

Improved and Generalized Vine Line Detection on Aerial Images Using Asymmetrical Neural Networks and ML Subclassifiers

Jérôme Treboux¹ and Rolf Ingold² and Dominique Genoud¹

1- University of Applied Sciences and Arts Western Switzerland (HES-SO)
Sierre, Switzerland

2- University of Fribourg - Department of Informatics
Fribourg, Switzerland

Abstract. It is widely accepted that deep neural networks are very efficient for detecting objects in images. They reach their limit when detecting multiple instances of long lines in low-resolution images. We present an original methodology for the recognition of vine lines in low-resolution satellite images. The method consists in combining an asymmetrical neural network with a sub-classifier. We first compare a traditional U-Net architecture with an asymmetrical U-Net architecture designed for precision agriculture. We then highlight the significant improvement in vine line detection when a Random Forest is added after the customized U-Net. This methodology addresses the complex task of dissociating vine lines from other agricultural objects. As a result, our experiments improve the precision from 0.83 to 0.94 over our optimized neural network.

1 Introduction

Machine Learning (ML) is widely used for image recognition. Deep Learning (DL) algorithms such as AlexNet Neural Network [1], Convolutional Neural Networks (CNN) [2], VGG-16 Neural Networks [3], DeepLabv3 [4] and U-Net [5] have demonstrated their performance in the classification and segmentation of objects in images.

The algorithms trained on huge datasets allow accurate recognition of objects in images. Many research groups focus on the detection of small objects in high resolution images such as [6]. However, the performances of the algorithms are limited when the segmentation task concerns the detection of multiple instances of the same object crossing a complete image, as for example the segmentation of vine lines in high resolution images taken with a drone. This challenge is amplified when dealing with low resolution satellite images because of the narrow spacing between the vine lines.

Moreover, in precision agriculture, the distinction between objects is complicated. Colors are similar (e.g. trees, leaves) and shapes are alike (e.g. bush lines and vine lines). The trained model must not be sensitive to the orientation of the vine lines.

This paper compares the accuracy of a U-Net, often used for object segmentation in images, and an asymmetrical U-Net combined with Random Forest (RF) to extract additional information to refine the classification. Our previous

experiments have shown that traditional algorithms and simple neural networks do not achieve sufficient precision [7]. The experiments are based on a problematic of precision agriculture: the detection of vine lines in low-resolution aerial images.

2 State-of-the-Art

Deep neural networks such as ResNet [8] or Inception-v4 [9] are highly efficient for image detection and segmentation. They have respectively a top-5 error of 5.7% and 4.2% for the classification and detection of objects in the images of the ImageNet dataset [8][9].

The U-Net neural network is used for image segmentation. It is widely used in the biomedical field for cell detection in medical images for example [10]. The FCN-8 is partly composed of the U-Net and allows a segmentation of buildings, cars and nature in high-resolution aerial images with an efficiency of 80% [11].

To address the problem of line detection, multiple approaches are proposed: 1) For problems such as detecting the edges of objects (e.g. documents) in images, Hough Transformation can be very efficient [12]. 2) In more complex cases, such as the detection of power lines in images, various neural networks are used to determine the areas containing lines in the image [13]. However, line segmentation at pixel level is very often missing, which does not allow a precise analysis of the line itself. Other researches on the detection of vine lines have interesting results but are dependent and limited by the colors, as for example in [14].

Even at the state of the art, the precision of vine line recognition with neural networks was around 70% [7]. So, to refine the classification, we added a post-processing layer based on random forest that are used for classification but also for the detection of objects in images, such as for face recognition [15]. They are frequently used for the extraction of specific parameters as well as for the selection of the most important ones. They allow to reduce the dimensions by selecting the most important parameters and converge more quickly and with less information than neural networks [16].

3 Dataset and Data Processing

The dataset is created with two different data sources. The first dataset is created with a drone flying over the alpine vineyards of Switzerland. The drone is programmed to take a high-resolution RGB photo every 5 meters. Each image contains meta-information such as the altitude, the speed and the geolocation of the drone. The second dataset is created with low-resolution satellite images acquired in the same region.

