizability of our model, a factor not accounted for in the comparison. The reasoning behind combining the CIMO-IPB dataset and the DIRT dataset was to increase the diversity and robustness of our model. By combining images collected under different situations, our method develops a more comprehensive grasp of the olive fruit fly's depiction. In comparison to the method that only uses the DIRT dataset, our mixed dataset approach has significant advantages in terms of improved generalizability and adaptability across varied settings. The suggested approach by Mamdouh and Khattab (Mamdouh and Khattab, 2021) achieve a slightly higher precision than our method, 84% versus 79%, while it surpasses it on the other metrics, due to the specific characteristics of the DIRT dataset.

In summary, the alternative technique that focuses entirely on the DIRT dataset produces great results, particularly in terms of recall and mAP. This is most likely owing to the dataset's narrow focus. Our technique, on the other hand, which mixes two datasets with differing conditions, maintains competitive performance across all criteria. This shows that it has the ability to perform well in a variety of real-world circumstances. The decision to mix the DIRT dataset and the CIMO-IPB dataset exemplifies an intentional effort to improve the model's adaptability and versatility, as evidenced by the acceptable results obtained.

During our research, we noticed that there is a lack of publicly available datasets containing the olive fruit fly. Due to the limited availability of datasets for our target species, we explored related studies on different insects, to potentially adapt methodologies from these works. Comparing the results between studies is not significant, due to their different morphologies and features. Our work was impacted by the lack of comprehensive datasets, but our tailored approach was shown to be effective for the detection of the olive fruit fly.

5 CONCLUSION AND FUTURE WORK

To summarize, the purpose of this study was to assess two alternative approaches for identifying and categorizing olive fruit flies, which pose a significant risk to olive cultivation in the European Union. The methodology consisted of augmenting the dataset and fine-tuning YOLOv5 and YOLOv8 models using the augmented data. The experimental results showed that both models effectively detected the olive fruit fly, but YOLOv8 obtained superior results, in terms of the evaluation metrics.

In the future, the fine-tuned models, which include domain knowledge to help improve the methodologies, are going to be integrated on a web-based information and management system, which will receive the images collected from the field, by a robotic smart trap developed by our team, and return the image with the detection results, to aid farmers in making informed decisions.

ACKNOWLEDGEMENTS

This work is co-financed by Component 5 - Capitalization and Business Innovation of core funding for Technology and Innovation Centres (CTI), integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026

REFERENCES

Bjerge, K., Mann, H. M., and Høye, T. T. (2022). Real-time insect tracking and monitoring with computer vision and deep learning. *Remote Sensing in Ecology and Conservation*, 8:315–327.

Caselli, A. and Petacchi, R. (2021). Climate change and major pests of mediterranean olive orchards: Are we ready to face the global heating? *Insects*, 12(9).

Diller, Y., Shamsian, A., Shaked, B., Altman, Y., Danziger, B. C., Manrakhan, A., Serfontein, L., Bali, E., Wernicke, M., Egartner, A., Colacci, M., Sciarretta, A., Chechik, G., Alchanatis, V., Papadopoulos, N. T., and Nestel, D. (2022). A real-time remote surveillance system for fruit flies of economic importance: sensitivity and image analysis. *Journal of Pest Science*, 1:1–12.

Eurostat, E. (2019). Statistics explained. (Accessed on 2/02/2023).

He, K., Zhang, X., Ren, S., and Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 346–361, Cham. Springer International Publishing.

Hong, S.-J., Kim, S.-Y., Kim, E., Lee, C.-H., Lee, J.-S., Lee, D.-S., Bang, J., and Kim, G. (2020). Moth detection from pheromone trap images using deep learning object detectors. *Agriculture*, 10(5):170.

Jocher, G. (2020). YOLOv5 by ultralytics. Version 7.0.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). YOLO by Ultralytics.

Kalamatianos, R., Karydis, I., Doukakis, D., and Avlonitis, M. (2018). Dirt: The dacus image recognition toolkit. *Journal of Imaging 2018, Vol. 4, Page 129*, 4:129.

Le, A. D., Pham, D. A., Pham, D. T., and Vo, H. B. (2021). Alerttrap: A study on object detection in remote insects trap monitoring system using on-the-edge deep learning platform.

Mamdouh, N. and Khattab, A. (2021). Yolo-based deep learning framework for olive fruit fly detection and counting. *IEEE Access*, 9.

Nazir, N., Bilal, S., Bhat, K., Shah, T., Badri, Z., Bhat, F., Wani, T., Mugal, M., Parveen, S., and Dorjey, S. (2018). Effect of climate change on plant diseases. *International Journal of Current Microbiology and Applied Sciences*, 7:250–256.

Pereira, J. A. (2023). Yellow sticky traps dataset _ olive fly (Bactrocera Oleae).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement.

Remboski, T. B., de Souza, W. D., de Aguiar, M. S., and Ferreira, P. R. (2018). Identification of fruit fly in intelligent traps using techniques of digital image processing and machine learning. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, page 260–267, New York, NY, USA. Association for Computing Machinery.

Shaked, B., Amore, A., Ioannou, C., Valdés, F., Alorda, B., Papanastasiou, S., Goldshtein, E., Shenderey, C., Leza, M., Pontikakos, C., and et al. (2017). Electronic traps for detection and population monitoring of adult fruit flies (diptera: Tephritidae). *Journal of Applied Entomology*, 142(1-2):43–51.

Tetila, E. C., MacHado, B. B., Menezes, G. V., Belete, N. A. D. S., Astolfi, G., and Pistori, H. (2020). A deep-learning approach for automatic counting of soybean insect pests. *IEEE Geoscience and Remote Sensing Letters*, 17:1837–1841.

Tzutalin (2015). Labelimg.

Uzun, Y., Tolun, M. R., Eyyuboglu, H. T., and Sari, F. (2022). An intelligent system for detecting mediterranean fruit fly [medfly; ceratitis capitata (wiedemann)]. *Journal of Agricultural Engineering*, 53.

Wang, C.-Y., Liao, H.-Y. M., Yeh, I.-H., Wu, Y.-H., Chen, P.-Y., and Hsieh, J.-W. (2019). Cspnet: A new backbone that can enhance learning capability of cnn.

Wang, K., Liew, J. H., Zou, Y., Zhou, D., and Feng, J. (2020). Panet: Few-shot image semantic segmentation with prototype alignment.