Data augmentation for pathogen segmentation in vinewood fluorescence microscopy images

Julie Munsch[1,2], Sonia Ouali[1], Jean-Baptiste Courbot[1], Romain Pierron[3] and Olivier Haeberlé[1]

[1]IRIMAS, UR 7499, Université de Haute-Alsace, Mulhouse, France
[2]Eiffage Énergie Systèmes – Clemessy, Mulhouse, France
[3]LVBE, UR 3991, Université de Haute-Alsace, Colmar, France

Abstract: In this paper, we address the problem of segmentation of pathogens within fluorescence microscopy images. To our knowledge, the quantification from such images is an original problem. As a consequence, there is no available database to rely upon in order to use supervised machine learning techniques. In this paper, we provide a workaround by creating realistic images containing the desired filamentary pattern and variable blur effect. Numerical results show the interest of this data augmentation technique, especially on images corresponding to a difficult segmentation.

***Keywords*:** data augmentation, image segmentation, fluorescence microscopy deep learning, machine learning

Grapevine trunk diseases are a significant global issue for vinegrowers, causing 13% of vineyards in France to be unproductive, resulting in annual losses of about 1 billion euros [1]. Esca is one of the oldest known grapevine diseases, causing wood deterioration or dieback. The behavior of its pathogens is poorly understood, and no cure exist. Understanding the colonization process is essential for developing treatments. Inoculation experiments, observed via fluorescence microscopy (see Fig. 1), help study pathogen behavior, and image segmentation techniques can quantify pathogen presence in grapevine wood.

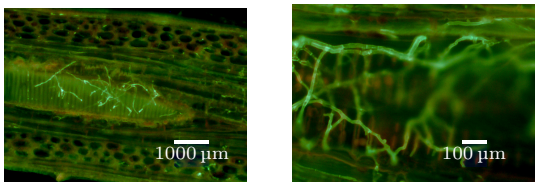

Figure 1: Examples of pathogen in vinewood images observed with a fluorescence microscope, at x10 magnification on the left and x40 on the right. The pathogen fluorescence appears as green-yellow filaments, while wood auto-fluorescence appears as yellow-brown.

Creating a database of microscopic grapevine images with expert segmentation is a time-intensive process, resulting in a small dataset. This limited size can hinder the use of robust supervised learning methods, making data augmentation a crucial step. In general, data augmentation techniques are a helpful tool to enlarge databases. This is in particular the case for medical images, for which the patient set is often limited [2]. When handling images, basic augmentation relies on some simple transform set, such as rotation, flip, and dilation. These techniques might be improved using advanced deep learning approaches, such as adversarial training or neural style transfer (see [3] for a review). Those are mostly based on the image textures, and do not incorporate a model for the image formation process. We can make a similar observation regarding segmentation in microscopy images, as highlighted in [4]. Hence, there is a lack for a data augmentation technique that specifically accounts for the image formation model. This is particularly striking when handling the case of images of fungi colonizing grapevine's wood, which can be affected by a varying blur depending on the region of the image. In this work, we propose to generate synthetic images that mimic real fluorescent microscopy grapevine images in order to train supervised algorithms (see Fig. 2).

In this paper, we propose a data augmentation technique dedicated to fluorescence microscopy images. Then, we use three datasets for training : (**A**) Synthetic images (see Fig.2), (**B1**) Real images together with expert-labeled ground truth and (**C**) Mixed synthetic images and real lower/higher quality images. We use three datasets for testing : (**B1**) Real lower quality images, (**B**) Mixed real lower/higher/fungus-free quality images and (**B3**) Real higher quality images. To segment our images, we use two models, the random forest algorithm [5] and an implemented convolutional network U-Net [6].

Prior to using the segmentation methods, we extract relevant features as a pre-processing. The latter were identified in a preliminary random forest-based study using the Ilastik software [7]. The

features were sorted along their Gini importance [8], and the 14 first features were retained (see Table 1). In addition, preliminary experiments have shown that including classical image filters (namely Gabor, Sobel, Roberts, Scharr, and Prewitt filters) also improves the segmentation. Then, for any 3-channel RGB input image, the pre-processed version contains 57 channels.