Each image is manually labeled with a mask containing white lines. These lines represent the vine lines. The images are divided into training, validation and test datasets. Each image is divided into patches [7]. The corresponding labelings are also divided into patches. Finally, to increase the data presented

to the NN, a data augmentation is performed on the training dataset. The specifications for the data augmentation are detailed by [7].

The patch size for our experiments is 145x145 pixels for the original image. The labeling is divided into patches of 72x72 pixels. These sizes are optimal for the problem of vine line detection and defined during our recent experiments [7].

4 Methodology and Experiments

To determine the optimal segmentation and classification algorithms for our problem, the following methodology is applied during the experiments: first, the datasets required for training, validation and testing are prepared. The original images and annotations are divided into **patches**. The original U-Net [5] and thereafter the asymmetrical U-Net proposed in this research are trained and validated. Then, to improve the detection of vine lines, a random forest is implemented after the asymmetrical U-Net. Finally, the results are calculated and they are statistically compared on the test images.

4.1 Labelling using an Asymetrical U-Net

As a baseline, the U-Net neural network is implemented with its original structure.

For a more precise recognition of vine lines in low-resolution satellite map images, the U-Net architecture is adapted to create an **asymetrical structure**. It analyses **patches** of an optimal size required for the experiment. The Figure 1 illustrates this difference with a larger input image size than the output. This asymetrical network take in account the surrounding for the segmentation.

The network is configured and trained to detect two classes: **vines and non-vines** (See the Figure 1).



(a) Asymmetrical U-Net Architecture (b) Original patch (left) and labelled patch (right).

Fig. 1: U-Net Architecture and example of patches used for our experiments

4.2 Classification Improvement using a Random Forest (RF)

To reduce pixel detection errors around vine lines, the edge detection algorithm proposed by [17] is applied. The algorithm determines the surroundness rela-

tions among the borders of a binary image. The output of the algorithm is one rectangle per vine line, containing the detected pixels as a vine. Each rectangle is then filtered based on its dimensions to eliminate false detections.

To improve the detection of the vine lines and to capture the longitudinal shapes, we propose to use a Random Forest. The output image of the asymmetrical U-Net is used as input for the RF. Each of the rectangles of the image is labeled as vine / non-vine. Then, each rectangle is passed to the RF to train the detection model. The settings used are the dimensions of the rectangle and the characteristics Haralick, Tamura and First Order Statistics [7]. The output of the RF provides a final detection and classification of the vine lines.

Finally, the classification validation is done on images containing lines of vines with varied orientations. It's executed according to the process described in the Section 4, from the original image, through asymmetrical U-Net, edge detection, filtering and finally the RF.

5 Results

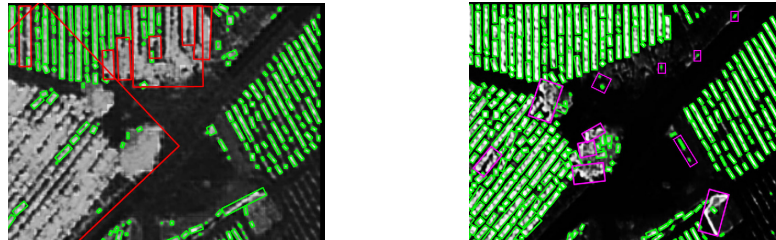
The results of the classification of U-Net, asymmetrical U-Net and the impact of the RF on the classification are compared in this chapter. The models are evaluated with the precision, the recall and the IoU (Intersection over Union). The IoU is used to determine the accuracy of a recognition based on its location [18]. We defined a threshold of 0.75 for IoU in our experiments to calculate precision and recall. This threshold (IoU) is the optimum value obtained in previous detection experiments that were not yet using Random Forest [7]. The results of the experiments are presented in the Table 1.

Our results include the standard error (SE) calculated with the equation $SE = Z_{\alpha} \sqrt{\frac{p(1-p)}{n}}$ [7], where Z_{α} is the confidence level, p is the precision and n is the number of data.

