Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 15th International Scientific and Practical Conference. Volume II, 38-42

Optimal Size of Agricultural Dataset for YOLOv8 Training

Ilmārs Apeināns
Institute of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
ilmars.apeinans@rta.lv

Abstract. **The smart farming solutions are mainly based on the application of convolutional neural networks for object detection tasks. The number of open datasets is restricted in the agricultural domain. Therefore, it is required to find the answer to the question: how big a dataset must be collected to train a convolutional neural network for object detection tasks? To solve this task, the YOLOv8 framework was selected for the experiment. Three datasets were prepared: MinneApples, PFruitlets640 and mosaic dataset using both previously named datasets. 100 images were selected for testing. Other images were used to create training datasets, which had the size from 100 until 1000 images with step 100 images. Training was repeated 10 times with each size of dataset. The experiment showed that the increase of dataset from 100 to 500 images provides an accuracy growth up to 15.48% mAP@0.5, but from 600 to 1000 images - only 2.98% mAP@0.5. This study experimentally proves that the dataset size equal to 500 images is the most efficient. Meanwhile, the experiment with the mosaic dataset shows constant accuracy improvement. Therefore, it is more advisable to collect different classes with 500 images than one large dataset. This study will be interesting not only for smart farming experts as well as for all machine learning experts.**

Keywords: artificial intelligence, deep learning, precision farming, YOLOv8.

I. INTRODUCTION

Precision farming is a management strategy with the goal to improve productivity, resource usage efficiency, quality and profitability of the food industry and sustainability in the agricultural sector [1]. To achieve the goal of precision farming, a wide range of tools can be applied starting from sensors to acquire information in real time until artificial intelligence (AI), which can be used to make assessments of data and automatic decisions based on aquired data. If the sensors and IoT are applied to collect data, then AI can be used to generalize them, find correlations and automatically analyze acquired data. The accuracy of the artificial neuron networks highly depends on the amount of data, which was used during the training process. In the scope of this article, experiments were made with the goal to find the optimal size of agricultural dataset for object detection using YOLOv8 architecture.

YOLOv8 [2] (You Only Look Once version 8) is an object detection architecture, which is a continuation of the YOLO family models, which are known for their speed and accuracy. One of improvements in YOLOv8 architecture is its enhancement for small object detection [3], which are quite often use-case in agricultural datasets. The detectable objects may be quite small, such as flowers, pests, small fruits like cherries [4]. The authors of the article "Small pests detection in field crops using deep learning object detection" [5] compared several versions of YOLO architectures. Their experiment showed that YOLOv8 provided the best results of 84.7% mAP.

The property of YOLOv8 is its ability to perform object detection in real time. As well as, YOLOv8 can work with several classes of objects at the same time and mark placement of detected objects. That is essential for smart agrobot development, because manual imagining is a cost-ineffective approach.

As stated before, in many cases the variation of cultivars may affect training results. For example, apples have more than 7500 different cultivars [6]. Same applies if we look at the regional scale as well. Each global region will have different fruits and vegetables cultivars compared to another region in the world. Considering the seasonal limits, when a dataset can be collected, the resulting amount of available datasets that can be used for training is highly limited.

Currently a large number of challenges still exist in the way of implementing AI solutions in agriculture such as insufficient research or security from cyber-attacks, or dependency on technology in general [7]. One of mentioned challenges relevant to this article is stakeholders' and owners' requirements for the high precision of AI and solution suitability to their farm ecosystems. To overcome this challenge it is important to study more about the rapid development of artificial intelligence. The big part of preparing an AI model for

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2024vol2.8041

usage in real-life situations is related to a dataset collection and its preprocessing. Based on that, the task to find out the optimal size of dataset for the agriculture domain is a good starting point to solve the defined challenge.

If the situation with general object detection in agriculture is salvageable with an increased number of specialists for collecting general images for datasets, then the situation with more specific objects is much more complex due to limited expert number or sample availability. E.g. there is a lack of datasets in the field of plant diseases [8]. The disease has several stages of maturing, and for the different types of fruits and vegetables, the disease will be visually different.

