*Agronomy Research* **22**(1), 168–179, 2024
https://doi.org/10.15159/AR.24.012

# Importance of mosaic augmentation for agricultural image dataset

S. Kodors[1,*], M. Sondors[1], I. Apeinans[1], I. Zarembo[1], G. Lacis[2], E. Rubauskis[2] and K. Karklina[2]

[1]Rezekne Academy of Technologies, Faculty of Engineering, Institute of Engineering, Atbrivosanas Str. 115, LV-4601 Rezekne, Latvia
[2]Institute of Horticulture (LatHort), Graudu Str. 1, LV-3701 Cerini, Krimunu pagasts, Dobeles novads, Latvia
[*]Correspondence: sergejs.kodors@rta.lv

Received: August 24th, 2023; Accepted: February 6th, 2024; Published: February 23rd, 2024

**Abstract.** The yield estimation using artificial intelligence is based on object detection algorithms. Firstly, the object detection algorithms identify the number of fruits on images, then tree fruit load is predicted using regression algorithms. YOLO is a popular convolution neural network architecture for object detection tasks. It is sufficiently well studied for fruit yield estimation. However, the experiments are traditionally restricted to only one specific fruit category and growing season. This is a big shortcoming for the smart solutions like agro-drones, which must automatically complete yield monitoring of the most popular fruit species in commercial orchards. Therefore, the modern studies related to yield estimation increasingly raise attention to multi-stage, multi-state and multi-specie detection tasks. The multi-stage datasets can be described as a collection of multiple sub-datasets, e.g. flowers, fruitlets and fruits. The multi-state dataset can contain classes like mature, immature or damaged fruits. Meanwhile, the multi-specie dataset contains images with representatives of multiple cultures. However, if classic object-detection tasks like urban or indoor object detection have multiple classes presented in one image, then yield estimation datasets usually have images with only one class presented on them. Therefore, an image shuffle or mosaic augmentation are the intuitive training strategies of YOLO for object detection working with a collection of multiple single class datasets. We applied the YOLOv5m model to test both strategies, which were verified on three datasets: apple fruits (MinneApple), pear fruits (Pear640) and pear fruitlets (PFruitlet640). Our experiment showed that mosaic augmentation improves mAP@0.5:0.95 better than simple image shuffle. The mean difference between both strategies is equal to 0.0438.

**Key words:** augmentation, deep learning, yield estimation, object detection, precision horticulture.

## INTRODUCTION

Fruit growing is an important branch of agriculture, an important sector of the economy, and provides a significant part of a healthy diet. It currently faces a number of challenges: climate change, new diseases and pests, public demand for less pesticide use, and competitive production. These contradictory tasks require effective decision-making and prognosis tools based on knowledge-intensive smart fruit-growing solutions and diverse orchard management information. One of the prerequisites for successful decision-making is the availability of up-to-date and accurate orchard monitoring data (Kodors et al., 2021). Regular collection of such data requires involvement of appropriately skilled human resources, which is not always possible in terms of time and costs, especially in small farms. Therefore, the automation of these processes is relevant, including drone-based imaging, recognition and tracking of different stages of fruit development for yield prediction modelling (Moravec et al., 2017).

The automatic fruit yield forecasting consists of three stages: 1) yield estimation; 2) tree fruit load prediction; 3) and yield prediction. The yield estimation is object detection task, which estimates the number of fruits visible on images. However, the actual number of fruits is different than visible on images, because some fruits can be occluded by leaves or counted multiple times on different photos. Therefore, the actual tree load must be predicted by regression algorithms. For example, Vijayakumar et al. (2023) presented combination of YOLO and regression algorithms for citrus load prediction (Vijayakumar et al., 2023), meanwhile, the similar combination of methods was applied by for blueberry load prediction (MacEachern et al., 2023). The yield prediction is based on the application of regression algorithms too, only the fruit load is forecasted by the number of fruits calculated in the previous season or development stage. For example, Cheng et al. (2017) applied simple back propagation neural networks to complete early fruit yield prediction within one season and for next year (Cheng et al., 2017).

Traditionally the experiments are restricted to only one specific fruit category and growing season. This is a big shortcoming for the smart solutions like agro-drones, which must automatically complete yield monitoring of the most popular fruit species in commercial orchards. Therefore, the modern studies related to yield estimation increasingly raise attention to multi-stage, multi-state and multi-specie detection tasks. The multi-stage datasets can be described as a collection of multiple sub-datasets, e.g. flowers, fruitlets and fruits. The multi-state dataset can contain classes like mature, immature or damaged fruits. Meanwhile, the multi-specie dataset contains images with representatives of multiple cultures.

