Strawberry yield monitoring based on a convolutional neural network using high-resolution aerial orthoimages

Alexey Kutyrev[*], and *Filippov* Rostislav

Federal Scientific Agroengineering Center VIM, 109428 Moscow, Russia

Abstract. This article presents the results of studies comparing the quality of work of two modern models of convolutional neural networks YOLOv7 and YOLOv8 used to monitor the yield of strawberries. To do this, we used the transfer method of machine learning models on a set of collected data consisting of four classes of development of generative formations of strawberry. As a result of the study, we obtained a data set that contained images of flowers, ovaries, mature and not mature berries. To ensure the balance of classes in the dataset, the Oversampling method was used, which included the generation of new images by applying various operations, such as resizing the image, normalizing brightness and contrast, converting images by rotating them by a certain angle and reflection, random noise addition, Gaussian blur. To collect data (images) in the field, a DJI Phantom 2 quadrocopter with a DJI Zenmuse Gimbal suspension and a GoPro HD HERO3 camera was used. To assess the quality of the YOLOv7 and YOLOv8 models when recognizing specified classes, well-known metrics were used that estimate the proportion of objects found that are really objects of a given class, such as Precision, Recall and mAP. Analysis of the results showed that the mAP metric for all classes of the YOLOv7 convolutional neural network model was 0,6, and the YOLOv8 model was 0,762. Analysis of the test sample images showed that the average absolute percentage error of image recognition of all classes by the YOLOv7 and YOLOv8 models was 9,2%. The most difficult to recognize was class the ovary of strawberries, the average absolute percentage error of which was 13,2%. In further studies, the use of high-resolution stereo cameras is recommended, which will further improve the accuracy of monitoring potential yields due to the possibility of determining the dimensional parameters of strawberry fruits and constructing 3D models of elevation maps using photogrammetry.

1 Introduction

Currently, strawberry (Fragaria x ananassa) remains one of the most popular crops in berry growing. It is characterized by high nutritional value, high yield.

The yield of garden strawberries can be predicted by the number of flowers, the number of ovaries and the number of unripe berries, for this purpose regular visual inspections of

---

[*] Corresponding author: alexeykutyrev@gmail.com

bushes and counting of generative parts of bushes. Depending on the age of the plantation, the results of the calculation are compared with the yield data for previous years. Such a calculation requires a lot of time and is too laborious, and the result may not always be accurate and has many assumptions [1,2]. Early monitoring of strawberry yield is necessary for proper planning of labor costs, it is necessary to accurately predict the future volume. It is proposed to improve the quality of collecting and counting generative components of garden dugout plants by quickly and accurately obtaining and processing digital data in automatic mode using machine learning algorithms such as convolutional neural networks (CNN) [3].

The introduction of digital monitoring of plantings based on the use of machine learning and computer vision technologies will optimize the process of cultivating strawberries, reduce losses and improve the quality of marketable products, and increase the efficiency of the use of material and labor resources [4,5].

The purpose of the research is to compare the quality of the work of YOLOv7 and YOLOv8 convolutional neural network models for monitoring the yield of strawberries, counting generative parts of plants in images using Transfer learning.

2 Materials and methods

The YOLOv7 convolutional neural network (You Only Look Once) was used to recognize the generative parts of strawberries in the images. The YOLOv7 model allows you to process images at up to 244 frames per second on the GPU (GPU) or up to 5 frames per second on the CPU (CPU). The YOLOv7 model used contains 227 layers, including 7 CSPDarknet53 blocks, 4 SP (Spatial Pyramid Pooling) blocks and 3 YOLO layers. The model provides high accuracy of object recognition through the use of machine learning techniques, such as training on a large amount of data (Big Data).

The YOLOv8 convolutional neural network was used to compare the accuracy of recognition of generative parts of strawberries in the images. The YOLOv8 model has an even higher image processing speed - up to 300 frames per second on the GPU and up to 7 frames per second on the CPU. YOLOv8 uses an architecture similar to YOLOv7, but with improved layers and learning algorithms. The YOLOv8 model contains 365 layers, including 8 blocks of CSPDarknet53, 5 blocks of SPP (Spatial Pyramid Pooling) and 3 layers of YOLO.The research uses the Transfer Learning method, which consists in adapting a pre-trained YOLOv7 model on the COCO (Common Objects in Context) dastaset for training on new data. This method allows us to use the knowledge of the model obtained in solving one problem to solve another narrower problem, which allows us to speed up the learning process for recognizing objects in a new area or with new characteristics and increase the accuracy of object recognition [5-9].

