Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 196 (2022) 101–106

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

Grapevine Segmentation in RGB Images using Deep Learning

Gabriel A. Carneiro[a], Rafaela Magalhães[a], Alexandre Neto[a], Joaquim J. Sousa[a,b,*] and António Cunha[a,b]

[a]*University of Trás-Os-Montes e Alto Douro, Vila Real, 5000-801, Portugal*
[b]*INESC-TEC – INESC Technology and Science, Porto 4200-465, Portugal*

Abstract

Wine is the most important product from the Douro Region, in Portugal. Ampelographs are disappearing, and farmers need new solutions to identify grapevine varieties to ensure high-quality standards. The development of methodology capable of automatically identify grapevine are in need. In the scenario, deep learning based methods are emerging as the state-of-art in grapevines classification tasks. In previous work, we verify the deep learning models would benefit from focus classification patches in leaves images areas. Deep learning segmentation methods can be used to find grapevine leaves areas.
This paper presents a methodology to segment grapevines images automatically based on the U-net model. A private dataset was used, composed of 733 grapevines images frames extracted from 236 videos collected in a natural environment. The trained model obtained a Dice of 95.6% and an Intersection over Union of 91.6%, results that fully satisfy the need of localise grapevine leaves.

Peer-review under responsibility of the scientific committee of the CENTERIS –International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

Keywords: segmentation; grapevines; deep learning

* Corresponding author.
E-mail address: jjsousa@utad.pt

Peer-review under responsibility of the scientific committee of the CENTERIS –International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021
10.1016/j.procs.2021.11.078

1. Introduction

Wine is the most important product from the Douro Region, in Portugal being the port wine internationally known. To ensure high-quality standards, the wines produced in this region are carefully controlled. The grape variety utilized is one of the most relevant factors in the wine production chain, directly influencing the product's authenticity and classification [1]. To maintain wines' quality, uniqueness and exclusivity in the Douro Demarcated Region (DDR), only specific varieties are authorized and, thus, identifying them is crucial for control activities and quality assurance, as well as for regulating production.

One accurate approach to identify grape varieties is Ampelography, a task based in visual analysis. However, like any visual task, it depends on who is doing it, inserting subjectiveness into the process. Yet, it can be exposed to interference from environmental, cultural and genetic conditions, which may introduce uncertainty into the identification process [2,3].

Deep Learning (DL) is now the state-of-the-art in most image classification tasks, representing the greatest research's field in this area, achieving comparable or even better results than humans. In the literature, some works using DL can be found for the task of plant identification [4], including few published in the grapevine context [5–7]. In Adão et al. [7] a deep learning model was used for classifying six grapevine varieties from DDR based in RGB images of a single leaf photographed against white background, achieving ~100% of accuracy. In some tests carried out by our group, we show that it is possible to classify 12 grapevine varieties utilizing RGB images acquired in a natural environment, achieving ~0.9 in F1 score. Furthermore, we presented explanations about predictions utilizing the LIME framework, as a result, we can verify that sometimes the model uses the background to make decisions about the classification. The background does not have any information useful to the grapevine classification task and, to ensure confidence in the model results, this should be avoided.

Segmentation is an effective way to separate the background from the grapevine area. With the rise of DL segmentation frameworks like U-Net [8], SegNet [9] and Mask R-CNN [10], a bigger improvement in the results was achieved. Many works that use DL for leaf segmentation can be found in the literature. Ward et al. [11] used Mask R-CNN to segment leaf instances using real and synthetic images of Arabidopsis plants. Shi et al. [12] used the Mask R-CNN to do instance and semantic segmentation of 2D and 3D plant images. In the grapevines context, Zabawa et al. [13] count berries with a traditional U-shaped decoder-encoder architecture, utilizing as a backbone the MobileNetV2, focusing on creating a lightweight framework, while Santos et al. [14] apply the Mask-RCNN to brunch segmentation and tracking in videos. Note that both works remain in fruit context, discarding the leaf information.

The development of methodology capable of automatically identify grapevine are in need. In the scenario, deep learning based methods are emerging as the state-of-art in grapevines classification tasks. In previous work, we verify the deep learning models would benefit from focus classification patches in leaves images areas. Deep learning segmentation methods can be used to find grapevine leaves areas.

In this paper, we present a methodology to do grapevine leaf semantic segmentation. The main idea is to avoid background interferences in deep learning classification models in development in our group.