The object of our study is the automation of yield estimation using unmanned aerial vehicles (UAV) and artificial intelligence (AI) in the commercial orchards. In this article we focus on training a neural network to recognize different fruits and their different development stages and states.

The automatic yield estimation using computer vision is sufficiently well studied. For example, Wang et al. (2022) applied modified YOLOv5 architecture for litchi fruit detection and obtained accuracy 92.4% mAP. Meanwhile, Lyu et al. (2022) experimented with yield estimation of green citrus achieving 98.23% mAP@0.5 by using YOLOv5-CS architecture. Many different types of fruits and berries can be mentioned as the object of yield estimation study: tomatoes (Liu et al., 2020), strawberries (Chen et al., 2019), pears (Parico & Ahamed, 2021), etc. However traditionally these studies are scoped by only one specie.

If we are speaking about practical application of unmanned ground vehicles (UGV) or UAV for automatic yield estimation, prediction or harvesting, single-specie and single-stage restrictions are strongly unattractive product features, therefore CNN must be trained to detect many classes.

The intuitive solutions to train convolution neural network (CNN) on a collection of multiple single class datasets are two strategies: image shuffle or mosaic augmentation. The goal of the study is to identify the most suitable strategy for CNN training on collections with multiple single class datasets: image shuffle or mosaic augmentation.

The following objectives are defined to achieve the study goal:

1. Prepare datasets for CNN training.
2. Train CNN using two strategies: image shuffle and mosaic augmentation.
3. Compare obtained accuracies.

A short summary of findings, which presents the novelty of our study:

• PFruitlet640 dataset was collected and annotated for the experiment. PFruitlet640 contains images of pear fruitlets. This dataset was required to evaluate the multi-stage object detection with visually similar objects like a pair of pear fruits (Pear640) and pear fruitlets (PFruitlet640). PFruitlet640 is published in Kaggle repository under CC-BY licence (Web, a).

• Our experiment showed that mosaic training strategy is more suitable for the agricultural image datasets. It showed 4.38% better mAP@50:95 results than image shuffle strategy.

The experiment was completed using the YOLOv5m model, which was selected in our previous experiment (Kodors et al., 2023).

## BACKGROUND

Two strategies of CNN training on multiple single class datasets are discussed in this article: image shuffle and mosaic augmentation (see Fig. 1). Image shuffle strategy considers new dataset creation when images of A and B datasets are mixed one after another. Mosaic augmentation: when images of A and B datasets are split in parts and combined as new images.

The transfer-learning and augmentation disciplines are the most suitable for the theoretical analysis of both solutions because the transfer-learning is related to training on multiple datasets, but the augmentation - image generation by modifying existing images.

For example, Ngiam et al. (2018) found that transfer-learning using fine-grained classified datasets provides better performance. It can be concluded that both strategies, image shuffle and mosaic augmentation, must improve object detection results. Additionally, Ngiam et al. (2018) mention that multiple different categories in datasets provide better results. Therefore, the combination of visually different objects like red apples and green pears must provide better results than visually similar objects like pears and pear fruitlets. Interesting study was completed by Rotshtein et al. (2004), they completed an experiment morphing the photos of Marilyn Monroe and Margaret

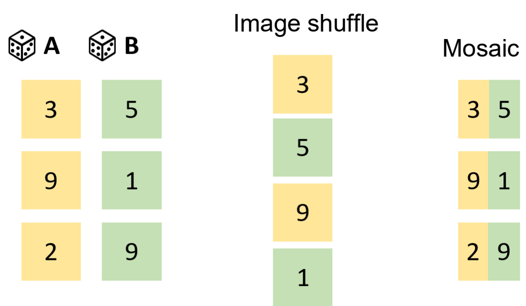


**Figure 1.** Two training strategies: image shuffle and mosaic augmentation.

Thatcher. They experimentally showed that it is harder for people to recognize morphed photos and the mistakes depend on the amount of the added foreign person visual features. Considering that, Chu et al. (2020) mention: 'when there is simply no sufficient data for the tail classes to recover their underlying distribution, the problem of finding an optimal decision boundary becomes ill-defined. In this scenario, it becomes extremely difficult to guess the location of the decision boundary without recovering the distribution first'. So, the increase of visually similar content can stabilise the object recognition finding more correct hyperplanes. Meanwhile, Barman et al. (2019) mention that transfer-learning requires less data than training from scratch. But Sun et al. (2017) has shown that the performance of deep learning models is logarithmically related to the number of training samples (Sun et al., 2017). It means if the number of categories is increased the number of images per class can be decreased. The similar conclusion is mentioned by Li et al. (2023), that collection and production of training samples require a high cost, but augmentation provides low-cost data (Li et al., 2023). Therefore, it is more economically interesting to collect multiple single class datasets than to solve simply one-class problem.

