Intelligent vineyard monitoring using YOLOv7

Pavel Kuznetsov[1,2], *Dmitry* Voronin[1,2,*], and *Dmitriy* Kotelnikov[1,2]

[1]Federal State Budget Scientific Institution "All-Russian National Research Institute of Viticulture and Winemaking "Magarach", 31 Kirova str., Yalta, Russia
[2]Sevastopol State University, 33 Universitetskaya str., Sevastopol, Russia

Abstract. The article discusses the technology for automated neural network monitoring of the vineyard's physiological condition. The proposed solution is based on the integrated use of convolutional neural network method and machine vision technologies. The training of the YOLOv7 neural network was implemented in the Python environment using the PyTorch framework and the OpenCV computer vision library. The dataset consisting of 6320 images of grape leaves (including healthy and diseased ones) has been used for neural network training. The obtained results showed that the detection accuracy is at least 91%. Visualization of monitoring results has been carried out using heatmap, allowing to obtain information about vineyard physiological condition in dynamics. The proposed mathematical model allows to calculate the monitored vineyard's area made by one complex per day.

1 Introduction

Agriculture is one of the most important sectors of the economy, which reduce dependence on other countries, provides food and resources for the industrial sector. Thus, more and more attention is paid to agricultural development [1, 3, 5]. The introduction of digital technologies in the agricultural industry allows not only to automate and systematize technologies for monitoring, collecting and producing agricultural products, but also to minimize and optimize human labor, which will ultimately have a comprehensive impact on the entire industry [1, 3, 5]. In particular, these technologies make it possible to automate the processes of collecting and analyzing information on growing grapes, as well as controlling and optimizing the processes of cultivating and caring for grapevines through the implementation of effective monitoring technologies [1, 3, 4-7, 9, 10]. The development and reduction in cost of unmanned aerial vehicle technologies have made it possible to increase effectiveness of local monitoring and control for agricultural cultivation purposes [1, 3, 5]. Solving the above-described problems requires the creation of a scientific and methodological foundation for the implementation of intelligent adaptive automated monitoring when solving problems in the fields of viticulture and winemaking. Timely detection of grapes damage signs made by diseases is a very actively developing area of scientific research [1, 4, 6, 7, 9, 10]. The proposed solution is based on the integrated use of convolutional neural network method and machine vision technologies.

* Corresponding author: dima@voronins.com

2 Materials and methods

The primary goal of the present study is to develop the method, technique and algorithm for vineyard-scale detecting visible symptomatic reactions on grape leaves using convolutional neural network method. We suggest to use technique based on application of an automated device for monitoring vineyards enhanced by neural network method for detecting objects. Comprehensive analysis of implementation of a high-performance computing module for neural network processing of video data showed that the most promising solutions are the following:

1. Allocation of high-performance computing module directly on board of the unmanned aerial vehicle. This variant allows processing the video stream from the unmanned aerial vehicle's camera directly during the flight [2, 8]. However, this variant negatively affects the weight and size characteristics and power consumption of the unmanned aerial vehicle. In addition, this variant limits the flight speed due to the low processing speed of video stream frames. It is possible to increase the speed of flight and processing of the video stream by implementing a high-performance computing device based on programmable logic device, but this solution is rather expensive.
2. Performing neural network data analysis on a stationary high-performance computing device. In this variant the unmanned aerial vehicle is used only for collecting video data and does not directly participate in data processing. This variant will significantly reduce the cost and speed up data collection. This is due to the fact that the unmanned aerial vehicle will not carry additional load. We carried out a variant analysis, as a result of which the most popular neural network models were analyzed for the speed and precision of diseased grape leaves detection.
3. To assess the quality of the selected neural network models and compare different algorithms, the following metrics (quality criteria) were used: Complete Intersection over Union (CIoU), Precision, Recall (completeness), Average Precision, Mean Average Precision (average AP), F1 score. The analysis of the variants described above showed that the second variant is optimal for automated vineyard-scale monitoring of physiological condition.

In accordance with the technology, the unmanned aerial vehicle must fly over each row of the vineyard at least three times in order to capture grape leaves from three different views: right side, left side, top. The video footage is transferred to a high-performance computing device with an installed program for automatic classification of diseased grape leaves. The program automatically counts a number of diseased grape leaves in the frame. In accordance with the obtained values, visualization is performed in the form of a heat map, which is presented as a separate layer in GIS. On the map the color of each point corresponds to the number of diseased grape leaves counted by the neural network and is linked to the coordinates where the unmanned aerial vehicle was shooting the corresponding frame. This procedure is necessary for synchronizing the video file and the log of the GPS tracker installed on the unmanned aerial vehicle.

