Feature Pyramid Network based Proximal Vine Canopy Segmentation ⋆

Szilárd Molnár *, **Barna Keresztes** ** and **Levente Tamás** *

* *Technical University of Cluj-Napoca, 400114 Cluj-Napoca, Romania*
e-mail: {Szilard.Molnar, Levente.Tamas}@aut.utcluj.ro
** *Université de Bordeaux, France*
e-mail: barna.keresztes@ims-bordeaux.fr

Abstract: With the widening of the Agriculture 4.0 era, the use of autonomous robots in the agriculture field is becoming a priority. The key component of such an autonomous, often multi-robot system is the perception of the environment, which is based on 2D and 3D cameras. A base processing part of the 2D images is the segmentation of different zones in the images. This is the case also in the vineyards where in order to process complex plant canopies, segmenting the parts of the image containing the area of interest is a part of the pre-processing chain. In this work, we present a Feature Pyramid Network-based grape canopy segmentation method, which has great potential to create a segmentation mask, containing only the leaves and fruits of interest. We conducted our tests in different vineyards and we also obtained the above state-of-the-art segmentation results on public and custom datasets.

Keywords: control in agriculture, robotics, precision viticulture, and computer vision.

1. INTRODUCTION

The paradigm shift in the era of precision agriculture toward the use of autonomous robots in the field yields a challenging task for the field of robotics research: the well-known algorithms in a controlled environment must be adapted to a wide variety of agronomic fields. This is valid also in viticulture, where relevant progress was made toward the employment of autonomous robots.

Viticulture is the science of the growing, cultivation, and harvesting of grapevines (*Vitis vinifera*), and it has a large amount of market share, as the average yearly wine production around the world is around 30 million tonnes (Castriota (2020)). In the last, almost thirty years, a great deal of work has been published around computer vision-assisted vine growing, or *precision viticulture* (Tillett (1993); Nuske et al. (2011); Dey et al. (2012); Seng et al. (2018)).

In precision viticulture, different types of vehicles with a wide variety of sensors are used, Figure 1. While ground vehicles provide proximal, highly accurate data, remote sensing applications, based on unmanned aerial vehicles such as drones or satellites, reveal other important plant health characteristics based on reflectance indices (Jackson et al. (1988); Hall et al. (2002)). Canopy segmentation is present in both cases, yet different challenges arise. In remote sensing, the canopies are relatively easy to differentiate, because of changes in texture and color, while
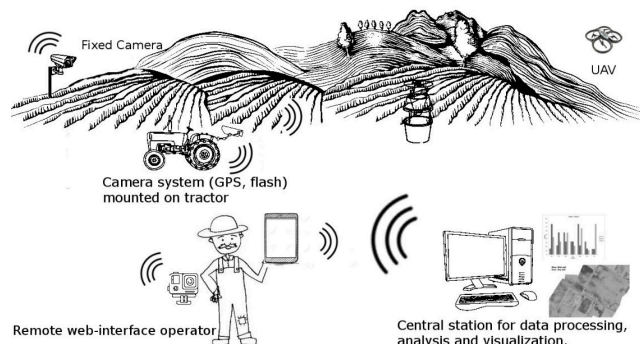

Fig. 1. Proposed viticulture system with ground and aerial robots

in proximal sensing, environmental objects and similar plant canopies provide difficulties.

Plant canopies are extremely complex structures, more so if the various environmental elements are also present, yet in *precision agriculture*, analyzing them is a crucial task. This analysis usually starts with segmenting only the regions of interest, avoiding excess computational resources for separately analyzing the background objects and other unimportant plants. Many computer vision applications build upon segmented canopies, including yield estimation (Ghiani et al. (2021)), orchard row navigation (Aghi et al. (2021)), or disease detection (Gutiérrez et al. (2021)).

Although the trend is to use ground vehicles in proximal sensing, unmanned aerial vehicles in proximal sensing are rare, but not entirely absent (Di Gennaro et al. (2019)) in this field. The main problem is due to the high risk of collision with the plants, but with an adequate navigation system, a drone can provide additional benefits over ground vehicles, such as independence of ground

clearance or inclination, while providing an adjustable viewpoint.