It must be told that the authors of YOLOv4 (Bochkovskiy et al., 2020) mention mosaic augmentation as important element of improvement combination to achieve better performance (Bochkovskiy et al., 2020). Meanwhile, Li et al. (2023) presented Dynamic Mosaic algorithm and Multi-Type Data Augmentation (MDTA) strategy, which were tested and compared with simple mosaic using YOLOv5s and Pascal VOC dataset (Li et al., 2023). The experiment of Li et al. (2023) showed that simple mosaic improved mAP by 7.81%, Dynamic Mosaic - by 8.54%, but MDTA provided additional 3.68% for Dynamic Mosaic. Summers & Dinneen (2019) experimented with different mosaic types, they identified importance of another transformation, which must be applied to the mosaic parts; that was considered by Li et al. (2023) as MDTA strategy.

Considering multiple stage recognition, the most popular subject is plant disease detection. For example, Cruz et al. (2022) trained YOLOv5 models for strawberry disease detection (gray mold, leaf spots, powdery mildew, anthracnose fruit rot, blossom blight), which were integrated to edge computing solution (Cruz et al., 2022). Liu et al. (2023) trained YOLOv5s model to detect disease 'brown rot' on tomatoes. The authors achieved accuracy 89.8% mAP@0.5 (Liu et al., 2023). Tian et al. (2019) applied YOLOv3-dense for the detection of infected apples (Tian et al., 2019).

Speaking about existing studies with mosaic augmentation in agriculture, Dulal et al. (2022) compared the YOLOv5 training strategies with and without mosaic augmentation. By developing a cattle identification solution, they showed that mosaic strategy improves object detection accuracy. Considering similar studies, Ge et al. (2022) developed UGV solution for tomato yield estimation in a greenhouse. Their solution was able to detect tomato-fruit development in the multi-stages. They applied the YOLO-Deepsort network and highlighted the importance of mosaic augmentation. Phan et al. (2023) applied YOLOv5m for tomato multi-state recognition: immature tomato, ripe tomato and damaged tomato. The augmentation description mentions only geometric transformations. However, they obtained 0.97% accuracy, which can be associated with image shuffle.

## MATERIALS AND METHODS

### Object detection model YOLOv5m

The object detection solution YOLO was firstly presented by Redmon et al. (2016) in the publication 'You Only Look Once: Unified, Real-Time Object Detection' (Redmon et al., 2016). The main idea of YOLO was to replace slow post-processing classification of bounding-boxes in older solutions like R-CNN. From limitations of YOLOv1 Redmon et al. (2016) mention struggles with small objects that appear in groups, such as flocks of birds. In YOLOv2 (YOLO9000), Redmon & Farhadi (2017) introduced anchor box architecture, which allowed to detect many objects inside the same grid cell. The anchor boxes of YOLOv2 were calculated with a k-means clustering algorithm. Meanwhile, the 5$^{th}$ generation of YOLO architecture (YOLOv5) provides auto-learning of anchor boxes from training datasets.

The YOLOv5m is a medium-sized version of the YOLOv5 architecture. YOLOv5 framework is available in GitHub repository (Web, b). YOLOv5 was not published in any scientific publication, all related documentation is available in GitHub Wiki of Ultralytics project.

The YOLOv5m model was applied, because it is the most efficient and compatible according to our previous experiments (Kodors et al., 2023). The YOLOv5m model is less GPU intensive than for example the YOLOv5l model, while YOLOv5m provides worse results than YOLOv5l considering Ultralytics experiments (Web, b). However, our experiments showed that YOLOv5l does not provide sufficient increase of accuracy for yield estimation tasks (Kodors et al., 2023). Phan et al. (2023) completed the similar experiments with the yield estimation, and they selected YOLOv5m as the most optimal model too. We have not tuned (changed) YOLOv5m architecture and applied the default model available in yolov5m.yaml, that was marked v6.0.

### Experiment datasets

Three datasets were used for the experiment: MinneApple, Pear640 and PFruitlet640 (see Fig. 2). MinneApple and Pear640 are existing image datasets, which are specially prepared for yield estimation studies, which were firstly presented in the articles of Häni et al. (2020) and Kodors et al. (2023). Both datasets are developed under research projects, have good annotation quality and are sufficiently large. Meanwhile, PFruitlet640 is a novel dataset collected and annotated by our team, which is firstly presented in this publication.

