

# Grapevine Plant Image Dataset for Pruning

Kyriakos D. Apostolidis, Theofanis Kalampokas, Theodore P. Pachidis \*  and Vassilis G. Kaburlasos 

Human-Machines Interaction Laboratory (HUMAIN-Lab), Department of Computer Science, International Hellenic University, Agios Loukas, 65404 Kavala, Greece

\* Correspondence: pated@cs.ihu.gr; Tel.: +30-2510-462-281

**Abstract:** Grapevine pruning is conducted during winter, and it is a very important and expensive task for wine producers managing their vineyard. During grapevine pruning every year, the past year's canes should be removed and should provide the possibility for new canes to grow and produce grapes. It is a difficult procedure, and it is not yet fully automated. However, some attempts have been made by the research community. Based on the literature, grapevine pruning automation is approximated with the help of computer vision and image processing methods. Despite the attempts that have been made to automate grapevine pruning, the task remains hard for the abovementioned domains. The reason for this is that several challenges such as cane overlapping or complex backgrounds appear. Additionally, there is no public image dataset for this problem which makes it difficult for the research community to approach it. Motivated by the above facts, an image dataset is proposed for grapevine canes' segmentation for a pruning task. An experimental analysis is also conducted in the proposed dataset, achieving a 67% IoU and 78% F1 score in grapevine cane semantic segmentation with the U-net model.

**Dataset:** <https://github.com/humain-lab/Buds-Dataset>.

**Dataset License:** Creative Commons Attribution 4.0 International.

**Keywords:** computer vision; image processing; grapevine pruning



**Citation:** Apostolidis, K.D.; Kalampokas, T.; Pachidis, T.P.; Kaburlasos, V.G. Grapevine Plant Image Dataset for Pruning. *Data* **2022**, *7*, 110. <https://doi.org/10.3390/data7080110>

Academic Editor: Joaquín Torres-Sospedra

Received: 12 June 2022

Accepted: 5 August 2022

Published: 9 August 2022

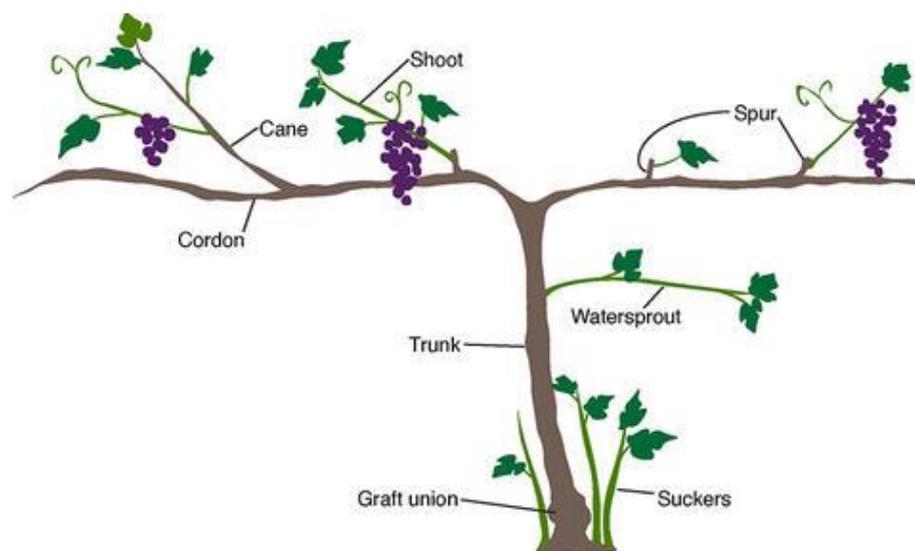
**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Summary

Grapevine pruning poses a difficult problem if a specific automation solution is required. From the agricultural domain, it is known that a grapevine is separated into different parts. Based on this knowledge, several attempts by computer vision and the image processing research community have been made to solve the problem through image data processing, with the support of stereoscopic cameras [1,2] and 3D laser scanners [3,4]. Most of these are non-invasive approaches, except [4] when a robotic system is used for grapevine pruning, based on image processing and computer vision methods. Other proposed publications are based on RGB images without the addition of distance estimation or other real-world attributes. These methods, with the support of real-world attributes from the scene, aim to decompose the grapevine from given images [5–8] to achieve a pruning cutting points estimation. The grapevine structure consists of three basic parts, (a) the trunk, which is the larger woody part of the plant, (b) the cordons, which are thinner than trunks and of which there are usually two in each plant, and (c) the canes starting on the top of cordons, which are the thinner woody parts of the plant and the tallest, as presented in Figure 1.