Deep learning-based solutions are promising for canopy segmentation. The canopy contains both larger and smaller leaves of different shapes and colors under a variety of illumination types and viewpoints. Therefore, analyzing the image on different levels of detail can be an advantage. An example of such a model is the *Feature Pyramid Network (FPN)* architecture (Lin et al. (2017)). The use of this model was inspired by the work of Molnár et al. (2021), where surface normals were estimated using different support sizes, similar to that of vine leaves. having different sizes and orientations in the images.

In this paper, we present a closed-proximity image canopy segmentation method in vineyards using the FPN architecture. The intuition for using this type of architecture is its multi-scale observation capability of it. The dataset used for training and validation is based on ground vehicle and drone-mounted image capture. After describing the method, we provide a comparison of our method with another highly-rated segmentation method. We provide the code and data of our method [1].

2. RELATED WORK

Canopy segmentation is required to reduce computational costs of detecting a leaf with a disease (Musci et al. (2020)) or estimating the size of a grape bunch (Ghiani et al. (2021)), others use canopy segmentation for navigation (Aghi et al. (2021)). The most popular canopy segmentation techniques are based on color segmentation (although usually only in remote sensing applications (Su et al. (2016))), or based on 3D data acquired directly from a depth sensor (Aghi et al. (2021); Peng et al. (2022)), a LiDAR (Nehme et al. (2021)), or reconstructing the 3D point cloud from monocular or stereo RGB images (Milella et al. (2019); Jurado et al. (2020)).

Concerning the methods used, different approaches exist depending on the type of data and area of interest. For other agricultural products, such as apple trees, segmentation was performed using ResNet-18 by Zhang et al. (2019). Examples of segmentation for remote sensing applications are *U-NET* (Tamvakis et al. (2022); Carneiro et al. (2021)), *SegNet* (Badrinarayanan et al. (2017)) or *ModSegnet* (Ganaye et al. (2018)). The articles available using these methods in vineyard application are listed in various surveys, for example, in the work of Barros et al. (2022). KNN-based image segmentation is used (Rangel et al. (2016)), similarly to k-means segmentation (Kaur et al. (2018)), although, these methods work with single leaf shots, captured in laboratory environmental conditions. Abdelghafour et al. (2019) and Abdelghafour et al. (2020) estimate the likelihood of the local properties to classify the pixels using *Local Structure Tensor* (Bigun et al. (1991)) and HSV color thresholding (Otsu (1979)). However, the use of high-power flash eradicates the difference between day-time images and night-time images, creating an unfair advantage against methods with less complex setups.

Furthermore, the *MobileNetV3* architecture is used by Aghi et al. (2021) to combine depth information with RGB images. A popular architecture for segmentation or detection is called *Mask R-CNN* (He et al. (2017)), used by Ghiani et al. (2021); Santos et al. (2020), or *Faster R-CNN* (Ren et al. (2017)) used by Ouattara et al. (2020) and Musci et al. (2020), who combined it with *Random Forest* algorithm (Breiman (2001)).

Feature Pyramid Networks (FPNs) (Lin et al. (2017)) are the basis of *Mask R-CNN*, yet vanilla FPNs are rarely used for segmentation tasks. Hence, we consider using the base FPN architecture for our canopy segmentation algorithm.

3. MATERIALS AND METHODS

3.1 Dataset

Our starting point for the creation of the dataset was the work of Aghi et al. (2021). They created the dataset for unmanned ground vehicle navigation, therefore, the viewpoint of their images, captured in northern Italy, is relatively low between the rows, similar to our approach. The dataset contains 500 images (Figure 2a), additionally, a binary mask (Figure 2b) image representing the segmented canopy. The provided binary masks were generated using data from an RGB-D camera 3D scanning.

We considered that this dataset presents vine rows without huge variations; therefore, we created our additional dataset. The acquisition device was a DJI Mini 2 drone equipped with a 4K RGB camera mounted on a gimbal. Using a drone as a platform helped us achieve different viewpoints and viewing angles by adjusting the onboard gimbal camera mount and flying at 1) half height of the vine plant - gimbal 0° (Figure 4a); 2) the top of the vine plant - gimbal 0° (Figure 4b); 3) half a meter higher than the vine plant - gimbal 20° (Figure 4c); 4) 2 meters higher than the vine plant - gimbal 45° (Figure 4d); 5) half height of the vine plant - camera directly oriented towards the grapevine (Figure 4e).

