2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) | 978-1-6654-5305-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/CSDE56538.2022.10089303

An Evaluation of the Generalization Capabilities of Machine Learning Models for Vine Line Detection

Jérôme Treboux
Institute of Informatics
HES-SO Valais
Sierre, Switzerland
jerome.treboux@hevs.ch

Aurore Pittet
Institute of Informatics
HES-SO Valais
Sierre, Switzerland
aurore.pittet@gmail.com

Dominique Genoud
Institute of Informatics
HES-SO Valais
Sierre, Switzerland
dominique.genoud@hevs.ch

Abstract—**Precision agriculture can optimize the production of agricultural crops by analyzing aerial images with varying resolutions and acquired from different sources. It is widely accepted that machine learning (ML) model, especially deep neural networks (DNN), are very efficient for image segmentation. DNNs have been used to segment complex texture and planting structures, such as vine lines, due to their variations in shape, color and orientation. However existing DNNs reach their limits to segment aerial images with varying resolution and multiple instance of vine lines crossing a entire image. In this paper, we present an improvement of the generalization capabilities of ML models to segment vine lines in satellite images. An approach from a previous works that combine neural networks and other classifiers allow us to improve the classification and generalize the models that increase the f-score by 17%.**

Index Terms—**Machine Learning, Neural Network, Deep Neural Network, Decision Tree Ensemble, Image Analysis, Vine Line Detection, Vineyards, Image Segmentation, Transfer Learning, Precision Agriculture**

I. Introduction

According to the report of the World Government Summit [1] and confirmed by the Food and Agriculture Organization of the United Nations (FAO) [2], food production must be increased by 70% by 2050. The estimated food shortage is due to high population growth, excessive demands on natural resources, declining productivity due to climate change and food waste. To address these issues, precision agriculture has emerged and is integrating new technologies such as drones and Artificial Intelligence (AI) [1]. The combination of these technologies enables, among other things, the detection of agricultural areas and objects as well as autonomous navigation by detecting obstacles.

Image segmentation is an image analysis task that assigns each pixel of an image to a class and produces an output with the detected segments [3]. It has been recently applied to precision agriculture, mainly to spot farming areas on aerial images acquired with a drone [4] and to determine the location of the crop to treat with products (e.g. fertilizers or pesticides) [5]. However, aerial images collection is not limited to drones; many sources provide satellite images with different resolution. Therefore, machine learning (ML) models should have the ability to be generalized so that they are insensitive to resolution variation. Among the numerous techniques used for image segmentation, Convolutional Neural Network (CNN) such as AlexNet [6], VGG-16 [7] or U-Net [8], have obtained high performance for this task.

Vine lines detection combines the complexity of line detection in images with varying natural conditions [9], such as variation of their orientation and shape and the change of color due to season and lighting [10]. Therefore, vineyards are considered as a difficult plantation structure to analyze with ML models.

Existing state-of-the-art ML models are not able to segment a difficult plantation structure or objects crossing an entire image, such as vine lines. Furthermore, existing models for precision agriculture are often trained on specific dataset and are not able to be generalized [11]–[13].

This paper presents an evaluation of the models' abilities to analyze images with different resolution. Two types of images used : aerial images acquired with a *drone* and satellite images from *swisstopo*. The detailed experiments are based on a model trained to detect vine lines using the *drone* dataset. They are evaluated to segment vine lines in satellite images from *swisstopo* dataset. The original task of the model is the same; but the input is varying. The work is based on two previous researches [14], [15] that developed original approaches to segment vine lines in aerial images. The first approach [14] combines a U-Net that segment an image with a Decision Tree Ensemble (DTE) that determines the class of the segments. The second approach [15] improves the previous one by combining an asymmetrical architecture of a U-Net and a Random Forest (RF). Finally, an evaluation of generalization improvement is presented with classifiers retraining. The performance of the models are evaluated with the precision, recall and f-score.

The rest of this paper presents the state-of-the-art related to our field of study, the dataset and the data processing used during the experiments , the methodology and the metrics, the experiments and their results and finally the conclusion.

