Neural network for identifying apple fruits on the crown of a tree

Igor Smirnov¹, Alexey Kuttyrev^{1,*}, and Nikolay Kiktev^{2,3}

¹Federal Scientific Agroengineering Center VIM, 1-st Institutsky proezd, 5, Moscow, 109428, Russia

²National University of Life and Environmental Sciences of Ukraine, Heroev Oborony str., 15, Kyiv, 03041, Ukraine

³Taras Shevchenko National University of Kyiv, Volodymyrs'ka str., 64/13, Kyiv, 01601, Ukraine

Abstract. To identify the fruits on the crowns of trees and count their number, a software and hardware complex (PAC) based on a technical vision system and a recurrent neural network of deep learning has been developed. The created neural network and class allocation algorithms allow the software and hardware complex to function stably in the conditions of industrial horticultural plantings, regardless of the size and interference of foliage, to determine the color of the fruit surface, to identify the presence of diseases and defects of fruits. The developed package provides digital monitoring of both photographic materials and video streams in online mode.

1 Introduction

There are many methods for identifying both weeds and cultivated plants. Technical vision systems, neural networks, and various spectrometry methods are used in plant recognition [1,2]. For this purpose, the morphological characteristics of plants obtained during their linear measurements are used. In the production of agricultural products in the field, the work of recognition systems that exist today is complicated by the variety of climatic conditions and light conditions. During the growing season, a large variety of the appearance of the same plant variety can be observed within the field, which also significantly complicates the process of recognizing them. Studies of scientists from various countries in the field of recognition of plants and fruits found that determining the coordinates of fruits only by color tone, using computer vision systems, is insufficient [3,4,5]. This is due to the large color palette of the fruit and the high light sensitivity of the matrix (ISO) cameras. You can get more information you need with less effort by using neural networks. Automated accounting and control of the number of fruits on the crown of the tree is one of the most important unsolved problems, which will significantly increase the efficiency of horticulture production.

In an article by Belgian scientists [1], a multispectral vision system was developed that includes four wavelength ranges in the visible / near-IR range. Multispectral images of healthy and defective fruits were obtained, allowing to cover the entire color variability of

* Corresponding author: alexeykuttyrev@gmail.com

this two-color apple variety. In the work of Ukrainian researchers [6], the technology of image recognition using standard and stress indices is described, applied to determine the diseases of rapeseed leaves when shooting with a UAV in the optical range using the RGB camera and the Slanrange complex. The affected leaves have an abnormal color, in the optical range the most informative are the red and green channels. For image recognition, it is also possible to use the method of clustering objects of fuzzy sets using evolutionary technologies [7].

The use of a neural network of back propagation and machine vision for automatic sorting of apples is described in an article by Indian researchers [8]. The neural network of back propagation trained by the authors makes it possible to classify an apple. Two sets of variables are used for training: the first set is an independent variable that is a parameter of the quality of the apple at the surface level, the second set is a dependent variable that represents the quality of the apple.

The article [9] presents a method for classifying Indian vegetables (leafy and wild) based on machine vision and a multi-layer neural network, which classifies vegetables based on their physical parameters - color and texture features. The feature vector consists of the combined elements of the RGB color and the GLCM texture, which contribute to the classification. Achieved an average classification accuracy of 95.4%.

Austrian researchers have proposed a way to classify hyperspectral images using modern convolutional neural networks pre-trained for RGB image data [2], and applied it to classify fruits and vegetables.

The authors of this article conducted and published research on the creation of robotic devices for collecting strawberries [10,11], on methods for recognizing the coordinates of berries [12], on 3D modeling of fruit capture devices, design of sensors and actuators for an intelligent robotic apple collection system [13].

2 Materials and methods

Based on the results of research by well-known scientists on the use of neural networks in agriculture [2, 8], a recurrent neural network of deep learning was chosen for the optimal speed of recognition of apple fruits on the tree crown [14]. By the type of training-with a teacher, by the type of setting – dynamic, by the type of input information – analog, by the type of problem to be solved – classifying. The principle of operation of such a neural network is to divide (segment) the analyzed photo into classes and select specific objects (disease, apple, branch, etc.). The chosen neural network is by design one of the best models available for solving most "perception problems" (such as image classification).

To implement the learning process of the developed neural network, the Python programming language and the Spyder development environment were chosen, and the PyTorch framework was used. The architecture of deep convolutional neural networks – MASK-RCNN-was chosen, which made it possible to find an apple on the frame, select it pixel-by-pixel, and with a probability of more than 0.9 determine whether it belongs to one of the classes. The classes of apples were selected:

- healthy apples – red, green;
- apples with diseases – rotten, bitter pitiness, hail, mechanical, scab, sawfly, fruit moth.