We captured the images from two different vineyards in the propriety of the University of Agricultural Sciences and Veterinary Medicine, located in the Transylvania region of Romania, during September and October of 2022, in different weather conditions. Most of our images are from the vineyard located in Cluj-Napoca (see Figure 3 for a satellite image), which is a 4 ha vineyard mainly for R&D purposes. Another vineyard for dataset creation is located in the village of Apoldu de Sus, Sibiu county, which covers about 65 ha of area. For training our method, since the drone used does not have a depth sensor, the binary masks of our dataset were obtained by manually masking the areas of interest. An example of manual masking can be seen in Figure 5.

We completed the dataset of Aghi et al. (2021) with 100 of our images. For training, we used a 1 to 5 ratio of validation and training images. Furthermore, since the pictures were of variable sizes, we resized all the photos to 640 by 480, which also helps to reduce computational costs.

---

[1] https://github.com/molnarszilard/canopy_segmentation

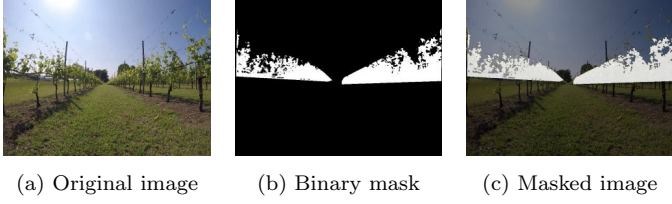(a) Original image   (b) Binary mask   (c) Masked image

Fig. 2. Example of ground robot-based image dataset from northern Italy (Aghi et al. (2021)) for masking vine canopy.

Fig. 3. Satellite view of the main vineyard for our dataset acquisition, located in Cluj-Napoca where we captured our proximal images.
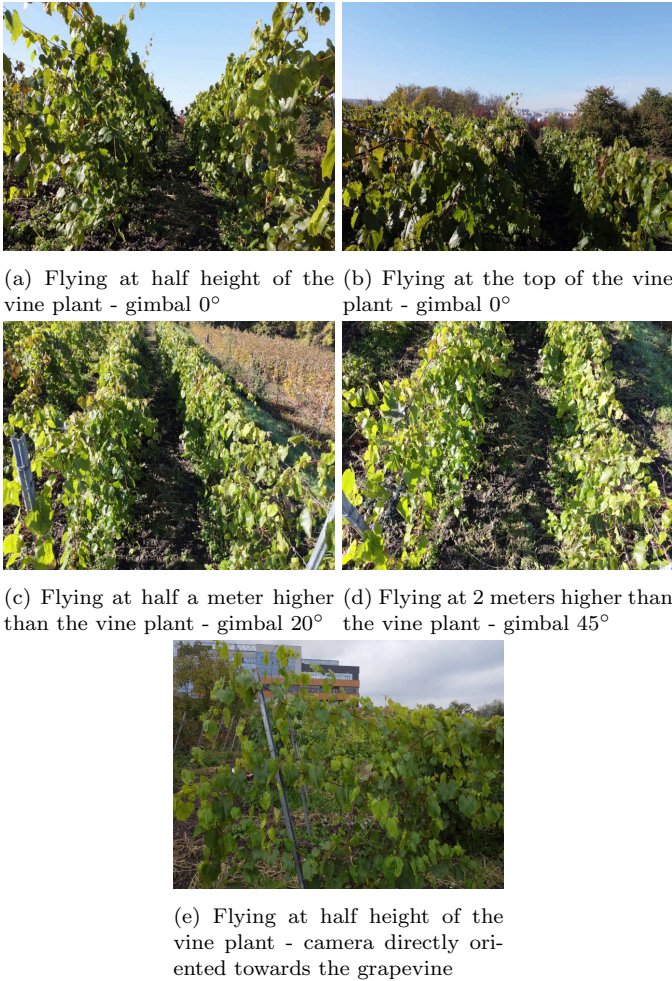

(a) Flying at half height of the vine plant - gimbal 0°

(b) Flying at the top of the vine plant - gimbal 0°

(c) Flying at half a meter higher than the vine plant - gimbal 20°

(d) Flying at 2 meters higher than the vine plant - gimbal 45°

(e) Flying at half height of the vine plant - camera directly oriented towards the grapevine

Fig. 4. Comparing different viewpoints captured with the 4K camera of the DJI Mini 2 drone.