2. Methods

In this section, we will describe our methods to segment grapevine leaf areas, following the pipeline shown in Figure 1.

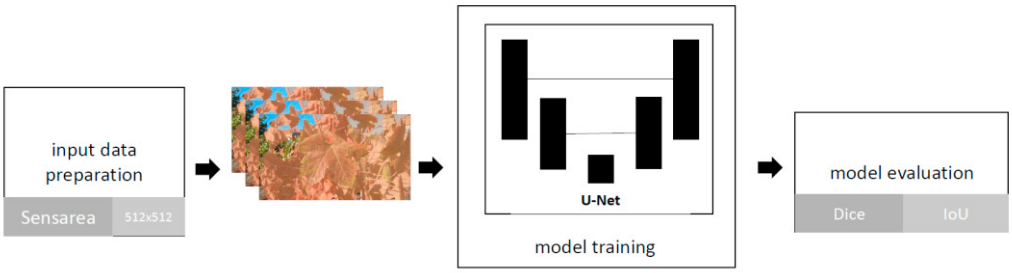

Figure 1. Methodology Pipeline.

First, we annotate input data frames using the Sensarea tool [15] and prepare the dataset for model training. Then, we trained the U-Net model with transfer-learning approach. Finally, evaluate the model and conclude the defined goal.

2.1. Dataset

The dataset was created from 236 videos recorded weekly in the research farm of the University of Trás-os-Montes and Alto Douro (UTAD) between May and September of 2017. Two smartphones are utilized, resulting in movies recorded with size 1080 x 1920 at 25 FPS. The camera was positioned between 20 and 40 centimetres far from the plant, moving closer and far while recording. The duration of the movies is irregular, varying between 16 and 6 seconds, so movies recorded closer to September have a smaller duration compared with others recorded in May.

We used the Sensarea [15] tool to annotate the full videos. The Sensarea allows semi-automatic tracking of a region in a video, reducing the time needed to create the dataset, but the software has some limitations. The polygons generated by Sensarea cannot contain holes in the middle, and automatic tracking often does not work correctly, requiring manual correction. The idea was to annotate polygons that contain grapevines leaves, creating masks with two classes: ‘grapevine’ and ‘background’. Note that the selected polygons do not need to be highly accurate as our objective is to identify areas that contain grapevine leaves roughly. In Fig. 2, we can see a mask example.

Figure 2: Example of mask obtained from Sensarea’s semi-automatic tracking used to train the U-net model. The shadow area represents region with grapevine leaves.

After the annotation, we create the dataset with 733 images that were manually chosen from the videos and divided into training, validation and test datasets, with 513 (70%), 110 (15%) and 110 (15%) images, respectively. This way,

we avoided blurred frames and the existence of several frames with just one class in the datasets. Furthermore, the images were resized to 512×512 pixels.

The dataset was composed of images from 5 DDR allowed grapevines varieties: Codega, Moscatel Galego, Rabigato, Tinta Roriz, and Tinto Cao, but the sample quantity per grapevines class varies.

2.2. Model Architecture and Training

Considering the good overall results presented in the literature for segmentation tasks, we decided to use a U-Net [8] based model with an Inception-ResNet v2 encoder [16] to segment the grapevine leaves. This is composed of two phases: the encoder and the decoder phases. The first phase encodes the input image into lower dimension feature spaces utilizing convolutions and max-pooling downsampling. The second phase consists of the feature maps upsampling and concatenation, which enables the localization of more precise features, outputting a segmented image with the same size as the input image.

The Inception-ResNet v2 encoder was initialized with weights pre-trained on ImageNet [17]. Firstly, the encoder layers were frozen, and only the decoder layers were trained on our training set (70% of our dataset) during 50 epochs, using the Adam optimizer with a learning rate of $10^{-3}$and a batch size of 4. After this, the model was unfrozen and trained for 150 epochs. If the Dice value did not improve after 2 epochs — once learning stagnates — the learning rate was reduced by a factor of 0.95. This is done to prevent the model from stagnating at a minimum.

2.3. Loss and Metrics

We used the binary cross-entropy loss to train the model, combining the regular binary cross-entropy (BCE) loss with Sigmoid activation.