Apples and pears belong to the multi-specie problem, but pear fruitlets and pears - to the multi-stage problem. Additionally, we use knowledge results from our previous study (Kodors et al., 2023), which provides baselines for YOLOv5m accuracy in the case of each dataset independently.

**MinneApple** is a dataset of apple tree photographs (Fig. 2, a), which was collected and annotated by Häni et al. (2020). The dataset details can be found in the paper of Häni et al. (2020), who completed similar experiments and achieved mAP@0.5 77.5% by applying method Faster RCNN. We cropped the images to size of 640×640 px, which are suitable with a YOLOv5m input layer. This was done to save the original image resolution, because the images contained some examples of apples with size 25×25 px, which could disappear after image resizing.

**Pear640** is a dataset of pear fruits (Fig. 2, b) specially prepared for YOLO training with input size 640×640 px. The collection of fruit images was obtained at the end of August (105 days after full bloom) prior to the harvest. The dataset is freely available in Kaggle repository under CC BY 4.0 license. The dataset details can be found in the paper of Kodors et al. (2023).

**PFruitlet640** is a dataset of pear fruitlet instances (Fig. 2, c), which was collected for this experiment due to the shortage of the similar dataset. The digital images of pear fruitlets were collected in the experimental site of the Institute of Horticulture (LatHort) with cultivars 'Suvenirs', 'Vasarine Sviestine' and 'Mramornaya' on seedling rootstocks 'Kazraushu' with planting distances 4×5 m (500 trees per 1 ha). (Krimūnu parish, Dobeles district, Latvia: 56.610169, 23.305956). The collection of fruitlet images of 'Suvenirs', 'Vasarine Sviestine' and 'Mramornaya' was done at the beginning of August (79 days after full bloom).

a) b) c)

**Figure 2.** Image examples of datasets: a) MinneApple; b) Pear640; c) PFruitlet640.

The collection of digital images was carried out using a photo camera of mobile device Huawei P 40: 50 MP Ultra Vision Camera (Wide Angle, f/1.9 aperture) + 16 MP Ultra-Wide Angle Camera (f/2.2 aperture) + 8 MP Telephoto Camera (f/2.4 aperture, OIS), the image size: 3,000×4000 px; 5.0 MP.

The collection of images was carried out in field conditions in 2022, in the orchard at the distance from the tree planting point 2.5 m (middle of alleyway). The whole canopy of trees was photographed as separate objects. The images were taken in the front of a tree (a tree trunk, a planting point), perpendicularly the tree row from the west side of rows (the rows of pear trees oriented from north to south) before noon (10:00–12:00) at natural light conditions.

The dataset is available in Kaggle repository under CC BY 4.0 license (Web, a).

### Experiment design

The experiment was completed in three stages (see Fig. 3): 1st and 2nd stages prepare datasets for CNN training, but 3rd stage trains CNN using two strategies: image shuffle and mosaic augmentation.

**1st stage:** each dataset (MinneApple, Pear640 and PFruitlet640) were split into training, validation and testing subdatasets. 20% of each dataset was taken out to create the testing baseline to compare the results of trained YOLOv5m models.

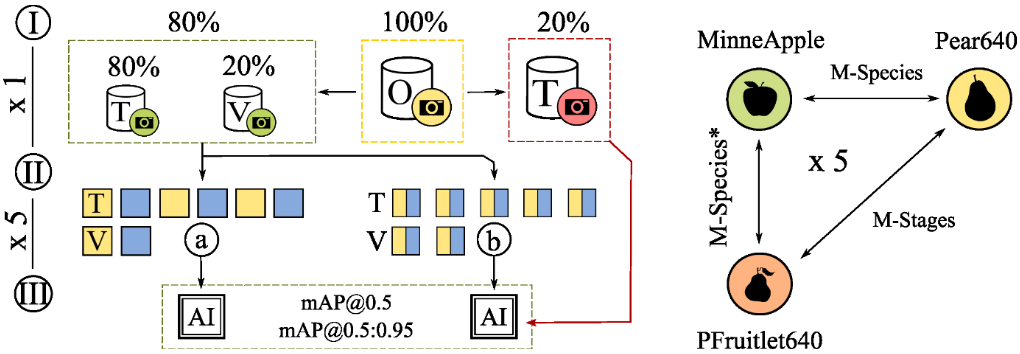


**Figure 3.** Experiment design: I – each dataset is split on training, validation and testing subdatasets; II – two datasets are joint using strategies: (a) image shuffle and (b) mosaic augmentation; III – YOLOv5m model is trained using each strategy and verified on testing dataset.

