YOLOv5 Deep Neural Network for Quince and Raspberry Detection on RGB Images

1st Kaspars Sudars
EDI - Institute of Electronics and Computer Science
Riga, Latvia
sudars@edi.lv

2nd Ivars Namatēvs
EDI - Institute of Electronics and Computer Science
Riga, Latvia
ivars.namatevs@edi.lv

3rd Jānis Judvaitis
EDI - Institute of Electronics and Computer Science
Riga, Latvia
janis.judvaitis@edi.lv

4th Rihards Balašs
EDI - Institute of Electronics and Computer Science
Riga, Latvia
rihards.balass@edi.lv

5th Artūrs Ņikuļins
EDI - Institute of Electronics and Computer Science
Riga, Latvia
arturs.nikulins@edi.lv

6th Astile Peter
EDI - Institute of Electronics and Computer Science
Riga, Latvia
astile.peter@edi.lv

7th Sarmīte Strautiņa
Institute of Horticulture
Riga, Latvia
sarmite.strautina@llu.lv

8th Edīte Kaufmane
Institute of Horticulture
Riga, Latvia
edite.kaufmane@llu.lv

9th Ieva Kalniņa
Institute of Horticulture
Riga, Latvia
ieva.kalnina@llu.lv

***Abstract*—Object detection based on deep learning can be widely used in all kinds of agricultural applications. In this paper, we present a deep neural network (DNN) model for quince and raspberry detection on RGB images. The trained DNN model is based on YOLOv5 architecture and it has 7 berry classes related to the berry development stage. YOLOv5 provides sufficiently good performance and precision trade-off. It is useful in the process of quince and raspberry phenotyping for the agriculture experts, where the yield and berry size parameters have to be estimated. Using our DNN model we have shown that it is possible to achieve a mean Average Precision close to 80.9 % and in some cases (Average Precision) close to 95 % for some classes. The DNN model is trained on labeled data gathered during the AKFEN project. The developed raspberry and quince detector is freely available at the GIT repository [1]. Further, the research on Sensor Networks, Wireless Systems, 3D point cloud processing and multi-spectral image processing has to be carried out leading to high-throughput phenotyping.**

***Index Terms*—Deep learning, Deep neural networks, YOLOv5, Precise agriculture, Yield estimation**

I. Introduction

In precise agriculture, high-throughput phenotyping is targeted leading to more productive berry varieties and larger yields. It requires multi-disciplinary research and advances in Sensor Networks, Wireless Systems, Image processing, 3D point cloud processing and multi-spectral image processing. In this paper, we use Artificial Intelligence and present a deep neural network (DNN) model for quince and raspberry detection in RGB images. It is based on YOLOv5 architecture and trained on labeled data sets - QuinceSet and RaspberrySet - created for this purpose during the AKFEN project [2], [3]. The developed model can be used for yield estimation and it has a value to phenotype plants in precise agriculture. The developed code is available as an open project at the GIT repository [1].

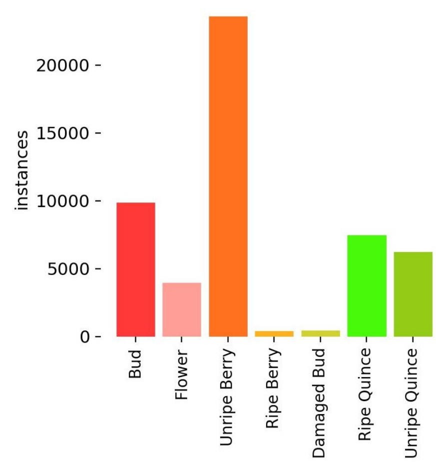

Fig. 1: Object count in merged data set for each class.

Object detection primary based on DNN is crucial for many applications in computer vision and now is widely used also in tasks related to agriculture. Two primary benchmarks of the DNN model are model inference speed and model accuracy. We address these issues by using the YOLOv5 deep neural network, which is considered to provide good overall performance and accuracy [6].