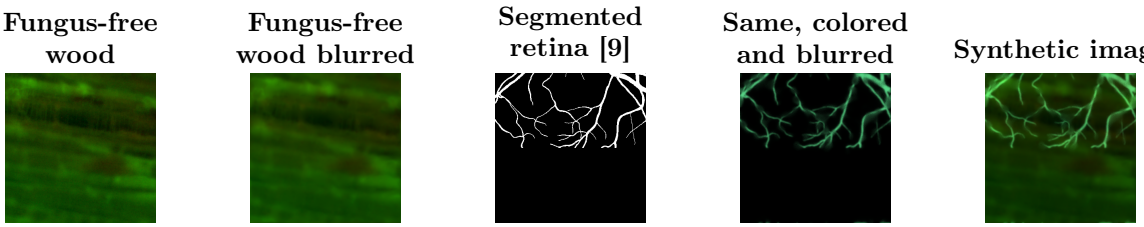

Figure 2: Illustration of the synthetic image formation process

Identity	Hessian	Gaussian $\sigma = \{3; 5; 7\}$	Gradient Magnitude $\sigma = 2$
Difference of Gaussian (σ_i, σ_j) $= \{(1, 3.5); (1, 12); (1, 30)\}$	Laplacian of Gaussian $\sigma = \{0.5; 1.6; 3\}$	Eigenvalues of Tensor Structure $\sigma = 0.7$	Hessian of Gaussian Eigenvalues $\sigma = 3.5$

Table 1: Features retained based on their Gini importance.

So, we test how their use impacts segmentation results, depending on the choice of the training and testing databases. The obtained results for the cleaner images show that supervised segmentation is feasible even on a small database, *i.e.* without augmentation. For lower-quality images, the addition of the synthetic images was indeed helpful, leading to noticeable accuracy improvements with an average 1.57% improvement of accuracy compared with the dataset train real lower quality images (see Fig. 3).

This work stems a few perspectives on the topic, as it could be generalized to other segmentation problems in fluorescence microscopy, as well as other imaging techniques.

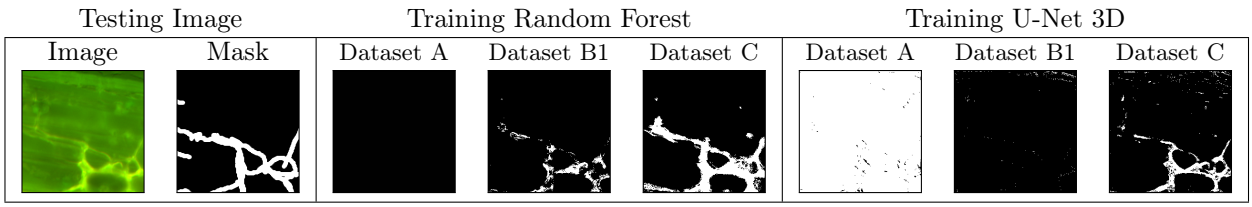

Figure 3: Example of results on real lower quality image.

References

[1] Giovanni Del Frari et al. "Characterization of the wood mycobiome of Vitis vinifera in a vineyard affected by esca. Spatial distribution of fungal communities and their putative relation with leaf symptoms". In: *Frontiers in Plant Science* 10 (2019), p. 910.

[2] Phillip Chlap et al. "A review of medical image data augmentation techniques for deep learning applications". In: *Journal of Medical Imaging and Radiation Oncology* 65.5 (2021), pp. 545–563.

[3] Connor Shorten and Taghi M Khoshgoftaar. "A survey on image data augmentation for deep learning". In: *Journal of big data* 6.1 (2019), pp. 1–48.

[4] Jingchao Ma et al. "Review of Image Augmentation Used in Deep Learning-Based Material Microscopic Image Segmentation". In: *Applied Sciences* 13.11 (2023), p. 6478.

[5] Leo Breiman. "Random forests". In: *Machine learning* 45 (2001), pp. 5–32.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.

[7] Stuart Berg et al. "Ilastik: interactive machine learning for (bio) image analysis". In: *Nature methods* 16.12 (2019), pp. 1226–1232.

[8] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[9] Grand Challenge. *DRIVE: Digital Retinal Images for Vessel Extraction*. Consulté le 21 janvier^{er}, 2024. URL: https://drive.grand-challenge.org/.