

An algorithm for crops segmentation in UAV images based on U-Net CNN model: Application to Sugarbeets plants

Khalid EL Amraoui^{*1}, Ayoub Ezzaki¹, Abdelkrim abanay¹, Mouataz Lghoul¹, Majid Hadri¹, Aziz Amari¹, and Lhoussaine Masmoudi¹

¹LCS Laboratory, Physics Dept. faculty of Science, Mohammed 5 University in Rabat, Ibn Battouta Street, 10000, Rabat, Morocco

Abstract. In recent years, Digital Agriculture (DA) has been widely developed using new technologies and computer vision technics. Drones and Machine learning have proved their efficiency in the optimization of the agricultural management. In this paper we propose an algorithm based on U-Net CNN Model to crops segmentation in UAV images. The algorithm patches the input images into several 256×256 sub-images before creating a mask (ground-truth) that will be fed into a U-Net Model for training. A set of experimentation has been done on real UAV images of Sugarbeets crops, where the mean intersection over Union (MIoU) and the Segmentation accuracy (SA) metrics are adopted to evaluate its performances against other algorithms used in the literature. The proposed algorithm show a good segmentation accuracy compared to three well-known algorithms for UAV image segmentation.

1 Introduction

Digital agriculture (DA) has been widely developed during the last decades, where the computer vision algorithms, and specially image segmentation, takes a great place in several DA applications.

Nowdays, Deep learning (DL) is presented as the most performance state-of-art systems for image classification. In the last decade, several Deep Convolutional Neural networks has been proposed, such as: AlexNet [1], GoogleNet [2], DenseNet [3], etc. The main idea of Deep learning is to automatically build a data representation through the learning phase, thus avoiding human intervention. We therefore speak of learning by representation. A Deep Learning algorithm will learn increasingly complex hierarchical representations of data. [4].

The use of Deep learning in DA has found a success for two reasons: i) the big amount of available data in DA (Aerial images). ii) The complexity of some DA tasks (disease detection, harvest estimation) against the performance of classical computer vision algorithms. The authors in [5] presented a DL architecture for real-time tomato disease and pest recognition based on three main families of detectors: Region-based Fully Convolutional Network (R-FCN), Faster Region-based Convolutional Neural Network (Faster R-CNN) and Single Shot Multibox Detector (SSD) and after that combining these detectors with a deep feature extractors. Another work presented in [6] proposed a CNN model for detection of mango fruit in trees canopies and the estimation of fruit load. In [7] the authors proposed a supervised deep learning-based classification called transfer learning for soil

segmentation. And finally, the work presented in [8], [9], use quantum superposition laws and machine learning for agricultural image segmentation

In this paper, we present an algorithm for aerial agricultural image segmentation based on U-Net CNN Model. The proposed algorithm can be divided into four major parts:

- 1- Image patching.
- 2- Mask creation.
- 3- U-Net model training and patch prediction
- 4- Reconstruction.

In order to show the performance of our proposal, a set of experimentation is adopted using agricultural images taking by a UAV and using the mean intersection over Union (MIoU) metrics and the Segmentation accuracy (SA) metrics to quantify its performance.

The paper is organised as follow: The next section is dedicated to the used materials. We detailed the proposed method in the section 3. After that, a set of experimentation is conducted in the section 4 and finally a conclusion is presented in the sixth section.

2 Materials

Our objective is the plant segmentation from aerial images. For that we used the database presented in [10] of aerial images taken by an UAV (Fig. 1). It contains 474 images taken by a DJI MATRICE 100 UAV drone for the Field A, and 194 images taken by DJI PHANTOM 4 UAV for the field B (Table 2). In our model, we used the images from the field A. Table 1 Represents the details of the used Database.