As a result, 2 classes of healthy apples and 7 classes of apples with diseases were identified. To train the model, the TensorFlow Object Detection API machine learning libraries, GPU computing libraries, and image and graph libraries are used. To collect photos during training, several Nikon D3500 AF-S 18-140 VR cameras were used, the Nikon Nikkor AF-P DX F 18-55 mm lens, at a distance of 0.2 m, 0.5 m and 1.0 meter, from angles that overlap each other. More than 3000 photos of the specified classes of apples were taken.

As a result, 2 classes of healthy apples and 7 classes of apples with diseases were identified. To train the model, the TensorFlow Object Detection API machine learning

libraries, GPU computing libraries, and image and graph libraries are used. To collect photos during training, several Nikon D3500 AF-S 18-140 VR cameras were used, the Nikon Nikkor AF-P DX F 18-55 mm lens, at a distance of 0.2 m, 0.5 m and 1.0 meter, from angles that overlap each other. More than 3000 photos of the specified classes of apples were taken.

At the first stage of the analysis, the neural network checks the image for the presence of an object and selects it in a frame. To do this, we use the efficient YOLO (You only look once) algorithm, which allows you to select objects in the image.

In the second stage of the analysis, the neural network determines the exact boundaries of the object. Algorithms for step-by-step reduction of image quality are used to search for known dependencies (distinctive features or patterns of the desired object in the image). The image is convoluted step by step from layer to layer by mixing neighboring pixels, depending on the task to the size of 2x1 pixels. To search for objects and their distinctive features, a neural network is trained using a prepared data set of the desired object. To prepare the sample for training, in the first approach, it was decided to divide the apples into 2 classes: the apple and the background, and make a markup of the photos. The open source program VGG ImageAnnotator (Fig. 1) was chosen as the markup.

Fig. 1. The process of marking up data and highlighting image classes in VGG Image Annotator.

The program allows you to get the outline of the selected apples and classify them as apples, for use in the learning process, while saving the selection results in a JSON file. The script works as follows. The files of the trained deep convolutional neural network are moved to a separate folder, then the "SegmentedImages" folder is created and files with the name IMG_0001. png are moved to it, after which the script is run. After running the script, a deep convolutional neural network model is run, a photo named IMG_0001.png is uploaded, and sent to the network. After processing the image, the neural network creates masks that most likely contain apples, saves these masks, and applies a common mask to the uploaded image IMG_0001. png, then saves it in the same folder. All coordinates and labels of the created frames for each photo are saved in a separate XML file. After processing and marking up the photos, a dataset is prepared with the division of all photos and label files into a training and test sample. The captured data set is grouped evenly by class. Each of the masks is a picture format .PNG, which has the dimension of a photo with markup (Fig. 2).

Fig. 2. Converting JSON markup to masks.

The results of the developed deep learning neural network in the laboratory are presented in Figure 3.

Fig. 3. Collecting a data set of apple images and processing the data obtained: a - shooting distance of 0.2 m, b - shooting distance of 0.5 m, c - shooting distance of 1 m.

A Basler ace 1920-155uc camera with a GigE interface and a Sony IMX174 CMOS sensor with a frequency of 164 frames per second was used for field research. The camera matrix has a resolution of up to 1920 x 1200 pixels, a resolution of 2.3 megapixels. To measure the illumination, a Radex Lupin luxmeter (Quarta Rad, Russia) was used, with a relative measurement error of 10%.

3 Results and discussion

As a result of the conducted research, a software and hardware complex was developed, including a computer vision system with a neural network for recognizing apple fruits. To configure and verify the calculated parameters of the PAC, an analysis of its operation on an industrial plantation of an apple orchard was carried out. To do this, a Selecline tripod with

a Basler camera is installed in a row of apple trees of the Northern Sinap variety at a distance of 0.3 m to 2 m from the tree (Fig. 4).

Fig. 4. Conducting a field experiment on an industrial apple plantation.

To avoid size estimation errors caused by partially hidden apples, the study considered only fully visible apples (at least on one side of the canopy). The apples used in the experiment were marked on the trees using the ZED Stereo 3D camera. This is done to ensure that the same apples were used to compare the number of fruits determined by the recognition system and the number measured manually (Fig. 5).

Fig. 5. Model of an apple tree of the Northern Synap variety used in the experiment (150 fruits excluding partially hidden ones).

The results of the experiment are presented in Table 1.

Table 1. The obtained results of a field experiment developed by PAC.

