POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Advanced Image Segmentation for Automatic Estimation of Exposed Leaf Area in Vineyards

Laurea Magistrale in Agricultural Engineering

Author: Jan Ferfolja

Advisor: Prof. Sara Mandelli

Co-advisor: Dott. Edoardo Daniele Cannas

Academic year: 2023-2024

---

1. Introduction

In a world increasingly relying on data-driven decisions, data analysis has become central to every company or business. This is true also for the agricultural sector. Even if digitalization and automation had a harder time spreading in this sector compared to other industries, it is now the driving force of growth and innovation. In this rapidly evolving environment, the company Perleuve, an agronomic consultancy specializing in high-quality wine production, faces an inconsistency problem in gathering data for the computation of its patented index, the Bigot Index (BI). The BI is used to evaluate the quality of the vineyard and the potential of the grapes harvested that year. One of the crucial parameters for determining the BI is the Exposed Leaf Area - *Exposed Leaf Area* (ELA). The ELA is intended as the percentage of the total area given to a single plant to grow its canopy, which is effectively covered by the canopy.

2. Objectives

The primary objective of this thesis is to develop an automated pipeline for the accurate and consistent estimation of ELA using image segmentation techniques. This solution aims to minimize inconsistencies in data collection caused by biases of agronomists evaluating the ELA starting from a dataset of 1456 images. Each image in the dataset came with various tags; the most important for our work is the ELA value, estimated by the agronomists. An important note is that the dataset provided by the company was not created with this thesis or task in mind, and the images were taken with only the purpose of saving an image to attach to a data point in the company's archive. From the initial dataset, the goal was to establish a standard way to take the images in the field, making them suitable for image segmentation and, more importantly, ELA calculation. After the characteristics of the images were decided, the first step was to remove all non-compliant images and, on the remaining images, develop a pipeline that could output a percentage value that represents the area effectively occupied by leaves over the area available to the vines. In other words, our goal is to develop a pipeline P such that, given an image of a vine plant $\mathbf{I}$, it outputs an estimate of the ELA

$$\hat{\text{ELA}} = P(\mathbf{I}). \tag{1}$$

The analysis focuses on the canopy's shape and, more importantly, its density. A critical aspect of the evaluation of ELA is detecting gaps within the canopy. After identifying the canopy area and classifying leaf pixels, the ELA is calculated as the ratio of leaf pixels to the total number of pixels within the canopy area.

3. Methodology

Starting from the dataset provided by the company, the first step was to look at the images and get a rough idea of the dataset. From this first inspection it was clear that many of the images where not suitable for the task as they were not taken perpendicularly on the row, but at an angle, making it nearly impossible to extrapolate the ELA, since the holes in the canopy, where not visible. The first step was, thus, to clean the dataset that the company provided and remove the images that were not taken perpendicularly to the vine row. This was done using part of the auto-rectification pipeline used by [1]. Through this was calculated the vanishing point of the images and based on the distance of the point from the center, was decided to keep or not the image. The developed pipeline is summarized in figure 1.

In the dataset, after filtering the data samples, we noticed two distinct categories of images, i.e., i) images previously cropped by the agronomists to contain only the canopy of the vines and ii) images that were not cropped and depicted the full plant, with the trunk, ground, sky, and various background objects like hills or houses. To account for this difference, the images were treated slightly differently throughout the pipeline and will be referred to as Cropped and Full Plant images. The next step in the pipeline is segmentation, i.e., to create a mask of the vine canopy. We developed two approaches:

1. **Color-based segmentation**: Uses the HSV color space to differentiate vine leaves from the background. It is very simple but, in some cases, a very effective segmentation method. The method consists in selecting pixels with particular values in the three HSV channels, in our case the most important was the first channel, which maps the color of the pixel. The other two channels, which map the saturation and brightness, are helpful in capturing all the various shades of the first selected color. Using the HSV color map is thus much more convenient than using RGB, where the color, brightness, and saturation information are equally spread across the three channels.
2. **Deep learning segmentation**: Implements the Segment Anything Model (SAM) [2], taking inspiration from [3], a transformer-based approach, for automated object detection. The Visutal Transformer (ViT) checkpoint that was used in this work is the largest of those provided, the Visual Transformer-Huge (ViT-H) (ViT-H). The model takes an image of any size as input, divides the input into processed patches, and then combines them in the final segmented images. The masks of the segmented objects are then output as a list of individual masks.