* Corresponding author: khalid.elamraoui@um5r.ac.ma

Table 1. Unmanned aerial vehicle (UAV) sensor and image parameters of the field A database

Field A	
Drone	DJI MATRICE 100
Resolution ($H \times V$)	4000 $\times$ 2250
FOV ($H^\circ \times V^\circ$)	73.7 $\times$ 53.1
Focal length (mm)	8.8
ground sampling distance (mm)	4
Sensor size ($H \times V$ mm)	13.2 $\times$ 8.8
Pixel size (μm)	2.41 $\times$ 2.41
Image format	JPG
Dynamic range	8 bit
Number of images	474
Image type	RGB

Table 2. Unmanned aerial vehicle (UAV) sensor and image parameters of the field B database

Field B	
Drone	DJI PHANTOM 4
Resolution ($H \times V$)	3840 $\times$ 2250
FOV ($H^\circ \times V^\circ$)	73.7 $\times$ 53.1
Focal length (mm)	8.8
ground sampling distance (mm)	9
Sensor size ($H \times V$ mm)	13.2 $\times$ 8.8
Pixel size (μm)	2.41 $\times$ 2.41
Image format	JPG
Dynamic range	8 bit
Number of images	194
Image type	RGB

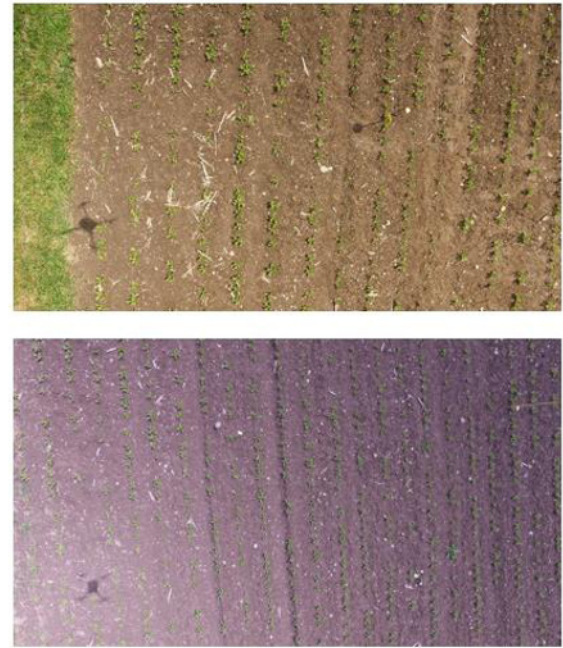


Fig. 1. Example of images from the used Database

3 Proposed Method

As mentioned before, the proposed method is composed of five major steps. In this section we present the five parts of proposed algorithm with a theoretical background of U-Net CNN model.

Our main goal is the segmentation and localisation of plants in UAV images. For that, we start with patching each image of the databases into 256×256 sub-images. After that, each sub-images is used to create a mask that will stand as a ground truth for the training model.

3.1 Image patching and labelling

The first step of our proposal consists of patching the input images of the database into several sub-images (144 sub-images for each 4K image). Each sub-image is labelled based on its position in the original image (Fig. 2). These labels will be used in the reconstruction process.

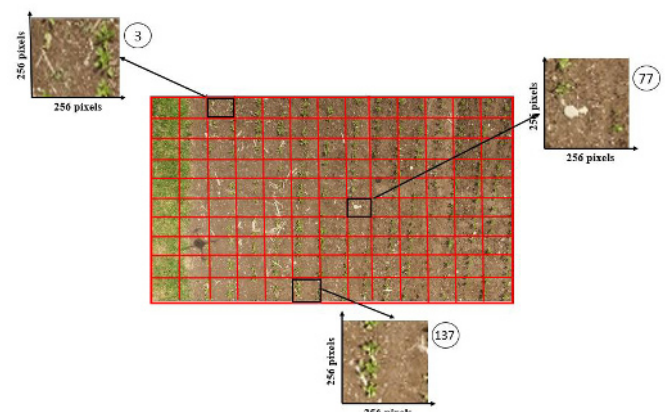


Fig. 2. Image patching and labelling.