To assess the vineyard-scale physiological condition, it is not enough to implement the Object Detection procedure. This is due to the fact that the detection of the same leaves will be carried out on several frames of the video sequence and will depend on external environmental factors and the unmanned aerial vehicle's flight mode. Thus, the count of diseased leaves will not be correct. To solve this problem and, consequently, to improve the quality of monitoring, an object tracking method should be additionally applied (Object Tracking). The integrated use of Object Detection and Object Tracking technologies allows to ignore already detected objects, which significantly reduces the number of repeated and false positives. The Object Tracking technology is based on SORT (Simple Online and

Realtime Tracking) or Deep SORT algorithms, which are used to track detected objects. As part of the study, the Deep SORT algorithm is used, since it allows you to identify previously detected objects even after they have been lost from the frame for a long time. This feature of the Deep SORT algorithm is achieved through the use of two mathematical methods - the Mahalanobis distance and the Kalman filter. The Mahalanobis distance is used to determine the similarities between known and unknown weights of objects detected by the neural network. Kalman filter is often used to eliminate noise and emissions in previously defined weighting factors.

2.1 Dataset creation

In order to solve the problem, it is advisable to use photos of grape leaves as initial data when forming a dataset. The labeled dataset allows to train neural network in order to recognize and classify objects of interest within images with prefixed values. In light of the automated monitoring, it is recommended to use unmanned aerial vehicle video recording, it is also advisable to use storyboarded video materials of flying around the vineyard rows as a data set for training neural network. At the same time, practice of video filming has shown that when creating a dataset, it is necessary to take into account features associated with the actual operation of the unmanned aerial vehicle:

1. When flying in rows, it is necessary that video recording of grape plants should be carried out by an unmanned aerial vehicle camera at a distance of one to two meters at a camera installation angle of 90° to 105° in the horizontal plane.
2. When flying an unmanned aerial vehicle directly over a row, video recording of grape plants should be carried out at a height of no more than three meters at an angle of 90° to 100° in the vertical plane.
3. When recording video, the automatic exposure function must be turned off in the unmanned aerial vehicle's camera. This procedure is necessary to preserve the details in the light and dark areas of the image under different lighting conditions. Video recording of grape plants should be carried out on a clear day with a wind speed of no more than 4 m/s.

Training neural network on the generated dataset requires its preliminary preparation, called markup, or image annotation. This process allows to attach metadata to each dataset image that carries information about the properties of objects (class names, object location on the image, etc.). The main complexity of this procedure is inevitable manual marking of all objects in the images.

An expert needs to highlight the objects of interest on image. The correctness of object recognition by the neural network will significantly depend on the quality of the annotation. In view of this, it is necessary to fully select all objects of interest on the photo. If necessary, objects are periodically omitted or incorrectly selected, the neural network will not be able to identify all the patterns required or will identify them incorrectly. During processing, the neural network will independently find patterns in the intensity of pixel color channels, their alternation, etc. Labeling (annotation) of images of grape leaves was carried out using the LabelImg tool (https://github.com/tzutalin/labelImg).

The YOLOv7 neural network was trained in the Python environment using the PyTorch 1.13.1 framework and the OpenCV 4.7.0.72 computer vision library. The following parameters were used to train the neural network model: number of epochs was 150, batch-size was 7, input image size 640×640, optimizer - stochastic gradient descent (SGD). Training was performed using the CUDA 11.6 hardware-software architecture of parallel computing and cuDNN 8.9.1 library for neural network training.

2.2 Test Bench

To test the technology in the vineyard, a DJI Phantom 4 quadcopter unmanned aerial vehicle was used. This unmanned aerial vehicle has an average flight duration of 30 minutes. It is equipped with a video camera mounted on a gyrostabilized suspension. Camera specifications: sensor – 1/2.3'' CMOS, 12.4 × 106 effective pixels; lens - FOV 94° 20 mm (35 mm format equivalent) f/2.8. In the experiment, the video recording mode (FHD 1920 × 1080, 24 fps) was used. The test bench was equipped with a high-performance computing device based on the NVIDIA GeForce RTX2080 GPU, Intel Core i5-8400 CPU, 16Gb RAM has been used. To implement an interactive map, a developed GPS tracker was attached to the unmanned aerial vehicle. The block diagram of the GPS tracker is shown in Figure 1. ESP8266 is used as the main microcontroller. To obtain geospatial information, the measuring module has a GPS receiver based on the NEO-6M-0-001 chip based on the Ublox NEO-6M STM chip. This module is a stand-alone GPS device with a high-performance u-blox 6 positioning processor. To communicate with the microcontroller, a UART (TTL) interface is used with a supported baud rate from 4800 to 230400 baud, 9600 baud by default. The log with geotags is recorded on a microSD memory card. For this, a specialized microSD card module is used, which is connected to the microcontroller via the SPI interface.