**2nd stage:** the training and validation datasets were preprocessed using two strategies: image shuffle and mosaic augmentation. Our mosaic augmentation is different from YOLOv5 framework - it joins two parts of images, which belong to different classes, taking the left and right vertical sides of images (see Fig. 4). The Python script was written to create balanced datasets with equal proportion of each category, that was achieved by image repeating of smaller datasets.

**Figure 4.** Example of mosaic augmentation applied in experiment (two images joint together).

**3rd stage:** the YOLOv5m model was trained using the datasets preprocessed in the 2nd stage. The length of training time was 300 epochs. The default augmentation of YOLOv5 framework was modified by setting mosaic and mix-up augmentation to 0. The 2nd and 3rd stages were repeated 5 times to collect statistics for box-plot diagrams and visual comparison of obtained results.The experiment was conducted on an NVIDIA RTX 2070 GPU, which provided sufficient memory and performance for this experiment.

**Accuracy comparison**

Two accuracy parameters of YOLOv5 framework were applied for the quality comparison: mAP@0.5 and mAP@0.5:0.95.

$$AP_k = \int_0^1 p(r)dr, \tag{1}$$

where $p$ – a precision; $r$ – a recall; $AP_k$ – an average precision of the $k$-th class.

$$mAP = \sum_{k=1}^{Q} \frac{AP_k}{Q}, \tag{2}$$

where $mAP$ – a mean average precision for $Q$ classes.

More math details can be found, for example, in the work of Wang et al. (2022).

Considering our experiment, we measured the trained YOLOv5m models on the testing datasets with only one class, therefore mAP (Eq. 2) is equal to AP (Eq. 1). In our study, the more important parameter is the difference between @0.5 and @0.5:095. The mAP@0.5 is a mean average precision for objects with Intersection over Union (IoU) greater than 0.5. Meanwhile, mAP@0.5:0.95 is means of mAP over different IoU thresholds, from 0.5 to 0.95 with step 0.05. In our case, mAP@0.5 shows the accuracy improvement, but mAP@0.5:0.95 depicts the stable object detection and classification.

## RESULTS AND DISCUSSIONS

The experiment results of YOLOv5m training are depicted in Fig. 5. Looking at the results of trained YOLOv5m models, it is required to say that the results should be analysed independently for each combination of datasets, then all separate results must be summarised as the final conclusion that helps to evaluate the best training strategy to improve the accuracy of YOLOv5m.

Analysing the results of Multi-Species (see Fig. 5), where two different species of fruits were used for CNN training, it can be seen that mAP@0.5 results do not differ among image shuffle, mosaic augmentation and even the baseline results obtained in the previous study (Kodors et al., 2023). But there is a significant difference of 5.6% at mAP@0.5:0.95 for the Pear640 dataset, where the results of mosaic augmentation are significantly greater than image shuffle and the baseline results. Even the MinneApple dataset had better results with mosaic augmentation showing accuracy improvement of 2.6%. Additionally, the box-plots of mosaic augmentation results are very narrow, showing stable accuracy results regardless of the case of training.

Moving to the next case, Multi-Stage presents the results (see Fig. 5) where the same fruit in the different growth stages was looked at. The experiment shows the similar results to the previous case (Multi-Species): the lack of difference in the case of mAP@0.5 and the mAP@0.5:0.95 improvement of 6.2% in the case of mosaic augmentation for the Pear640 dataset. And the improvement of 3.9% can be seen with the PearFruitlets dataset using mosaic augmentation as well.

Analysing the last case, Multi-Species* (see Fig. 5) was the dataset where mature apples and pear fruitlets were used for the experiment. The mAP@0.5 results had insignificant differences between image shuffle and mosaic augmentation, similarly to the previous cases. However, looking at mAP@0.5:0.95 results, similarly to the previous cases, mosaic augmentation provides significant improvement of mAP@0.5:0.95 accuracy: PearFruitlets – +3.8% and MinneApple – +4.2%; compared to image shuffle strategy.

To summarise the accuracy difference among three dataset combinations, the accuracy difference was calculated using Eq. 3, the results are depicted in Table 1:

$$\Delta x = (x_1 - x_2), \qquad (3)$$

where $x_1$ is the result accuracy of mosaic augmentation strategy; $x_2$ – of image shuffle.