Table 1: Results obtained on a low-resolution satellite imagery for vine line detection.

	Precision	Recall
Altered U-Net	0.711 \pm 0.06	0.239
Asymmetrical U-Net	0.830 \pm 0.05	0.784
Asymmetrical U-Net with RF subclassifier	0.942 \pm 0.03	0.784

The results presented in the Table 1 show that our asymmetrical U-Net has a significantly better precision compared to a U-Net with a standard structure [5]. This improvement is confirmed on the low-resolution satellite map imagery. The use of an asymmetrical neural network allows to improve the precision from 0.711 to 0.830 compared to a standard U-Net for a vine line segmentation task. The use of a random forest to refine the detection of the lines allows to obtain



(a) U-Net

(b) Asymmetrical U-Net combined with a RF

Fig. 2: Visualization of the results of the classification of the vine lines. Green: True Positive. Red : False Positive. Purple : False Positive corrected with the RF.

a precision of 0.942. The recall remains practically unchanged, since the RF is mainly used to remove **False Positives** (e.g. bushes or trees). The RF allows to extract additional information in order to capture a linear shape.

Figure 2 visualizes the detection of the vine lines. The white areas are the pixels detected as a vine line with our U-Net. The green rectangles point out the vine lines. The red rectangles highlight the bad detection. The purple rectangles identify the classifications corrected by the random forest. These visualizations allow to highlight the improvements made with an asymmetrical U-Net but also with the addition of a RF for the final decision.

6 Conclusion

The use of a Random Forest allows to improve the final decision making of a neural network and improve significantly the classification of vine lines in low-resolution satellite images. The combination of an asymmetrical U-Net and a RF improves the precision from 0.830 to 0.942. This combination makes the model robust and generalized. It reduces false detections and therefore reduces the noise generated by the bad detection of agricultural objects. The integration of a Random Forest after a neural network allows to extract additional information and quickly converge on a more accurate segmentation and classification of vine lines. Moreover, the use of a RF allows to adapt the classification problem without having to re-train the neural networks from scratch, reducing considerably the training time [7]. It also enables to quickly filter specific objects

The next steps consist of a complete detection of the vine lines, allowing to omit potential false detections present in the middle of a vine line. This detection will allow to generate automatic flight plans for drones to treat the vines autonomously. We also plan to explore other algorithms similar to RF in order to cross-validate our results.

References

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [2] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449, 2017.
- [3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [4] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [6] Kang Tong, Yiquan Wu, and Fei Zhou. Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing*, 97:103910, 2020.
- [7] Jérôme Treboux, Dominique Genoud, and Rolf Ingold. Improved line detection in images using neural networks and dte subclassifiers. In *2021 9th European Workshop on Visual Information Processing (EUVIP)*, pages 1–6. IEEE, 2021.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [9] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.
- [10] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1):67–70, 2019.
- [11] Jamie Sherrah. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv preprint arXiv:1606.02585*, 2016.
- [12] Bilal Iqbal, Waheed Iqbal, Nazar Khan, Arif Mahmood, and Abdelkarim Erradi. Canny edge detection and hough transform for high resolution video streams using hadoop and spark. *Cluster Computing*, 23(1):397–408, 2020.
- [13] Rabab Abdelfattah, Xiaofeng Wang, and Song 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.
- [14] Mohamed Kerkech, Adel Hafiane, and Raphael Canals. Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in uav images. *Computers and electronics in agriculture*, 155:237–243, 2018.
- [15] Emir Kremic and Abdulhamit Subasi. Performance of random forest and svm in face recognition. *Int. Arab J. Inf. Technol.*, 13(2):287–293, 2016.
- [16] Jung Hwan Cho and Pradeep U Kurup. Decision tree approach for classification and dimensionality reduction of electronic nose data. *Sensors and Actuators B: Chemical*, 160(1):542–548, 2011.
- [17] Satoshi Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing*, 30(1):32–46, 1985.
- [18] Hamid Reza Tofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.