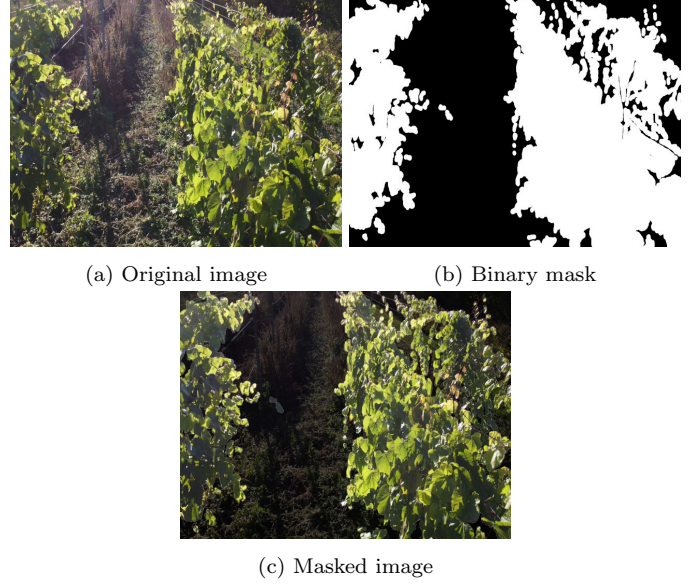(a) Original image   (b) Binary mask

(c) Masked image

Fig. 5. Example of the manual masking of our dataset.

3.2 Feature Pyramid Network

As we mentioned previously, Feature Pyramid Network is used for precision agriculture as the backbone of *Mask R-CNN*, but not by itself. The idea of using FPNs in canopy segmentation came from a previous work Molnár et al. (2021), on 3D data analysis. In this work, an FPN-based normal estimation and smoothing algorithm was considered for different-sized surfaces, similarly, as in the case of a vine canopy different leaves are present. This analogy convinced us to reuse this architecture as the basis of our canopy segmentation. Although the normal estimation and smoothing methods were based on depth images, the adaptation of the RGB images is straightforward. The model uses ResNet-101 as the starting pretrained weights.

The architecture, based on PyTorch, can be seen in Figure 6, receives an RGB image, and outputs a mask. This mask is not binary (containing only black and white pixels), but the output of a sigmoid layer, which is converted to the required binary mask by applying thresholding. As a training goal, our objective is to increase the *Intersection over Union (IoU)* between the ground truth mask and the predicted mask, Equation (1), therefore the loss is $1 - IoU$:

$$IoU = \frac{\mathbf{X}_{GT} \cdot \mathbf{X}_{pred}}{\sum_{i,j} (\mathbf{X}_{GT}^{ij} + \mathbf{X}_{pred}^{ij} - \mathbf{X}_{GT}^{ij} * \mathbf{X}_{pred}^{ij})} \quad (1)$$

where $\mathbf{X}_{GT}$ and $\mathbf{X}_{pred}$ are the ground truth and the prediction masks respectively.

The training of the model was performed using an Nvidia A100 GPU, with 10 epochs and a batch size of 4, applying Adam optimization. The model contains approximately 36 million trainable parameters, therefore, using the convention of multiplying the parameter number by 4 due to the data type, the model size is around 140MB. However, since the relatively large image size, training with a batch size of 4 requires 11GB of memory.

Additionally, we experimented with different model sizes, when we added a level above the current architecture or removed one level, to verify the model performance. In
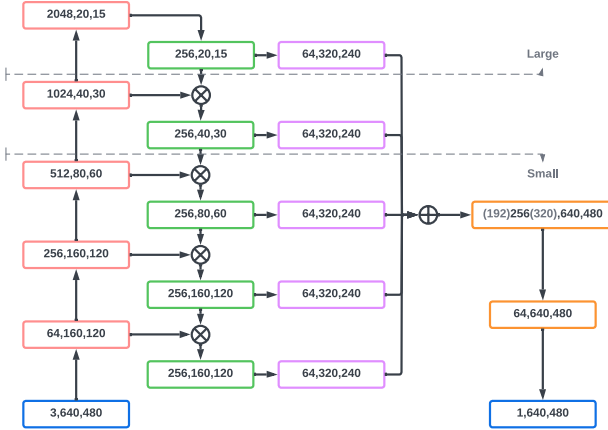
Fig. 6. The FPN architecture in our work without including the batch size, we represent the channel size, the width, and the height of the data. ⊕ represents concatenation, and ⊗ represents element-by-element addition. The drawing does not include other operators, such as convolution or upsampling. We marked the cuts where the sub-architectures differ (small, base, large).

the first case, the model contained 62 million trainable parameters, with 250MB of model size, while in the second case, the model contained 6 million trainable parameters, with 24MB of model size.