3.2 Mask creation

The first step of our proposal is the mask generation. The idea here is to automatically segment the database images to be used as ground-truth. For that we used a color-based manual segmentation using Hue Saturation Value HSV color space, the Hue matrix of the input images is extracted by thresholding the green elements of the image (Fig. 3). An example of the Mask generation result is presented at the Fig. 4

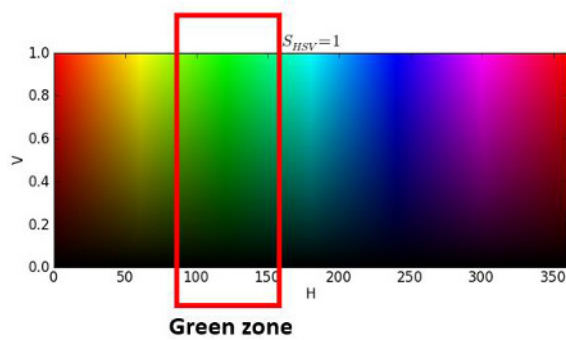


Fig. 3. Extracting Green zone from the HSV color space.



Fig. 4. Example of Mask creation of the used images.

3.3 U-Net Model training and patch prediction

U-Net is a semantic segmentation architecture [11]. It has two paths: one that contracts and one that expands. The convolutional network's contracting path follows the standard architecture. It comprises of two 3x3 convolutions (unpadded convolutions) that are applied repeatedly, each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for down sampling. We quadruple the number of feature channels with each down-sampling step. Every step in the expansive path involves up-sampling the feature map, a 2x2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the proportionally cropped feature map from the contracting path, and two 3x3 convolutions. each of which is followed by a ReLU. Due to the loss of boundary pixels in every convolution, cropping is required. A 1x1 convolution is employed at the final layer to convert each 64-component feature vector to the desired number of classes. The network comprises a total of 23 convolutional layers (Fig.5).

After the training process, we used the trained model to predict on the patches (Fig.6).

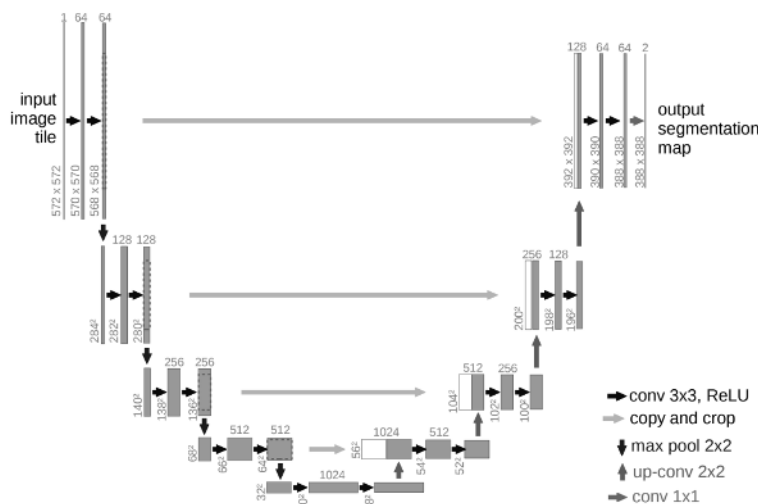


Fig. 5. U-Net CNN Model architecture



Fig. 6. Example of patch prediction using the trained model.

3.4 Image reconstruction

The objective of the reconstruction is to remount segmented sub-images to the original segmented high-resolution images. For that we used the etiquette given to each sub-images to reconstruct the original segmented 4K image, as shown at the Fig. 7.