II. State-of-the-Art

Among existing CNNs, we present the U-Net [8] that is often used for image segmentation and one of the most used for aerial image analysis. Furthermore, this CNN can be used with transfer learning reducing the amount of data necessary to retrain the network for a new task [16].

978-1-6654-5305-9/22/$31.00 ©2022 European Union

The U-Net is an auto-encoder based CNN [8]. The encoder aims at detecting objects in image by applying convolutions. The decoder aims to locate objects in image by applying transposed convolutions. The U-Net has been successfully used in numerous fields such as rice detection in image [17] and satellite image analysis [18].

Lu D. et al. [19] demonstrate the impact of image resolution on image analysis, in this case, bamboo detection. The performance of the models decreases in proportion to the resolution. The paper by Sozzi et al. [20] demonstrates the difficulty of detecting vineyard areas smaller than 0.5 hectares in satellite images. This paper highlights the problem of generalization of the models when the resolution of the images varies.

Various studies focuses on vine lines detection in aerial images. The first study of Comba L. et al. [21] presents a precision of 59.8%, a recall of 70.6% and a f-score of 32.4%. We note in this study a large number of false detection. Two more studies focus on ML models for image segmentation, more specifically for disease detection in vineyards. The result is the segmentation of vine lines with disease. Their models achieve an accuracy of 92% [13]. But they include infrared images and depth (e.g. the difference in height between the vine and the ground) to train the models which do not allow to generalize the models to other images from various open access aerial image databases.

Transfer learning is used to retrain an existing model for a new task [16]. With Transfer learning, the knowledge from a model is used to retrain a model for a related task. It is often used for image analysis [22], by fine-tuning a pre-trained CNNs. The amount of data needed to retrain a model using transfer learning is drastically reduced.

III. Dataset and Data Processing

A. Dataset

The first dataset used in this research is composed by images acquired with a drone. We refer to this dataset as *drone dataset*. It is used to train the original models from [14] and [15]. It is composed of RGB aerial images with a resolution between 60 and 80 pixels per meter. The number of vine lines per images is between 29 and 191. The width of a vine line in average is between 30 and 75 pixels. The size of an image is 4000 x 3000 pixels.

The second dataset used in this research is composed by images acquired on the platform Swisstopo[1]. We refer to this dataset as *Swisstopo dataset*. It is composed of RGB images acquired with a plane every three years. The images have different colors due to the variation of light conditions and seasons. The images have a resolution between 5 and 15 pixels per meter and a varying size. The width of a vine line in average is between 2 and 30 pixels. The size of an image is 10000 x 10000 pixels.

Figure 1 shows an example of an image from the *drone* et *Swisstopo* datasets. Figure 2 is a zoom on two images of these two datasets, highlighting the variation of the resolution.

[1] Swiss Federal Office of Topography (https://www.swisstopo.admin.ch)

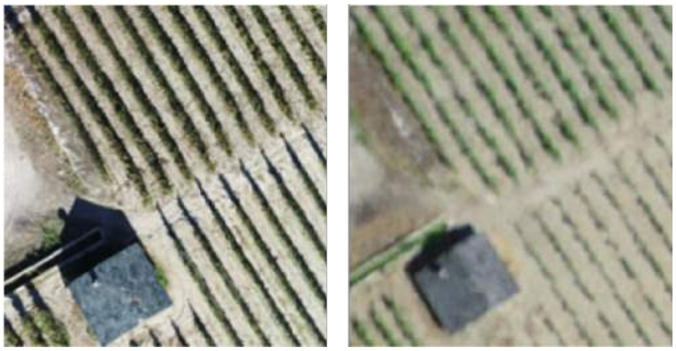


Fig. 1. Images from our dataset extracted from each of the two data sources, drone image (left) and swisstopo image (right), representing the same geographical location.

