Image-based analysis of yield parameters in viticulture

Laura Zabawa[a], Anna Kicherer[b], Lasse Klingbeil[a], Reinhard Töpfer[b], Ribana Roscher[c], Heiner Kuhlmann[a]

[a] *Bonn University, Department of Geodesy, Institute of Geodesy and Geoinformation*
[b] *Julius Kühn-Institut, Federal Research Centre of Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof*
[c] *Bonn University, Remote Sensing Group, Institute of Geodesy and Geoinformation*

Abstract

Yield estimation is of great interest in viticulture, since an early estimation could influence management decisions of winegrowers. The current practice involves destructive sampling of small sets in the field and a subsequent detailed analysis in the laboratory. The results are extrapolated to the field and only approximate the actual conditions. Therefore, research in recent years focused on sensor-based systems mounted on field vehicles since they offer a fast, accurate and robust data acquisition. However many works stop after detecting fruits, rarely the actual yield estimation is tackled. We present a novel yield estimation pipeline that uses images captured by a multi-camera system. The system is mounted on a field phenotyping platform called Phenoliner, which has been built from a modified grapevine harvester. We use a neural network whose output is used to count berries in single images. In contrast to other existing methods we take the step from the single vine image processing to the plant level. The information of multiple images is used to acquire a count on plant level and the approach is extended to the processing based on the whole row. The acquired berry counts are

used as input for the yield estimation, and we explore the limitations and potentials of our pipeline. We identify the variability of the leaf occlusion as the main limiting factor, but nonetheless we achieve a mean absolute yield prediction error of 26% for plants in the vertical shoot positioned system. We evaluate each described stage comprehensively in this study.
Keywords: Deep Learning, Semantic Segmentation, Geoinformation, Viticulture, Yield Estimation

1. Introduction

Grapevine is one of the ecologically most valuable crops. But in contrast to many other agricultural crops, like wheat (Ray, Mueller, West and Foley, 2013) or maize, the goal is not yield enhancement. The focus in viticulture is on quality aspects and the optimization of resources, since a larger amount of grape bunches and berries leads to a decrease in wine quality (Howell, 2001).

In general, guided management and breeding decisions in agriculture require reliable, objective and exhaustive information about phenotypic traits of the relevant crop. Especially yield estimations depend on information like the number, size and weight of crops. The large scale acquisition of these phenotypic data and their respective automatic analysis is called high-throughput phenotyping (Araus and Cairns, 2014) and the recent developments in this research field are mainly driven by the development of sensors, like thermal-, RGB (red/green/blue)-, RGBD (r/g/b/depth)- and multi-spectral cameras, as well as laser scanners. The main advantages of these automatic procedures are the coverage of large field portions, as well as objectivity, repeatability, and high quality results.

Challenges arise especially for perennial crops like apples or grapevine, where phenotyping has to be performed in the field under sometimes difficult terrain situations. Traditionally these procedures include simple visual screening, counting of yield components in the field or even harvesting samples and weighing them in the lab (Alercia et al. (2009), Lorenz et al. (1995), Bloesch and Viret (2008)). The collected samples or information need to be investigated by skilled experts (Nuske et al., 2014) but these procedures are nonetheless time consuming, error prone and highly subjective. Another drawback is that information are extrapolated from small samples to large fields. Therefore recent research focuses on objective, sensor-based approaches to enable a fast and reliable high throughput phenotyping for perennial crops, including viticulture (Matese and Gennaro, 2015).

For this purpose, different sensors, data acquisition strategies and even platforms were developed in several studies. From handheld sensors (Kurtulmus, Lee and Vardar, 2011, Linker, Cohen and Naor, 2012, Dorj, Lee and Yun, 2017), over small (Gao et al., 2020, Hemming et al., 2014) to big phenotyping vehicles (Kicherer et al., 2017, Gan et al., 2020), up to drones (Kalantar, Edan, Gur and Klapp, 2020, Gennaro, Toscano, Cinat, Berton and Matese, 2019). The used sensors include RGB (Nuske et al., 2014, Zabawa et al., 2020), (Aquino, Millan, Diago and Tardaguila, 2018), multi- or hyperspectral cameras (Bendel et al., 2020), over RGBD-sensors (Kurtser, Ringdahl, Rotstein, Berenstein and Edan, 2020) to laser scanners (Tagarakis, Koundouras, Fountas and Gemtos, 2018).

Many works, not only in viticulture, tackle the problem of detecting and localizing fruit as a first step for yield estimation. In some cases the detection

is straightforward if the fruit has a very distinct color, for example orange citrus fruit (Dorj, Lee and Yun, 2017) or red grapes (Diago et al., 2012, Silver and Monga, 2019) in front of green canopy. In other cases the detection is more challenging, especially for green fruits in front of green canopy (Kurtser, Ringdahl, Rotstein, Berenstein and Edan, 2020, Kurtulmus, Lee and Vardar, 2011, Linker, Cohen and Naor, 2012, Stein, Bargoti and Underwood, 2016). For the latter case, many different strategies were developed, either by using only RGB images or with the help of additional sensors. For example Wachs, Stern, Burks and Alchanatis (2010) detect green apples using multimodal data consisting of RGB and thermal images while Linker, Cohen and Naor (2012) use only RGB images. Gené-Mola et al. (2020) on the other hand detect apples with a LIDAR mounted on mobile platform. Gan, Lee, Alchanatis and Abd-Elrahman (2020) use only thermal images in combination with a water spraying system to detect green citrus fruit. Nuske, Achar, Bates, Narasimhan and Singh (2011) and Roscher et al. (2014) detect single green grapevine berries as circular objects in images by using a Hough- or radial-symmetry-transform respectively. Diago et al. (2012) use the Mahalanobis distance for the detection of berry and leaf pixels. Kicherer et al. (2015) developed a field phenotyping platform which was later used by Rose et al. (2016) to reconstruct 3D point-clouds from image sequences for the detection of single grapevine berries. Nyarko et al. (2018) identify fruits as convex surfaces.

With the rise of deep learning methods new approaches based only on RGB images emerged. The applications for images include the detection of grapevine influorescenses (Rudolph, Herzog, Töpfer and Steinhage, 2018)

or berries (Aquino et al., 2016, 2018, Zabawa et al., 2020, Cecotti et al., 2020) with convolutional neural networks (CNNs) or the adaptation of a crowd-counting algorithm (Coviello, Cristoforetti, Jurman and Furlanello, 2020). Others use a state-of-the-art instance segmentation network, the Mask-RCNN, for the detection and tracking of grapevine berries (Nellithimaru and Kantor, 2019, Santos, de Souza, dos Santos and Avila, 2020). For a further review about different sensors and algorithms, which were used for fruit detection and localization we refer the reader to Gongal, Amatya, Karkee, Zhang and Lewis (2015). For deeper insights in the usage of computer vision, especially stereo methods in fruit picking systems, we refer the reader to Tang et al. (2020).