The metrics used to evaluate the performance of the model were the Dice coefficient and the Intersection over Union (IoU). The Dice coefficient measures the similarity between the predicted and ground truth segmentations. So, a bigger Dice coefficient indicates higher similarity. The Dice coefficient is calculated using Equation 1.

$$Dice = \frac{2*True\ Positive}{2*True\ Positive+False\ Positive+False\ Negative} \quad (1)$$

As for the IoU, the area of overlap is the area between the predicted segmentation and the ground truth divided by the area of their union. The higher the IoU coefficient, the more perfect the overlap segmentation is. We calculate the IoU using Equation 2.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (2)$$

3. Results and Discussion

At the end of testing, the Dice and IoU coefficient values on our test set were 95.6% and 91.6%, respectively, which allowed the identification of grapevine areas in the images. Note that, as we said before, the dataset has rough segmentations.

For the results analysis, we considered six situations that we found to be interesting to see how the model behaves. The first situation is when the image has a lot of background, including another grapevine (see Figure 3 (a)). In this type of images, the model performs well in separating both classes. The model's performance is still high in the inverse situation, where images have more leaves than background, Figure 3 (b). In the cases where leaves and background areas are even, the model behaves very well, too, Figure 3 (c). The fourth case is when the image has one portion of background inside of the leaf region, Figure 3 (d). Here the model does not catch the background portion entirely inside the grapevine leaves area. This is because when we used Sensarea to simplify the annotation process, we used just one layer, so if the background is inside the grapevine region, we kept it inside the polygon. This means that the small background area inside the grapevine area in the ground truth is considered as grapevine class, thus the small

portions of background predicted by the model contributed to down the IoU and Dice metrics for this sample. The fifth case is when there are two regions of the main grapevine in the same image. This can be seen in Figure 3 (e). In order to simplify the annotation process, when we found this kind of situation, we connected both polygons making one. Even so, the model could predict more than one region of a specific class and was able to accurately separate the background of the two grape regions without considering other grapevines in the background. This way, better results than annotation were obtained since the model learned well the differences between leaves and background.

As we extracted frames from videos to create this dataset, there are blurred frames in our dataset. In this situation, the model's performance depends on how much the image is blurred. There are cases, like Figure 3 (f), where the performance is acceptable, but the metrics are down. The explanation for that lies in the fact that when Sensarea handles with blurred images, the generated masks can be error-prone.

Considering the model's performance, it can be concluded that it is possible to separate grapevine leaves and background in frames of videos. The trained model can be used to improve the results of differ

ent tasks, like the classification of grapevine varieties and diseases.

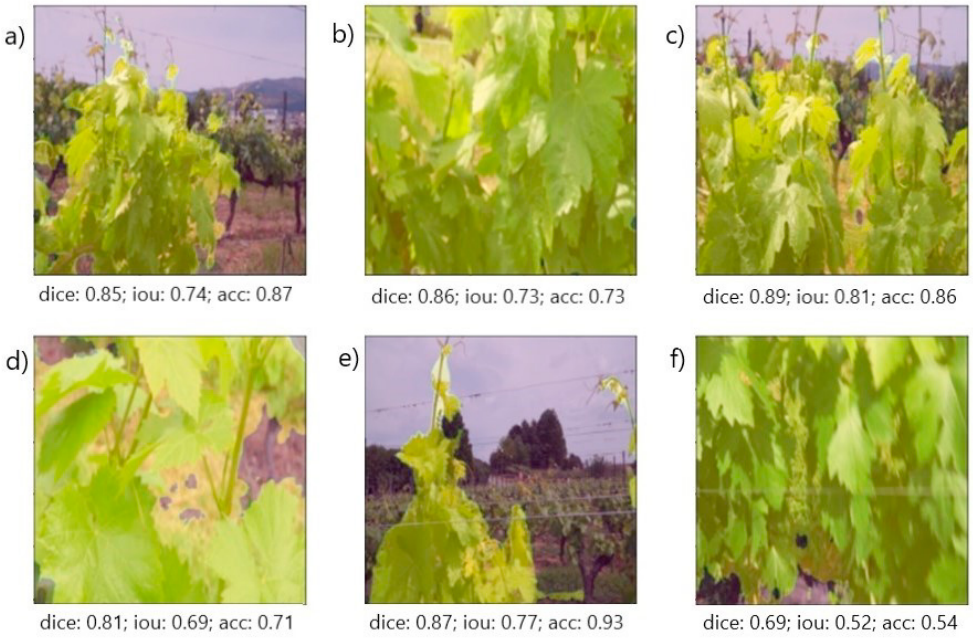

Figure 3. Illustrative example segmentation results. The yellow shadow areas represent the predicted grapevine class. Below each sample the metrics Dice (dice), Intersection over Union (Iou) and Accuracy are presented.