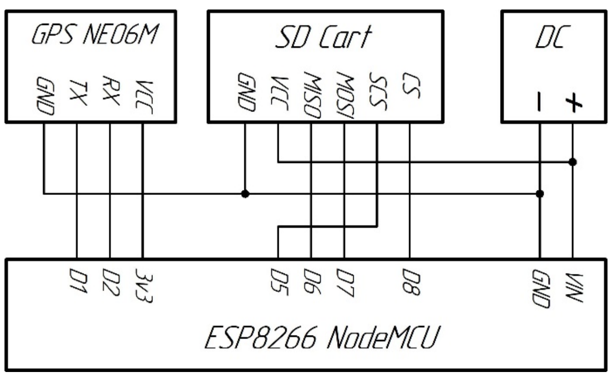

Fig. 1. Structural diagram of a GPS tracker.

Testing of the technology was carried out on the vineyard of the JSC "Agrofirma Chernomorets", which is situated in Crimea Republic, Bakhchisaraysky district, Uglovoye village (Figure 2). Area of the plot is 72 ha, cultivar is black Pinot, planting date is 2007. Planting scheme is 3x3 (0.3) m, formation is one-sided cord, high stem, free-hanging shoot, rootstock is Berlandier x Riparia Cabernet 5BB. Non-covered, drip irrigation system. Soil types in plots are ordinary black earth mycelar-carbonate foothills. The humus content of the upper layers is 2.9–3.6%. Total nitrogen content ranges from 0.21% to 0.3%. Hydrolyzable nitrogen content is 5 - 11mg/100gr, which indicates the high availability of mobile nitrogen. The phosphorus content ranges between 0.07% and 0.16%. Mobile phosphorus content ranges from 0.5 to 6 mg/100 grams. The total potassium content in carbonate-rich chernozem ranges from 1.1% to 2.6%, and the mobile content ranges from 16 to 43 mg/100 grams. The absorption capacity in the upper horizons equals 32–39 mg-eq. The profile of micellar-carbonate chernozems was leached from water-soluble salts to a depth of 150–200 cm and more. Salinization at these depths is sulfate-calcium.

Fig. 2. Testing of the technology on the vineyard of the JSC "Agrofirma Chernomorets".

3 Results and discussion

The study demonstrates how UAVs enhanced with convolutional neural network can be used to monitor the physiological condition of vineyards (Figure 3).

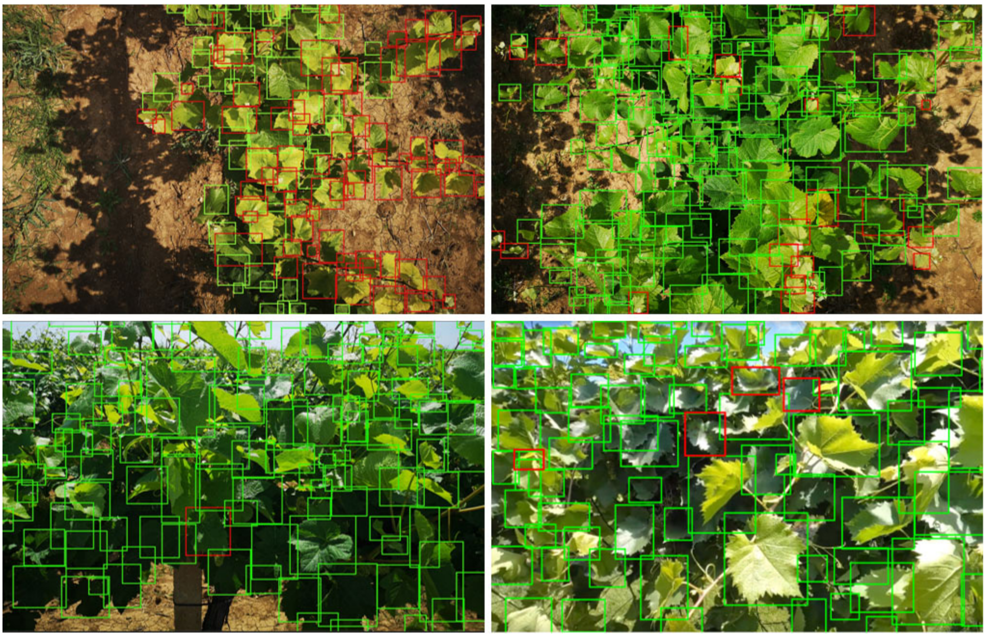

Fig. 3. Results of detection of diseased grape leaves.