Many of these works stop after detecting the fruits, only rarely the subsequent problem of yield estimation is actually addressed. Even the definition of yield ranges from a comparison of predicted and actual fruit count (Bargoti and Underwood, 2017, Dorj, Lee and Yun, 2017) to the investigation of the harvest weight with respect to estimated weight (Kalantar, Edan, Gur and Klapp, 2020). Anderson, Walsh and Wulfsohn (2021) give a systematic overview of recent advances on the forecasting of yields for tree fruits.

In viticulture, different parameters can be used at different points in time to forecast yield with a yield function (de la Fuente, Linares, Baeza, Miranda and Lissarrague, 2015, Clingeleffer, Dunn, Krstic and Martin, 2001, Coviello, Cristoforetti, Jurman and Furlanello, 2020). These parameters include cluster number, size, volume, length and width (Kicherer, Roscher, Herzog, Förstner and Töpfer, 2014, Ivorra, Sánchez, Camarasa, Diago and Tardáguila, 2015, Hacking, Poona, Manzan and Poblete-Echeverría, 2020,

Gennaro, Toscano, Cinat, Berton and Matese, 2019), (Kurtser, Ringdahl, Rotstein, Berenstein and Edan, 2020, Santos, de Souza, dos Santos and Avila, 2020). Usually a weight factor needs to be taken into account, respectively an average bunch weight or the weight of single berries, which are collected manually and averaged over many years. Other works estimate the fruit weight directly from the image (Kalantar, Edan, Gur and Klapp, 2020, Silver and Monga, 2019) or correlate the number of detected objects with the harvested yield at the end of the season (Nuske, Achar, Bates, Narasimhan and Singh, 2011). One example for a complex approach which is based on a combination of counting and manual sampling is presented by (Coviello, Cristoforetti, Jurman and Furlanello, 2020). They describe the yield initially as a function of the number of vines per surface unit, the number of grape bunches per vine and the average grape bunch weight. Due to practical limitations, they substitute the average grape bunch weight with the number of berries per bunch and the average berry weight. Others on the other hand propose a minimalistic approach inspired by Clingeleffer, Dunn, Krstic and Martin (2001). They estimate the yield at harvest as a function of the berry number and the average berry weight (Nuske et al., 2014, Aquino et al., 2018), since 90% of the yield variation are caused by the berry number per vine, while the remaining 10% are caused by the average berry weight (Clingeleffer, Dunn, Krstic and Martin, 2001). For a detailed review regarding yield estimation we refer the reader to Darwin, Dharmaraj, Prince, Popescu and Hemanth (2021) and Koirala, Walsh, Wang and McCarthy (2019).

One of the limiting factor for all automatic yield estimation approaches is the visibility of the investigated yield components. This could mean for

example, that grapes or berries are occluded in the images due to leaves. One idea to minimise the occlusion is the usage of active airflow to move canopy out of the way (Gené-Mola et al., 2020). (Koirala, Walsh, Wang and McCarthy) used a dual viewpoint approach for mango fruit load estimation and estimated occlusion correction factors based on manual counts. Other approaches, especially for sweet pepper harvesting in horticulture, try to decrease the effect of occlusions by investigating the benefits of multiple but static viewpoints (Hemming, Ruizendaal, Hofstee and Henten, 2014, Harel, van Essen, Parmet and Edan, 2020). Wang, Walsh and Koirala (2019) developed a multi-viewpoint approach based on 10 fps video sequences for the counting of mango. They tracked detected fruits in neighbouring frames with the Hungarian Algorithm and predicted their positions in following frames using a Kalman-Filter. Kurtser and Edan (2018a,b) focused on a dynamic planning of viewpoints for a robotic arm, which can guide harvesting decisions. Gao et al. (2020) on the other hand classified the detected apples into different occlusion classes to decide on different picking strategies.

In this paper, we present and evaluate novel data analysis pipeline for automated grapevine yield estimation. This pipeline is based on a method for counting of berries in single images, which has been presented by the authors in an earlier publication (Zabawa et al., 2020). This method has been extended to be used on multiple overlapping images, which are taken from a mobile multi-camera system. The pipeline considers vertical and horizontal overlap between the images and predicts the yield of full plants or even rows. With the collected image data in combination with reference yield data, acquired over several years, we focused on the estimation of yield

and the main challenges in this process. We identified the visibility of the used yield components, the berry number, as the key factor influencing the yield estimation and performed leaf occlusion experiments with the goal of improving the yield estimate. Lastly, we perform a yield estimation for the vertical shoot positioned (VSP) system based on the number of detected berries and an average berry weight which we collected manually over three years. In addition, we put the results into perspective with regards to other image-based and classical approaches.

2. Materials and Methods

Our pipeline covers various steps from image acquisition to the estimation of yield in kg. In the following, we will first present the field phenotyping platform and the sensors which were used for the data acquisition. Furthermore, we will give detailed information regarding collected data and algorithms for the evaluation. We start with a brief introduction regarding single image data, explain the approach for handling the vertical overlap on the plant level and the way we handle the horizontal overlap, and end with the yield estimation procedure.

2.1. Sensor System

The data acquisition process was performed with the Phenoliner (Kicherer et al., 2017), a field phenotyping platform. The platform itself is a modified grapevine harvester from ERO Gerätebau (Niderkumbd, Germany), the ERO-Grapeliner SF200. The harvesting equipment, including the shaking unit, destemmer and grape tank were removed and a multi-camera system was installed in the so called 'tunnel'. The platform was further equipped

with an artificial lighting system as well as a diffuse back wall. To ensure a precise positioning, a real time kinematic (RTK)-GNSS System (SPS852, Trimble ®, Sunnyvale, CA USA) was put on top of the Phenoliner. It is able to achieve an accuracy of 2 cm.

(a) Phenoliner (b) Camera System

Figure 1: 1a shows the phenotyping platform Phenoliner (Kicherer et al., 2017). It is based on a grapevine harvester from ERO Gerätebau, where the harvesting equipment was removed and camera systems, lighting units and diffusing background were installed. The camera system used can be seen in 1b and consist of five cameras which deliver overlapping images of the canopy. The vertical cameras are positioned with 35 cm between each camera resulting in a maximum distance between the outer two cameras of approximately 70 cm. This leads to a vertical canopy coverage of 1.2 m.

The camera system consists of five cameras, four RGB (DALSA Genie NanoC2590, Teledyne DALSA Inc., Waterloo, ON, Canada) and one NIR (DALSA Genie NanoM2590-NIR, Teledyne DALSA Inc., Waterloo, ON, Canada) camera. Three of the RGB cameras are positioned on top of each

other, each with a distance of 35 cm between each. The lowest camera is part of the lower camera system as well, together with the forth RGB camera it encloses the NIR camera. These cameras are positioned directly beside each other, leading to a baseline of 7 cm each. The cameras have a 5.1 Megapixel sensor and 12 mm lenses, each image has dimensions of 2592×2048 pixels, leading to a real world resolution of 0.3 mm. In our experiments we used the images captured with the three vertical cameras. We didn't use the two remaining cameras (RGB and NIR), but included them in the above-description for the sake of completeness. The combination of the GNSS receiver on top of the Phenoliner, the data from a dual axis inclinometer and a previously performed system calibration, we were able to determine the absolute position of each camera with an accuracy of about 3cm.

