Workshop Proceedings of the 19th International Conference on Intelligent Environments (IE2023)
G. Bekaroo et al. (Eds.)

doi:10.3233/AISE230025

Deep Learning for Post-Harvest Grape Diseases Detection

Lucas Mohimont [a,1]
[a] *Université de Reims Champagne-Ardenne, LICIIS, 51687 Reims Cedex 2, France*

Abstract. Post-harvest fruit grading is a necessary step to avoid disease related loss in quality. This is relevant in the context of the Champagne industry where grapes can not be manipulated by machines to avoid crushing. Our team have been developing a computer vision based solution to automate this process. In this paper, our main contribution is the usage of a PSPnet segmentation model for real-time visible symptoms detection with a IoU score of 58%. The associated classification score reach 95%, which improved our previous work. We also study a MobileNet-V2 model's ability to discriminate between different grape diseases in ideal condition.

Keywords. classification, segmentation, deep learning, fruit grading, grape disease

1. Introduction

The impact of diseases is one of the main problems in agriculture. Diseases are directly responsible for yield and quality loss in many cultivars. The Food and Agriculture Organization of the United Nations estimates that plant diseases and invasive insects are responsible for 20% to 40% worldwide yield loss [1]. Disease detection is therefore an important problem in the field of Smart Agriculture. Recent advances in computer vision allow for in-field symptom detection of plant disease. The development of new models is made easier with deep convolutional neural networks that can directly process images. Automatic plant disease detection in field conditions is difficult because there are many challenging factors such as background, natural lighting, symptoms variations, plant phenological stage, etc. Post-harvest detection is therefore still applied during a grading process. Fruits and vegetable grading also includes the detection of other defects such as bruises or lack of maturity.

In this paper, our research is focused on post-harvest grape grading before the pressing stage. Winemakers use preventive pesticide spraying to limit disease proliferation. However, infected grapes can still be harvested if needed. For this reason, a grading process is applied before the pressing to sort the grapes into quality categories. This is necessary to limit the quality loss induced by infected grapes in the press. Our goal is the automation of this process with computer vision. Our approach is based on deep learning with semantic segmentation models for automatic grape disease detection. This paper is based on our previous work [2]. In this paper, we are focused on the detection of grape

[1] Corresponding Author: Lucas Mohimont E-mail: lucas.mohimont@univ-reims.fr

diseases with two models: (1) a semantic segmentation model for disease detection and (2) a model to classify the type of diseases (gray mold, acid rot, powdery mildew). The paper is organized as follows: related works are presented in Section 2, the image acquisition protocol, the datasets, and the proposed methods are presented in Section 3; then the classification and segmentation results are presented in Section 4, before presenting our conclusions and future works.

2. Related Works

Automated vine disease detection is a dynamic sub-field of AI research in agriculture because it would allow for large-scale detection of diseases on the field or after harvest. This work is mainly performed by human operators, whose detection of grape diseases is often biased [3]. In comparison, computer vision over-performs humans in a single grape leaf disease classification task [4], all while ensuring an unbiased evaluation.

Plant disease detection from RGB images is a difficult task. Many difficulties have been described in detail in the work of Barbedo [5]. Building a representative dataset is a complex task because there are many sources of variations (phenological stages, disease severity, weather and lighting conditions, etc.). This also leads to heavily unbalanced classes because the symptoms represent a small part of the images.

Practical in-field disease detection is a more difficult task for multiple reasons. The images contain more complex scenes, whole vines instead of a single leaf for example, with natural background and lighting. Multiple deep learning-based models have been proposed recently. Popular object detection architectures, Faster R-CNN [6], R-FCN [7], and SSD [8], were successfully applied to tomato symptoms detection on greenhouse images. A single model reached 83% mean Average Precision on leaf and gray mold, canker, plague, miner, low-temperature symptoms, mildew, whitefly pest, and nutritional excess [9]. This type of model can be optimized for in-field real-time prediction, which would allow for efficient large-scale processing.