4. EVALUATION

For the evaluation, we calculate the precision, which is the relationship between the correct pixels and the total pixels; in addition, we calculate the percentage of false positive (FP) and false negative (FN) pixels for the predicted mask and *IoU*. To check the efficiency, we measure the runtime while the data is processed in the neural network.

We also compare our method with other segmentation methods, the *Mask R-CNN* (He et al. (2017)) and *MobileNetV3*, proposed by Aghi et al. (2021). Initially, *Mask R-CNN* requires that the masks be described as polygons in a JSON file. However, our method works with mask images, and describing this mask using polygons reduces the accuracy. Therefore, we slightly altered the original *Mask R-CNN* code to read mask images, instead of JSON files.

In our experience, this alteration improved the precision of the evaluation by approximately 2%. Previously we mentioned that the size of our model is 140MB, while the size of the *Mask R-CNN* is 290MB. Regarding *MobileNetV3*, its astonishingly small model size of only 4MB is promising for embedded applications.

The results of the evaluation and comparison between our method *Mask R-CNN* and *MobileNetV3* are organized in Table 1. The visual comparison is provided in Figure 7 (best view in color). Here we can observe that the *Mask R-CNN* loses accuracy around the edges of the canopy. Our method follows the curvature of the canopy with a smoother line.

Furthermore, we can see that our method masks out most of the ground vegetation, which is a positive aspect of this method. If we check the differences between our base model, the large version, and the small version, we observe, that the large version does not provide sufficiently higher accuracy, than the base version, while being almost twice the size. Meanwhile, the small version trades accuracy for speed and model size. Despite the fact that the medium version of our method is not the best in any of the categories, we decided that the combination of the results is the best in this case.

Concerning runtime without pre- or postprocessing, our method, with approximately 18 ms per image, is 10 times faster than *Mask R-CNN* and 4 times faster than *MobileNetV3*, despite the lower model size in the latter case. We think that the poorer performance of *MobileNetV3* is due to the small network, which cannot learn the complex nature of the canopy.

Similar behavior can be seen with our smaller model. However, the accuracy for this case can be increased, in our test by at least 5%, by limiting the environment to only one vineyard. Additionally, Aghi et al. (2021) proposed this method, for row detection and navigation, which does not require high accuracy.

Table 1. Comparing the results of our method (with different model sizes), *Mask R-CNN*, and *MobileNetV3* listing the accuracy, the percentage of false positives(FP), and false negatives(FN), the *IoU*, and the runtime.

	Acc[%]	FP[%]	FN[%]	*IoU*[%]	Time[s]
OwnL	**94.7**	3.36	**1.95**	**77.78**	0.022
Own	94.26	3.08	2.66	76.91	0.018
OwnS	92.93	4.3	2.77	73.88	**0.005**
MRCNN	92.71	5.17	2.11	73.16	0.177
MNetV3	87.02	**2.28**	10.7	48.27	0.072

Moreover, we tested the runtime of our method, on different devices from server-grade GPUs, such as Nvidia A100 and Tesla T4 (through Google Colab), to embedded devices, such as Jetson Xavier NX tested in 20W 6 Core mode. To show the advantages of a GPU, we also run the test on three CPUs. The results are shown in Table 2.

Table 2. Comparing the runtime of our method on different devices(in seconds).

Device	Time[s]
Nvidia RTX3080 (10GB)	0.012
Nvidia A100 (40GB)	0.018
Nvidia TeslaT4 (16GB)	0.019
Jetson Xavier NX	0.085
Intel®Core™ i9-10900K	0.811
Intel®Xeon®Gold 6226R	0.934
Intel®Core™ i7-6700K	1.579

For further evaluation, we captured a few images in a challenging darker environment, when the leaves were extremely brown and sparse, for which our method was not trained. These images were captured with our drone, in the same way as the other training images, or with the same setup used by Abdelghafour et al. (2021), using a high-power flash. In Figure 8 we present an example for both cases, presenting the original image alongside the masked

(a) The predicted mask of our method.