2.2. Data

We collected several datasets in three consecutive years, 2018 - 2020. The data collection was performed in experimental vineyard plots at the JKI Geilweilerhof located in Siebeldingen, Germany (49°21.747'N, 8°04.678'E). The datasets include image sequences covering whole rows of vines and reference data corresponding to selected plants in each row. Tab. 1 summarises the collected data regarding the different varieties and training systems. For each variety we give the number of observed plants and the number of reference plants. The reference plants remained the same over the three years.

2.2.1. Image data

Three different grapevine (*vitis vinifera*) varieties were observed with the Phenoliner over these three years: Riesling, Felicia and Regent (see Fig.

Table 1: Overview of the collected data over the years 2018 to 2020. We observed three different varieties and two training systems. In each row 10 plants (one meter of canopy) were harvested manually at the end of the season as a yield reference. The observed reference plants were the same each year.

Years	**Variety**	**# Rows**		**# Plants**	**# Reference plants**
		SMPH	**VSP**		(10 each row)
2018 - 2020	Riesling	4	3	175	70
2018 - 2020	Felicia	2	2	387	40
2018 - 2020	Regent	2	2	205	40

2). The image acquisition process involved driving through each row and collecting images approximately every 4 cm. Each variety was grown in two different training systems, the vertical shoot positioned system (VSP) and the semi minimal pruned hedge (SMPH). Each training system poses different challenges which have to be acknowledged. The VSP is a traditional system, with only one main branch. In one side of the canopy defoliation is performed in the fruit-zone after flowering. The grape bunches are compact, with a homogeneous berry size, and occur mainly in the lower part of the canopy in the fruit zone. The SMPH on the other hand features many branches and a large amount of leaves, since the canopy is not reduced. The grape bunches are scattered across the whole canopy but appear mainly in the upper part of the canopy. The bunches itself are loose and have inhomogeneously sized berries.

From this extensive image database 59 images were labelled manually on berry instance level using a image editing software. For more details we refer

(a) Riesling (b) Felicia (c) Regent

Figure 2: Images showing the three different varieties observed in the experiments. Riesling and Felicia are green varieties, while Regent is a red one. All images were recorded with the Phenoliner at the beginning of June, shortly after thinning.

the reader to Zabawa et al. (2020). These images in combination with the annotations were used as training data for our berry detection pipeline.

2.2.2. Reference data yield

The reference data for the yield estimation experiment were collected from the reference plants (see last column of Table 1). Corresponding to each reference plant one meter of canopy was harvested and weighed manually at the end of the season from 2018 - 2020. Yield estimation was restricted to Riesling.

2.2.3. Reference data berry weight

Destructive samples were collected from the remaining vines, which did not overlap with the reference vines. These exemplary grape bunches were collected at three points in time, or precisely at three different BBCH stages (Alercia et al., 2009) (see Fig. 3), to monitor the plant performance over the season. The first observed stage was BBCH 75, with pea sized berries. At BBCH 79 the majority of berries are touching each other. The last stage was

(a) BBCH75 (b) BBCH79 (c) BBCH89

Figure 3: The images show the same Riesling plant observed in the different BBCH stages. From left to right the berry size increased and the color changed.

BBCH 89, where the grapes are ready for harvest.

At each of these points in time ten grape bunches were collected per row and investigated in detail in the laboratory. Each bunch was weighed, measured manually and observed with a handheld scanner (Rist et al., 2018), delivering grape bunch characteristics like bunch length and width, berry number and weight, average berry size per bunch, as well as the bunch volume. To use an representative measure, the berry weight measures at the end of the season were averaged over the three years, for each variety and training system. This value was used as the weight factor in the yield estimation.

2.2.4. Leaf occlusion

To account for the percentage of occluded berries, we conducted two leaf occlusion experiments in 2019 and 2020. In both years, Riesling plants were observed in both training systems, VSP and SMPH. We collected data with the Phenoliner twice, first with full leaf occlusion and at the second time with removed leaves. We observed 35 plants in the VSP (5 in 2019 and 30 in SMPH) and 28 in the SMPH (10 in 2019 and 18 in 2020).

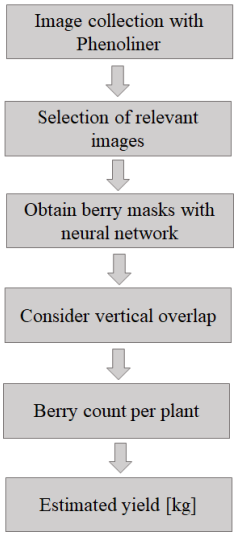


Figure 4: Pipeline illustrating the steps from image acquisition to the count on plant level.

2.3. Yield estimation pipeline

We propose an image-based approach for the detection, counting, and yield estimation of grapevine berries in overlapping image sequences (see Fig. 4). In the following, we will explain the single steps of the pipeline. We briefly explain the handling of single images which was already discussed in detail in Zabawa et al. (2020). This approach was extended to include the handling of vertically overlapping images and taking the information from single images to the plant level. Moreover, we will discuss the image selection process from the horizontal image sequence and explain our approach for yield estimation.

2.3.1. Detection of grapevine berries in single images

The detection of single berries was realised with a convolutional neural network (CNN) (Long, Shelhamer and T.Darell, 2015) which performs a

semantic segmentation following the approaches presented in Zabawa et al. (2020). The network has an encoder-decoder structure, the encoder backbone consists of an MobilenetV2 (Sandler, Howard, Zhu, Zhmoginov and Chen, 2018) and the decoder head of a DeepLabV3+ (Chen, Zhu, Papandreou, Schroff and Adam, 2018). The network performs a semantic segmentation with the classes 'berry', 'edge' and 'background' (see Fig. 6a). Each single berry is surrounded by an edge, enabling the identification of single object instances, which is usually realised with computation-intensive instance segmentation approaches. On top of the CNN, we apply a post-processing step to reduce the amount of misclassifications by using knowledge of the berry geometry and quality aspects of our predictions. To count the number of berries per image, we discard the class 'edge' and count the remaining objects of the class 'berry' with a connected component algorithm. For more details we refer the reader to Zabawa et al. (2020).

For this work, the hyper-parameters were not changed, but we added additional images to the training dataset, increasing the amount of training data to 59 images from the data described in Sec. 2.2.1, covering more variations over the years 2018 to 2020. The 59 images with pixel-wise annotations contain 38 images of plants trained in the SMPH, and 21 in the VSP. We annotated more images showing SMPH plants, since the conditions are more challenging. 30 images show Riesling, 18 Felicia and 11 Regent. The annotated images did not show the mentioned reference plants which were observed for yield estimation purposes. For detailed information regarding the data split for the training process we refer the reader to Zabawa et al. (2020).