To collect a set of data (images), a DJI Phantom 2 quadrocopter with a DJI Zenmuse Gimbal suspension with a GoPro HD HERO3 Edition camera (CHDHX-301), a 12MP matrix resolution, a 4096x2160 Pixels video resolution (4K Ultra HD), a fixed camera focal length of 2.77 mm (wide-angle lens) was used. The strawberry variety used in research is Zenga-Zenga. The flight route of the quadcopter included movement over rows of strawberries in a shuttle way, at an altitude of no more than 2 meters. With the help of a quadrocopter, a data set of 2000 images was collected. The illumination during the shooting ranged from 60000 lux to 110000 lux.

To prepare a data set for training models of the YOLOv7 and YOLOv8 neural networks, the annotation of the obtained images using the Roboflow service was carried out. The annotation of the images was carried out by experts by selecting objects in a rectangular frame and classifying them (choosing the class to which the object in the frame belongs) (Fig. 1).
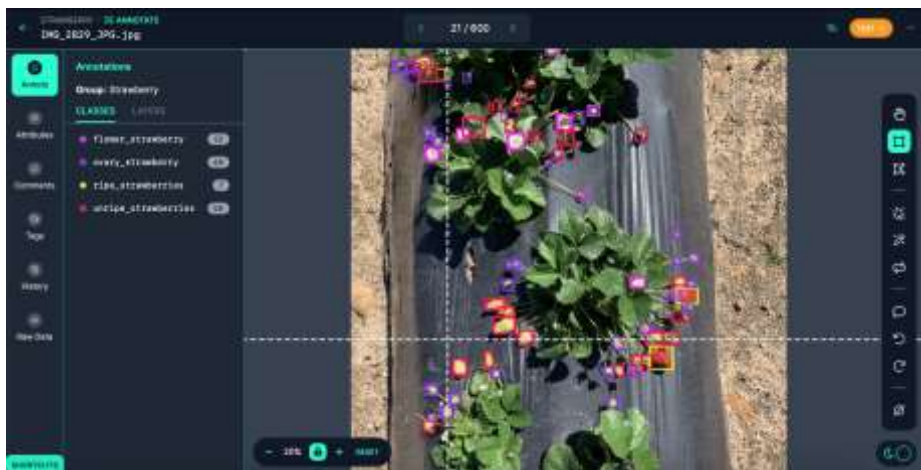
Fig. 1. Annotating images in the Roboflow service.

There are 4 classes of generative strawberry formations for neural network training, flowers (class «flower_strawberry»), ovary (class «ovary_strawberry»), immature berry (class «unripe_strawberry»), mature berry (class «ripe_strawberry»). Generative formations with white petals were assigned to the class of flowers, generative formations with a rudiment of a fetus no larger than 1 cm in length were assigned to the class of flowers, not ripe fruit with a length of more than one centimeter of green, white and white-pink flowers were assigned to the class of ovary, fruits completely colored in pink or red colors. The JSON (JavaScript Object Notation) format is used to store markup data on images of strawberry plants.

To ensure the balance of classes in the data set used, the Oversampling method (artificial increase in the sample) was applied by creating new examples based on existing data. The online Roboflow service of synthetic data generation (synthetic data generation) was used. Data augmentation included such operations as horizontal and vertical reflection (flip: horizontal, vertical), rotation by an angle selected between -15° and +15° (rotation: between -15° and +15°), random addition of noise, introduction of changes to image pixels in the amount of up to 5% (noise: up to 5% of pixels) and adding Gaussian blur (blur:10 px) [10,11].

To visualize the annotation, evaluate the distribution of classes and determine the relationships between them in the data set used, Annotations hitmap is built (Fig.2).
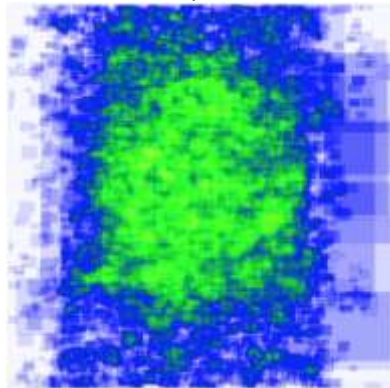

Fig. 2. Annotations hitmap of all specified classes.