(b) FPs and FNs of the predicted mask of our method.

(c) The predicted mask of the *Mask R-CNN* method.

(d) FPs and FNs of the *Mask R-CNN* method.

(e) The predicted mask of the *MobileNetV3* method.
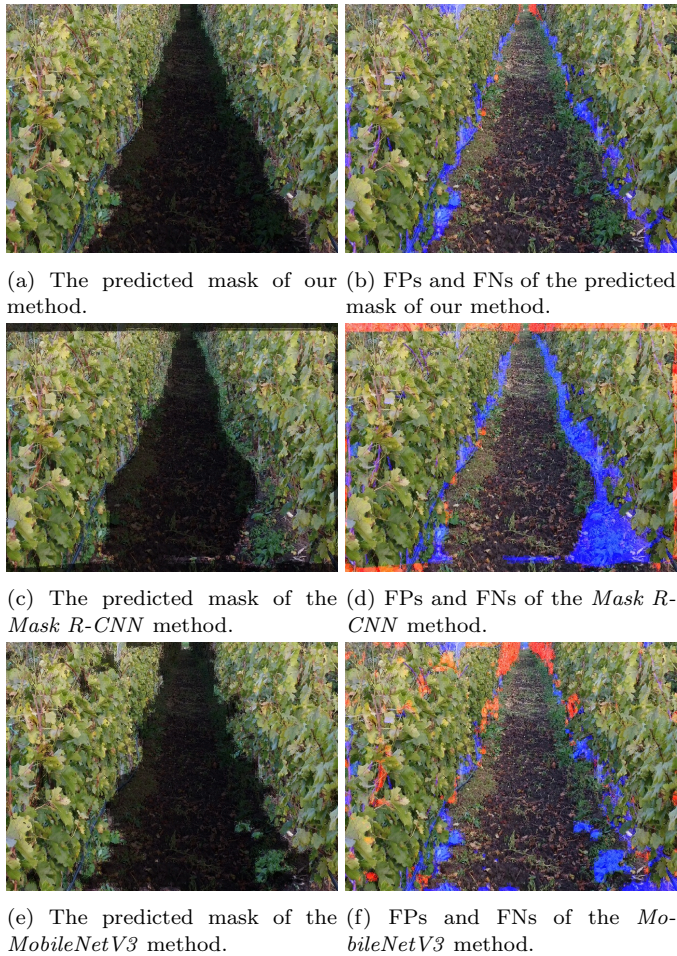
(f) FPs and FNs of the *MobileNetV3* method.

Fig. 7. Comparing the predicted masks of the three methods by masking the canopy, and highlighting the false positive (Blue) and false negative (Red) areas.

(a) RGB image with flash.

(b) The masked image.

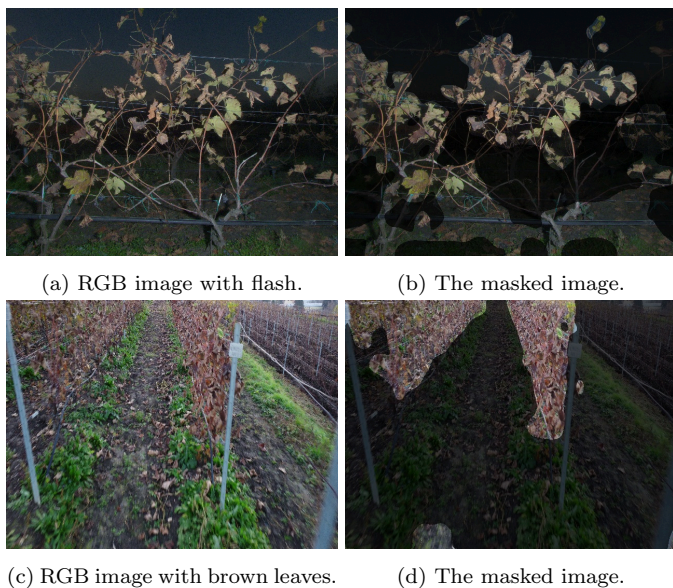(c) RGB image with brown leaves.