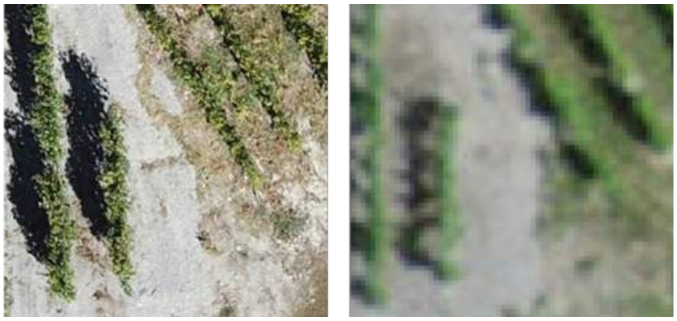


Fig. 2. Difference in resolution between an image taken with a drone at a distance of 50m from the ground (left) and an image extracted from Swisstopo (right). This figure shows also shadow between the vine lines.

B. Data Processing

Our models are trained based on a supervised learning [23], that needs labeled samples. Therefore, we applied a manual per-pixel labeling to each image. Labeled images are used as the ground truth to evaluate the ML models performance. Each dataset are split in three subsets : train set, validation set and test set. A summary of the reference name for each dataset is presented in Table I.

TABLE I
This table explains the division of the dataset and the names to refer to. Two datasets are used : Drone and Swisstopo. Each dataset is split into three subsets: train, validation and test.

	Train set	Validation set	Test set
Drone dataset	Drone train set	Drone validation set	Drone test set
Swisstopo dataset	Swiss train set	Swiss validation set	Swiss test set

We then applied data augmentation [24], due to the small amount of images available, to the images of the train sets. We applied 2 types of augmentations to the images and their ground truth: (1) horizontal and vertical flips, and (2) a rotation between $-90°$ to $90°$ every $10°$. We also applied a downsampling to the images and ground truth of the train set from the *drone dataset* to obtain a generalized dataset.

This data augmentation increases the variation of vine lines orientation and therefore will match the ground variability.

IV. Methodology and Metrics

A. Methdology

Our methodology aims at dividing an image into tiles, using a U-Net to segment each tile, combining the segmented tiles to obtain the segmentation of the original image, extracting each vine line and defining their class with a classifier: vine or other. This methodology is summarized in Figure 5.

We evaluate the performances of our models on the number of correctly detected vine lines. We first create groups of pixels of the same class (e.g. vine pixels), commonly referred to as connected component. For each connected component, we find its minimal rectangle area, called bounding box, characterized by a center, an angle, a height and a length. Figure 3 shows an example of a connected component and its bounding box.

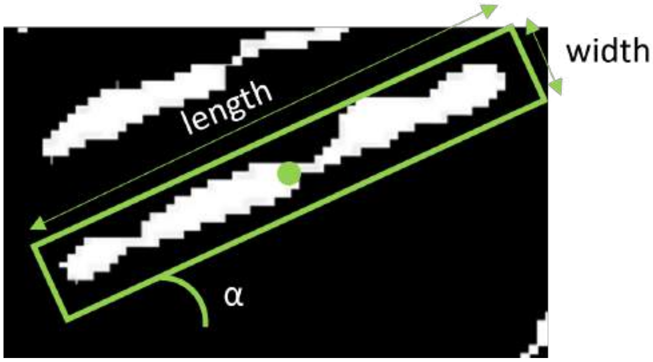
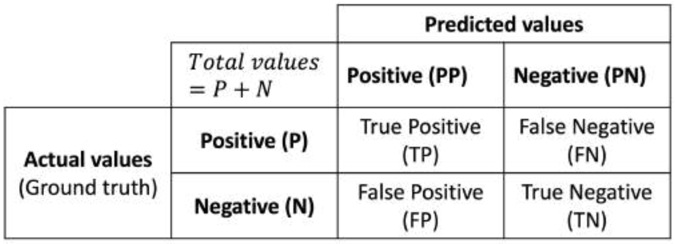


Fig. 3. Example of a bounding box for a vine line. the bounding box is the green rectangle, with a center, a angle (α), a height and a width.

We then use these bounding boxes to determine if a detected vine line corresponds to its ground truth. For each bounding box on the segmented image, we compute the Intersection over Union (IoU) with all the bounding boxes of the ground truth. A detected bounding box that is included within a ground truth bounding box is considered as correctly detected as it is a part of the vine line. For all other bounding boxes, we set the threshold for the IoU to consider a vine line correctly detected to 0.75. This threshold is the optimum value obtained in previous detection experiments [14]. A detected bounding box with all the computed IoU under 0.75 is a false positive. A ground truth bounding box with no corresponding detection is a false negative.