**Figure 1.** Grapevine plant anatomy [9].

Some basic visual characteristics of the plant are the fact that the trunk is always vertical, the cordons are horizontal, and the canes are in a random direction with a vertical formation. In the winter, during grapevine pruning, the shoots, spurs, suckers, and water sprouts do not exist. As mentioned above, the basic concept of pruning is to remove the last year's canes from the cordons and let the new canes grow to give new grapes [10]. This leads to the conclusion that canes are the object of interest in this specific problem. Grapevine pruning via computer vision and image processing has been proven to be a very challenging problem, since canes are increasing in length, in random directions, and very near each other. This structure creates a high overlap between objects of interest, and, with the addition of a complex background and noise from the environment, image analysis is very difficult, especially for an invasive approach where a robotic system must take action. Another challenge in this problem is the complex background, where a grapevine image will contain not only the foreground plant but also the background, especially during winter, when the grapevine has no leaves at all. Finally, each winery applies different pruning strategies, such as pruning some canes completely and just shortening the rest. Different pruning strategies add some constraints to the problem, which makes it more complex than simple. A feasible solution is the usage of stereoscopic cameras, where, based on an RGB image and depth map (real distance values), the foreground and background could be separated in a given image [2]. The above issues justify why most studies in the literature focus on non-invasive approaches [1–3,5–8], where background and foreground algorithms, object detection [7], and segmentation [6] deep learning algorithms have been used without pruning points' estimation. In invasive approaches [4], a closed environment surrounds the plant to be pruned with the support of robotic arms and computer vision methods based on 3D data processing. Despite the pruning, canes should also be removed, and in a fully automated scenario, the pruning strategy might change from the classical approach, since robotic manipulators with the support of cameras cannot mimic humans completely. Achieving this through image analysis means that all canes should be segmented in order to extract the whole canes' bodies for cutting point estimations on them. Semantic segmentation is a very popular method in robotic systems that interact in a wild-free environment. With this approach to grapevine pruning, automation methods can achieve cutting point estimations or feature extraction mechanisms, since the whole area of interest is segmented. Pruning could be characterized as a problem for which it is very hard to find robust solutions, since vineyard plants do not have more than a 3 m distance from each other and have a very complex structure. Based on the above issues, the objectives of this study are:

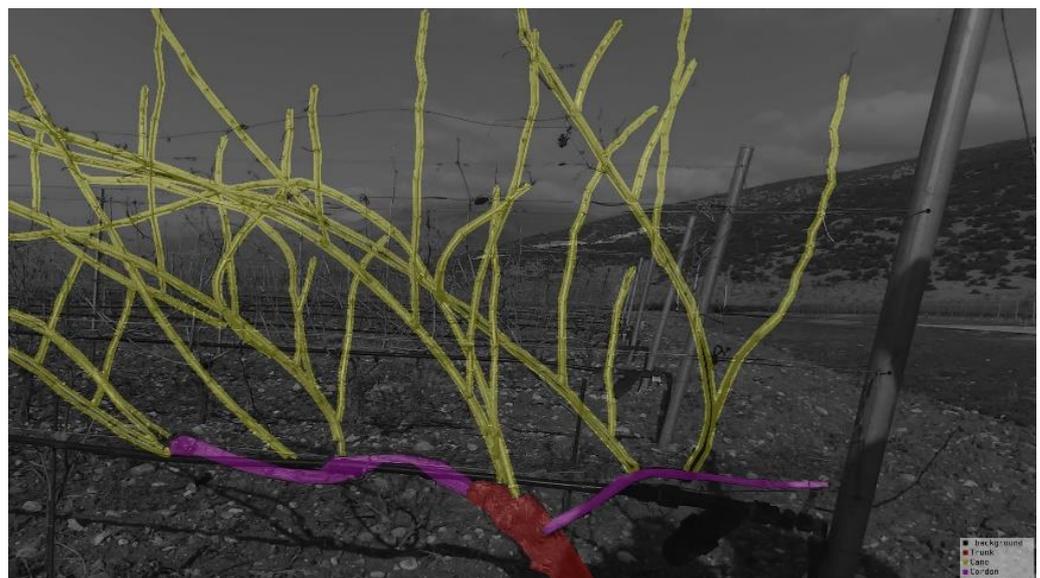
- To provide to the computer vision and image processing research community a dataset that will motivate further research on the automation of the very difficult work of pruning in vineyards, for example by accurately estimating cutting points.
- To propose a dataset of images collected from a grapevine during the pruning period, with the complexity and difficulties of the free environment of a vineyard and corresponding to the real conditions of pruning.
- To justify the value of the proposed dataset with the application of a semantic segmentation model to segment and deconstruct the basic pieces of grapevine plants.