Precision farming and artificial intelligence usage in farming is hindered by real life elements, mainly, by the low number of datasets which can be used for training. Fruits and vegetables are usually harvested once per year, therefore the time frame to collect images for the dataset is limited. If by any chance a time period is missed or errors are made during data collection, the next opportunity is only the next period of maturity of the studied fruits or vegetables. After images are collected, then a dataset is prepared for training. The annotation of images is done together with an image discarding, which are damaged, blur or contain other visual distortions. In summary, the preparation of an agricultural dataset is a monotonic and time-consuming process. As a result of that the question appeared "How many images in the dataset is enough to train an efficient object detection prototype?".

Summarizing the previously stated, the collection and annotation of a large training dataset is time consuming process, which is limited by time frame, when it is possible to photo the flowering, maturing or any other phase of growth of fruits or vegetables, that may be only a week or even less. If time to collect data is missed, the next chance may be only provided in the next season. Considering these factors, the study questions are "Is it more efficient to collect more, but smaller datasets for training is better?" or "What is the optimal size of dataset for the agricultural domain?".

The aim of the study is to identify the optimal size of agricultural dataset for the CNN training using the YOLOv8 framework.

The experiment was performed using three datasets: two datasets were publicly available datasets and the third was generated using vertical mosaic augmentation. The mosaic augmentation represents the merging of 2 or more images into a single image. The importance of mosaic augmentation was mentioned in an article about YOLOv4 architecture [9], where it was proven that mosaic augmentation improves average accuracy of trained CNN. The additional influence of mosaic augmentation is an artificial increase of training dataset. And mosaic augmentation was added to all later YOLO family architectures.

II. Materials and Methods

Datasets: in the experiment three datasets were used for training, validation and testing. Two public datasets were applied: MinneApple [10] and PFruitlets640 [11]. The third dataset was generated from two mentioned datasets.

The first dataset which was selected for the experiment is MinneApple. It consists of 1308 images and over 41000 annotations. The dataset contains images of apples (see Fig. 1). The images were cropped into size of 640x640 without any alterations to bounding boxes to preserve the dataset compatibility with other experiments.

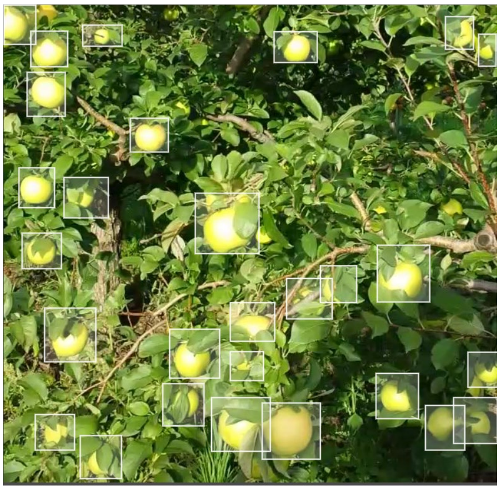

Fig. 1 MinneApples [10]

PFruitlets640 consists of 1455 annotated images of pear fruitlets (see Fig. 2). The images are already prepared for training using YOLOv8, therefore the images have size equal to 640x640. The object of annotation in this dataset is pear fruitlet, which visually looks similarly to apples from MinneApple dataset.

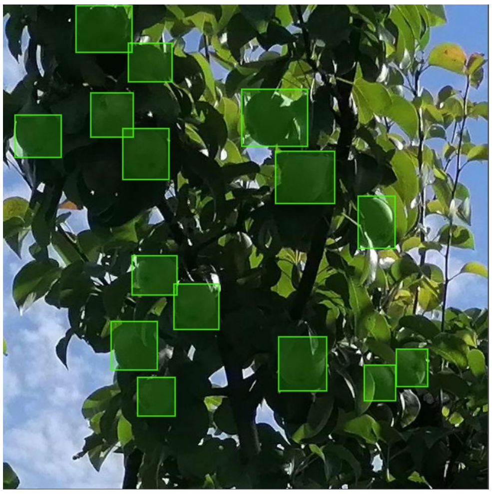

Fig. 2 PFruitlets640 [11]

The mosaic dataset was created using MinneApples and PFruitlets640 datasets. It consists of 4544 images with bounding boxes. To create this dataset, original datasets, MinneApples and Pfruitlets640, were automatically cut and merged by vertical axis in the middle of the images (see Fig. 3). The labels consist of two classes: apples and pear fruitlets.