B. Metrics

We used three of the most used metrics for object detection, described below, to evaluate the performance of our models [25]. These metrics can be computed using a confusion matrix that helps to interpret the quality of a classification (see Figure 4).



		Predicted values	
	Total values $= P + N$	**Positive (PP)**	**Negative (PN)**
Actual values (Ground truth)	**Positive (P)**	True Positive (TP)	False Negative (FN)
	Negative (N)	False Positive (FP)	True Negative (TN)

Fig. 4. Confusion matrix for two classes.

The first metric is the precision, also called Positive Predictive Value (PPV), which is the proportion of object correctly identified compared to all positive objects [25]. The precision returns a value between 0.0 and 1.0. A perfect precision score of 1.0 means that all objects predicted as positive are positive in the ground truth. Precision is computed as

$$Precision = \frac{TP}{TP + FP}. \quad (1)$$

The second metric is the recall, also called True Positive Rate (TPR), which is the proportion of object correctly identified compared to the total of positive objects in the ground truth [25]. The recall returns a value between 0.0 and 1.0. A recall of 1.0 means that all positive values from the ground truth were predicted as positive. Recall is computed as

$$Recall = \frac{TP}{TP + FN}. \quad (2)$$

The false positive rate (FPR) is the probability of false rejection of a negative value. Precision and recall depend on a threshold corresponding to the probability of the prediction [26]. The threshold can be the Equal Error Rate (EER), that reflects the equality of the TPR and FPR [27], or determined based on the result of the probability density function [28].

The third metric is the f-score (F_1) that is the harmonic mean of the precision and the recall [25]. The f-score returns a value between 0.0 and 1.0. A F_1 of 1.0 indicates a perfect precision and recall. We compute the f-score as follow :

$$F_1 = 2 \times \frac{PPV \times TPR}{PPV + TPR}. \quad (3)$$

Finally, to determine if the models are significantly different, we computed the Standard Error (SE). To compute the SE, we use the following Equation [15] :

$$SE = Z_\alpha \sqrt{\frac{p(1-p)}{n}}, \quad (4)$$

where Z_α is the confidence level, p is the precision and n is the number of data.

V. Experiments

We conducted four experiments to evaluate the ability of our models to be generalized to new images. The first two experiments are conducted with the approach using a U-Net and a DTE [14], the next two experiments are conducted with the approach using an asymmetrical architecture of the U-Net and a RF [15].