Ideally, winemakers want to avoid putting diseased grapes in the pressing batches. In practice, diseased grapes' incidence may be too high and could represent a big proportion of the yield. Therefore, post-harvest disease quantification is necessary for grapes grading: healthy grapes and diseased grapes are separated to keep a good wine quality. There are few published research work about post-harvest grape grading from images. The authors of Vazquez-Fernandez et al. proposed a segmentation algorithm based on Gabor filters and pixel classification from raw pixels neighborhoods for grape batch segmentation (ripen grapes, green grapes, rotten grapes, leaves, etc) [10]. A classification success rate of 94% was obtained with multi-layer perceptron at the pixel level. The proposed method used artificial lighting with a simple industrial background to avoid confusion. It does not take the low probability of diseased grapes into account during the prediction, this can result in many false positives. Another published work proposed an automatic stem and leaves detection method from hyperspectral images [11].

Our two contributions are (1) the binary segmentation of visible symptoms with a PSPNet model and (2) the classification of different types of grape diseases with a MobileNet model. This method would ideally be applied directly in the pressing site without specific calibration and without costly changes in the industrial process.

3. Materials and Methods

The presented work has been carried out on three image collection campaigns from 2019 to 2021. This section presents the developed acquisition protocol used in 2019, 2020, and 2021, the collected datasets, the proposed methods, and the evaluation metrics.

3.1. Image acquisition

An image acquisition protocol was defined in the summer of 2019 to collect images for supervised learning. This protocol was designed to be applied directly to the existing industrial setting without too much interference. Data collection was performed in Vranken-Pommery Monopole's wine presses during the weighting stage. The existing industrial process is the following: (1) pallets containing one to four levels of four grape crates are delivered on the press site; (2) each pallet is moved to the weighting site to collect relevant information (weight, variety, parcels); (3) a quality grade is attributed to the whole pallet after a quick visual inspection by an employee (potential alcoholic content is also estimated during this stage); (4) the pallet is then moved to the docks before pressing.

The quality grade depends on multiple factors like the presence of diseases (powdery and downy mildew, gray mold, acid rot), leaves, soil, etc. It is assumed that the visible area on the upper level of the pallet is representative of the whole pallet. The visual inspection is subjective to the human employee who can be biased by fatigue, visual impairment, etc. Automating this inspection is the goal of the project. GoPro Hero Black 7 cameras, with 4000x3000 resolution, were attached over the weighting sites to obtain images of the visible area of the pallets. The camera's WiFi connection is used with a computer to associate the images with the properties of the pallets (quality grade, variety, weight, alcoholic content). The installation was performed by our team and the data collection was performed by Vranken-Pommery's employees.

Data collection campaigns were carried on in two wine presses in 2019 and 2020 in Tours-sur-Marne and Merrey-sur-Arce in France. A third site in Saudoy was added to the campaign in 2021. The industrial process was not changed. The two first sites are in semi-open areas and can be affected by the sunlight and multiple pallets can be visible in the images. Artificial lighting was only used during the 2021 campaign in Tours-sur-Marne. The resulting datasets were cleaned, few images were removed (redundant takes or dark images taken near sun fall). Each visible crate of the weighted pallet is labeled for binary segmentation with LabelMe polygonal mask [12]. Most images contain 4 crates. Therefore, the segmentation allows us to process each crate separately to detect the disease. Each crate was extracted from the images and infected grapes were labeled for binary segmentation. For classification, images are split into two classes: class 0 for the healthy crates and class 1 for the crates that contain infected grapes (gray mold, powdery mildew, acid rot). Other diseases such as downy mildew and dry grapes were not labeled.

3.2. Methods

3.2.1. Diseases Segmentation

Image segmentation is the grouping of image pixels into a small number of categories. This is similar to classification and clustering in the pixel space. Nowadays, neural network architecture based on CNN were adapted to semantic segmentation. The segmentation is therefore not based on subjective criteria, such as thresholds. The segmentation is semantic because it is able to understand, to some extent, the content of the scene. This allows for more complex image segmentation tasks. Classifiers can be used for segmentation by classifying every pixel with a sliding window. This is however inefficient (too much redundancy) and can lead to limited results (the input is too small). A better method was introduced with the Fully Convolutional Network (FCN) [13]. FCN replaces the dense layer with convolutional and up-sampling layers. In this manner, a dense prediction can be achieved for the whole input image. Many semantic segmentation models have been proposed since FCN. Those include popular models such as U-Net [14], with a symmetrical architecture, or lightweight models such as CGNet [15].