## 2. Data Description

The proposed dataset contains 100 image samples from grapevine plants during the winter and vine pruning season. Each image is encoded in PNG with a resolution of  $1920 \times 1080$  and RGB color space. For each image sample, a hand-annotated mask is produced with the same resolution, and it contains marked pixel areas that point to a specific target class. Each marked pixel area has its unique color attribute and corresponds to each target class, whereas the rest of the image is colored in black and points to the background class. In Figure 2, a data sample is presented with the RGB image on the left (Figure 2a) and the marked pixel areas for each target class on the right (Figure 2b). In Figure 3, a mixed image between the RGB image (Figure 2a) in grayscale and its image mask (Figure 2b) is presented.



**Figure 2.** Image data sample of grapevine plant showing: (a) RGB image of the plant and (b) annotated mask of RGB image.

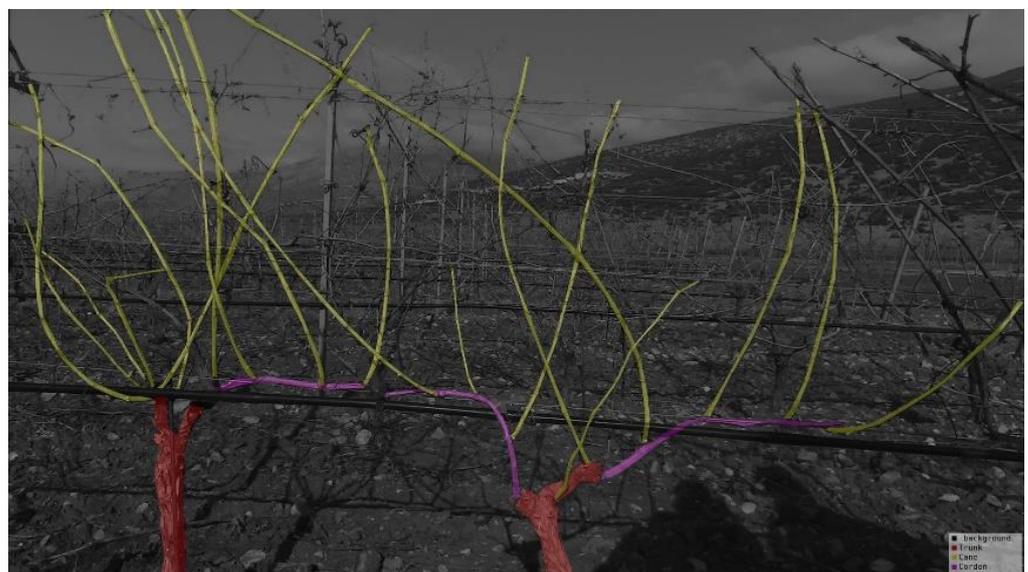


**Figure 3.** Mixed image between RGB image in grayscale and its image mask of Figure 2.

The data samples' structure is connected to the semantic segmentation methods, which is a classification approach for each pixel in a given image. The proposed dataset solves a 4-class problem (trunk, cordon, cane, background), where each datum consists of two images, as presented in Figures 2 and 4. In Figure 2a or Figure 4a in particular, the challenges presented in the previous section are visually justified. It is clear that the background creates a deformation between the target area of interest and the remaining image content. A basic approach with semantic segmentation solutions would be a CNN architecture as a backbone for feature extraction in a pair of images and a fully connected layer on top to apply the classification at the pixel level of the given images. In Figure 5, a mixed image between the RGB image (Figure 4a) in grayscale and its image mask (Figure 4b) is also presented.