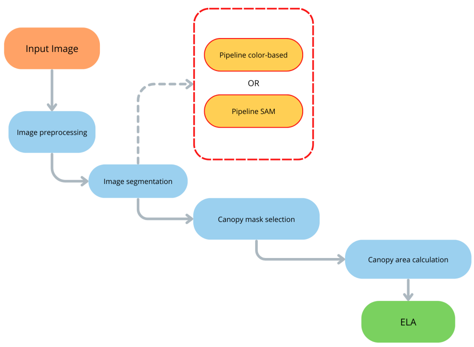


Figure 1: Block diagram of the pipeline, with the two different approaches used for segmentation.

3.1. Color-based approach

In the case of cropped images, after the color-segmentation stage, all the masked pixels were assumed to be part of the canopy and were used to calculate the ELA, assuming the bounding box to be the size of the whole image. The precise formula for the $\hat{\text{ELA}}$ is:

$$\hat{\text{ELA}} = \frac{\text{Number of white pixels}}{\text{Total number of pixels}} \tag{2}$$

where the Total number of pixels is the number of all the pixels in the bounding box. For the Color-based approach with the Full Plant images, once the binary mask that segmented the green parts of the image was created, the mask representing the canopy was selected by taking the biggest pixel island and putting all other pixels to 0. This way, the final binary mask obtained was assumed to represent the canopy, and from here, we calculated the bounding box calculation around the white pixel island. Finally, we calculated the percentage of white pixels over all the pixels in the bounding box similarly to eq. (2).

3.2. SAM-based approach

After segmentation, the model outputs a list of masks containing the segmented objects. From this list, selecting the mask that depicts the canopy is necessary. For the cropped images, to obtain the mask of the canopy, all the generated masks were checked to see if they contained leaves or not, similar to the approach from [3], using HSV colorspace. All the masks representing green pixels were merged into one mask that became the canopy estimate. From this point, the ELA was calculated similarly for the color-based approach. For the Full Plant images, we used the following criteria: since the canopy represents the largest part of most images and is primarily composed of green pixels, we started with the mask containing the highest number of green pixels. We then checked whether at least 40% of those pixels were green to avoid selecting masks of the ground or sky. If the mask did not meet this criterion, we moved on to the next largest mask and continued until the correct canopy mask was identified. Once the canopy mask was selected, the next step was to define the bounding box of the entire area available for the plant to grow its canopy. The bounding box was defined by creating a rectangle with the full width of the image and a height equal to that of the canopy mask. Once the bounding box was defined, the last step in the pipeline was to calculate the ELA by dividing the number of pixels that masked the canopy by the total pixels in the bounding box.
To improve the performance of the SAM, some changes were made to the initial hyperparameters described in table 1. The first change from the original configuration was to remove the initial resize that halved the image dimensions since computation time was not a concern in our case. The second adjustment was to reduce the Points per Side value, which influences the number of patches that the ViT model divides the image into at the beginning of the process. In our case, reducing this value increases the size of the patches and reduces their number. The third and last hyperparameter tweaked was the Min Mask Area, which defines, as its name implies, the minimum number of pixels that the generated masks need to have to make it to the output list of masks.

4. Results

To give a numerical evaluation to the calculated ELA, calculated with both the color-based and SAM-based pipeline, the estimated value was compared with the reference value given by the agronomists. From now on, it will be called Difference between the pipeline guess and the reference value given by the agronomists (Delta). More formally:

$$\Delta_{\mathrm{ELA}} = \mathrm{ELA} - \hat{\mathrm{ELA}}, \tag{3}$$

where ELA is the reference value given by the agronomists, and $\hat{\mathrm{ELA}}$ the outcome of our pipeline as formulated in eq. (1). To summarize the results of the thesis, in figure 2 we report the histogram of the Delta from across the best SAM runs for Full plant images, the most realistic scenarios. The best guess in each case was defined as the one with the lowest Delta. By selecting only the lowest Delta, we eliminate potentially outlier guesses from the other three runs, which may have resulted from unsuccessful segmentation, and retain only the most credible estimate. The distribution of the best guesses results in a much more tight bell shape, with the peak around the 0.0 value. This indicates that in the case of the full plant images, the second pipeline tends to agree with the estimations of the agronomists. There is still a group of images around 0.6-0.8 away from the estimated values that represent the problematic cases that were not properly segmented in all four of the runs.

