# Perennial plant density assessment using UAV images and neural networks

Milan Grujev[1[0009-0002-3211-6844]], Miloš Ilić[2[0000-0002-9696-2874]],

Aleksandar Milosavljević[1[(0000-0003-3057-7727]] and Petar Spalević[2[0000-0002-6867-7259]],

[1] Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia

[2] Faculty of Technical sciences, University of Pristina, Knjaza Milosa 7, 38220 Kosovska Mitrovica, Serbia

**Abstract.** One of the main challenges in primary food production is maintenance of the crops, and assuring optimal yield. However, very often, due to various reasons, plant density can be reduced which leads to reduced yield and longevity of the crops. First step in solving this problem is to address assessment of the actual plant density, which can be automated using images obtained via UAVs, and artificial intelligence methods. In this paper we present software solution for plant density assessment on agricultural land using neural networks on images obtained from UAVs. The proposed software solution covers the entire workflow, starting from training the model on the training set of images, and ending with the use of the trained model. This solution will provide support in decision making regarding condition of the plantations and further agrotechnical measures.

**Keywords:** Plant density, Convolutional neural networks, UAV imagery, Image processing, Instance segmentation

## 1 Introduction

Climate changes in recent years have led to unfavorable conditions for growing plants. By recording production plantations and counting plants at an early stage, further growing plans for these plants can be determined. For example, by applying this solution, a decision can be made about the profitability of further investment. This is very important if plantation contains perennial plants. In the same time, it is more complex to observe current state, and predict possible problems on perennial plantations.

Main motivation of our research is creation of solution that will provide multi user support. For example, proposed solution can be used both by farmers and government agriculture experts. From the farmer's point of view, the proposed solution can be used to define further steps in growing plants, as well as to predicted time of harvesting and amount of harvested products. Viewed from the point of view of state experts, the condition of agricultural crops on large areas can be monitored in this way, which can provide precise information about expected yields.

## 2 Literature review

In the literature, there are few survey studies that investigated the applications of either UAV or various deep learning algorithms in several fields, including agriculture. Some of the surveys investigate the use of UAVs and deep learning algorithms in agriculture. In order to create as much as better solution proposal, we do some research on existing solutions, and in continuation are listed some of them.

Velumani et al. used RGB images taken from UAVs to estimate maize plant density [1]. Authors in this study explored the impact of image ground sampling distance (GSD) on the performances of maize plant detection at three-to-five leaves stage using Faster-RCNN object detection algorithm. Data were collected at high resolution (GSD $\approx$ 0.3 cm) over six contrasted sites were used for model training. Two additional sites with images acquired both at high and low (GSD $\approx$ 0.6 cm) resolutions were used to evaluate the model performances. Results show that Faster-RCNN achieved very good plant detection and counting (rRMSE = 0.08) performances when native high-resolution images are used both for training and validation. Results show some significant improvement (rRMSE = 0:22) compared to bicubic upsampling approach, while still far below the performances achieved over the native high-resolution images.

Bouguettaya et al. presented different object-based and pixel-based algorithms that can be used in order to help researchers and farmers to choose the right classifier according to the targeted crop, the used camera sensors, among other parameters [2]. The choice of the right model is a fundamental key that should improve the accuracy, speed, and reliability to classify different crop types. Authors presented different CNN architectures helping to realize the crop classification task. Approach like this can be used to classify and calculate plant density on the agriculture land.

Pantazi et al. presented detection and mapping of Silybum marianum using a hierarchical self-organising map [3]. The images are collected by using a multispectral camera (green-red-NIR) mounted on a fixed wing UAVs. In that way high-resolution images were created. The Supervised Kohonen Network (SKN), Counter-propagation Artificial Neural Network (CP-ANN) and XY-Fusion network (XY-F) were used to identify the S. marianum among other vegetation in a field. As input features to the classifiers, the three spectral bands of Red, Green, Near Infrared (NIR) and the texture layer were used. Authors achieved accuracy level of 98.64% when they use SKN, 98.87% with the use of ANN, and 98.64% with the use of XY-F.

Liu et al. defined tree counting as end-to-end density regression, which uses a remote-sensing image as an input and generates a tree density map as an output. For this purpose, the authors created a deep neural network to aggregate multiple decoding paths to extract hierarchical features at different encoding stages, in order to merge tree features of multiple scales in remote-sensing images. A hybrid loss function is used to effectively guide the network training and enhance the model ability. Authors used a tree count dataset containing 2400 sample pairs for training and validation and achieved results with a relative mean absolute error (rMAE) of 16,72%, a root mean squared error (RMSE) of 77,96 and R2 of 0.96 [4].