Repeatability of measurements	The identified number of apple fruits on the crown of the tree when using the PACK, pieces	The actual number of apple fruits on the crown of the tree, pieces	Average absolute percentage measurement error, %
1	136	150	10,3
2	134		11,9
3	136		10,3
4	133		12,8
5	131		14,5
Maximum, pcs	136		10,3
Minimum, pcs	131		14,5
Standard deviation, pcs	2,25	0	1,9
Average value, pcs	134	150	11,9

It was found that the accuracy of estimating the number of apples on the tree crown compared to the true value measured manually was at least 89.3%. With the help of the PAC, under changing climatic conditions, an average of 134 apples on the crown of the tree were determined, with their true value of 150 pieces. The average absolute percentage error was 11.9% with a five-fold repetition of the measurements.

The main errors in estimating the number of apples using the PAC are related to the segmentation of low-resolution images. This is due to the detection of only partial areas of apples, or the erroneous perception of the environment and background as areas of apples, which led to inaccuracies in identifying apples in the images. In addition, inaccuracies in the estimation of the number of apples are due to the low resolution of the images. These errors in estimating the number of apples per tree crown can potentially be reduced by increasing the camera resolution to 3840×2160 pixels.

4 Conclusions

As a result of the research, a technical vision system with a neural network was developed to estimate the number of apples on the crowns of trees. As a result of the field experiment, it was found that the errors of the developed PAC in estimating the number of fruits were mainly caused by inaccurate image segmentation, as well as the low resolution of the camera used. It is revealed that the convolutional recurrent deep learning network is the most suitable neural network for the tasks of identifying apple fruits, since its use makes it possible to recognize the contour of the fruit with high accuracy in conditions of changing climatic parameters. This is in demand when implementing digital technologies in the field.

The developed software and hardware complex based on the created neural network will allow for digital monitoring of both photographic materials and video streams in online mode. Using the created neural network and class allocation algorithms, the developed PAC will be able to:

- function sustainably in industrial plantings, regardless of the size and interference of foliage;
- determine the color of the surface of the fruit and the size of the fruit;
- identify the presence of diseases and defects of the fruit with a probability of at least 99%.

This is possible as a result of the incremental expansion of the dataset during the operation of the complex and the gradual evolution of the solution by training the network in the process of working on new data. The developed neural network will expand the functionality of the

PAC not only for monitoring the yield of fruit crops, but also for robotic fruit harvesting. When harvesting apples, it is carried out:

- determining the coordinates of each fruit or part of it;
- returning the coordinates of the center of the fetus and its contour to the controller of the manipulator device;
- indication of the areas of defects or diseases on the fruit.

References

1. O. Kleynen, V. Leemans, M.F. Destain, Journal of Food Engineering, **69**, 41 (2005)
2. J. Steinbrener, K. Posch, R. Leitner, Computers and Electronics in Agriculture, **162**, 364 (2019)
3. N. Kumar, P.N. Belhumeur, A. Biswas, D.W. Jacobs, W.J. Kress, I.C. Lopez, J.V.B. Soares, Part of the Lecture Notes in Computer Science book series, **7573**, 516 (2012)
4. S. Kaur, P. Kaur, Journal of Multimedia Information System, **6(2)**, 49 (2019)
5. R. Mia, J. Mia, A. Majumder, S. Supriya, T. Habib, International Journal of Engineering and Advanced Technology, **9(1)**, 2810 (2019)
6. M. Dolia, V. Lysenko, N. Pasichnyk, O. Opryshko, D. Komarchuk, V. Miroshnyk, T. Lendiel, Journal International Conference on Advanced Information and Communications Technologies 469 (2019)
7. H. Hnatiienko, O. Suprun, CEUR Workshop Proceedings, **2318**, 129 (2018)
8. A.K. Bhatt, D. Pant, AI & Soc, **30**, 45 (2015)
9. M. Madgi, A. Danti, International Journal of Computer Engineering and Applications, **12(3)**, 17 (2018)
10. D.O. Khort, A.I. Kutyrev, R.A. Filippov, N.A. Kiktev, D.S. Komarchuk, Science and Technology PIC S&T, 869 (2019)
11. D.O. Khort, A.I. Kutyrev, R.A. Filippov, R.V. Vershinin, E3S Web of Conferences, **01045** (2020)
12. D.O. Khort, A.I. Kutyrev, I.G. Smirnov, V.V. Osypenko, N.A. Kiktev, Communications in Computer and Information Science, **1158**, (2020)
13. D.O. Khort, A.I. Kutyrev, D.S. Pupin, N.A. Kiktev, Information technology and security, **20**, 154, (2020)
14. S.A. Marhon, C.J.F. Cameron, S.C. Kremer, Handbook on Neural Information Processing, **49**, 29 (2013)