(d) The masked image.

Fig. 8. Predicting vine canopy in challenging conditions.

image. As can be observed, the proposed method performs well even in these challenging, unseen environments as well.

5. CONCLUSION

In this work, we have presented the completion of a dataset to mask the vine canopy, proposed a method for canopy segmentation based on the feature pyramid network, and compared our method with other highly rated segmentation and detection networks named *Mask R-CNN* and *MobileNetV3*. We concluded that our basic FPN-based canopy segmentation method outperforms each of them, both in accuracy and in runtime.

For future work, we plan to improve the ground truth masks in our dataset, because now we masked small areas around the leaves, that do not represent the canopy. A possible optimization step would be to include HSV color-space thresholding in the training. Additionally, our method has the potential to be implemented for other plant species.

REFERENCES

Abdelghafour, F., Keresztes, B., Deshayes, A., Germain, C., and Costa, J.P.D. (2021). An annotated image dataset of downy mildew symptoms on Merlot grape variety. *Data in Brief*, 37, 107250.

Abdelghafour, F., Keresztes, B., Germain, C., and Da Costa, J. (2020). In Field Detection of Downy Mildew Symptoms with Proximal Colour Imaging. *Sensors*, 20(16), 4380.

Abdelghafour, F., Rosu, R., Keresztes, B., Germain, C., and Costa, J.D. (2019). A Bayesian framework for joint structure and colour based pixel-wise classification of grapevine proximal images. *Computers and Electronics in Agriculture*, 158, 345–357.

Aghi, D., Cerrato, S., Mazzia, V., and Chiaberge, M. (2021). Deep Semantic Segmentation at the Edge for Autonomous Navigation in Vineyard Rows. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2021, Prague, Czech Republic, September 27 - October 1, 2021*, 3421–3428. IEEE.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495.

Barros, T., Conde, P., Gonçalves, G., Premebida, C., Monteiro, M., Ferreira, C.S.S., and Nunes, U.J. (2022). Multispectral vineyard segmentation: A deep learning comparison study. *Computers and Electronics in Agriculture*, 195, 106782.

Bigun, J., Granlund, G.H., and Wiklund, J. (1991). Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 775–790.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

Carneiro, G.A., Magalhães, R., Neto, A., Sousa, J.J., and Cunha, A. (2021). Grapevine Segmentation in RGB Images using Deep Learning. In M.M. Cruz-Cunha, R. Martinho, R. Rijo, D. Domingos, and E. Peres (eds.), *CENTERIS 2021 - International Conference on ENTERprise Information Systems / ProjMAN 2021 - International Conference on Project MANagement*

/ *HCist 2021 - International Conference on Health and Social Care Information Systems and Technologies 2021, Braga, Portugal*, volume 196 of *Procedia Computer Science*, 101–106. Elsevier.

Castriota, S. (2020). *Wine Economics*. The MIT Press.

Dey, D., Mummert, L., and Sukthankar, R. (2012). Classification of Plant Structures from Uncalibrated Image Sequences. In *IEEE Workshop on Applications of Computer Vision, WACV 2012, Breckenridge, CO, USA, January 9-11, 2012*, 329–336. IEEE Computer Society.

Di Gennaro, S.F., Toscano, P., Cinat, P., Berton, A., and Matese, A. (2019). A Low-Cost and Unsupervised Image Recognition Methodology for Yield Estimation in a Vineyard. *Frontiers in Plant Science*, 10, 559.

Ganaye, P.A., Sdika, M., and Benoit-Cattin, H. (2018). Semi-supervised Learning for Segmentation Under Semantic Constraint. In A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, 595–602. Springer International Publishing.

Ghiani, L., Sassu, A., Palumbo, F., Mercenaro, L., and Gambella, F. (2021). In-Field Automatic Detection of Grape Bunches under a Totally Uncontrolled Environment. *Sensors*, 21(11), 3908.

Gutiérrez, S., Hernández, I., Ceballos, S., Barrio, I., Díez-Navajas, A.M., and Tardaguila, J. (2021). Deep learning for the differentiation of downy mildew and spider mite in grapevine under field conditions. *Computers and Electronics in Agriculture*, 182, 105991.