Table 2: Overview of the annotated images which were used to train the CNN for the berry detection, with respect to training system and variety.

	Riesling	**Felicia**	**Regent**
VSP	15	4	2
SMPH	15	14	9

2.3.2. Counting of berries in vertically overlapping images

Our camera system covers 1.5 m height of canopy by providing three vertically overlapping images. In order to prevent a distorted count, objects which are present more than once were identified in overlapping regions. realise this by matching grape bunches from one image to the other one using the normalised correlation coefficient. To ensure a sufficient similarity between the found patches, we set a threshold of 0.3 for the matched patches and only allow the best result, if the matching value is higher than the threshold. The single steps of the full pipeline are illustrated in Fig. 5.

In detail, we extract image patches around objects of interest based on the CNN predictions and match them to another, overlapping image. In contrast to the counting process, where the object level is of interest, i.e. single berries, the object level is insufficient for the matching of image patches because individual berries are too small and look very similar to each other. Therefore, we join the classes 'edge' and 'berry' to a new mask, which represents information corresponding to the bunch level (see Fig. 6b). In the following, we denote this as 'bunch mask'. To simplify the matching problem, we mask the original image with the bunch mask (example can be seen in

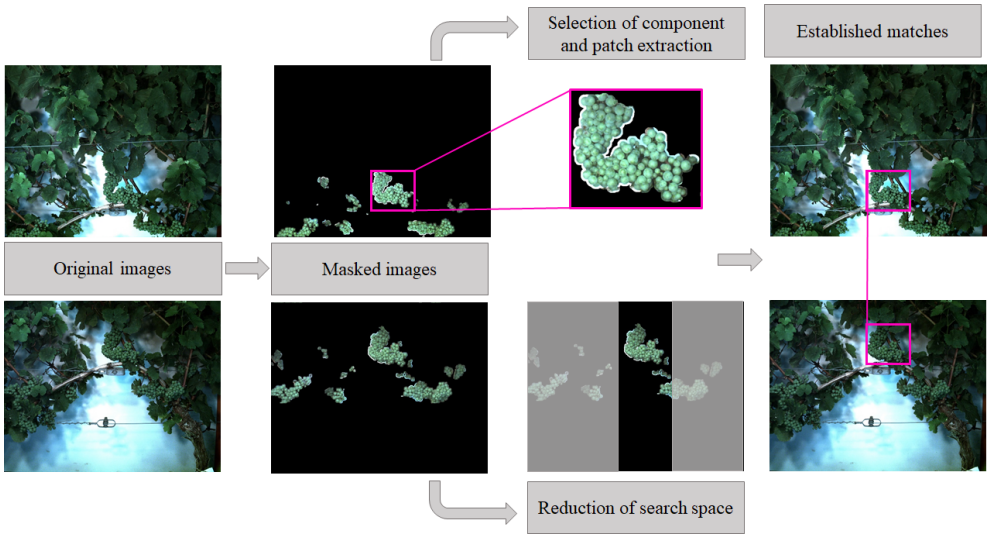
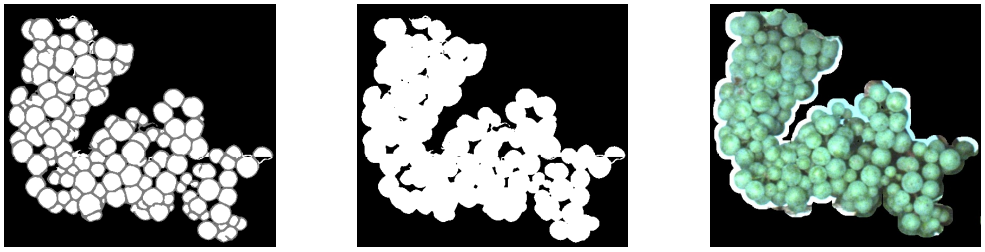


Figure 5: Pipeline for the patch matching process. The input are two images, which are masked with the bunch mask. A patch is extracted in the upper image and the search space in the lower image is reduced due to an approximately known image orientation. The corresponding bunches are identified after correlating the patch with the image.

Fig. 6b), to identify regions which contain berries. These masks are used to remove the unnecessary canopy information. The result is an image, where every pixel is black except the berry regions (see Fig. 6c). In the following, we denote this masked image as the 'berry region image'.

For the VSP the grape bunches are compact and single berries are mostly joined in grape clusters. In contrast to this, images showing the SMPH training system contain many single berries since the grape bunch structure is very loose. Therefore, we dilate the boundaries of each object three times to join single berries with neighboring ones. For the VSP, we perform the dilation operation only once.

However, since this can lead to joint objects which can cover up to half of



(a) Berry-edge mask (b) Bunch mask (c) Masked image patch

Figure 6: Fig. 6a shows an extract from the berry-edge mask which is the output of the neural network (Zabawa et al., 2020). White shows pixels corresponding to the class 'berry' while grey pixels are the class 'edge'. For the matching of image patches into another image we join both classes (see Fig. 6b) and use these to mask the original images (see Fig. 6c). This reduces the patch matching problem to regions containing berries.

the image, we divide large bounding boxes into smaller ones. The threshold to divide bounding boxes is chosen manually and corresponds to an average grape bunch size, in our case 400 pixels in vertical and horizontal direction.

For the matching process, we divide the three vertically overlapping regions into two parts each. The three overlapping images are denoted as top, middle and bottom image. The lower half of the top image is matched to the upper half of the middle image. The lower half of the middle image is matched to the upper part of the bottom image. There is no image region where all three images overlap.

Since the cameras are mounted on top of each other and the intrinsic and extrinsic camera calibration parameters are known, we can reduce the search space in vertical and horizontal direction. Since the cameras are vertically aligned, we only accept matches approximately in the same columns of the

image with a small offset of 20 pixels to the left and the right. Furthermore we set a depth threshold for valid matches. The depth or more precisely the distance between camera and detected object is computed using the known calibration and the disparity from the matched patch pairs. The minimal allowed depth is set to 0.3 m, while the maximal allowed depth is 1.3 m. The maximum value is chosen based on the maximum available space in the measurement chamber.

2.3.3. Processing of a whole row

Due to the movement of the sensor platform, we are also able to cover the whole row vertically. In the described experiments we set the system to capture as many images as possible, which was about one set of images every 4cm. This large amount of images allows various different evaluation and interpretation strategies and also enables the 3D reconstruction of the canopy using structure from motion methods (Rose et al., 2016). The latter, however, is beyond the scope of this paper.

To deal with the horizontal overlap of the images, it would be possible to use the same approach as described in Sec. 2.3.2, including the processing of all images taken in a row and the identification of recurring matches in subsequent images. Due to the high overlap between the images, this approach would be very time consuming. Instead, we decided to perform an image selection based on the GNSS coordinates, optimizing for maximal coverage and minimal overlap.