**Figure 4.** Image data sample of grapevine plants skeletons showing: (a) RGB image of the plants' skeletons and (b) annotated mask of RGB image.



**Figure 5.** Mixed image between RGB image in grayscale and its image mask of Figure 4.

A huge challenge, for example, in these approaches is handling false-positive estimations of the model, where areas classified in a target class but belonging to the background class are segmented. This is one of the challenges that motivated computer vision researchers to propose and implement a huge number of CNN architectures and image processing methods in order to achieve robust solutions to semantic segmentation approaches. An important factor that determines the appearance of false positives in segmentation model estimations is the complexity of the image content, which includes the targeted area of interest. The aforementioned issue plays a significant role in invasive robotic systems because these perform based on these estimations. For this reason, only one model is not enough for a robust solution, and, consequently, supplementary methods are vital to solid

estimations. A possible solution might be a fusion of foreground-background algorithms and semantic segmentation models under one learning process, where the foreground-background algorithms would provide supplementary support to the segmentation models in order to emphasize the foreground features and produce better-segmented areas. Returning to the proposed dataset, the number of data samples could be characterized as small, but it contains high values of object instances in each image. In Table 1, the number of object instances from each class is presented for the whole dataset.

**Table 1.** The number of object instances for each target class.

| Trunks | Cordons | Canes |
|--------|---------|-------|
| 128    | 241     | 1316  |

Despite the small number of data samples, the number of object instances is high enough to support image processing or a computer vision and deep learning method, such as semantic segmentation with the support of augmentation methods, in order to produce a huge amount of images for training deep learning model architectures. Additionally, many images contain more than one grapevine plant. The imbalance between the object instances that appear in the dataset can be a problem for any methodology, since the objects of each class have a uniqueness in texture, color, and geometric features, along with the characteristic location of each object inside the image content. The selected classes emerged based on the age and mass of each plant part, where the trunk was first, followed by the cordon and the cane.

### 3. Methods

For the construction of the dataset, images were collected from a vineyard in the Greek city of Drama. Drama is an area in which many wineries are intensively active, and its selection has a special value. In Table 2, data concerning the location, the structure, and the density of grapevines in the vineyard are presented.

**Table 2.** Vineyard structure information of Pavlidis winery.

| Location                    | Lat: 41.20101042956404,<br>Long: 23.95247417327117 |
|-----------------------------|--|
| Posts' height               | 3–3.5 m  |
| Posts' width                | 7.62 cm  |
| Wires' diameter             | 3 mm   |
| Irrigation pipes' diameter  | 16 mm  |
| Plants' distance (same row) | 1.2–2 m  |
| Plants' distance (row–row)  | 2.2 m  |
| Corridor width              | 2.2 m  |

Since the dataset is proposed for pruning, the objects of interest are the canes, which provide enough to be processed by a deep learning model. Creating a dataset is important for providing a variety of instances to avoid overfitting. However, the work of pruning has some particularities. For example, pruning is mainly carried out in the morning with daylight. Additionally, the robot should see the plants vertically and not at an angle. That is why all images are captured vertically and with daylight. To obtain diversity in the dataset, images were taken from different distances in order for the model to learn to be scale-invariant. Scale invariance in this problem is very important, allowing one to discriminate areas that might belong to other corridors. Additionally, different distances are taken in order to accommodate robotic system scenarios where a camera is mounted on

the end effector of a robotic manipulator along with the cutting tool. In this scenario, the object of interest might change in scale in camera images based on robotic arm movements.

### 3.1. Obtaining the Dataset

The device that was used to collect the image data was a ZED Mini 3D camera with a ZED SDK installed and the appropriate software application for image capturing and file exportation. Table 3 provides details about the camera's settings. The images were hand-annotated with free polygon shapes. Each shape was characterized with a class name, and for each image, a JSON file was produced. Next, with each image and corresponding JSON file, which contained the annotated areas with the appropriate class names, the image masks were produced. The RGB images were kept the same without any further processing, maintaining the image content as it was collected.