The process of argumentation of images allowed to increase the sample size to 6000 thousand images. As a result of the conducted research , the set of data obtained was divided into a training set – 4200 pcs. images (70%), validation – 1200 pcs. images (20%) and test – 600 pcs. images (10%). To train the YOLOv7 and YOLOv8 models on the created data sample, 250 epochs were used, each of which consisted of several iterations (batch). The size of iterations (batch size) for each epoch was 16. The number of epochs was determined experimentally, taking into account the size of the data sample, the complexity of the model used, the architecture of the model and other parameters. Updating of model parameters (including weight coefficients) after each epoch was performed automatically in the process of training the model using an optimization algorithm (stochastic gradient descent, SGD), which adjusts the weights of the model based on the gradients of the loss function (model error) and applies them to update the weights.

To assess the quality of the YOLOv7 and YOLOv8 models when recognizing objects, the well-known metrics Precision (accuracy), Recall (completeness) and AP (average accuracy) were used, indicators were evaluated when recognizing both individual classes (binary classification, Binary Classification) and for all classes on average (multiclass classification, Multi-Class Classification) (Table.1) [12-14].

Table 1. Metrics used to analyze the quality of the YOLOv7 and YOLOv8 models when recognizing specified classes.

Metric	Formula
When recognizing individual classes	
Precision	$\dfrac{tp_i}{tp_i + fp_i}$
Recall	$\dfrac{tp_i}{tp_i + fn_i}$
AP	$\displaystyle\int_{r=0}^{1} Precision_i(r)dr$
When recognizing all the specified classes	
Average recall	$\dfrac{\Sigma_{i=1}^{n} Precision_i}{n}$
Average precision	$\dfrac{\Sigma_{i=1}^{n} Recall_i}{n}$
mAP (mean average precision)	$\dfrac{(AP_1 + AP_2 + \cdots + AP_n)}{n}$

The Precision metric shows the proportion of correctly recognized class i objects relative to all objects that the model assigned to this class. The higher the accuracy, the fewer false positive results, i.e. objects that the model incorrectly assigned to class i. Recall shows what proportion of objects of this class out of the total number of objects the algorithm was able to recognize. The average precision metric is used to evaluate the performance of the object classification and recognition algorithm. It is determined by finding the area under the Precision-Recall curve, which is based on the results of the algorithm. To analyze the average AP for all classes, the metric mAP (mean average precision) was used, which allows you to evaluate the overall quality of the algorithm for all classes of objects.

To train the neural network and conduct research, a computer system equipped with an Intel Core i9-10900X processor with 10 cores and 20 virtual threads was used. To speed up the learning process, two NVIDIA GeForce RTX 2080 Ti graphics cards were used, capable of processing a large amount of data in parallel. The GIGABYTE X299 UD4 Pro motherboard was used to ensure high system performance. An Intel PCI-E 1Tb 660P SSD drive was used to store data and speed up the data loading process. The amount of RAM of the system was 32GB, Kingston DDR4 DIMM modules were used.

3 Results and discussion

As a result of the conducted research, the YOLOv7 and YOLOv8 models were trained. The results of recognition of the specified classes in the images are shown in Figure 3.

Fig. 3. Recognition results of the specified classes in images model YOLOv7 and model YOLOv8.

In Figure 4, the Precision-Recall graphs show the obtained dependencies of accuracy and completeness when changing the threshold (a measure of similarity between an object and classes). The constructed curves allow us to evaluate the quality of image classification using the YOLOv7 and YOLOv8 models, depending on the selected threshold.
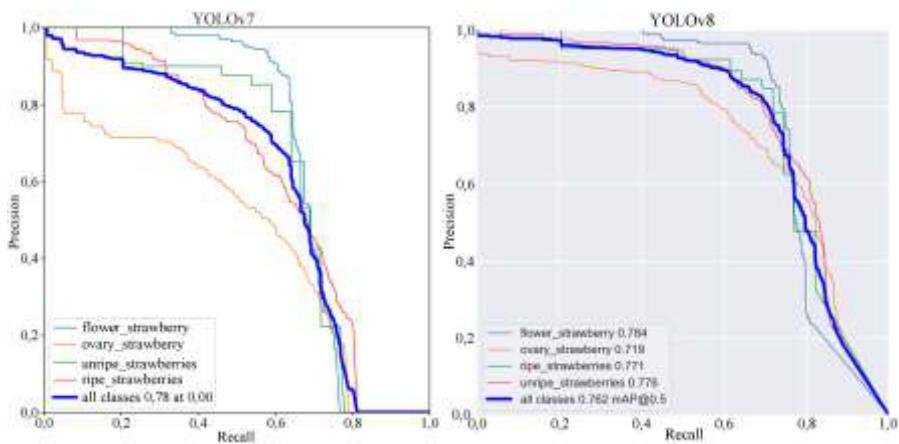

Fig. 4. Precision-Recall Curves model YOLOv7 and model YOLOv8.