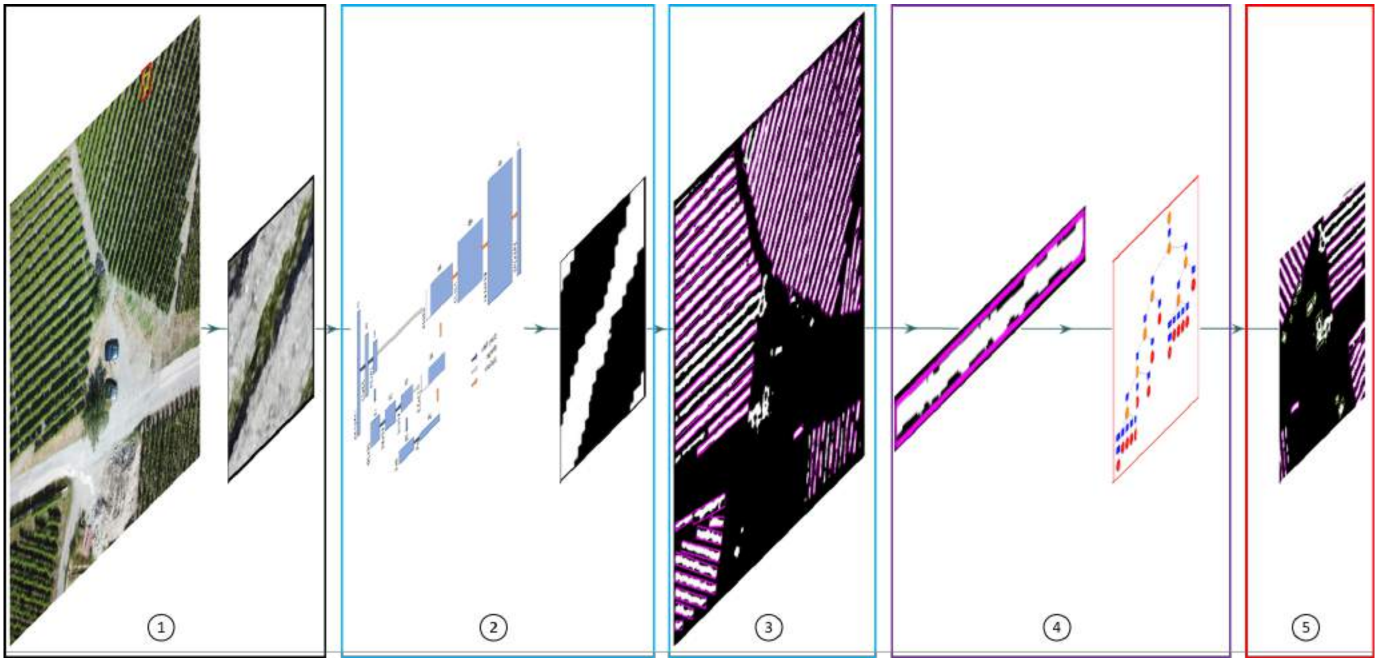


Fig. 5. Our original methodology for vine line detection. (1) Original image is divided into tiles. (2) Segmentation of vine lines with the U-Net. (3) Connected component and minimum rectangle area. (4) Extraction of the vine lines and their features to train a DTE. (5) Result produced by the DTE.

A. *U-Net and RF*

The first two experiments are based on the approach of [14] that combines a U-Net and a DTE. The model has an input of 145x145 pixels and an output of 145x145 pixels. Therefore, we divided the images into tiles using a sliding window with a size of 145x145 pixels and a stride of 145x145 pixels.

The U-Net first segments the tile. We then combine all segmented tiles to reconstruct the original image. For each bounding box, we extract its size (length and height) and five First Order Statistics [29], described below, computed for each RGB histogram from the portion of the original image.

- The mean describes the center of the histogram.
- The variance describes the distance of a value from the mean.
- The standard deviation measures the dispersion of the distribution.
- The skewness measures the symmetry of the distribution.
- The kurtosis measures the intensity of the peak of the histogram.

We already used successfully these features during the original experiment. These features are used with the DTE to confirm the class of each bounding box.

1) Experiment 1: For the first experiment, we evaluated the performance of a model trained with the images of the *drone train set* to predict vine lines in the images of the *swiss test set*.

2) Experiment 2: The second experiment aims to improve the performance we obtained with the Experiment 1. We use the first part of the model, the U-Net, trained with images of the *drone train set*. We retrained the DTE using the images from the *swiss train set*. A new DTE model is generated using the newly extracted features.The overall approach, that combines the pre-trained U-Net and the retrained DTE, is evaluated using the images from the *swiss test set*.

B. *Asymmetrical U-Net and RF*

The last two experiments are based on the approach of [15] that combines an asymmetrical architecture of a U-Net (see Figure 6) and a RF. The asymmetry of the U-Net improves the robustness of the model to the variation of the orientation and colors of the vine lines [15]. The model has an input of 145x145 pixels and an output of 72x72 pixels. The output is the center of the tiles. Therefore, we divided the images into patches using a sliding window with a size of 145x145 pixels and a stride of 72x72 pixels.

The asymmetrical U-Net segments the tile. We then combine all segmented tiles to reconstruct the original image. For each bounding box, we extract the same features as described in the previous experiments.

1) Experiment 3: For the third experiment, we evaluated the performance of the asymmatrical U-Net trained with the images of the *drone train set* to predict vine lines in the images of the *swiss test set*.