**Table 3.** Specifications of the ZED Mini 3D camera.

| ZED Mini 3D   |   |
|---------------|---|
| Resolution    | 1920 × 1080                                 |
| FOV           | 90° horizontal, 60° vertical, 100° diagonal |
| Aperture      | f/2.0                                       |
| Sensor format | 16:9  |
| Sensor size   | 1/3"  |

The ZED Mini camera exposes all camera settings to the user, such as brightness, contrast, hue, saturation, gamma, sharpness, white balance, exposure, and gain. Each parameter can be adjusted manually, and changes are applied to both sensors without supporting individual adjustment. The camera settings are presented in Table 4.

**Table 4.** ZED Mini camera settings.

| ZED Mini 3D   |      |
|---------------|------|
| Brightness    | 4/8  |
| Contrast      | 4/8  |
| Hue           | 0/11 |
| Saturation    | 4/8  |
| White Balance | auto |
| Gain          | auto |
| Exposure      | auto |

### 3.2. Dataset Validation

To evaluate the dataset (Table 5), we implemented a U-Net [11] model in TensorFlow [12] with a segmentation-models library [13]. We used the U-Net model with ResNet34 [14] (Table 6) as the backbone, which was pre-trained in ImageNet. Additionally, data augmentation techniques were applied in order to increase the data. Specifically, we applied horizontal flip and 15 and −15 degrees rotation because they do not change the structure of a vineyard plant. The train and validation splits were 90% and 10%, respectively. The training process was done in Google Colab in order to deploy a 12 GB RAM GPU. The implementation achieved 67% IoU and 78% F1 Score (Equations (1) and (2)), with parameters that are presented in Table 7. Figures 6 and 7 present two results from the trained U-Net.

$$\text{IoU} = \frac{\text{Area of Intersection of two masks}}{\text{Area of Union of two masks}} \quad (1)$$

$$\text{F1 Score} = \frac{\text{True Positives}}{\text{True Positives} + \frac{1}{2}(\text{False Positives} + \text{False Negatives})} \quad (2)$$

**Table 5.** Amount of data before and after data augmentation.

| Augmentation | Train Data | Test Data |
|--------------|------------|-----------|
| No           | 90         | 10        |
| Yes          | 360        | 40        |

**Table 6.** Details of used models.

| Model    | Convolutional Layers | Parameters (Millions) |
|----------|----------------------|-----------------------|
| ResNet34 | 34                   | 21,797                |
| U-Net    | 23                   | 7765                  |

**Table 7.** Parameters used for training.

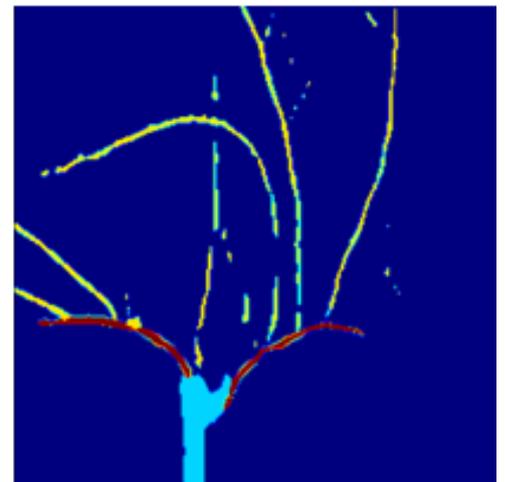
|               |          |
|---------------|----------|
| Learning Rate | 0.0001   |
| Batch Size    | 8        |
| Optimizer     | Adam     |
| Backbone      | ResNet34 |
| Epochs        | 50       |



(a)