Analysis of the Precision-Recall graph allowed us to determine the optimal classification threshold for the YOLOv7 – 0,58 model and the YOLOv8 – 0,62 model, which provides the best ratio between Precision and Recall for the specified classes. Precision-Confidence and

Recall-Confidence curves were constructed to assess the reliability of model predictions and their compliance with the actual values of classes (Fig.5). Precision-Confidence curves allowed us to estimate the accuracy of the model's predictions depending on the confidence with which it makes its predictions, whereas Recall-Confidence curves allowed us to estimate the completeness of the model's predictions for a positive class at different confidence values.
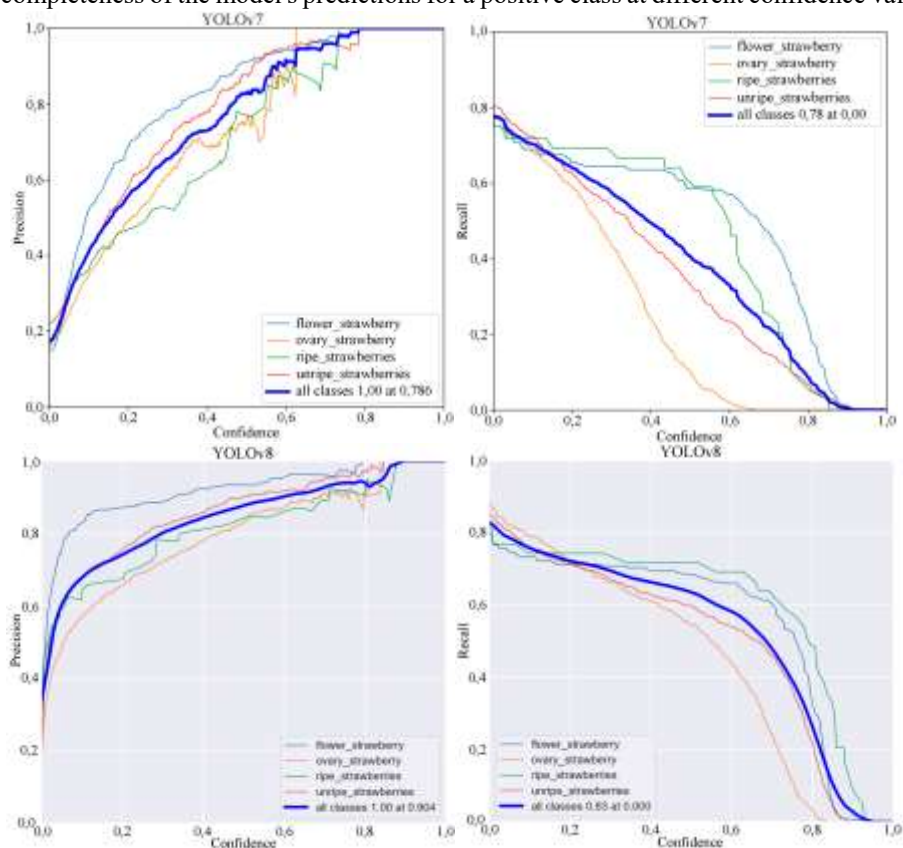

Fig. 5. Curves for evaluating the confidence of model predictions and determining their correspondence to the true values of the classes.

The results of calculating metrics for recognizing individual classes and for all classes on average are presented in Table 2.

Table 2. Results of calculations of metrics of binary and multiclass classification of models YOLOv7 и YOLOv8.

Classifier	Precision	Recall	mAP@0,5
Model YOLOv7			
Strawberry flowers / «flower_strawberry»	0,757	0,646	0,679
Strawberry ovary / «ovary_strawberry»	0,574	0,492	0,47
Unripe strawberry berry / «unripe_strawberry»	0,669	0,564	0,612
Mature strawberry berry / «ripe_strawberry»	0,523	0,692	0,64
Average (all classes)	0,631	0,598	0,6

Continuation of Table 2.