Hall, A., Lamb, D.W., Holzapfel, B., and Louis, J. (2002). Optical remote sensing applications in viticulture - a review. *Australian journal of grape and wine research*, 8(1), 36–47.

He, K., Gkioxari, G., Dollár, P., and Girshick, R.B. (2017). Mask R-CNN. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, 2980–2988. IEEE Computer Society.

Jackson, R.D., Kustas, W.P., and Choudhury, B.J. (1988). A Reexamination of the Crop Water Stress Index. *Irrigation Science*, 9(4), 309–317.

Jurado, J.M., Pádua, L., Feito, F.R., and Sousa, J.J. (2020). Automatic Grapevine Trunk Detection on UAV-Based Point Cloud. *Remote Sensing*, 12(18), 3043.

Kaur, S., Pandey, S., and Goel, S. (2018). Semi-automatic leaf disease detection and classification system for soybean culture. *IET Image Processing*, 12(6), 1038–1048.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2117–2125.

Milella, A., Marani, R., Petitti, A., and Reina, G. (2019). In-field high throughput grapevine phenotyping with a consumer-grade depth camera. *Computers and Electronics in Agriculture*, 156, 293–306.

Molnár, S., Kelényi, B., and Tamás, L. (2021). Feature Pyramid Network Based Efficient Normal Estimation and Filtering for Time-of-Flight Depth Cameras. *Sensors*, 21(18), 6257.

Musci, M.A., Persello, C., and Lingua, A.M. (2020). UAV Images and Deep-learning Algorithms for Detecting Flavescence Doree Disease in Grapevine Orchards. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 43.

Nehme, H., Aubry, C., Solatges, T., Savatier, X., Rossi, R., and Boutteau, R. (2021). LiDAR-based Structure Tracking for Agricultural Robots: Application to Autonomous Navigation in Vineyards. *Journal of Intelligent & Robotic Systems*, 103(4), 61.

Nuske, S., Achar, S., Bates, T., Narasimhan, S., and Singh, S. (2011). Yield Estimation in Vineyards by Visual Grape Detection. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2011, San Francisco, CA, USA, September 25-30, 2011*, 2352–2358. IEEE.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.

Ouattara, I., Hyyti, H., and Visala, A. (2020). Drone based Mapping and Identification of Young Spruce Stand for Semiautonomous Cleaning. *IFAC-PapersOnLine*, 53(2), 15777–15783.

Peng, C., Fei, Z., and Vougioukas, S.G. (2022). Depth camera based row-end detection and headland manuvering in orchard navigation without GNSS. In *30th Mediterranean Conference on Control and Automation, MED 2022, Vouliagmeni, Greece, June 28 - July 1, 2022*, 538–544. IEEE.

Rangel, B.M.S., Fernández, M.A.A., Murillo, J.C., Ortega, J.C.P., and Arreguín, J.M.R. (2016). KNN-based image segmentation for grapevine potassium deficiency diagnosis. In *2016 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, 48–53. IEEE.

Ren, S., He, K., Girshick, R.B., and Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.

Santos, T.T., de Souza, L.L., dos Santos, A.A., and Avila, S. (2020). Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Computers and Electronics in Agriculture*, 170, 105247.

Seng, K., Ang, L., Schmidtke, L., and Rogiers, S. (2018). Computer Vision and Machine Learning for Viticulture Technology. *IEEE Access*, 6, 67494–67510.

Su, B., Jinru, X., Chunyu, X., Fang, Y., Song, Y., and Fuentes, S. (2016). Digital surface model applied to unmanned aerial vehicle based photogrammetry to assess potential biotic or abiotic effects on grapevine canopies. *International Journal of Agricultural and Biological Engineering*, 9(6), 119–130.

Tamvakis, P.N., Kiourt, C., Solomou, A.D., Ioannakis, G., and Tsirliganis, N.C. (2022). Semantic Image Segmentation with Deep Learning for Vine Leaf Phenotyping. *Computing Research Repository*, abs/2210.13296.

Tillett, N.D. (1993). Robotic manipulators in horticulture: a review. *Journal of Agricultural Engineering Research*, 55(2), 89–105.

Zhang, X., Fu, L., Karkee, M., D. Whiting, M., and Zhang, Q. (2019). Canopy Segmentation Using ResNet for Mechanical Harvesting of Apples. *IFAC-PapersOnLine*, 52(30), 300–305.