For this, we utilise the camera calibration and the approximate distance between the camera and the canopy. Using this knowledge, we are able to calculate the distance between camera positions which is needed to acquire

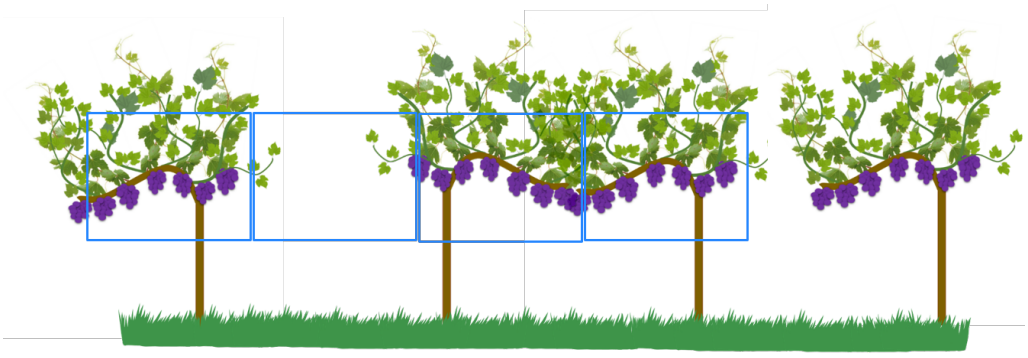


Figure 7: Image selection based on the geo-referenced image coordinates with the goal of minimal overlap between selected images.

non-overlapping images as can be seen in Fig. 7. For our sensor system a distance of 0.8 m between camera positions ensures a complete coverage of the row while using the minimal amount of necessary images. Since the image acquisition is approximately every 4 cm, in some cases grape bunches are visible in the edges of two consecutive images. These overlapping regions are not critical, since the bunch is visible from two different sides and the observed information is therefore independent.

2.4. Yield estimation

As mentioned in Sec. 1, various approaches exist for the estimation of yield ranging from destructive methods, which require the removal and investigation of samples, to counting of certain yield components in combination with reference values collected over several seasons (Clingeleffer, Dunn, Krstic and Martin, 2001, Coviello, Cristoforetti, Jurman and Furlanello, 2020). These approaches can be used at different points in time to forecast yield with a yield function (de la Fuente, Linares, Baeza, Miranda and Lissarrague,

2015). Usually, weight factors like an average bunch or berry weight were collected over several years and taken into account. A prominent approach was presented by Coviello, Cristoforetti, Jurman and Furlanello (2020) which, however, relied on manual sampling. They described the yield initially as a function of the number of vines per surface unit N^v, the number of grape bunches per vine N^{gb} and the average grape bunch weight W^{gb}:

$$Y = N^v \cdot N^{gb} \cdot W^{gb} \tag{1}$$

The average grape bunch weight can be replaced by the number of berries per bunch N^{bb} and the average berry weight W^b:

$$Y = N^v \cdot N^{gb} \cdot N^{bb} \cdot W^b \tag{2}$$

We follow Nuske et al. (2014) who proposed a minimalistic approach inspired by Clingeleffer, Dunn, Krstic and Martin (2001). They expressed the yield at harvest as a function of the berry number N^b and an average berry weight W^b.

$$Y = N^b \cdot W^b \tag{3}$$

In our case, the total number of berries N^b cannot be measured directly. Since only a portion of the berries N^{vis} is visible in images, we have to take the following aspects into account: berries which are occluded since they are on the other side of the grape bunch; whole bunches which are on the other side of the canopy; and bunches which are covered by leaves. We take theses

invisible berries into account with a visibility factor P^{vis}:

$$Y = (N^{vis} + P^{vis} * N^{vis}) \cdot W^b \tag{4}$$

2.5. Evaluation measures

We applied various measures for the evaluation of our methods. We investigated the counting of berries in overlapping images and the yield estimate by computing the coefficient of determination (R^2), which describes the correlation between the reference count and the output of our pipeline.

Furthermore we compute the mean absolute error (MAE), defined as

$$\text{MAE} = \frac{1}{N} \cdot \sum_{n=1}^{N} \left| C_n - C_n^{\text{ref}} \right|, \tag{5}$$

as well as the mean squared error (MSE), defined as

$$\text{MSE} = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} \left(C_n - C_n^{\text{ref}} \right)^2}. \tag{6}$$

The variable C_n denotes the estimated berry count while C_i^{ref} represents the manually acquired reference for image n. The MAE represents a measure of accuracy while the MSE gives information about the robustness. Besides the absolute value we also add the measures in percent with respect to the mean reference to provide a better interpretability.

3. Results

The counting of grapevine berries in single images was evaluated extensively in Zabawa et al. (2020). Therefore we performed the following ex-

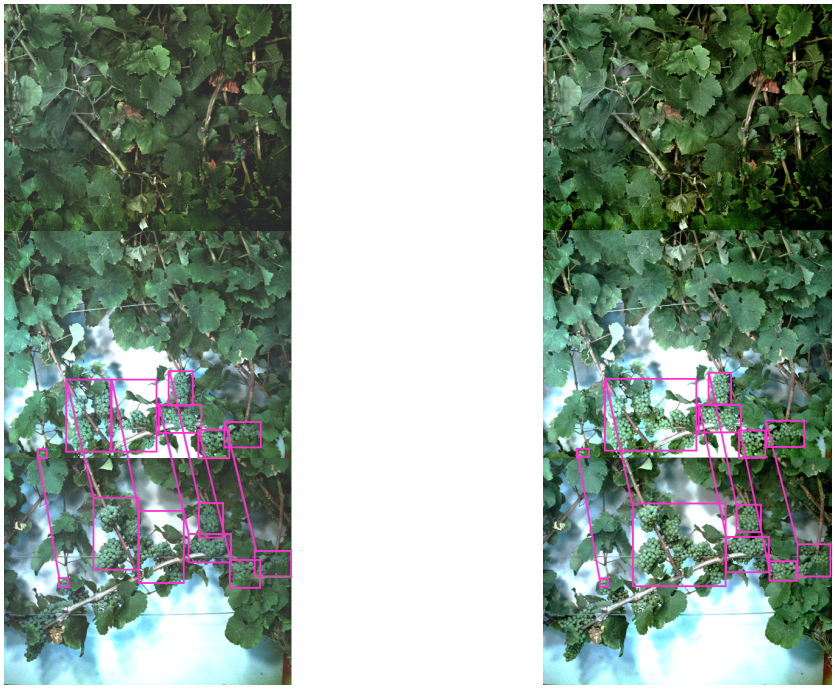
periments to evaluate our pipeline with respect to the performance on plant level:

1. evaluation of vertically overlapping images,
2. empirical investigation of repeatability,
3. investigation of leaf occlusion with respect to different training systems, and
4. yield estimation.