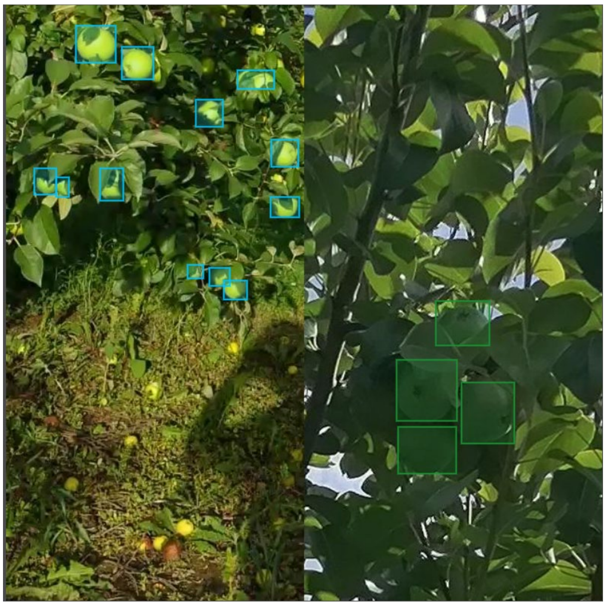

Fig. 3.Mosaic dataset

CNN training: YOLOv8n model [2] was applied in this experiment.The experiment was conducted on a GPU NVIDIA RTX 4070Ti.

Training was done using three datasets:

- PFruitlets640 - 1455 images;
- MinneApples - 1308 images;
- Mosaic dataset was created from two previous datasets - 4544 images.

100 images of each dataset were selected for the testing datasets for later usage after training, the rest of images was left for training and validation. In the experiment, it was important to test the precision on similar images changing the size of the training dataset.

The starting size of training and validation datasets were 100 training images and 30 validation images, which were selected using a random shuffle method in Python script. The YOLOv8n was trained 5 times. After finishing training, a new dataset was created with an increased number of images in the training dataset and validation set. Increases of datasets were +100 images in the training dataset and +30 images in the validation dataset. It was continued until the size of the training dataset and validation dataset reached 1000 training and 300 validation images. Same steps were repeated for all 3 used datasets, if that was possible. MinneApples dataset was not big enough to facilitate training steps with 1000 training and 300 validation images.

After finishing the training phase of the experiment, the testing phase was done by selecting the best trained model for each datasets, which were used to test training results on separated datasets with 100 images. The separate datasets were not used in the training or validation stages, they provided the most accurate testing results. The achieved results were used in discussion.

III. Results and Discussion

Three datasets were used to determine the optimal size of agricultural dataset for CNN training, validation and testing. If we look at achieved training results (see Tab. 1.), it can be seen that the larger size of dataset provides the better accuracy results. E.g. MinneApples dataset provided accuracy mAP@0.5 of 66.41% in the case of 100 training images. Trained models using PFruitlets640 dataset achieved the worst result at mAP@0.5 of 61.49% and trained models on Mosaic dataset created by combining both previously named datasets achieved mAP@0.5 of 69.78% at the same size of dataset.

The largest size of the dataset was selected to be 1000 images in the training dataset. Due to the limitations of MinneApples dataset, which contained only 1208 images, the accuracy results for 1000 images are not provided in Table 1. The maximal training size of MinneApples was 900 images, which provided mAP@0.5 of 83.05%. Models trained on PFruitlets640 achieved mAP@0.5 of 83.29% with a dataset size of 1000 images. At the same size of dataset (1000), Mosaic dataset's models achieved the best result of mAP@0.5 equal to 86.76%, which is the highest precision among 3 datasets used in the experiment (see Fig. 1).

If we look at the experiment results in Fig. 4, then on Mosaic dataset provided the best trained models in general, with the highest mAP@0.5 of 69.78% for training datasets size of 100 images and mAP@0.5 of 86.76% for training dataset size of 1000 images.