In our context, we have selected the PSPNet [16] model with an Inception-Resnet-V2 backbone because it gives good results in many of our datasets while having a reasonable size (about 3M parameters). It was trained on the 3300 images of healthy and diseased crates collected in 2019, 2020, and 2021. Data augmentation was used to reduce over-fitting. Our aim is to assess the semantic segmentation performances for grape disease detection (the first step toward disease quantification) and to understand if the model is able to differentiate the different types of diseases. The Umap algorithm [17] is used for this purpose because it could be useful for automatic disease types labeling. The evaluation was carried on with 5-fold cross-validation with the Intersection over Union segmentation score and classification metrics at the pixel level (F-score, precision, recall). We also evaluate the model's ability for image-level classification (either healthy or infected).

3.2.2. Diseases Classification

The images from the three campaigns (2019,2020 and 2021) enabled us to create a large dataset with 3300 images of crates in the "Healthy" class and 1600 in the "Disease" class (about 32.6%). These classes are based on the actually visible symptoms after image masks were produced with the LabelMe tool. However, this is still limited to binary classification/segmentation. This is insufficient to satisfy winemakers' needs because the quality loss depends on the type of disease affecting the grapes. In our previous work, a pre-trained MobileNet-V2 [18] model reached 92% F-score for binary classification [2]. In the current work, we used a similar model with the three following classes: gray mold, acid rot, and powdery mildew. Those labels were based on the quality grades given at the press during the weighting. We excluded healthy crates to evaluate the potential of grape disease discrimination in ideal conditions. The Umap algorithm was then used to visualize the activations of the neural network.

4. Results

4.1. Diseases Segmentation

This section presents our first segmentation models for post-harvest grape disease segmentation at the press. Every figure shown in this subsection was generated with the model of the first Cross-Validation split.

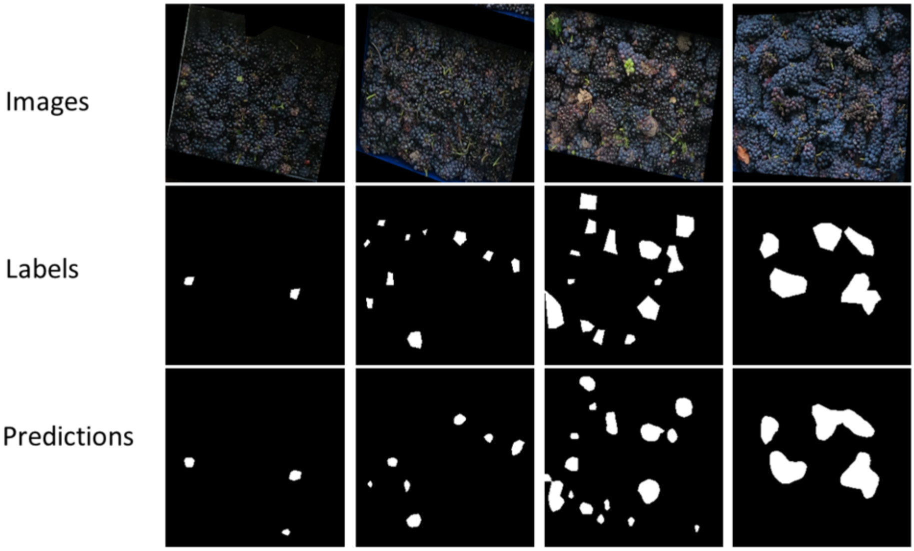

Figure 1. Examples of image segmentation with different levels of infected grape size.

Table 1. Disease segmentation performances

Split	IoU	Recall	Precision	F-score
Training	0.61	0.75	0.76	0.76
Validation	0.58	0.73	0.74	0.73

The segmentation model reached an average of 0.58 IoU with recall and precision rates of 0.73 and 0.74 on validation sets (Table 1. Data augmentation was successful in reducing the gap between training and validation performances, with only a 2-3% difference. The gap without data augmentation was bigger with a 10% difference on IoU. The two main sources of errors are the imprecision of the labeling and the small apparent size of the symptoms. Most images with symptoms that represent more than 20% of the grapes area were artificial crates taken during the 2020 campaign. Those images contain bigger infected clusters compared to the other images that contain many small infected parts. This is illustrated in Figure 1. In this figure, the first three images show multiple small symptoms, while the last image shows five infected clusters. Therefore, the errors due to label imprecision are systematically bigger for the "natural" images because there are more difficult and small symptoms to detect. The number of symptoms also makes the labeling process tedious.