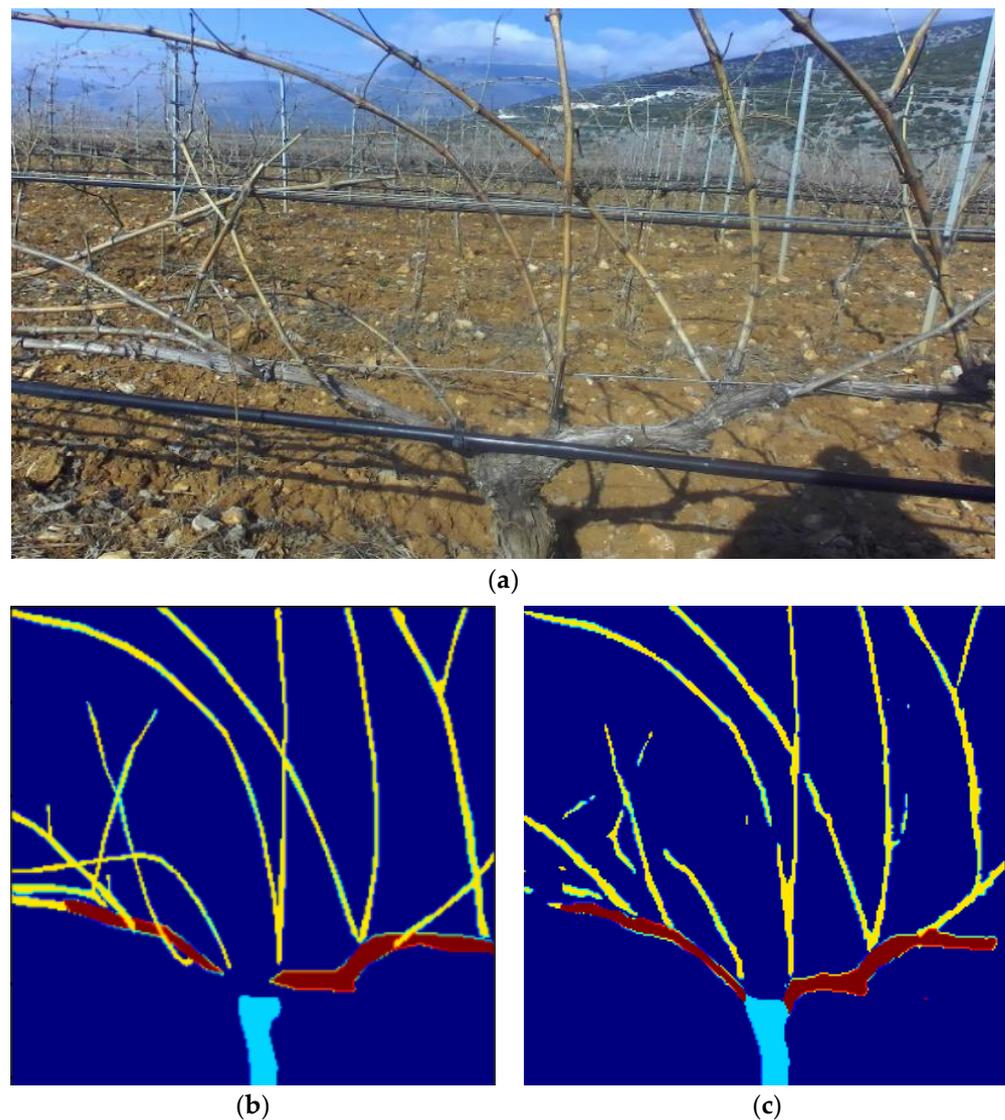


(b)



(c)

**Figure 6.** Results from the trained model: (a) Original Image, (b) Ground Truth, and (c) Model Prediction. (1st example).



**Figure 7.** Results from the trained model: (a) Original Image, (b) Ground Truth, and (c) Model Prediction. (2nd example).

With the above results, the applicability of the image dataset is justified. The predicted masks could be used for further processing to estimate cutting points at the start of the canes on top of cordons. Despite pruning, the above image dataset and application could be used for further plant analyses such as grape maturity growing.

#### 4. User Notes

The way that this dataset was created is intended to provide data for model development in a vineyard pruning task. Vineyard pruning based on robotic systems is a very difficult task. A proposed procedure for a vineyard pruning task is roughly as follows. A mobile robot stops in front of a vine in a preselected location. The vision system mounted on the robotic manipulator detects the desired area. The acquired images are processed by means of the trained models, and the specific cutting points are calculated. The system, based on the optimum practices, selects the cutting point. An initial calculation of the target location permits the execution of the path planning algorithm, which will create an initial path for the manipulator. The manipulator starts to move towards the target, following this path. The vision system for each cycle acquires an image and, after processing, provides the coordinates of the target with a better accuracy (visual servoing). When the tool arrives at the desired location, a cutting procedure is implemented. To extend this scenario, a

second collaborative mobile robot follows the previous one, and, with a second robotic manipulator and a properly designed gripper at the end-effector of it, it removes the cut branch. When the cutting procedure is completed for that specific vine, the mobile robots move to the next vine.