Table 1. Experiment results (mAP@0.5)

Dataset size	*MinneApples*	*Mosaic*	*PFruitlets640*
100	0.66414	0.69786	0.61492
200	0.71610	0.75534	0.67192
300	0.74449	0.80103	0.73231
400	0.75954	0.81256	0.78092
500	0.79957	0.83565	0.80640
600	0.80833	0.84265	0.82048
700	0.81587	0.85096	0.82126
800	0.81807	0.85503	0.82440
900	0.83060	0.86304	0.82760
1000	-	0.86763	0.83294
1000-500	0.03103	0.03198	0.02655
500-100	0.13542	0.13780	0.19148

The main goal of this study was to find out the optimal size of a dataset for agricultural model training. As such it is important to look at the difference of precision of models with different training dataset sizes. In Tab. 1, the rows "1000-500" and "500-100" depict the accuracy improvements, which were achieved increasing the datasets from 100 to 500 and from 500 to 1000. Based on the experiment results, it can be calculated that the increase in mAP@0.5 for dataset MinneApples from 100 to 500 images was 13.54%, while mAP@0.5 increased only by 3.1% in the range of dataset size from 500 to 1000 images. The similar results were achieved in the case of PFruitlets640 dataset, where the precision improvement, from 100 to 500 images, was 19.14% at mAP@0.5. Meanwhile, from 500 to 1000 images, the increase of mAP@0.5 for PFruitlets640 was only 2.65%. In the case of Mosaic dataset, the increase of precision from 100 to 500 images was 13.77% at mAP@0.5, while the increase was 3.19% in the range from 500 to 1000 images.

Based on previously looked data, it was shown that the models trained on Mosaic dataset provided the best results among all sizes of datasets that were used. Taking in consideration that, Mosaic dataset provided the best training results in general, it will be applied to calculate the mean increase of precision to compare it with the training results of MinneApples and PFruitlets640 datasets (see Tab. 2).

TABLE 2. ACCURACY IMPROVEMENT PROVIDED BY MOSAIC AUGMENTATION

Dataset size	*Mosaic-MinneApples*	*Mosaic-Pfruitlets640*	*Mean*
100	0.03372	0.08294	0.05833
200	0.03923	0.08342	0.06133
300	0.05655	0.06872	0.06263
400	0.05301	0.03164	0.04233
500	0.03609	0.02925	0.03267
600	0.03432	0.02217	0.02824
700	0.03509	0.02970	0.03239
800	0.03696	0.03063	0.03380
900	0.03245	0.03544	0.03394
1000	-	0.03468	0.03468

Data shown in Tab. 2 are calculated considering the fact that the Mosaic dataset provided the best training results. In the case of column “Mosaic minus MinneApples” of Tab. 2, where the accuracy improvements by mosaic augmentation are provided in comparison with MinneApples dataset, data shows an increase in the range from 3.24 % until 5.65%. Next column (Tab. 2, Mosaic minus Pfruitlets640) shows comparison of Mosaic dataset and PFruitlets640 dataset results. Increase of accuracy in the case of Mosaic dataset provides a wider range from 2.21% to 8.34%. Using the acquired data, the mean accuracy improvement was calculated (Tab. 2, Mean). The highest increase in accuracy is seen at the dataset sizes: 200 images (6.13% of mAP@0.5) and 300 images (6.26% of mAP@0.5). The improvement of accuracy is increasing up to 300 images and from this point the accuracy starts to decline (Fig. 5) until it reaches the lowest point of 2.82% mAP@0.5 at a dataset size of 600 images. After this point, accuracy increase per dataset size is stable around 3%.

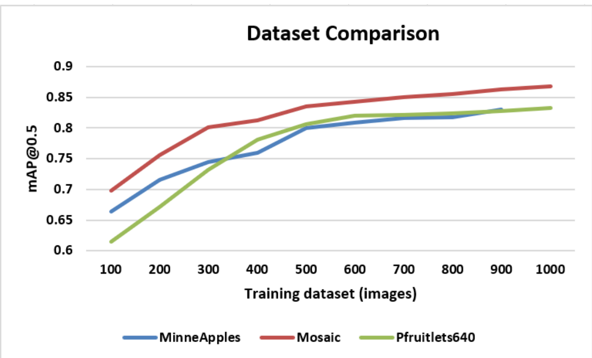

Fig. 4. Dataset accuracy comparison

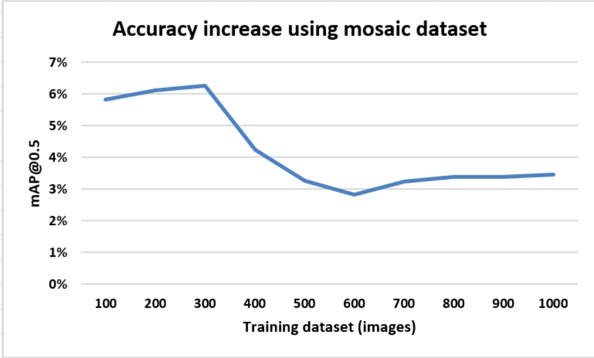

Fig. 5. Increase in accuracy using the mosaic augmentation

IV. CONCLUSIONS

In this article, the optimal size for agricultural dataset was experimentally identified for training object detection CNN. The experiment was completed by using the YOLOv8n model and several public datasets.