3.1. Vertical overlap

We verified our matching algorithm to the manual selection done by a human for 10 plants in the VSP and 10 trained in the SMPH system. Each plant was covered by three vertically overlapping images. Our pipeline extracts the number of berries per image by counting the connected components in the berry masks predicted by our neural network. The overlap is taken into account by matching image patches containing grape bunches from the overlapping regions (see Fig. 8a), leading to a count per image set. The reference data consisted of a manual berry count for each image and a manual identification of areas which were observed in two images (see Fig. 8b). Therefore, we jointly investigated the berry count and the vertical matching process.

For the VSP we achieved a correlation between manually and automatically counted berries per plant of 91.6%. The minimum number of berries per plant was 493 berries and the plant with the most berries showed 911 berries, the mean was 649. The MAE was 45 (6.9%) berries and the MSE 56 (8.6%).



(a) Automatic vertical matching (b) Manual vertical matching

Figure 8: Comparison of the matching for vertically overlapping images. The left image shows the manual reference matching while the right side shows the matching performed with our method.

For the SMPH the correlation was slightly worse with 89.0%. This is in line with the results for counting in single images in Zabawa et al. (2020). In general the counting of berries in the SMPH system is more difficult due to the inhomogeneous berry sizes. The number of berries per plant was between 1027 and 2066 berries with an average number of 1359 berries. The MAE was slightly lower compared to the VSP, with 86 (6.3%) berries while the MSE was 114 (8.1%) berries.

3.2. Repeatability and horizontal overlap

We conducted a further experiment to provide information, how repeatable the counting pipeline of our method is. Therefore we observed two rows, one of each training system, three times each. We evaluated the collected data from all six drives with our image selection process based on the camera positions, and the following counting on vertically overlapping image sets.

Table 3: Counted berries per row VSP

Drive	**# Selected images**	**# Berries**	**Mean**	**Standard deviation**
1	32	14399.0		
2	32	14322.0	14685.7	564.5
3	32	15336.0		

Table 4: Counted berries per row SMPH

Drive	**# Selected images**	**# Berries**	**Mean**	**Standard deviation**
1	32	19950.0		
2	34	21923.0	20530.0	1017.6
3	33	19935.0		

In both cases we achieved similar results for each of the three drives. For the VSP row we detected in average 14686 berries with a standard deviation of 564 berries (3.8%). The number of selected images was the same for each of the three drives.

(a) plant before leaf removal (b) plant after leaf removal

Figure 9: Both images show the same plant in the vertical shoot positioned system. Fig. 9a shows the plant before the leaf removal while Fig. 9b shows the plant without leaves. Only a small portion of the berries where covered by leaves.

For the second row, which was trained in the SMPH, we have different numbers of images which are selected per drive. During the second drive, problem with the GPS positioning occurred which influenced the image selection process. This led to a mean berry number of 20530 berries with a standard deviation of 1017 berries (5.0%).

3.3. Leaf occlusion

The estimation of yield is highly dependent on the number of visible yield components. In the VSP system, the influence of leaf occlusion is relatively small since most of the grape bunches are positioned at the lower part or even beneath the canopy in the so called fruiting zone (see Fig. 9). SMPH, on the other hand, has a high proportion of occluded berries with its bush-like canopy structure (see Fig. 10).

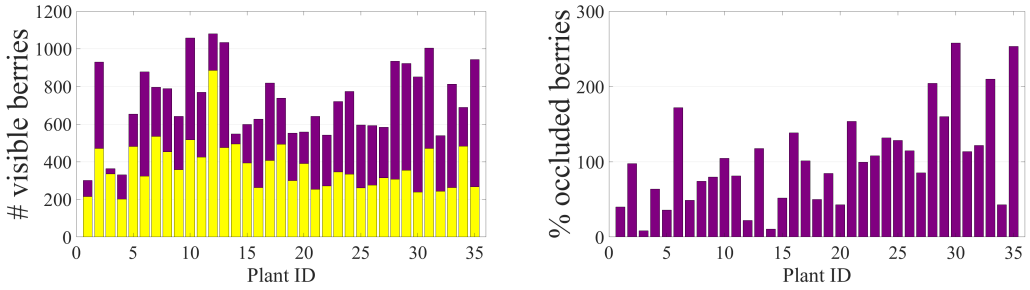
Between the two different training systems and both years, the percentage

(a) plant before leaf removal (b) plant after leaf removal

Figure 10: Both images show the same plant in the semi minimal pruned hedge system. Fig. 10a shows the plant before the leaf removal while Fig. 10b shows the plant without leaves. Most of the berries which can be seen on the right side are not visible on the left due to the leaf occlusion.

of covered berries was highly variable. The VSP featured less but larger berries in general. Since most of the grape bunches grew in the lower part of the canopy, the occlusion was low as well (see Fig. 11). In 2019, in average 49% ±33% more berries were visible after leaf removal. For the single plant, the occlusions ranged between 8% up to 97%. In 2020, the average occlusion was higher with 112% ±62%. But here, similar to 2019, the numbers between single plants varied between 11% to 250%. Details for the single plants can be found in Fig. 11. We use the mean occlusion rate over both years for the VSP (70% ±60%) as the visibility factor $P^{vis} = 0.7$ in the following yield estimation experiment.

On the contrary, the amount of invisible berries was large for plants in the SMPH system. Since the grape bunches were present all over the canopy,

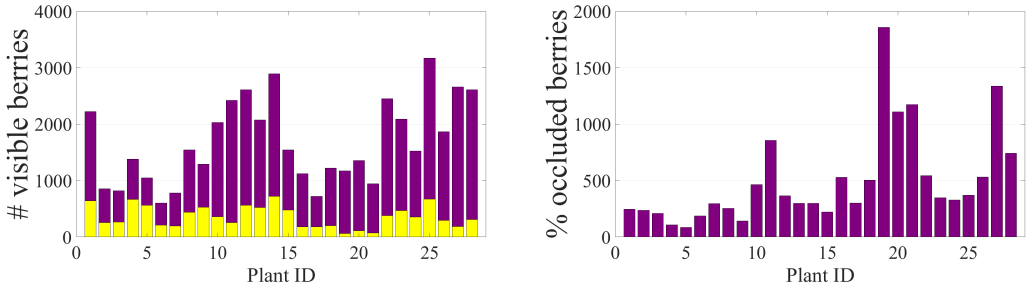


(a) Berry number (b) Percentage of occluded berries

Figure 11: Details of the leaf occlusion experiment for the VSP in 2019 and 2020. The left side shows the number of berries before leaf removal (yellow) and after the leaf removal (violet). The right side shows gain of visible berries in percent in violet respectively.

the occlusion level was high. In 2019 in average 223% ±108% more berries were visible after leaf removal. The spread ranges from 86% to up to 466% of covered berries. In 2020, in average even 446% ±440% more berries were visible. Here, the range was from 298% up to 1857% (see Fig. 12).