Khidher et al. presented combination deep learning techniques based on a convolutional neural network with drone image data as new technology for the classification of multiple vegetation densities. Authors presented a deep neural network algorithm to classify three types of density for forest trees (high, medium and low) on dataset collected by drone. The analysis is based on a DCNN algorithm to VGG19 and ResNet50 neural network and achieved accuracy 84,07% and 91,17% respectively [5].

Ochoa et al. presented a system for tree detection and segmentation that recognizes four different types of trees. The authors presented two neural network models for different task. For tree detection, authors used modified YOLO model with prediction grid. This model uses RGB image of 256x256x3 as input which is later split using a 5x5 grid. Every cell is responsible for recognizing one object. Every object is predicted as one or more bounding boxes with confidence factor and one-hot vector that represents the class of object, in this example, the tree species. The authors set the confidence factor value to 0.8. Dataset used for training process was splitted into training/validation/test sets in a ratio 60/5/35. Second part of this system is tree segmentation task. The segmentation model is based on modified SegNet model. The input size of segmentation model is also a 256x256x3 image. The proposed model consists of set of hierarchical convolutional modules to reduce size and gain more channels. The modules contains three to five convolutional layers with batch normalization, and at the end of module, it has one pooling layer and one activation layer. These compressed images presents a input for a set of up-sampling modules. Upsampling modules contains one up-sampling layer and followed by several convolutional layers, and the last is activation layer. Evaluation metrics that authors used for proposed model is F1 score and classification accuracy. Authors achieved classification level of 97.5% and value 0.89 as highest F1 score value.

## 3 Methodology

A large number of benefits brought by the use of digital technologies in agriculture have led to the increasing use of software solutions in agriculture. If we look at the available solutions and conducted research in the field of processing images obtained from agricultural areas, as well as the application of deep learning algorithms in order to make decisions, it can be concluded that the areas of application are diverse. However, the largest number of researches is based on images obtained from areas sown with cereals and vegetable crops. The application of deep learning methods is low if the images used as training and test data sets contain perennial plantations. Precisely for this reason, the goal of our work is focused on the creation of a software solution that will enable the processing of such images in the first place and their use for the purpose of counting plants and determining the percentage of cultivated plants per unit of the production area.

In order to handle this problem in the most appropriate way, our research began with the acquisition of images. This kind of approach will provide to us our dataset that will be used in future preprocessing and processing phase. For this purpose, we used the DJI Phantom 4 drone and recorded different agricultural areas in the resting phase (winter period) (see Fig. 1). Photographing during winter period provide visible different between fruit tree and background. The problem that arises here is the active and inactive weed community. This is especially noticeable on the images of one-year and two-year fruit plants, because of the fact that difference between fruit tree and weed is very small.

**Fig. 1.** DJI Phantom 4 and images from dataset

The recordings were made at different heights and distances from the plants themselves. The recordings also included plantations of different ages. In order to determine which of the deep learning algorithms will give the best results, the next phase in the research was the training and testing of different algorithms. In this way, a comparison of different algorithms applied to the same data set was created. Based on the test results, the best model that can be used practically was selected. This approach showed us the justification for further development of the software solution. The proposal of the software solution within the paper will be given from an architectural and technical point of view.

## 4 Solution and discussion

The general model of the system that we propose is given in the figure bellow (see Fig. 2). Input images block represents training and test images that are collected from the field. The preprocessing block is optional. This actually means that in some cases, in order to improve the accuracy, some of the image preprocessing methods can be applied. For example, we could applied contrast adjustment.

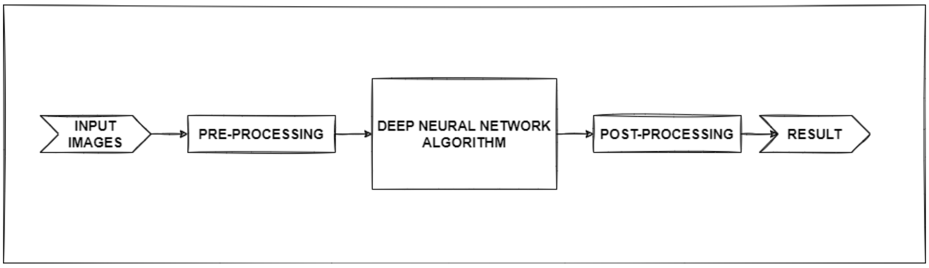


**Fig. 2.** General model of the proposed system

As can be seen from the image, the main part of the system is neural network algorithm. In order to select the best neural network, evaluation of different neural networks need to be done. Selected neural network will be used for segmentation and instance counting. For the process of evaluation of neural networks, we selected encoder-decoder networks (U-Net), as well as regional convolutional networks (RCNN). U-Net is a suitable network in the domain of segmentation, but it requires additional implementation of counting, i.e. detection of object instance in the image. As for RCNN, they are intended for instance segmentation.