For training the DNN model in a supervised way the labeled

Fig. 2: The DNN model detection results on test image.

data set is required. Therefore data processing is applied before the training. Data are taken from merging QuinceSet and RaspberrySet data sets [2], [3], which are both created during the AKFEN project [4]. It leads to representative data set, which is important for training robust and usable in real-life DNN model.

The trained DNN model provides up to 95 % average precision (AP) for some classes and 79.8 % mean average precision (mAP) for all classes on a validation set. It is sufficient for an application like a yield estimation in agriculture. Also, the trained model provides 7 object classes enabling to estimate of the berry development stage, which gives more detailed information about yield.

The following Sections of the paper are addressed to describe the usage of deep learning approaches in phenotyping and yield estimation, merged data set for training the DNN model, and obtained results.

II. Related work

Deep neural networks are widely used in applications related to agriculture. For instance, YOLO can be used for Crop Disease Detection [7]. The deep learning model based on YOLOv3 was trained using PlantVillage dataset available on Kaggle. Pre-trained darknet-53 convolutional weights were used for the training. The model can be used to automatic disease detection.

Automated weed detection with the help of convolution neural networks [8]. This paper uses Python and Tensorflow framework to implement the approach. The images captured using the phone and camera were labeled and fed into the neural network model. The accuracy of 98% was achieved using the transfer approach.

Convolutional Neural Network-based method to identify corn diseases [9]. In this paper, the images were pre-processed to remove the noise and the unwanted details in the sample images. The trained deep convolutional model detected various diseases associated with the corn plants.

YOLOv3 and YOLOv5 can detect Apples in the Orchards [10]. In this paper, the model detects the Apples with high accuracy without further complications. To improve the accuracy of detecting the Apples on a challenging background or environment, the images were pre-processed by thickening the borders, introducing slight blurs, and adaptive histogram alignment.

III. YOLOv5 deep neural network

Due to its good performance and accuracy, the YOLOv5 deep neural network architecture is chosen for the particular task. For object detection, it is supposed to be the latest model of the YOLO family. It supports the widely used Pytorch framework, which also leads to a preference for YOLOv5. It is introduced by Glenn Jocher using the Pytorch framework. The open-source code is available on GitHub [5], [6].

YOLOv5 is a real-time object detection framework in which images are only passed once through a fully convolutional neural network FCNN. YOLOv5 uses an evolutionary algorithm that takes the result of k-mean anchors and uses those as the initial starting point. Then all the anchors are passed through thousands of generations on the actual cost function on which the model will be trained. YOLOv5 also uses Auto-anchor, in which code will look at the anchors and compares them against your data. Then it will be compared to the threshold, and if they don't fill well, it will start training automatically using the same method. YOLOv5 is six times faster than YOLOv4 and much faster than EfficientDet provided by Google. The mean Average Precision offered by the YOLOv5 is also greater than Efficientdet. YOLOv5 can achieve fast detection at 140 FPS running on the Nvidia TESLA P100, while YOLOv4 can attain only 50 FPS [5], [10].

IV. Training and validation data

The merged data set is used for the training and validation of the DNN model. It is merged from two data sets and elaborated by agriculture experts in the Institute of Horticulture (located in Dobele, Latvia) during the AKFEN project. One data set used in the development of the DNN model is QuinceSet - the annotated dataset for Japanese quince, which consists of images of Japanese quince (Chaenomeles japonica) fruits taken at two phenological developmental stages and annotated for detection and phenotyping [2]. It contains 1515 high-resolution RGB images with the same number of annotated text files. A total of 17171 annotations were provided by the experts.

Another raspberry data set (RaspberySet) is similarly created and labeled by experts in the Institute of Horticulture during the AKFEN project [3]. It consists of 2039 RGB images and 46659 object instances in total. This set has 5 object classes and it will be published as an open data.