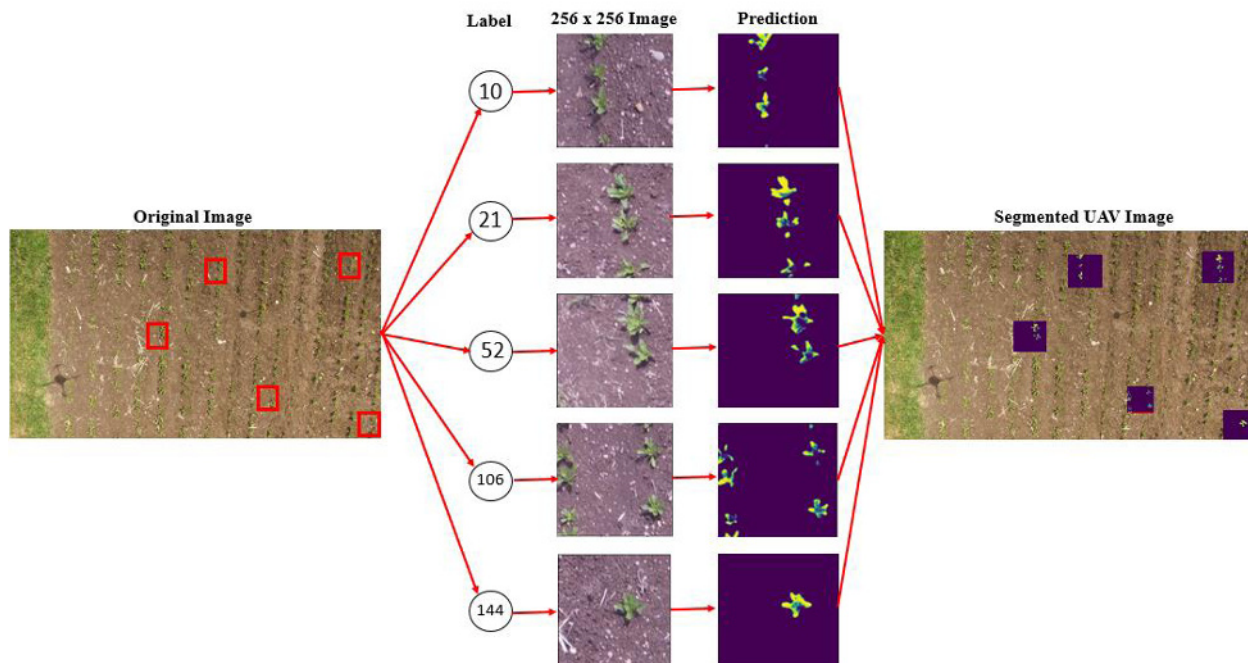


Fig. 7. Reconstruction process.

The proposed algorithm is summarized in the Fig. 8.