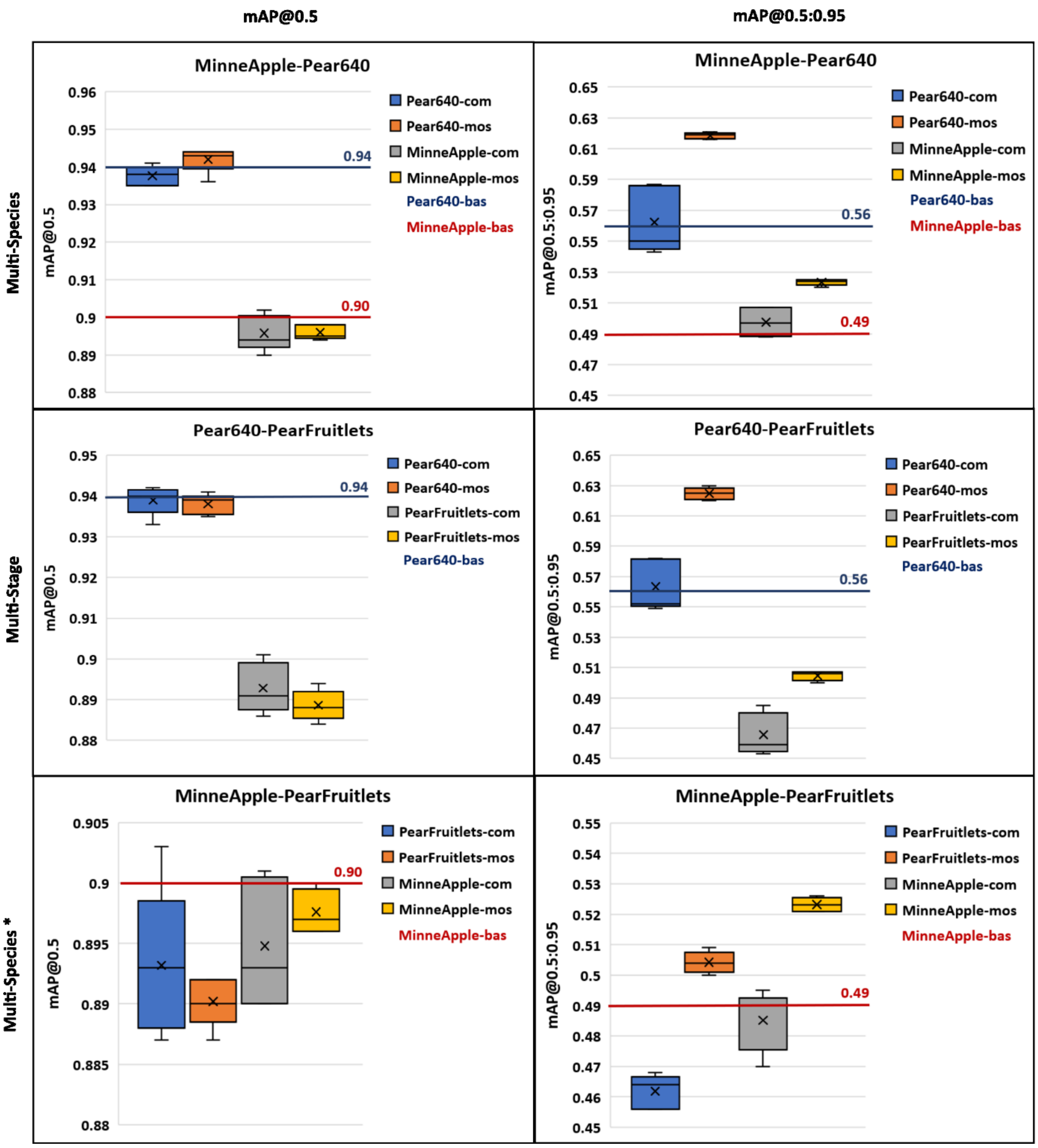


**Figure 5.** Experiment results of YOLOv5m training: *com* – image shuffle strategy; *mos* – mosaic augmentation strategy; *bas* – baselines obtained in previous study (Kodors et al., 2023).

In summary of all three cases, it can be seen that the mosaic augmentation made significant mAP@0.5:0.95 improvement of 4.38% (see Table 1) compared with image shuffle. Meanwhile, the mean value of mAP@0.5 difference is close to zero.

**Table 1.** Difference $\Delta x$ calculation between mosaic augmentation ($x_1$) and image shuffle ($x_2$)

| Combination | M-Species | | M-Stage | | M-Species* | | Mean |
|---|---|---|---|---|---|---|---|
| Category | Apple | Pear | Pear | PFruitlet | Apple | PFruitlet | |
| mAP@0.5 | 0.004 | 0.000 | -0.001 | -0.004 | -0.003 | 0.003 | -0.0001 |
| mAP@0.5:0.95 | 0.056 | 0.026 | 0.062 | 0.039 | 0.042 | 0.038 | 0.0438 |