The presented results indicate that it is difficult to determine a single factor which accounts for the number of berries which are covered by leaves. The covered berry number deviates highly within the plants of the same training systems and the two years. We underline our findings with Fig. 13 showing histograms of the percentage gain in berry visibility. For none of the training systems we can identify a single maximum or a specific distribution. For the VSP we can see that the distribution is wide spread between 0 to 175%. The SMPH shows an especially large variation in the factor for the leaf coverage up to 400% and many outliers in the direction of large values.



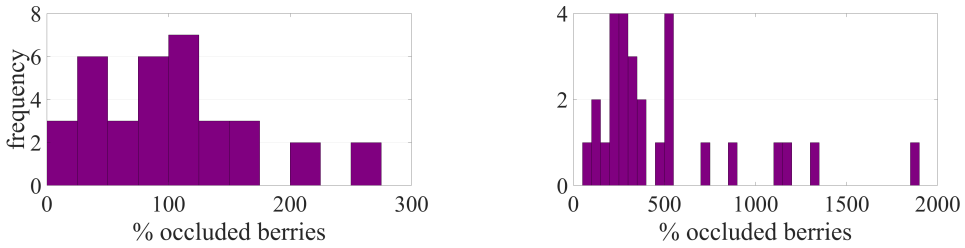
(a) Berry number (b) Percentage of occluded berries

Figure 12: Details of the leaf occlusion experiment for the SMPH in 2019 and 2020. The left side shows the number of berries before leaf removal (yellow) and after the leaf removal (violet). The right side shows gain of visible berries in percent in violet respectively.

3.4. Yield estimation

In the following, we correlated the number of visible berries with the yield at harvest and analysed the relationship between both. Fig. 14 and Fig. 15a show the number of visible berries in context with the yield per plant. We used data from three consecutive years to estimate a yield function. In each year we observed the same 20 plants in the VSP and the SMPH. As can be seen in Fig. 14, clusters representing the three years can be identified. Furthermore we can see that the relation between the number of visible berries and the yield seems to be nearly linear, with a slight plateau in the middle. For the SMPH no apparent relationship is observable, especially for the year 2019. This is also supported by the results of our leaf occlusion experiment, which shows a high variation of the leaf occlusion. Therefore we focus on the VSP in the following.

Beside the number of yield components, we need to take a weight fac-



(a) Histogram VSP (b) Histogram SMPH

Figure 13: Histogram of the percentage gain of visible berries after yield removal. In the VSP we observed 35 plants, with a maximal gain in berry visibility of 300%. Therefore we chose a bin size of 25%. For the SMPH we observed 28 plants, but the spread goes up to 1900%, we chose a larger bin size of 50%.

tor into account, in our case a mean berry weight which was averaged over three years. This weight factor was extracted from manual measurements, by weighing multiple grape bunches over three years and for both training systems. With this information we determined an average weight per berry at harvest time which poses the weight factor for our yield estimation. In our case the berry weight was estimated as 1.53 g and was very homogeneous over the different years with a standard deviation of 0.13 g. This corresponds to 8.4% of the estimated mean berry weight.

Fig. 16 shows three attempts to estimate the yield for 20 plants in the VSP. In a first attempt, we directly multiply the mean berry weight with the number of visible berries, which corresponds to a visibility factor $P^{vis} = 0$. This leads to an underestimation of the yield. We achieve an MAE of 1.19 kg (88.15%) and and MSE of 1.38 kg (101.62%). In a second attempt, we double the number of visible berries to account for the parts which are not seen by the camera ($P^{vis} = 1$). This decreases the MAE to 0.78 kg (28.95%)

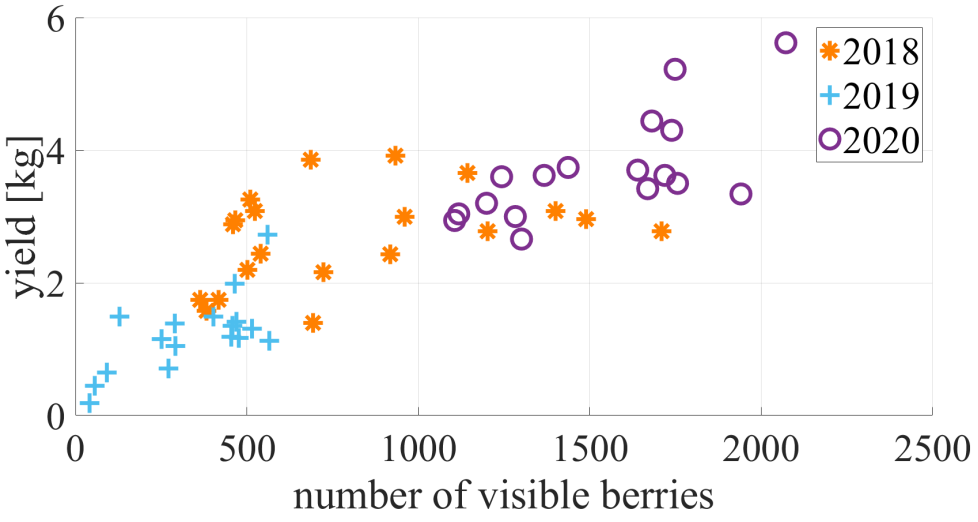
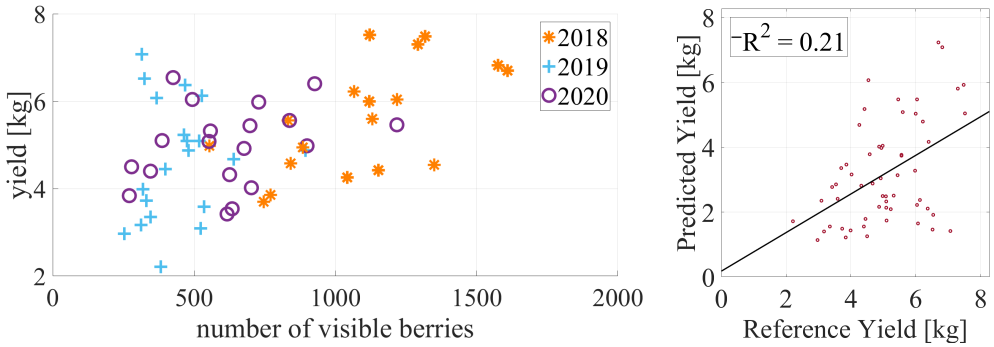


Figure 14: The plot shows the number of visible berries with respect to the yield at the end of the season in kg for the VSP. The orange stars show plants observed in 2018, blue plus in 2019 and violet circles in 2020.

and the MSE to 1.00 kg (37.06%). And lastly, we add 70% of the number of visible berries ($P^{vis} = 0.7$), following the leaf occlusion experiment, which yields a MAE of 0.67 kg (27.37%) and an MSE of 0.87 kg (36.05%). For all experiments the correlation between the predicted and estimated yield is the same, $R^2 = 0.69$. Since the berry weight is a constant factor with relation to the berry number, the spread of the results is not affected, only the MAE and MSE change.