Despite this problem, good performances were achieved without post-processing correction. Most false positives in the images of the "Healthy" class were due to labeling errors and small ambiguous areas of the images. False-positive pixels represent only 0.1% of the Healthy images. This can be easily corrected with noise removal with morphological operations. The segmentation models also reached better disease detection compared to the classifiers proposed in the previous section, with a 0.95 F-score compared to the 0.92 score of the pre-trained MobileNet-V2 model from our previous work(Table 2.

Table 2. Binary classification performances

Split	Recall	Precision	F-score
Training	0.96	0.95	0.95
Validation	0.96	0.94	0.95

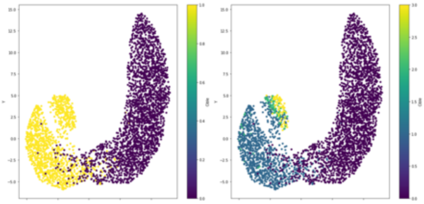

Figure 2. Umap visualization. On the left, the healthy class is purple, and the disease class is yellow. On the right, the disease class is divided into three sub-classes: blue for diseases within the 0-10% range, green for the 10-20%, and yellow for those above 20%.

Umap was applied to the activations of the convolutional layer that follows the Pyramid Pooling Module (similar results were obtained with activations from different layers of the model). We show in Figure 2 the two same images with different labels: (1) purple for the healthy class and yellow for the infected grapes and (2) the disease class is divided into three sub-classes of different symptoms sizes. We can see that the two classes are not perfectly separated. This is due to two reasons: (1) labeling imprecision and errors and (2) confusion between diseases that were not labeled (such as downy mildew or dry grapes). In the second picture, we can see that images with the highest symptom sizes are clustered together and well separated from the other images. This corresponds to the artificial images created in 2020.

To conclude this section, the proposed segmentation model was applied in real-time at the press during the 2022 harvest. The average inference time on a Jetson Nano 2GB board was about 3.5s for 1080p images with Onnx-runtime optimization.

4.2. *Diseases Classification*

The classification model was trained to differentiate grapes infected by gray mold from grapes infected by acid rot and powdery mildew. This is a biased preliminary result because most of our images of grapes infected with gray mold were taken during the original industrial process. Some of the other images were the artificial crates taken during the 2020 campaign. This was necessary because harvest quality was really good in 2020, there were almost no infected grapes arriving at the press. For this reason, the model easily reached over 99% F-score for the three classes.

This bias is perfectly illustrated in Figure 3. Every image of gray mold is in the same cluster in yellow while acid rot(in purple) and powdery mildew(in cyan) are in much smaller clusters. The collection of new images is necessary to obtain an unbiased disease discrimination model. This is difficult because gray mold's occurrence is way higher compared to others diseases.

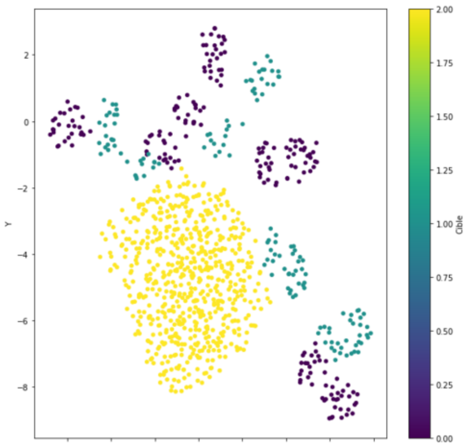

Figure 3. Umap visualization of the gray mold/other disease classification model. The gray mold class is represented in yellow, the acid rot in purple, and the powdery mildew in cyan.

5. Conclusion and Future Works

The proposed PSPNet segmentation model reach a score of 58% IoU and it was first deployed in a prototype during the 2022's campaign. A real-time inference of 3.5s/image was obtained and will be further improved with quantization and architecture optimization for the next season. Our next goal is the discrimination of different types of diseases. Our classification shows ideal results but it is heavily biased because of the artificial crates. Ideally, we want to automate the labeling process to reduce human labor and human mistakes. Future works include therefore the study of generative and clustering models for this purpose.