Considering to Li et al. (2023) experiment results, the simple mosaic increased mAP@0.5 of YOLOv5s by 7.81%, that is higher than in our case. However, Sun et al. (2017) mentioned about logarithmical relation to the number of training samples. Li et al. (2023) worked in the range 70%–80% of mAP@0.5, meanwhile, our YOLOv5m worked in the range 88%-95%. Therefore, considering to logarithmic relation, our study should show smaller accuracy increase than in the case of Li et al. (2023). Ngiam et al. (2018) and Li et al. (2023) mentioned that multiple different categories improve accuracy. Comparing with Li et al. (2023), we only mixed two classes in each combination. Therefore, it would be required to continue experiment to investigate dependence on the number of combined datasets for the yield estimation. Additionally, Barman et al. (2019) mentioned that transfer-learning requires less data than training from scratch. It is useful to verify the accuracy dependence from the number of images per each dataset, which are combined for the yield estimation. That must have stronger impact in the case of small datasets, considering to logarithmical relation detected by Sun et al. (2017).

## CONCLUSIONS

The goal of study was to evaluate image shuffle and mosaic augmentation strategies to select the most appropriate solution for CNN training on an agricultural image collection, which contain multiple single class datasets. The obtained results show that in all cases mAP@0.5 results are negligible as difference between image shuffle and mosaic augmentation strategies and it was close to 0 with mean difference value of -0.0001. Meanwhile, mAP@0.5:0.95 showed all results in favour of mosaic augmentation with mean difference value of 0.0438. Based on experiment results it can be concluded that the most appropriate strategy for the agricultural datasets with multiple single class sub-datasets is the usage of mosaic augmentation. Considering the baselines, which showed better results for one-class problem in the case of mAP@0.5, it can be concluded that mosaic augmentation is strongly required for YOLO CNN training to improve accuracy for the yield estimation. Additionally mosaic augmentation provides more stable accuracy results, which are training case independent.

ACKNOWLEDGEMENTS. This research is funded by the Latvian Council of Science, project ‘Development of autonomous unmanned aerial vehicles based decision-making system for smart fruit growing’, project No. lzp2021/1-0134.

## REFERENCES

Barman, R., Deshpande, S., Agarwal, S. & Inamdar, U. 2019. Transfer Learning for Small Dataset. *In proceedings of National Conference on Machine Learning*. Amity University Mumbai, Mumbai, India, pp. 132–137.

Bochkovskiy, A., Wang, C.Y. & Liao, H.Y.M. 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y. & He, Y. 2019. Strawberry Yield Prediction Based on a Deep Neural Network Using High-Resolution Aerial Orthoimages. *Remote Sens* **11**(13), 1584. doi: 10.3390/rs11131584

Cheng, H., Damerow, L., Sun, Y. & Blanke, M. 2017. Early Yield Prediction Using Image Analysis of Apple Fruit and Tree Canopy Features with Neural Networks. *J. Imaging* **3**(6). doi: 10.3390/jimaging3010006

Chu, P., Bian, X., Liu, S. & Ling, H. 2020. Feature Space Augmentation for Long-Tailed Data. In: *Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision - ECCV 2020. ECCV 2020. Lecture Notes in Computer Science,* vol. **12374**. Springer, Cham. doi: 10.1007/978-3-030-58526-6_41

Cruz, M., Mafra, S., Teixeira, E. & Figueiredo, F. 2022. Smart Strawberry Farming Using Edge Computing and IoT. *Sensors* **2022**(22), 5866. doi: 10.3390/s22155866

Dulal, R., Zheng, L., Kabir, M.A., McGrath, S., Medway, J., Swain, D. & Swain, W. 2022. Automatic Cattle Identification using YOLOv5 and Mosaic Augmentation: A Comparative Analysis. *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA),* 1–8. doi: 10.1109/DICTA56598.2022.10034585

Ge, Y., Lin, S., Zhang, Y., Li, Z., Cheng, H., Dong, J., Shao, S., Zhang, J., Qi, X. & Wu, Z. 2022. Tracking and Counting of Tomato at Different Growth Period Using an Improving YOLO-Deepsort Network for Inspection Robot. *Machines 2022* **10**, 489. doi: 10.3390/machines10060489

Häni, N., Roy, P. & Isler, V. 2020. MinneApple: A Benchmark Dataset for Apple Detection and Segmentation. *IEEE Robotics and Automation Letters* **5**(2), 852–858. doi: 10.1109/LRA.2020.2965061

Kodors, S., Lacis, G., Sokolova, O., Zhukovs, V., Apeinans, I. & Bartulsons, T. 2021. Apple scab detection using CNN and Transfer Learning. *Agronomy Research* **19**(2), 507–519. doi: 10.15159/AR.21.045