4. Conclusions

In this paper, we presented an approach to segment grapevine leaves using a U-Net deep learning model. To train the model, we use a manually built 733 image dataset extracted from 236 videos taken in vineyards. The trained model obtained a Dice of 95.6% and an Intersection over Union of 91.6%, results that fully satisfy the need of localising grapevine leaves.

To improve and make our results more consistent, in future work, we plan to increase the dataset with more frames and use patches instead of the entire image. Our hypothesis is that with this approach, the model will learn how to segment grapevine at different crop scale. To turn our results more comparable, we also plan to train a Mask R-CNN model as well, verifying the results for both models.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020.

References

[1] Moncayo S, Rosales JD, Izquierdo-Hornillos R, Anzano J, Caceres JO. Classification of red wine based on its protected designation of origin (PDO) using Laser-induced Breakdown Spectroscopy (LIBS). Talanta 2016;158:185–91. https://doi.org/https://doi.org/10.1016/j.talanta.2016.05.059.

[2] Garcia-Muñoz S, Muñoz-Organero G, de Andrés M, Cabello F. Ampelography - An old technique with future uses: the case of minor varieties of Vitis vinifera L. from the Balearic Islands. J Int Des Sci La Vigne Du Vin 2011;45:125–37. https://doi.org/10.20870/oeno-one.2011.45.3.1497.

[3] Tassie L. Vine identification--knowing what you have. Grape Wine Res Dev Corp Â€“Australian Gov GW RDC Innov Network, Greenhill Road Wayv 2010.

[4] Sobha P, Thomas P. Deep Learning for Plant Species Classification Survey, 2019, p. 1–6. https://doi.org/10.1109/ICAC347590.2019.9036796.

[5] Milella A, Marani R, Petitti A, Reina G. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. Comput Electron Agric 2019;156:293–306. https://doi.org/https://doi.org/10.1016/j.compag.2018.11.026.

[6] Fernandes AM, Utkin AB, Eiras-Dias J, Cunha J, Silvestre J, Melo-Pinto P. Grapevine variety identification using "Big Data" collected with miniaturized spectrometer combined with support vector machines and convolutional neural networks. Comput Electron Agric 2019;163:104855. https://doi.org/https://doi.org/10.1016/j.compag.2019.104855.

[7] Adão T, Pinho TM, Ferreira A, Sousa A, Pádua L, Sousa J, et al. Digital Ampelographer: A CNN Based Preliminary Approach. In: Moura Oliveira P, Novais P, Reis LP, editors. Prog. Artif. Intell., Cham: Springer International Publishing; 2019, p. 258–71.

[8] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR 2015;abs/1505.04597.

[9] Badrinarayanan V, Kendall A, Cipolla R. SegNet: {A} Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. CoRR 2015;abs/1511.00561.

[10] He K, Gkioxari G, Dollár P, Girshick RB. Mask {R-CNN}. CoRR 2017;abs/1703.06870.

[11] Ward D, Moghadam P, Hudson N. Deep Leaf Segmentation Using Synthetic Data. CoRR 2018;abs/1807.10931.

[12] Shi W, van de Zedde R, Jiang H, Kootstra G. Plant-part segmentation using deep learning and multi-view vision. Biosyst Eng 2019;187:81–95. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2019.08.014.

[13] Zabawa L, Kicherer A, Klingbeil L, Töpfer R, Kuhlmann H, Roscher R. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. ISPRS J Photogramm Remote Sens 2020;164:73–83. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.04.002.

[14] Santos TT, de Souza LL, dos Santos AA, Avila S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. Comput Electron Agric 2020;170:105247. https://doi.org/https://doi.org/10.1016/j.compag.2020.105247.

[15] Bertolino P. Sensarea: an Authoring Tool to Create Accurate Clickable Videos, 2012. https://doi.org/10.1109/CBMI.2012.6269804.

[16] Szegedy C, Ioffe S, Vanhoucke V. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. CoRR 2016;abs/1602.07261.

[17] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. 2009 IEEE Conf. Comput. Vis. pattern Recognit., 2009, p. 248–55.