**Author Contributions:** Conceptualization, T.P.P. and V.G.K.; methodology, K.D.A. and T.K.; software, K.D.A. and T.K.; validation, K.D.A., T.K. and T.P.P.; formal analysis, K.D.A. and T.K.; investigation, K.D.A. and T.K.; resources, T.P.P. and V.G.K.; data curation, K.D.A. and T.K.; writing—original draft preparation, K.D.A. and T.K.; writing—review and editing, T.P.P. and V.G.K.; visualization, K.D.A., T.K. and T.P.P.; supervision, T.P.P. and V.G.K.; project administration, V.G.K. and T.P.P.; funding acquisition, V.G.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge the support of this work by the project “Technology for Skillful Viniculture (SVtech)” (MIS 5046047), which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure” funded by the Operational Program “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available at <https://github.com/humain-lab/Buds-Dataset> under Creative Commons Attribution 4.0 International license.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Moreno, H.; Rueda-Ayala, V.; Ribeiro, A.; Bengochea-Guevara, J.; Lopez, J.; Peteinatos, G.; Valero, C.; Andújar, D. Evaluation of Vineyard Cropping Systems Using On-Board RGB-Depth Perception. *Sensors* **2020**, *20*, 6912. [CrossRef]
- Fernandes, M.; Scaldaferrri, A.; Fiameni, G.; Teng, T.; Gatti, M.; Poni, S.; Semini, C.; Caldwell, D.; Chen, F. Grapevine winter pruning automation: On potential pruning points detection through 2d plant modeling using grapevine segmentation. In Proceedings of the 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Jiaxing, China, 27–31 July 2021.
- Botterill, T.; Green, R.; Mills, S. Finding a vine’s structure by bottom-up parsing of cane edges. In Proceedings of the 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), Wellington, New Zealand, 27–29 November 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 112–117.
- Botterill, T.; Paulin, S.; Green, R.; Williams, S.; Lin, J.; Saxton, V.; Mills, S.; Chen, X.; Corbett-Davies, S. A Robot System for Pruning Grape Vines: A Robot System for Pruning Grape Vines. *J. Field Robot.* **2017**, *34*, 1100–1122. [CrossRef]
- Corbett-Davies, S.; Botterill, T.; Green, R.; Saxton, V. An expert system for automatically pruning vines. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand—IVCNZ ’12, Dunedin, New Zealand, 26–28 November 2012; ACM Press: Dunedin, New Zealand, 2012; pp. 55–60.
- Xu, S.; Xun, Y.; Jia, T.; Yang, Q. Detection method for the buds on winter vines based on computer vision. In Proceedings of the 2014 7th International Symposium on Computational Intelligence and Design, Hangzhou, China, 13–14 December 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 44–48.
- Gao, M.; Lu, T. Image processing and analysis for autonomous grapevine pruning. In Proceedings of the 2006 International Conference on Mechatronics and Automation, Luoyang, China, 25–28 June 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 922–927.
- Guadagna, P.; Frioni, T.; Chen, F.; Delmonte, A.I.; Teng, T.; Fernandes, M.; Scaldaferrri, A.; Semini, C.; Poni, S.; Gatti, M. Fine-tuning and testing of a deep learning algorithm for pruning regions detection in spur-pruned grapevines. In Proceedings of the Precision Agriculture ’21, Budapest, Hungary, 19 July 2021; Wageningen Academic Publishers: Budapest, Hungary, 2021; pp. 147–153.
- Growing Grapes in the Home Garden. Available online: <https://extension.umn.edu/fruit/growing-grapes-home-garden> (accessed on 16 April 2022).
- Hellman, E.W. Grapevine structure and function. In *Oregon Viticulture*; Hellman, E.W., Ed.; Oregon State University Press: Corvallis, OR, USA, 2003.
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
- Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI ’16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.

- 
13. GitHub—Qubvel/Segmentation\_Models: Segmentation Models with Pretrained Backbones. Keras and TensorFlow Keras. Available online: [https://github.com/qubvel/segmentation\\_models](https://github.com/qubvel/segmentation_models) (accessed on 29 July 2022).
  14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.