Kodors, S., Sondors, M., Lācis, G., Rubauskis, E., Apeināns, I. & Zarembo, I. 2023. RAPID PROTOTYPING OF PEAR DETECTION NEURAL NETWORK WITH YOLO ARCHITECTURE IN PHOTOGRAPHS. ENVIRONMENT. TECHNOLOGIES. RESOURCES, *Proceedings of the International Scientific and Practical Conference* **1**, 81–85. doi: 10.17770/etr2023vol1.7293

Li, Y., Cheng, R., Zhang, Ch., Chen, M., Liang, H. & Wang, Z. 2023. Dynamic Mosaic algorithm for data augmentation. *Mathematical Biosciences and Engineering* **20**(4), 7193–7216. doi: 10.3934/mbe.2023311

Liu, G., Nouaze, J.C., Touko Mbouembe, P.L. & Kim, J.H. 2020. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3. *Sensors* **20**(7), 2145. doi: 10.3390/s20072145

Liu, J., Wang, X., Zhu, Q. & Miao, W. 2023. Tomato brown rot disease detection using improved YOLOv5 with attention mechanism. *Frontiers in Plant Science* **14**. doi: 10.3389/fpls.2023

Lyu, S., Li, R., Zhao, Y., Li, Z., Fan, R. & Liu, S. 2022. Green Citrus Detection and Counting in Orchards Based on YOLOv5-CS and AI Edge System, *Sensors* **22**(2), 576. MDPI AG. doi: 10.3390/s22020576

MacEachern, C.B., Esau, T.J., Schumann, A.W., Hennessy, P.J. & Zaman, Q.U. 2023. Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks, *Smart Agricultural Technology* **3**. doi: 10.1016/j.atech.2022.100099

Moravec, D., Komárek, J., Kumhálová, J., Kroulík, M., Prošek, J. & Klápště, P., 2017. Digital elevation models as predictors of yield: Comparison of an UAV and other elevation data sources. *Agronomy Research* **15**(1), 249–255. Available at https://agronomy.emu.ee/wp-content/uploads/2017/03/Vol15Nr1_Moravec.pdf

Ngiam, J., Peng, D., Vasudevan, V., Kornblith, S., Le, Q. & Pang, R. 2018. Domain Adaptive Transfer Learning with Specialist Models. Available at https://arxiv.org/pdf/1811.07056.pdf

Parico, A.I.B. & Ahamed, T. 2021. Real Time Pear Fruit Detection and Counting Using YOLOv4 Models and Deep SORT. *Sensors* **21**, 4803. doi: 10.3390/s21144803

Phan, Q.-H., Nguyen, V.-T., Lien, C.-H., Duong, T.-P., Hou, M.T.-K. & Le, N.-B. 2023. Classification of Tomato Fruit Using Yolov5 and Convolutional Neural Network Models. *Plants* **12**(4), 790. MDPI AG. doi: 10.3390/plants12040790

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. 2016. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788. Available at https://arxiv.org/abs/1506.02640

Redmon, J. & Farhadi, A. 2017. YOLO9000: better, faster, stronger. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271. Available at https://arxiv.org/abs/1612.08242

Rotshtein, P., Henson, R., Treves, A., Driver, J. & Dolan, R. 2004. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat Neurosci* **8**, 107–113. doi: 10.1038/nn1370

Summers, C. & Dinneen, M.J. 2019. Improved mixed-example data augmentation. In: *2019 IEEE winter conference on applications of computer vision (WACV)*, pp. 1262–1270. Available at https://arxiv.org/abs/1805.11272

Sun, C., Shrivastava, A., Singh, S. & Gupta, A. 2017. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, In: *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 843–852. doi: 10.1109/ICCV.2017.97

Tian, Y., Yang, G., Wang, Zh., Li, E. & Liang, Z. 2019. Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense. *Journal of Sensors* **2019**. doi: 10.1155/2019/7630926

Vijayakumar, V., Ampatzidis, Y. & Costa, L. 2023. Tree-level citrus yield prediction utilizing ground and aerial machine vision and machine learning. *Smart Agricultural Technology* **3**. doi: 10.1016/j.atech.2022.100077

Wang, L., Zhao, Y., Xiong, Z., Wang, S., Li, Y. & Lan, Y. 2022. Fast and precise detection of litchi fruits for yield estimation based on the improved YOLOv5 model. *Frontiers in Plant Science* **13**, doi: 10.3389/fpls.2022.965425

Web (a) PFruitlet640 dataset.
Available at https://www.kaggle.com/datasets/projectlzp201910094/pfruitlet640

Web (b) Ultralytics, YOLOv5-7.0 GitHub repository. Available at https://github.com/ultralytics/yolov5