Classes of the merged data set are the following: (1) Raspberry Bud, (2) Raspberry Flower, (3) Unripe Raspberry, (4) Ripe Raspberry, (5) Damaged Raspberry Bud, (6) Unripe Quince (the fruits have reached 30-50 % of their final size) and (7) Ripe Quince (just before the fruits are yielded). Also,

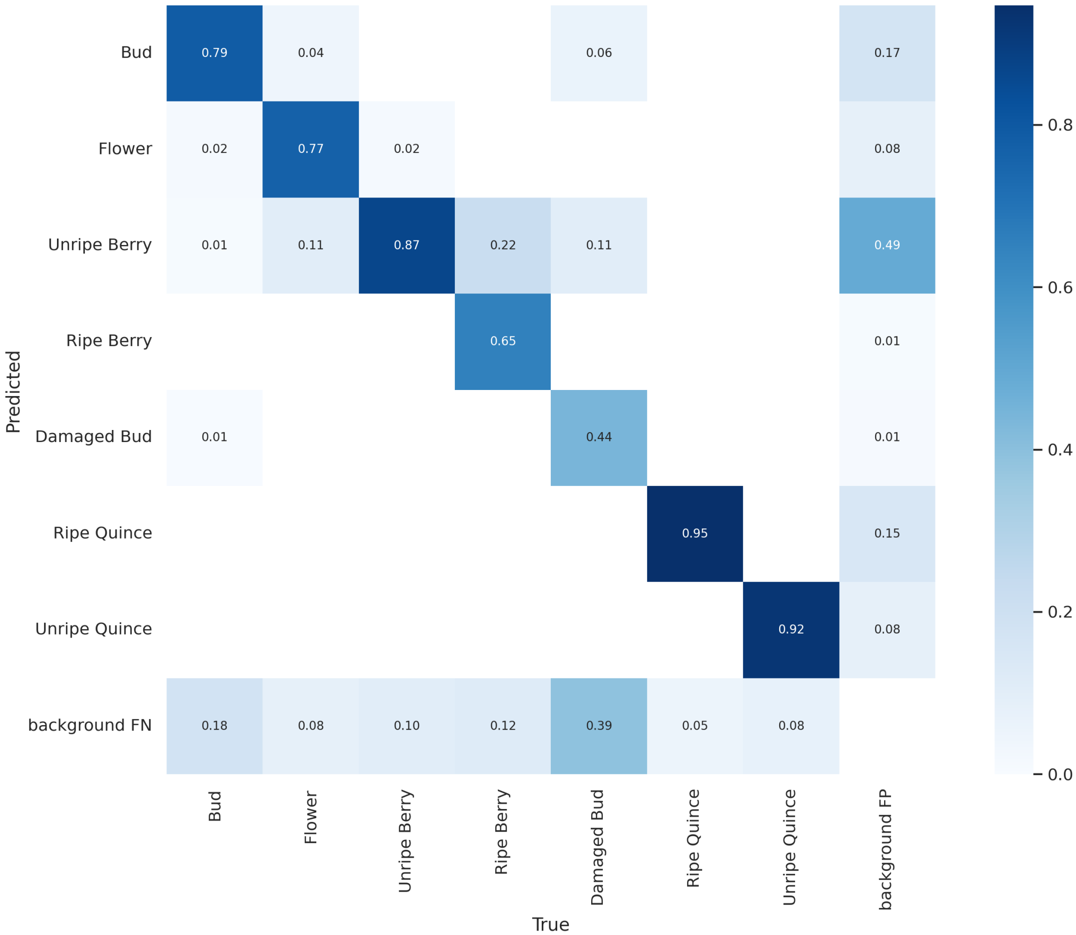

Fig. 3: Confusion matrix (confusion rate of the predicted and true object classes).

the total number of instance count for each class is shown in Figure 1. The proportion for the training set and validation set is 80:20.

V. Obtained Results

After labeled data pre-processing the DNN model based on YOLOv5 architecture can be trained and validated. Some examples of detection results at different development stages can be seen in Figure 2.