2) Experiment 4: With this last experiment, we improve the performance of the model from the experiment 3. We use the asymmetrical U-Net trained with images of the *drone train set* for vine line detection. We retrained the RF with the images from the *swiss train set*. A new model is generated using the new extracted features. The overall approach, that combines a pre-trained asymmetrical U-Net and a retrained RF is used to predict vine lines in images from the *swiss test set*.

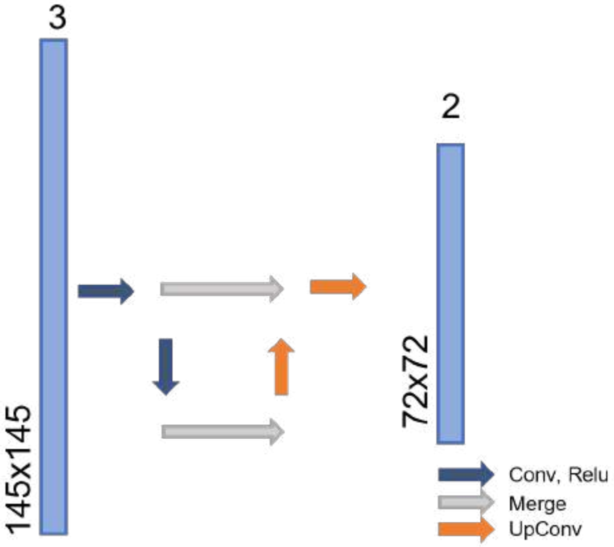


Fig. 6. Asymmetrical U-Net architecture. The input is a tile. The output is the segmentation of the center of the tile [15]

VI. Results

In this Section, we present the results we obtained with our four experiments. The results are summarized in Table II.

With the first experiment, using the U-Net trained on the *drone dataset*, we obtained a precision of 30% ±3 %, a recall of 93% and a f-score of 45% ±3 % using images from the *swisstopo dataset*. We note that vines are generally well detected but numerous other objects are detected as vine (false positive). The model is not generalized enough.

With the second experiment, using the same U-Net as in the first experiment but with a retraining of the classifier using images from the *swisstopo dataset*, we obtained a precision of 41% ±3 %, a recall of 93% and a f-score of 57% ±3 %. We improved significantly the precision by around 37%. We note that the recall remains unchanged as the DTE mainly acts as a filter by removing false positives. Figure 7 highlights the false positive and objects wrongly detected as vine.

With the third experiment based on an asymmetrical U-Net, we obtained a precision of 42% ±3 %, a recall of 95% and a f-score of 58% ±4 %. We slightly improved the precision, the recall and the f-score compare to the first experiment. Numerous other object are wrongly detected as vine.

With this last experiment, we significantly improved the precision and the f-score compared to the third experiment. We obtained a precision of 53% ±4 %, a recall of 95% and a f-score of 68% ±3 %. We improved significantly the precision by around 26% and the f-score by around 17%. We note that the recall remains unchanged as the RF mainly acts as a filter by removing false positives. Figure 8 highlights the RF acting as a filter by removing other objects such as a house (in purple).

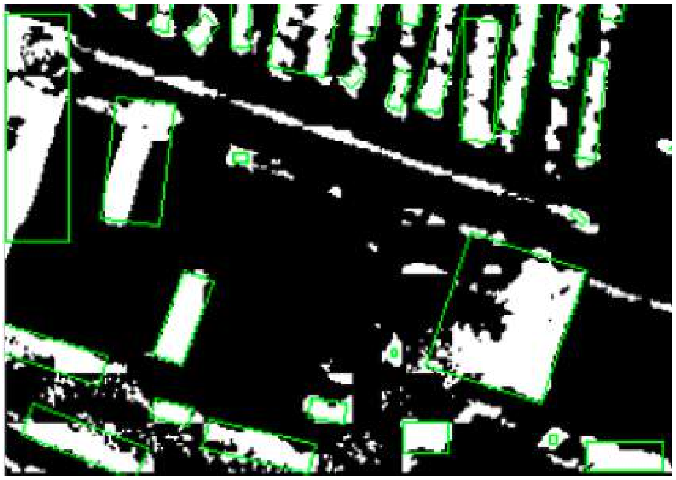


Fig. 7. Result of the image segmentation with the model trained by [14] applied on high-resolution images. Object detected as vine are in green. On the bottom left of the image, numerous other objects are detected as vine.