In this experiment of training diseased grape leaves using the YOLOv7, the mAP_curve, PR_curve, P_curve, R_curve and etc. metrics were used during the model training process. It is noticeable that the metrics describing the average value of mAP go into saturation. It indicates that the neural network has successfully trained on the prepared dataset. The trained and tested neural network was already capable for detecting typical diseases of the vineyard however, for greater clarity and ease of interpretation of the results, we decided to create an interactive map of the vineyard.

An interactive map allows one to display the detection results in the form of points on the map (geotags) with a photo and a number of detected disease foci, which may help the vineyard staff to locate the problem area. Also, displaying a photo with a detected problem will allow one to eliminate possible false positives at an early-stage triggering. If necessary, the final file with geotags can be loaded into the navigator to plot the route to the problem area. In the process of implementing the procedure for neural network processing of video materials, a log is formed containing the frame time and the number of diseased leaves detected by the neural network. Visualization of vineyard-scale physiological condition is made in the form of a heat map (Figure 4), the input data for which are synchronized logs of neural network processing and a GPS tracker (with combined timestamps). The limitations of the proposed methods restrict the possibility of effective monitoring of entire vineyards. Additionally, to implement these methods, mainly lightweight convolutional neural network models (MobileNet, ShuffleNet, YOLO-tiny, etc.) are utilized. The use of lightweight versions of convolutional neural network models is necessitated by the performance limitations of mobile computing devices. To overcome the constraints of the aforementioned methods, a new approach is required. It must efficiently detect and count leaves captured under different conditions and minimize the chance of recounting previously detected leaves.

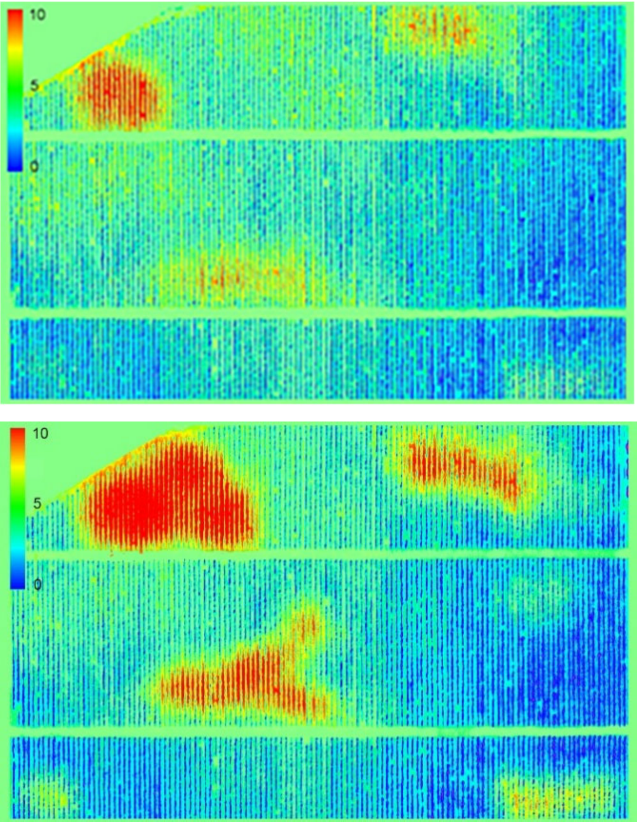

Fig. 4. Interactive heatmap of the vineyard-scale physiological condition obtained with a time interval of one month.

The solution to these challenges necessitates much more computationally intensive processes. In the present study we demonstrated a new technique and methods based on data collection and processing procedures to be performed on separate hardware devices (unmanned aerial vehicle and PC). This approach enables the utilization of stationary GPU-based computing devices to perform intensive computations.

Competitive distinction of the proposed approach is that it can evaluate not only individual leaves, but the whole vineyard. This approach is less accurate in determining the disease of specific vines, because the resolution of individual leaves on the analyzed image is lower, and not all leaves are in the frame, so some diseased leaves may be missed.

However, this approach allows tracking the disease spread dynamics and detecting disease outbreaks early. Additionally, several key findings can be highlighted from this work.