Class	Images	Labels	mAP@.5	mAP@.5:.95
all	722	12785	0.809	0.537
Bud	722	2276	0.828	0.467
Flower	722	952	0.825	0.484
Unripe Berry	722	5882	0.891	0.521
Ripe Berry	722	107	0.77	0.513
Damaged Bud	722	113	0.445	0.215
Ripe Quince	722	1932	0.954	0.765
Unripe Quince	722	1523	0.947	0.796

Fig. 4: Average Precision AP for each class in the validation set.

The trained DNN model can make classification mistakes and confuse some labels more frequently. For example, as it can be seen from Figure 3, the damaged buds are more likely to be confused with unripe berry (raspberry) or undamaged bud. It can be explained by the relatively small presence of the damaged buds in the whole data set (see Figure 1). Damaged buds are not represented in sufficient numbers in the data leading to confusion by the DNN model. The more even instance distribution among classes would make the data set more representative and it would improve the model's mean Average Precision. This issue has to be addressed by including more labeled data from these classes.

In Figure 5 there can be seen the DNN model convergence against data epochs from 0 to 125. It shows the parameters like bounding box loss, classification loss, and total object detection loss on training and validation sets. Also, the DNN model precision and recall are estimated during the training. The DNN model is trained using the Nvidia K40 graphic card for several hours. The inference for one image takes 0.02 seconds, which is sufficient for numerous real-time applications.

The Average Precision AP for each class can be seen in Figure 4. Also, it is shown the trained DNN model has a mean Average Precision mAP of 80.9% for objects intersecting more than 50%. In the case of highly intersecting objects (more than

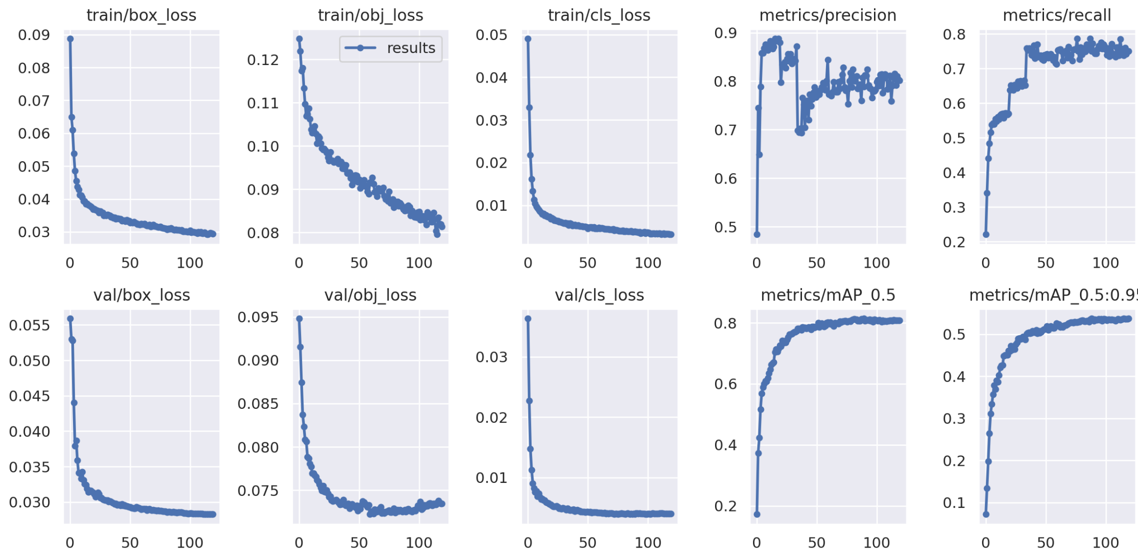

Fig. 5: The DNN model characteristics against epochs on training and validation sets.

95% of intersection), the DNN model provides mAP close to 53.7%.