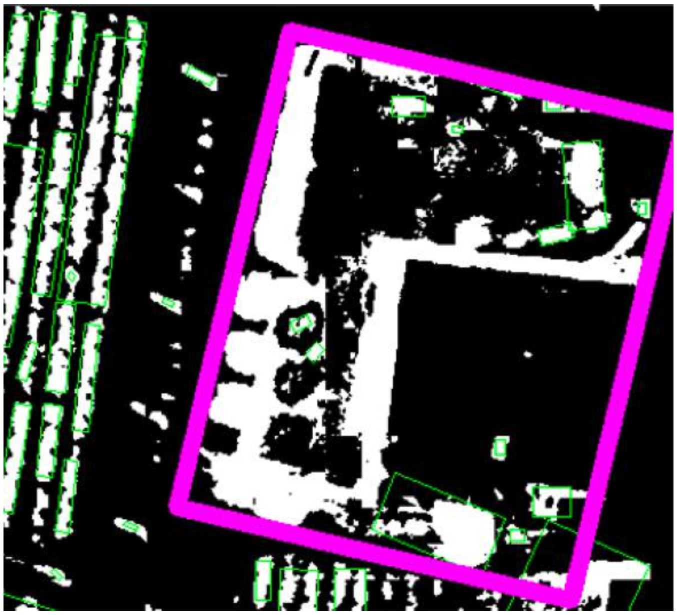


Fig. 8. Result of the image segmentation with the model trained by [14] applied on high-resolution images. Objects detected as vine are in green. Object originally segmented as vine and corrected by the RF is in purple.

TABLE II
This table compares the performances we obtained during the experiments. The models are evaluated using images from the *swiss test set*. The performances are presented with the precision, recall and f-score.

	Precision $\pm SE$	**Recall**	**F-score** $\pm SE$
Experiment 1	0.30 ±0.03	0.93	0.45 ±0.03
Experiment 2	0.41 ±0.03	0.93	0.57 ±0.03
Experiment 3	0.42 ±0.03	0.95	0.58 ±0.04
Experiment 4	**0.53** ±0.04	**0.95**	**0.68** ±0.03

VII. Conclusion

The detection of a complex plantation structure crossing an entire image is a difficult task for ML models. Existing models are sensitive to the variations of vine line orientation and color, and to the variations of image resolution. With our previous experiments [14], [15] we saw numerous examples of wrong classification of objects with similar shape as vine (e.g. bushes) or similar color as vine (e.g. trees).

In this paper, we present our original approach to detect vine lines using two steps machine leaning models. We first segment an image using a CNN and we then confirm or reject each detection with a classifier. We delimit vine lines from the segmented images by delimiting the minimum rectangle area of the connected components. These bounding boxes are used to determine if a detection matches a ground truth and to extract features from the original image to feed the classifier.

With our research, we confirm the ability of our asymmetrical architecture of the U-Net to be generalized to detect vine lines in high-resolution satellite images. We evaluated models trained using images from the *drone train set* with images from the *swiss test set* (see Table I). We significantly improved the f-score of the asymmetrical U-Net, compared to a U-Net, by around 28%. By retraining the Random Forest (RF) with images form the *swiss train set*, we further improved the f-score by around 17% to obtain a f-score of 68% (see Table II).

We note that the Random Forest acts as a filter by removing false positives, such as other agricultural objects. Indeed, the RF does not segment the image but confirms or rejects the decision of the image segmentation model. For each segment, features are used to define its class.

Our original approach could be generalized to other fields of study or other types of agriculture, such as tree detection in crops. There are many other practical use cases to evaluate the generalization capabilities of our models.

References

[1] M. De Clercq, A. Vats, and A. Biel, "Agriculture 4.0: The future of farming technology," *Proceedings of the World Government Summit, Dubai, UAE*, pp. 11–13, 2018.

[2] G. Sylvester, *E-agriculture in action: Drones for agriculture*. Food and Agriculture Organization ofn the United Nations and International . . . , 2018.