Both pre-processing and post-processing are an optional operation, and the formatting of the results can be included within pre-created templates. The goal of such formatting is to adapt the form of the obtained results to as many users as possible.

The advantages of using such a system can be multiple. With annual plants, the system can assist in making decisions about seeding or further processing. Furthermore, counting the trees of perennial plants can help in making various decisions and calculations such as how many trees need to be replanted, or, when it comes to older plantations, a decision can be made about the profitability of further processing of the plantation or whether it is time to clear it and establish a new one.

Based on photographed images in raw dataset training and test data sets were created. Images in raw dataset are high resolution images, and in order to create more appropriate one's those images are tiled in preprocessing phase. Dimensions of created tiled images were 128x128 pixels. In this way from the raw dataset we create 4800 images. From the total number of created images training and test datasets were created. Training dataset contained 4320 images, and test dataset contained 480 images. In this way we provide that training and test datasets contains different images. This was important because of model testing.

In the training phase from the total number of training images (4320) ten percentage of images are selected to be used in validation phase. In this way we provided that images that are used for validation are not previously used during the training of the model. Approach like this provide possibility for good validation during the training phase. In this way we could have precise results right after the training phase for each model.

The initial model training process was started with fifty epochs. After the each finished epoch early stopping analyze results was applied. This mechanism tries to stop training the model in case the model does not improve the results in the last 5 epochs. In this way, unnecessary further execution of epochs were interrupted. During the training of U-Net early stopping was hit on the thirty-fifth epoch. From the other side during the training of RCNN early stopping was hit on the thirty-one epoch.

After the training of both models was completed, the testing of them began. For the purposes of testing the created models, a pre-prepared test dataset was used. Testing of both models was done in the same way so that the obtained results could be compared. On the images below, test results for the U-Net and RCNN training models are presented, respectively (see Fig. 3. and Fig. 4). Images in the first row represent real images collect from the test dataset. Second row images represents ground truth masks created based on the real images from test dataset. Third row images represents masks created during the prediction phase. In this way we could compare differences between ground truth masks and predicted one's. In the same way, comparison of results obtained for these two models can be carried out.