Classifier	Precision	Recall	mAP@0,5
Model YOLOv8			
Strawberry flowers / «flower_strawberry»	0,815	0,709	0,762
Strawberry ovary / «ovary_strawberry»	0,741	0,642	0,719
Unripe strawberry berry / «unripe_strawberry»	0,83	0,656	0,776
Mature strawberry berry / «ripe_strawberry»	0,778	0,72	0,771
Average (all classes)	0,815	0,682	0,762

Analysis of the results showed that the mAP metric for all classes of the YOLOv7 convolutional neural network model was 0,6, and the YOLOv8 model was 0,762. Analysis of the test sample images showed that the average absolute percentage error of image recognition of all classes by the YOLOv7 and YOLOv8 models was 9,2%. The most difficult to recognize was the "ovary_strawberry" class, the ovary of strawberries, the average absolute percentage error of which was 13,2%. The analysis of the conducted field studies showed that with a decrease in the number of flowers in the rows of garden plantings, there is an increase in false-negative model recognition results. This is explained by the fact that with a decrease in the number of flowers, there is an increase in the number of ovaries, immature and mature fruits, as well as the number and size of leaves, which in turn leads to an increase in the overlap zones of objects, as a result of which it is difficult to recognize them and leads to errors in counting.

4 Conclusions

Analysis of the research results showed that monitoring the yield of strawberries on an industrial plantation carried out using a DJI Phantom 2 quadrocopter with a DJI Zenmuse Gimbal suspension with a GoPro HD HERO3 camera using convolutional neural network models YOLOv7 and YOLOv8 for processing the data obtained will allow quantitative accounting of flowers, ovaries, immature and mature berries with an average accuracy of at least 90% at a flight altitude of no more than 2 meters. A comparative analysis of the quality of the models for monitoring strawberry yields showed that the most accurate and productive model is YOLOv7, the mAP metric of which was 0,762.

Analysis of the obtained graphs, binary and multiclass classification metrics for evaluating the quality of the trained neural network models allowed us to set optimal settings, select the confidence threshold at which the model shows optimal accuracy and completeness, balanced with the number of recognized objects. The configurations (hyperparameters) of the machine learning algorithm of the model for recognizing the specified classes of strawberries are determined.: the learning rate is 0.01 LR (learning rate), the number of epochs is 250, the size of the mini-package (batch size) is 16.

In further studies, the use of high-resolution stereo cameras is recommended, which will further improve the accuracy of monitoring potential yields due to the possibility of determining the dimensional parameters of strawberry fruits and constructing 3D models of elevation maps using photogrammetry. Regular digital monitoring using machine learning methods and digital cameras will automatically generate tasks for ground vehicles or unmanned aerial vehicles, make optimal management decisions in real time when cultivating strawberries. The use of data obtained in the field for each component of plant productivity and the coefficients of the ratios between them will allow to determine the potential and

actual yield, as well as to manage this yield, guided by the biological productivity of various varieties of strawberries.

References

1. P. Taylor, G. Savini, D. Neri, F. Zucconi, N. Sugiyama, Int. J. Fruit Sci. **5**, 29-50 (2008)

2. Y. Chen, W. S. Lee, H. Gan, N. Peres, C. Fraisse, Y. Zhang, Y. He, Remote Sens. **11**, 1584 (2019)

3. D. O. Khort, A. I. Kutyrev, R. A. Filippov, R. V. Vershinin, E3S Web of Conferences **193**, 01045 (2020)

4. A. Kutyrev, N. Kiktev, O. Kalivoshko, R. Rakhmedov, CEUR Workshop Proceedings **3347**, 90-101 (2022)

5. D. O. Khort, A. I. Kutyrev, R. A. Filippov, S. V. Semichev, E3S Web of Conferences **262**, 01024 (2021)

6. E. Andrejko, S. Valley, Comput. Electron. Agric. **147**, 70-90 (2018)

7. I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, C. McCool Sensors **16**, 1222 (2016)

8. H. Zhou, J. Ou, P. Meng, J. Ye H. Tong, Z. Li, Horticulturae **9**, 400 (2023)

9. C. Zheng, A. Abd-Elrahman, V. Whitaker, Remote Sens **13**, 531 (2021)

10. D. Khort, A. Kutyrev A., N. Kiktev, T. Hutsol, S. Glowacki, M. Kuboń, T. Nurek, A. Rud, Z. Gródek-Szostak, Sensors **22(9)**, 3164 (2022)

11. H. Zheng, T. Cheng, M. Zhou, D. Li, X. Yao, Y. Tian, W. Cao, Y Zhu, Precis. Agric. **20**, 611-629 (2019)

12. A. E. Maxwell, T. A. Warner, L. A. Guillén, Remote Sensing **13(13)**, 2450 (2021)

13. C. Xie, C. Yang, Comput. Electron. Agric. **178**, 105731 (2020)

14. Z. Guan, A. Abd-Elrahman, Z. Fan, V. M. Whitaker, B. Wilkinson, Remote Sens. **163**, 171–186 (2020)