[3] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer vision, graphics, and image processing*, vol. 29, no. 1, pp. 100–132, 1985.

[4] C. Anderson, "Growing use of drones poised to transform agriculture," Online, Mar. 2014. [Online]. Available: https://eu.usatoday.com/story/money/business/2014/03/23/drones-agriculture-growth/6665561/

[5] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big data in smart farming–a review," *Agricultural systems*, vol. 153, pp. 69–80, 2017.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[9] R. Abdelfattah, X. Wang, and S. Wang, "Ttpla: An aerial-image dataset for detection and segmentation of transmission towers and power lines," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[10] J. Llorens, E. Gil, J. Llop, and M. Queraltó, "Georeferenced lidar 3d vine plantation map generation," *Sensors*, vol. 11, no. 6, pp. 6237–6256, 2011.

[11] M. Kerkech, A. Hafiane, and R. Canals, "Vine disease detection in uav multispectral images using optimized image registration and deep learning segmentation approach," *Computers and Electronics in Agriculture*, vol. 174, p. 105446, 2020.

[12] M. Kerkech, A. Hafiane, R. Canals, and F. Ros, "Vine disease detection by deep learning method combined with 3d depth information," in *International Conference on Image and Signal Processing*. Springer, 2020, pp. 82–90.

[13] M. Kerkech, A. Hafiane, and R. Canals, "Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in uav images," *Computers and electronics in agriculture*, vol. 155, pp. 237–243, 2018.

[14] J. Treboux, D. Genoud, and R. Ingold, "Improved line detection in images using neural networks and dte subclassifiers," in *2021 9th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2021, pp. 1–6.

[15] T. Jérôme, I. Rolf, and G. Dominique, "Improved and generalized vine line detection on aerial images using asymmetrical neural networks and ml subclassifiers," 2021.

[16] T. Wolf, V. Sanh, J. Chaumond, and C. Delangue, "Transfertransfo: A transfer learning approach for neural network based conversational agents," *arXiv preprint arXiv:1901.08149*, 2019.

[17] X. Zhao, Y. Yuan, M. Song, Y. Ding, F. Lin, D. Liang, and D. Zhang, "Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging," *Sensors*, vol. 19, no. 18, p. 3859, 2019.

[18] A. Soni, R. Koner, and V. G. K. Villuri, "M-unet: Modified u-net segmentation framework with satellite imagery," in *Proceedings of the Global AI Congress 2019*. Springer, 2020, pp. 47–59.

[19] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[20] M. Sozzi, A. Kayad, F. Marinello, J. Taylor, and B. Tisseyre, "Comparing vineyard imagery acquired from sentinel-2 and unmanned aerial vehicle (uav) platform," *Oeno One*, vol. 54, no. 2, pp. 189–197, 2020.

[21] L. Comba, P. Gay, J. Primicerio, and D. R. Aimonino, "Vineyard detection from unmanned aerial systems images," *computers and Electronics in Agriculture*, vol. 114, pp. 78–87, 2015.

[22] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowledge-Based Systems*, vol. 80, pp. 14–23, 2015.

[23] R. Sathya, A. Abraham *et al.*, "Comparison of supervised and unsupervised learning algorithms for pattern classification," *International Journal of Advanced Research in Artificial Intelligence*, vol. 2, no. 2, pp. 34–38, 2013.

[24] D. A. Van Dyk and X.-L. Meng, "The art of data augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.

[25] B. Özdemir, S. Aksoy, S. Eckert, M. Pesaresi, and D. Ehrlich, "Performance measures for object detection evaluation," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1128–1137, 2010.

[26] B. Ozenne, F. Subtil, and D. Maucort-Boulch, "The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases," *Journal of clinical epidemiology*, vol. 68, no. 8, pp. 855–859, 2015.

[27] E. Trentin, "Rejection and the equal error rate: Principles and a case study," 1995.

[28] J. Cook and V. Ramadas, "When to consult precision-recall curves," *The Stata Journal*, vol. 20, no. 1, pp. 131–148, 2020.

[29] N. Aggarwal and R. Agrawal, "First and second order statistics features for classification of magnetic resonance brain images," 2012.