First, the YOLOv7 model was shown to provide a good balance of accuracy and computational efficiency for grape leaf detection, with over 90% mAP. This indicates that state-of-the-art deep learning methods can successfully detect foliage under variable field conditions.

Second, the use of the Deep SORT algorithm for object tracking significantly reduced false positives by avoiding re-counting of leaves in overlapping frames. This step is critical for getting an accurate assessment of vineyard health over time.

The interactive heatmap visualization tool provides a user-friendly format to view results and identify problem areas that require attention. The capability to monitor a vineyard status dynamically over a growing season enables growers to respond promptly to biotic and abiotic stresses. In terms of scalability, the study found that a single UAV-computer setup could effectively survey 2.5 hectares per day. This paves the way for adoption by small to mid-sized vineyards. Some limitations should be noted.

The image dataset, while substantial at over 6000 images, was collected from a single vineyard. Expanding the diversity of grape varietals and diseases would help improve model robustness.

Data augmentation techniques could also be utilized to increase the number and variability of training images. In addition, optimizing flight patterns and image acquisition parameters could potentially increase the survey area covered per unmanned aerial vehicle sortie.

4 Conclusion

The developed technology of automated vineyard physiological condition monitoring based on unmanned aerial vehicle and object detection algorithm YOLOv7 aimed to ensure high yields of high-quality table and technical grapes by minimizing and optimizing human labor. As an indicator of the physiological state, images of grape leaves obtained with the help of UAVs were used. For automated classification of leaves, it is proposed to use deep learning convolutional neural network.

The results of testing the precision detection of diseased leaves by a trained neural network showed that the mAP value is less than 91%, which is sufficient to identify problem areas. Visualization of the vineyard-scale physiological condition is made in the form of a heatmap.

The proposed technology makes it possible to use stationary GPU-based computing devices to perform resource-intensive calculations and shows rather good results in diseased leaves detection even in hard shooting conditions: variable lighting, complex background, partial overlap of leaves.

The integrated use of Object Detection and Object Tracking technologies allows to ignore already detected objects, which significantly reduces the number of repeated and false positives.

Overall, this work demonstrates proof-of-concept for an intelligent UAV-enabled system to monitor vineyard health. The capacity to detect foliar abnormalities in a rapid, comprehensive and automated manner would be a valuable precision viticulture tool.

Future research could explore incorporating multispectral or thermal imagery to complement visual disease detection. The system could also be extended to estimate additional physiological parameters, such as crop yield. This study provides a foundation to build upon with deep learning and robotics technologies in agriculture.

Implementation of the technology into the production process of agro-industrial enterprises will effectively detect and promptly eliminate diseases in the early stages, which will positively affect the yield of products, as well as reduce the possible financial risks of the enterprise. In addition, the proposed solution is the basis for creating a decision support system to protect grapes from diseases and assessing biotic risks in vineyards. Monitoring biotic risks will allow winegrowers to prevent the spread of pests and diseases, thereby improving the sustainability and resilience of vineyards.

The work was carried out within the framework of the state task of the Ministry of Science and Higher Education of the Russian Federation (subject No. FNZM-2022-0010 Development of a methodology for intelligent automated monitoring for solving problems in the field of winemaking and viticulture).

References

1. M. Ammoniaci, S.-P. Kartsiotis, R. Perria, P. Storchi, Agriculture **11**, 201 (2021)
2. P.N. Kuznetsov, D.Y. Kotelnikov, J Phys Conf Ser **2094**, 052025 (2021)
3. E. Egorov, Z. Shadrina, G. Kochyan, BIO Web Conf **53**, 03002 (2022)
4. M. Fraiwan, E. Faouri, N. Khasawneh, Agriculture **12**, 1542 (2022)
5. B. Basso, J. Antle, Nat Sustain **3**, 254 (2020)
6. P. Kaur, S. Harnal, R. Tiwari, S. Upadhyay, S. Bhatia, A. Mashat, A.M. Alabdali, Sensors **22**, 575 (2022)
7. M. Kerkech, A. Hafiane, R. Canals, Comput Electron Agric **155**, 237 (2018)
8. P. Kuznetsov, D. Kotelnikov, L. Yuferev, V. Panchenko, V. Bolshev, M. Jasiński, A. Flah, Sustainability **14**, 11930 (2022)
9. M. Li, Z. Zhang, L. Lei, X. Wang, X. Guo, Sensors **20**, 4938 (2020)
10. J. Lin, X. Chen, R. Pan, T. Cao, J. Cai, Y. Chen, X. Peng, T. Cernava, X. Zhang, Agriculture **12**, 887 (2022)