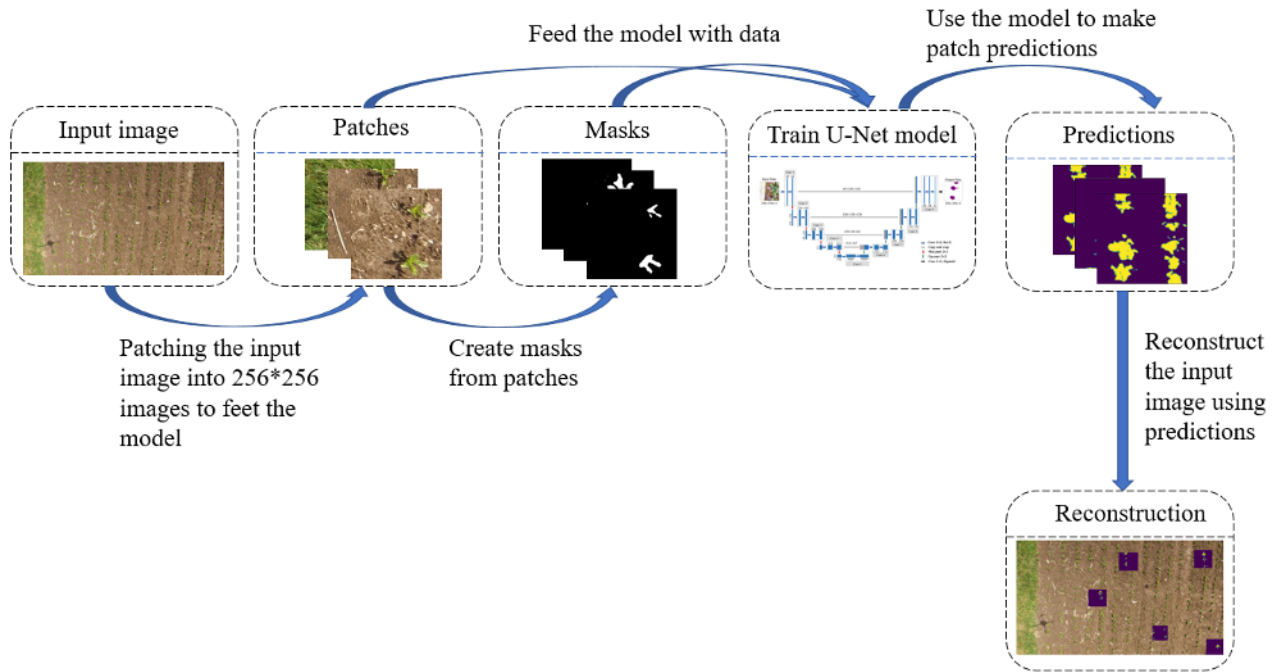


Fig. 8. Summary of the proposed algorithm

4 Experimentation

In order to test the performance of the proposed algorithm, a set of experimentations is done on agricultural aerial images.

4.1 Evaluation method

To evaluate the algorithm in term of segmentation accuracy, a quantified and measurable metrics are needed. We used the the mean intersection over Union (MIoU) and Segmentation accuracy (SA) metrics, defined as:

Segmentation Accuracy (SA): This criterion presents how well the segmentation is performed.

$$SA = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

With:

TP: True positive.
FP: False positive.
TN: True Negative.
FN: False negative.

The Mean Intersection over Union (MIoU): The MIoU present how is the prediction is similar to the ground truth.

$$MIoU = Mean \left(\frac{TP}{TP+FP+FN} \right) \quad (2)$$

4.2 Experimentations

In order to show the performance of our algorithm, we applied it to aerial images from the database presented in the section 2, and compared it with the method proposed by Zhun and al. for tobacco plant, which use three stage algorithms: i) plant segmentation using morphological operations. Ii) CNN training ii) Post-treatment operation [12]. The method proposed in [13] which used a CNN model for plants detection and counting, and the method proposed by Pacheco and al. that use an encoder-decoder architecture for classifying each pixel as non-crop or crop [14].

The Table 3 and the figure 10 shows the results of applying the proposed algorithm to the UAV images compared to other methods presented for the same purpose.

The figure 9 shows the test we did on aerial image from the database, field B, presented in section 2. We divide the image into small images of 256*256 and make the predictions. After we reconstruct the image into its original size using the predictions.

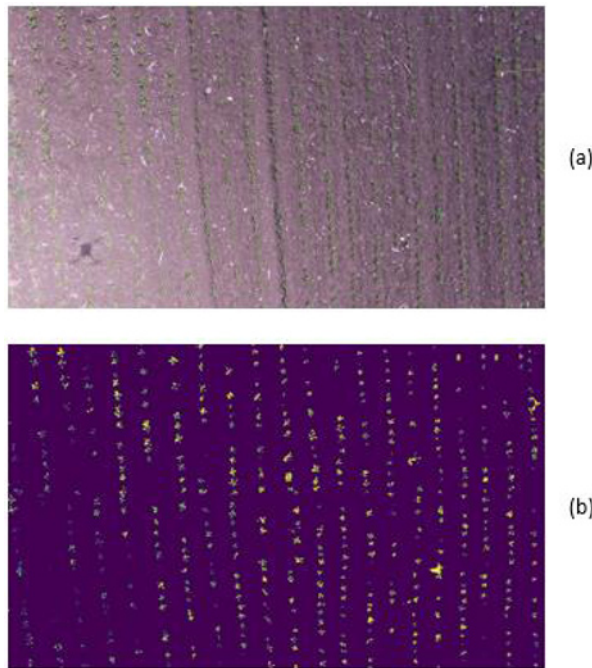


Fig. 9. (a) Original image, (b) Prediction

Table 3. The SA and MIoU values for the tested methods.