Acknowledgements

This work has been conducted funded by the "Edge AI Technologies for Optimised Performance Embedded Processing" project, received funding from KDT JU under grant agreement No 101097300. The KDT JU receives support from the European Union's Horizon Europe research and innovation program and Austria, Belgium, France, Greece, Italy, Latvia, Luxembourg, Netherlands, Norway.

We want to thank Vranken-Pommery Monopole, our partner in the EdgeAI project, for allowing image collection in their sites. We also thank the ROMEO Computing Center[2] of Université de Reims Champagne-Ardenne, whose Nvidia DGX-1 server allowed us to accelerate the training steps and compare several model approaches.

References

[1] Secretariat I, Gullino M, Albajes R, Al-Jboory I, Angelotti F, Chakraborty S, et al. Scientific review of the impact of climate change on plant pests. FAO on behalf of the IPPC Secretariat; 2021. Available from: DOI:10.4060/cb4769en.

[2] https://romeo.univ-reims.fr

[2] Mohimont L, Alin F, Gaveau N, Steffenel LA. Lite CNN Models for Real-Time Post-Harvest Grape Disease Detection. In: Workshops at 18th International Conference on Intelligent Environments (IE2022). IOS Press; 2022. p. 116-25.

[3] Hill GN, Evans KJ, Beresford RM, Dambergs RG. Comparison of methods for the quantification of botrytis bunch rot in white wine grapes. Australian Journal of Grape and Wine Research. 2014;20(3):432-41. Available from: DOI:10.1111/ajgw.12101.

[4] Cruz A, Ampatzidis Y, Pierro R, Materazzi A, Panattoni A, De Bellis L, et al. Detection of grapevine yellows symptoms in Vitis vinifera L. with artificial intelligence. Computers and Electronics in Agriculture. 2019;157:63-76. Available from: DOI:10.1016/j.compag.2018.12.028.

[5] Barbedo J. Factors influencing the use of deep learning for plant disease recognition. Biosystems Engineering. 2018 08;172. Available from: DOI:10.1016/j.biosystemseng.2018.05.013.

[6] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks; 2016.

[7] Dai J, Li Y, He K, Sun J. R-FCN: Object Detection via Region-Based Fully Convolutional Networks. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16. Red Hook, NY, USA: Curran Associates Inc.; 2016. p. 379–387.

[8] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: Single Shot MultiBox Detector. Lecture Notes in Computer Science. 2016:21–37. Available from: DOI:10.1007/978-3-319-46448-0_2.

[9] Fuentes A, Yoon S, Kim SC, Park DS. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. Sensors. 2017;17(9). Available from: https://www.mdpi.com/1424-8220/17/9/2022.

[10] Vazquez-Fernandez E, Dacal-Nieto A, Martin F, Formella A, Torres-Guijarro S, Gonzalez-Jorge H. A computer vision system for visual grape grading in wine cellars. In: International Conference on Computer Vision Systems. Springer; 2009. p. 335-44.

[11] Portalés C, Ribes-Gómez E. An image-based system to preliminary assess the quality of grape harvest batches on arrival at the winery. Computers in Industry. 2015;68:105-15. Available from: DOI:10.1016/j.compind.2014.12.010.

[12] Wada K. labelme: Image Polygonal Annotation with Python; 2016. https://github.com/wkentaro/labelme.

[13] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society; 2015. p. 3431-40. Available from: https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298965.

[14] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234-41.

[15] Wu T, Tang S, Zhang R, Cao J, Zhang Y. Cgnet: A light-weight context guided network for semantic segmentation. IEEE Transactions on Image Processing. 2020;30:1169-79.

[16] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. arXiv:161201105 [cs]. 2016 Dec. ArXiv: 1612.01105. Available from: http://arxiv.org/abs/1612.01105.

[17] McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv; 2018. Available from: https://arxiv.org/abs/1802.03426.

[18] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: IEEE Confe. on Computer Vision and Pattern Recognition (CVPR); 2018. .