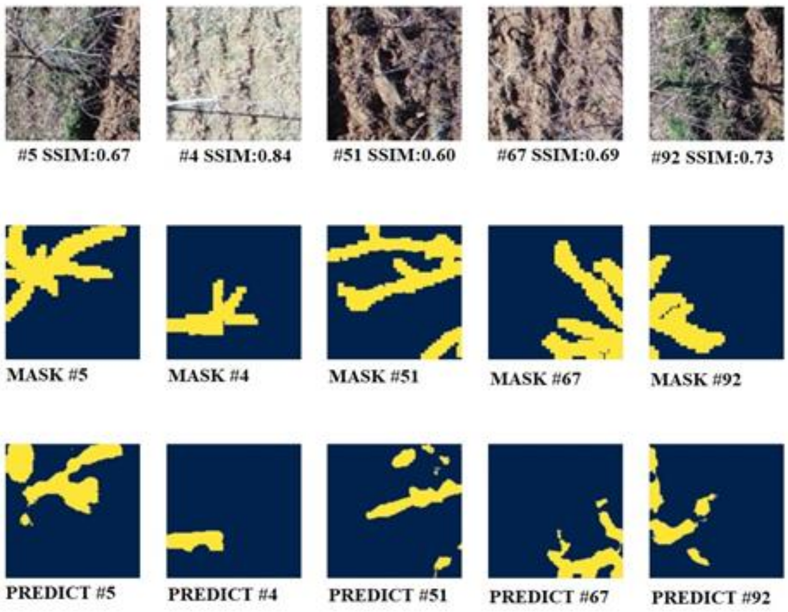


**Fig. 3.** U-Net Result

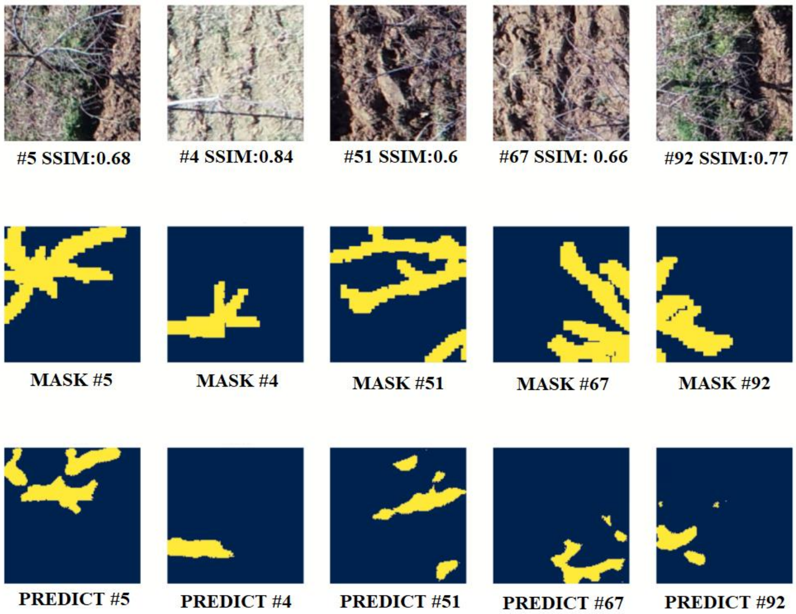


**Fig. 4.** RCNN Result

Comparison of created models was done based on obtained prediction results. For this purpose structural similarity index was used. The Structural Similarity Index (SSIM) is a perceptual metric that quantifies image quality degradation caused by processing. It is a full reference metric that requires two images from the same capture — a reference image and a processed image. This metric defines the level of similarity between two compared images. If we look at the results obtained for U-Net, it can be seen that depending on the complexity of the trees in the image as well as depending on their number, the values for the SSIM index range from 0.60 do the 0.84. From the other side results for RCNN showed that values for SSIM index are in the range from 0.60 to the 0.84. Average SSIM achieved with U-Net and RCNN is 0.74 and 0.75, respectively. The results obtained in this way show a slight difference in the SIM index of the models obtained in this way. This is particularly significant considering the fact that the models were trained and tested using the same datasets.

## 5 Conclusion

Monitoring the state of agricultural crops based on aerial images is one of the areas that is developing at a high speed. Data obtained in this way, processed with adequate algorithms, can enable the collection of a large number of precise data. Some of the data is reflected in the monitoring of the current state of health of plants, their abundance, the growth of plants in a certain period or the estimation of the number of trees per unit area. It is also possible to evaluate the habit of the crown in perennial plants in order to differentiate between physiologically strong plants and those that lag behind in development. Precisely for this reason, the focus of the conducted research was on the entire process from image creation to training and testing of the model and up to the comparison of the obtained prediction results.

Based on the obtained results, we can conclude that both approaches are equally good for this problem. The mutual similarity of ground truth masks and predicted images of about 75% is a good enough result for this kind of problem. The obtained results indicate that this system for tree segmentation can be used for the analysis of images obtained from plantations, but in conjunction with a system that will perform tree detection. When used in this way, the tree detection system will extract only those parts of the image that need to be segmented, that is, those on which there are trees. After tree detection, the segmentation process begins, which tells us about the state of the tree itself (number of branches, size, branching...).

Another advantage that comes with using a tree detection system is the improvement of the dataset itself, in the sense that the training of the segmentation model can only be done with images that contain trees. The idea for future research is to use such a system for tree segmentation in conjunction with a system for tree detection, so that the analysis of trees can provide as much useful information as possible and be as efficient as possible.

## References

1. Velumani, K., Lopez-Lozano, R., Madec, S., Guo, W., Gillet, J., Comar, A., Baret, F.: Estimates of maize plant density from UAV RGB images using faster-RCNN detection model:impact of the spatial resolution. Plant Phenomics 2021, 1–16 (2021).
2. Bouguettaya, A., Zarzour, H., Kechida, A., Taberkit, A.M.: Deep learning techniques to classify agricultural crops through UAV imagery: a review. Neural Computing and Applications 34(12), 9511–9536 (2022).
3. Pantazi, X.E., Tamouridou, A.A., Alexandridis, T., Lapopodi, L., Kashefi, J., Moshou, D.: Evaluation of hierarchical self-organising maps for weed mapping using UAS multispectralimager. Computers and Electronics in Agriculture 139, 224–230 (2017).
4. Liu, T., Yao, L., Qin, J., Lu, J., Lu, N., Zhou, C.: A deep neural network for the estimation of tree density based on high-spatial resolution image. IEEE Transactions on Geoscience and Remote Sensing 60, 1–11 (2021).
5. Khinder, A.M., Sehree, N.A.: Automatic trees density classification using deep learning of unmanned aerial vehicles images. International Journal of Mechanical Engineering 7(2), 3155–3136 (2022).
6. Ochoa, K. S., Guo, Z.: A framework for the management of agricultural resources with automated aerial imagery detection. Computer Electronic in Agriculture 162, 53–69 (2019).