The objective of this experiment conducted during the writing of this article was to determine the optimal size of agricultural dataset for YOLOv8n architecture training. To achieve this goal two public datasets, Pear640 and MinneApples, and Mosaic dataset, created by using both public datasets, were used in experiments. To determine the optimal dataset size, CNN models were trained using YOLOv8n on randomly selected images from starting datasets. Dataset size was in the range from 100 images until 1000 images in the dataset with step of size 100 images. Experiment results showed that the biggest increase in accuracy was achieved with dataset size of 500 images, for MinneApple dataset: 13.52% mAP@0.5, for PFruitlets640 dataset: 19.14% mAP@0.5 and for Mosaic dataset: 13.78% mAP@0.5. In comparison with dataset increase from 500 images to 1000 images accuracy increase was smaller, for MinneApple dataset: 3.1% mAP@0.5, for PFruitlets640 dataset: 2.65% mAP@0.5 and for Mosaic dataset: 3.19% mAP@0.5.

ACKNOWLEDGEMENT

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. lzp-2021/1-0134.

REFERENCES

[1] E. S. Ramil Brick, J. Holland, D. E. Anagnostou, K. Brown, and M. P. Y. Desmulliez, “A review of agroforestry, precision agriculture, and precision livestock farming—The case for a data-driven agroforestry strategy,” Frontiers in Sensors, vol. 3, Sep. 2022, doi: https://doi.org/10.3389/fsens.2022.998928.

[2] Ultralytics, YOLOv8 GitHub repository. [Online] Available: https://github.com/ultralytics/ultralytics [Accessed: Jan15, 2024]

[3] R. Bai, F. Shen, M. Wang, J. Lu, and Z. Zhang, “Improving Detection Capabilities of YOLOv8-n for Small Objects in Remote Sensing Imagery: Towards Better Precision with Simplified Model Complexity,” Jun. 2023, doi: https://doi.org/10.21203/rs.3.rs-3085871/v1.

[4] Y. Bai, J. Yu, S. Yang, and J. Ning, “An improved YOLO algorithm for detecting flowers and fruits on strawberry seedlings,” Biosystems Engineering, vol. 237, pp. 1–12, Jan. 2024, doi: https://doi.org/10.1016/j.biosystemseng.2023.11.008.

[5] S. Khalid, H. M. Oqaibi, M. Aqib, and Y. Hafeez, “Small Pests Detection in Field Crops Using Deep Learning Object Detection,” Sustainability, vol. 15, no. 8, p. 6815, Jan. 2023, doi: https://doi.org/10.3390/su15086815.

[6] A. Elzebroek and K. Wind, Eds., “Guide to Cultivated Plants,” Jul. 2008, doi: https://doi.org/10.1079/9781845933562.0000.

[7] I. Apeinans, L. Litavniece, S. Kodors, I. Zarembo, G. Lacis, and J. Deksne, “Smart Fruit Growing Through Digital Twin Paradigm: Systematic Review and Technology Gap Analysis,” Engineering Management in Production and Services, vol. 15, no. 4, pp. 128–143, Dec. 2023, doi: https://doi.org/10.2478/emj-2023-0033.

[8] M. Arsenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, “Solving Current Limitations of Deep Learning Based Approaches for Plant Disease Detection,” Symmetry, vol. 11, no. 7, p. 939, Jul. 2019, doi: https://doi.org/10.3390/sym11070939.

[9] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” Apr. 2020. Available: https://arxiv.org/pdf/2004.10934.pdf

[10] Nicolaihaeni. (n.d.). GitHub - nicolaihaeni/MinneApple: A Benchmark Dataset for Apple Detection and Segmentation. [Online]. Available: https://github.com/nicolaihaeni/MinneApple [Accessed Jan. 3, 2024].

[11] S. Kodors, M. Sondors, I. Apeinans, I. Zarembo, G. Lacis, E. Rubauskis and K. Karklina, “Importance of mosaic augmentation for agricultural image dataset”, 2024.