Method	SA(%)	MIoU(%)
Proposed method	97	91
Zhun and al.	92.8	-
Oscos and al.	95.1	-
Pacheco and al.	93.95	-

Fig. 10. Accuracy graph



5 Conclusion

We presented in this paper an algorithm for plant segmentation in Aerial images based on U-Net CNN Model. The proposed method starts with an image patching and labelling before using them for mask creating. The created masks are after that used as an input for the training model.

The proposed algorithm gives a segmentation accuracy of 97% against 92.8% by the method proposed by Pacheco and al., 92.8% by the method proposed by Zhun and al. and 95.1% by the method proposed by Oscos and al.

Our proposal can be optimized by adding a step for enhancing the quality of the patched images. Also a manual mask creating can increase the accuracy of the model by preventing false segmented masks.

As a perspective, we intend to add a super-resolution step to increase the quality of the sub-images using quantum principals.

Acknowledgment

The authors of this paper are thankful to the Ministry of Higher Education, Scientific Research and Innovation (MHESRI), the National Centre of Scientific and Technical Research of Morocco (CNRST) and Digital Development Agency of Morocco (ADD) for funding this project through AL-KHAWARIZMI program of Morocco.

References

1. A. Krizhevsky, I. Sutskever, et G. E. Hinton, *Adv. Neural Inf. Process. Syst.*, **vol. 25** (2012)
2. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, *Proc. IEEE Int. Conf. Comput. Vis.*, p. 1-9 (2015)
3. G. Huang, Z. Liu, L. van der Maaten, et K. Q. Weinberger, *Proc. IEEE Int. Conf. Comput. Vis.*, p. 4700-4708 (2017)
4. A. Koirala, K. B. Walsh, Z. Wang, et C. McCarthy, *Comput. Electron. Agric.*, **vol. 162**, p. 219-234 (juill. 2019)
5. A. Fuentes, S. Yoon, S. C. Kim, et D. S. Park, *J. Sens.*, **vol. 17**, n° 9, Art. n° 9 (sept. 2017)
6. A. Koirala, K. B. Walsh, Z. Wang, et C. McCarthy, *Precis. Agric.*, **vol. 20**, n° 6, p. 1107-1135 (déc. 2019)
7. C. Douarre, R. Schielein, C. Frindel, S. Gerth, et D. Rousseau, *Imaging*, **vol. 4**, n° 5, Art. n° 5 (mai 2018)
8. K. El Amraoui, A. Ezzaki, L. Masmoudi, M. Hadri, H. EL Belrhiti, M. El Ansari, A. Amari, *IRASET*, **vol. 2**, p. 1-7 (2022)
9. A. Ezzaki, M. Lhoussaine, M. E. Ansari, F.-A. Moreno, R. Zenouhi, et J. G. Jimenez, *Int. J. Electr. Comput. Eng.*, **vol. 10**, n° 2, Art. n° 2, pp1655-1666 (2020)
10. « UAV Sugarbeets 2015-16 Datasets – StachnissLab ». <https://www.ipb.uni-bonn.de/data/uav-sugarbeets-2015-16/> (consulté le 23 avril 2022).
11. O. Ronneberger, P. Fischer, et T. Brox, *MICCAI 2015*, Cham 2015, p. 234-241 (2015)
12. Z. Fan, J. Lu, M. Gong, H. Xie, et E. D. Goodman, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **vol. 11**, n° 3, p. 876-887 (2018)
13. L. P. Osco, M.D.S. de Arruda, D.N. Gonçalves, A. Dias, J. Batistoti, M. de Souza, F.D.G. Gomes, A.P.M. Ramos, L.A. de Castro Jorge, V. Liesenberg, J. Li, *ISPRS J. Photogramm. Remote Sens.*, **vol. 174**, p. 1-17 (2021)

14. J. Fuentes-Pacheco , *J. Torres-Olivares, E. Roman-Rangel, S. Cervantes, P. Juarez-Lopez, J. Hermosillo-Valadez, J.M. Rendón-Mancha, Remote Sens.*, **vol. 11**, nº 10, 1157-1174 (2019)