4.1. Qualitative assessment of performance

Given the inhomogeneity in the ELA values from the different data samples of the Perleuve dataset, we provide a qualitative evaluation of the two segmentation pipelines in the following. The color-based segmentation approach proved quite effective in images with a clear color separation between the ground, the sky, and the background of the canopy. This is, though, a very rare case in a real-world scenario. This approach, as expected, did not perform well in im-

Table 1: Comparison of Different Runs for SAM Mask Generation

Run	Resize	Points per Side	Min Mask Area	Adjustment goal
1	Yes (0.5×0.5)	32	200	Default settings
2	No	32	100	Reduced min area
3	No	16	200	Bigger grid
4	No	16	300	Increased min mask area

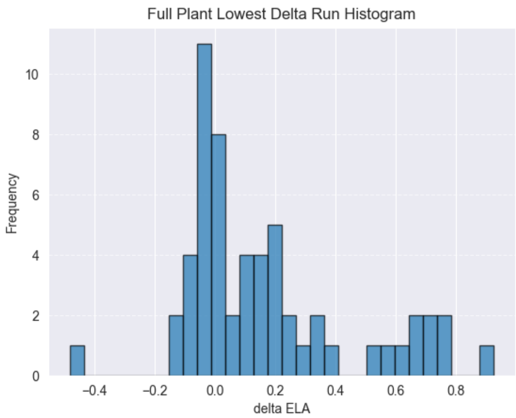
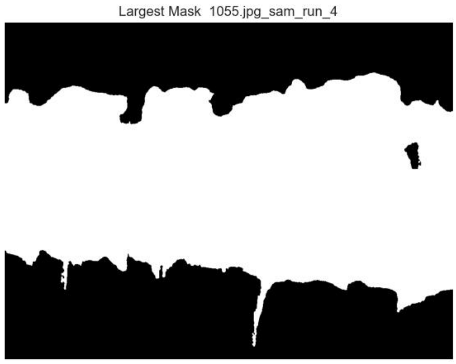


Figure 2: Histogram of the best guesses across the four SAM-runs on Full Plant images.

ages where there was grass on the ground, as it disrupted the segmentation and created inaccurate masks of the canopy. The second approach worked better in those more realistic scenarios as the SAM model could discern the grass from the leaves in most examples. In figure 3, we can see one of those examples, and in the next figure 4, four of the stages of the pipeline that uses SAM to segment the image, extract the mask of the canopy and then filter out of the mask only the leaves. In the last image, we displayed the final mask from with the ELA is calculated. A comparison with agronomists' manual assessments revealed that the automated methods provide a more standardized evaluation of ELA. However, challenges such as canopy occlusion and varying lighting conditions remain.

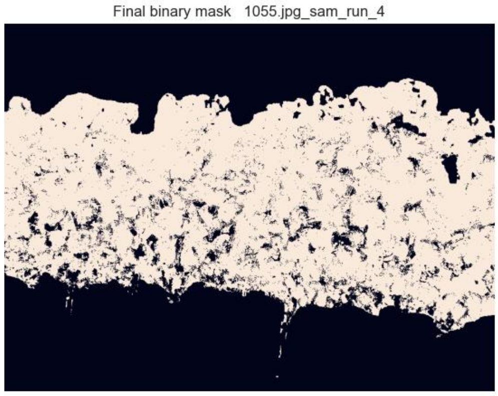
Figure 3: Image of a complex example where grass is on the ground.

(a) Plot of all masks generated by the SAM model



(b) 2 Biggest mask that has at least 40% of the pixels green

(c) Canopy mask after the removal of all non-green pixels with the HSV-based segmentation.



(d) Final binary mask from which is calculated the bounding box and then the ELA.

Figure 4: Example of four of the main steps in the pipeline with the SAM-based approach to segmentation.

5. Conclusions

In the wine industry, quality and reputation are crucial in determining a wine's market value. The Bigot Index (BI) provides a standardized way to assess vineyard quality, helping winemakers enhance their reputation and refine vineyard management decisions. Among the nine parameters of the BI, the Exposed Leaf Area (ELA) is the most challenging for agronomists to measure consistently.

This thesis explores a computer vision-based approach to improve ELA estimation, comparing two segmentation pipelines. The first method, color-based segmentation, performs well under controlled conditions but struggles in real-world scenarios. The second method leverages a pre-trained Vision Transformer (ViT) to generate segmentation masks, significantly improving performance in complex environments.

Testing multiple model runs of the second pipeline revealed that no single iteration consistently outperformed the others, suggesting that fine-tuning the model on a specialized dataset would enhance accuracy. Additionally, incorporating depth maps or automated plant measurements could further improve results.

In an industry increasingly driven by data-driven decision-making, this research demonstrates the potential of computer vision in vineyard assessment while highlighting the challenges of adapting AI tools to new agricultural applications.

References

[1] Krishnendu Chaudhury, Stephen DiVerdi, and Sergey Ioffe. Auto-rectification of user photos. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 3479–3483, 2014.

[2] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023.

[3] Dominic Williams, Fraser Macfarlane, and Avril Britten. Leaf only sam: A segment anything pipeline for zero-shot automated leaf segmentation, 2023.