The lowest AP has damaged raspberry buds due to relatively low numbers presented in the labeled data. Other issues are the available RGB image resolution and the image size of YOLOv5 input for the smaller objects, which can result in poorer AP if some vital information is missing necessary for distinguishing the difference between classes. In the case of larger objects like quinces, the DNN model provides better Average Precision. It is shown in Figure 4. For improving mAP the model can be retrained on larger YOLOv5 input sizes compromising performance.

VI. Conclusions

Object detection based on deep learning can be widely used in all kinds of agricultural applications. In the paper, the DNN model is shown for quince and raspberry detection on RGB images, which can be used for yield and berry size parameter estimation. This is especially useful in the process of berry phenotyping. We have shown that it is possible to achieve a mean Average Precision of about 80.9 % and in some cases close to 95 % for some classes. The trained DNN model is based on YOLOv5 architecture. Nevertheless, other DNN architectures can be considered for object detection. The DNN training process was based on the supervised method. For this reason, the labeled data set merged from two data sets - QuinceSet and RaspberrySet [2], [3]. Both are labeled by experts and obtained during the AKFEN project. The developed code (Raspberry and quince detector in images) is freely available as an open project at the GIT repository [1].

Further research has to lead to research and development (R&D) activities on Sensor Networks, Wireless Systems, 3D point cloud processing and multi-spectral image processing to help improve expert productivity and efficiency of the high-throughput phenotyping.

Acknowledgment

This work was funded by Latvian Council of Science project No. LZP-2020/1-0353 “Smart non-invasive phenotyping of raspberries and Japanese quinces using machine learning and hyperspectral and 3D imaging AKFEN”.

References

[1] Open software of the AKFEN object detector for the raspberries and quinces at GIT: https://pubgit.edi.lv/kaspars.sudars/akfen-object-detector

[2] E. Kaufmane, K. Sudars, I. Namatēvs, I. Kalniņa, J. Judvaitis, R. Balašs, S. Strautiņa, ”QuinceSet: Dataset of Annotated Japanese Quince Images for Object Detection”, Data in Brief, https://doi.org/10.1016/j.dib.2022.108332, 2022.

[3] S. Strautiņa, I. Kalniņa, E. Kaufmane, K. Sudars, I. Namatēvs, A. Ņikuļins, J. Judvaitis, R. Balašs, RaspberrySet: Dataset of Annotated Raspberry Images for Object Detection, Available online: https://zenodo.org/record/7014728

[4] Latvian Council of Science project No. LZP-2020/1-0353 ”Smart non-invasive phenotyping of raspberries and Japanese quinces using machine learning and hyperspectral and 3D imaging AKFEN”.

[5] YOLOv5 Git repository: https://github.com/ultralytics/yolov5

[6] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, TaoXie, J. Fang, imyhxy, K. Michael, Lorna, Abhiram V, D. Montes, J. Nadar, Laughing, tkianai, yxNONG, P. Skalski, Z. Wang, ... M. T. Minh, ”ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference (v6.1).”, Zenodo, https://doi.org/10.5281/zenodo.6222936, 2022.

[7] A. Morbekar, A. Parihar, and R. Jadhav, ”Crop Disease Detection Using YOLO,” International Conference for Emerging Technology(INCET), 2020.

[8] O. Tiwari, V. Goyal, P. Kumar, and S. Vij, “An experimental set up for utilizing convolutional neural network in automated weed detection,” 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), 2019.

[9] J. Tian, Y. Zhang, Y. Wang, C. Wang, S. Zhang, and T. Ren, ”A Method of Corn Disease Identification Based on Convolutional Neural Network”, 12th International Symposium on Computational Intelligence and Design (ISCID), 2019.

[10] T. Maleva, A. Kuznetsova, and V Soloviev, “Detetcting Apples in Orchards Using YOLOv3 and YOLOv5 in General and Close-Up Images”, International Symposium on Neural Networks, 2020.

[11] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, “YOLOX: Exceeding YOLO Series in 2021”, https://arxiv.org/abs/2107.08430