For SMPH, the relationship between the number of visible berries and yield is not as obvious, as indicated by the greater variation between the number of visible berries and the yield within years (see Fig. 15a). This is especially prominent for 2019, where the berry numbers are relatively small but the yield has a great range compared to the other two years. This further



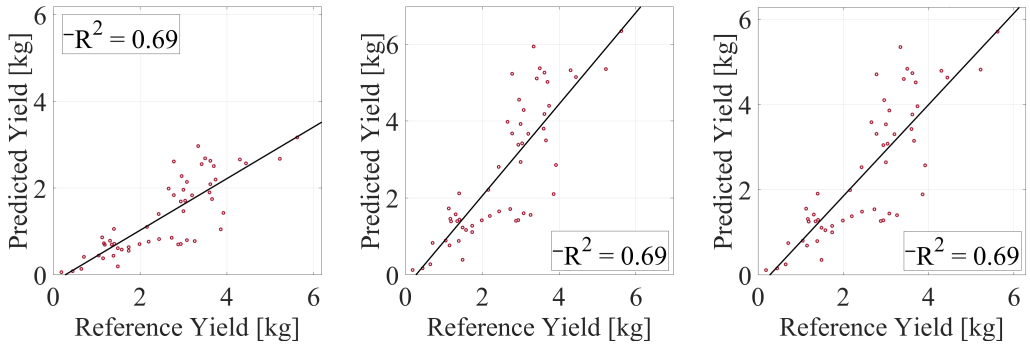
(a) berry number vs. yield (b) $6 \cdot N^{vis} \cdot W^b$

Figure 15: The left plot shows the number of visible berries with respect to the yield at the end of the season in kg for the SMPH. The orange stars show plants observed in 2018, blue plus in 2019 and violet circles in 2020. The right side shows the correlation between the estimated yield from the number of visible berries with respect to the reference yield.

strengthens our observation of the high variability of the leaf occlusion for the SMPH.

4. Discussion

Most berry detection approaches investigate plants trained in the VSP or similar training systems. Therefore the regions of interest are focused on the lower part of the canopy and were observed with a single camera, either handheld (Aquino, Diago, Millan and Tardaguila, 2016), stationary (Diago et al., 2012) or mounted on moving sensor platforms (Nuske et al., 2014), (Aquino, Millan, Diago and Tardaguila, 2018). Our system differs in this respect, since we covered most of the vertical canopy. This was motivated by the fact, that for the SMPH most berries grow in the upper part of the



(a) $N^{vis} \cdot W^b$ (b) $2 \cdot N^{vis} \cdot W^b$ (c) $(N^{vis} + 0.7 \cdot N^{vis}) \cdot W^b$

Figure 16: Correlation between the actual yield and the predicted yield. On the left we multiplied an average berry weight W^b with the number of visible berries N^{vis}. In the middle we doubled the number of visible berries to account for the non visible parts. On the right we added 70% of the number of visible berries.

canopy.

The vertical overlap was realised with the usage of three cameras which were positioned on top of each other. This is comparable to other works which investigated the influence of multiple viewpoints on the fruit detection (Hemming, Ruizendaal, Hofstee and Henten, 2014, Kurtser and Edan, 2018a). In our case we did not tackle an active viewpoint planning, but make use of the vertical camera positions to select the image which offers the best view of the crops. In our case this means that we select the matched image patch with the highest count of visible berries.

Besides a vertical overlap, the approaches working on field phenotyping platforms have different methods to handle redundant information in the horizontal direction, caused by the movement of the sensor platform. For example, (Aquino, Millan, Diago and Tardaguila, 2018) used a consumer

software to match image regions with approximately 50% overlap before the processing stage. (Nuske et al., 2014) on the other hand projected the berry candidates onto a local fruit wall. Both approaches worked well for simple geometries with little variations. In these cases the vines were partly defoliated and the grape bunches fully visible. We chose to select images based on the recording position, to achieve a minimal overlap between images. Especially the projection onto a local fruit wall would be difficult for the SMPH with it's loose and complex bunch structure in combination with the small berry size.

Another important aspect besides the vertical and horizontal coverage of the rows is the visibility of the investigated yield components. In our case, the state of occlusion when acquiring images in vineyards is a crucial point for the accuracy of berry number detection. Furthermore the occlusion is highly dependent on the training system. As we showed, plants trained in the SMPH system have a very high and variable rate of occlusion as they are not defoliated manually at any stage of the growing season. (Kliewer and Dokoozlian, 2005) used the leaf area/ crop weight ratio as a mean to investigate the wine quality. Diago et al. (2012) performed leaf removal experiments in order to evaluate different leaf areas and cluster exposures. Pereira, Morais and Reis (2018) performed a leaf segmentation with the later goal of a variety classification based on leaf characteristics. Gao et al. (2020) classified the images in different occlusion stages and used theses as a decision base for the fruit picking strategy of apples. Similar to this, Koirala et al. (2021) investigated 5 different methods for the estimation of mango tree fruit load. They found that the automatic estimation of an occlusion factor per

tree was necessary due to high variability between trees. This however was hard to achieve since no direct correlation between the number of partly occluded and fully visible fruits could be found. The further explored the potential of the direct estimation of fruit count per tree based on the count of fully exposed and partly occluded fruit, as well as the canopy area and achieved good results within one growing season, but failed to extend their approach to another season. We performed leaf occlusion experiments over two years. A yield estimation is highly dependent on the number of visible yield components, in our case the berry number. Therefore the ratio of invisible berries compared to the visible berries is a crucial factor. In the scope of our experiments we saw, that the ratio of invisible berries varied greatly between the two different training systems and even the years. The influence on the plants trained in the VSP was smaller, since a large portion of berries was already visible due to defoliation of the fruit zone after fruit set and the fact that most berries grew in the lower part of the canopy and were therefore not covered. The case was different for the plants in the SMPH. The variation within the training system was very large, ranging from 100% up to nearly 2000%. This is also reflected in the mean occlusion and the respective standard deviation. Both values are exceptionally large which also reflect on the large variability of the occlusion. This makes any attempts at estimating the yield based on the number of visible berries extremely challenging.

For VSP systems, on the other hand, there are several different defoliation stages done in viticulture practice. From no defoliation up to a complete defoliation of the fruit zone. In our study, a moderate defoliation of the fruit zone was performed after fruit set, leading to a partially defoliated fruit zone

due to regrowth, a method often used in German viticulture. Nuske et al. (2014) for instant used images with a fully defoliated fruit zone achieving an average error between 3% and 11% of total yield and thus, obtained more accurate results compared to our pipeline. Nevertheless it was very important to us to do the image acquisition under German viticulture standard conditions rather than under trail specific optimised image standardisation, in order to get an idea how this method could be transferred into an application used in viticulture practice.

Apart from the berry visibility the estimated weight factor plays a role for the yield estimation. We estimated the berry weight by averaging the measured berry weight over three years, resulting in a berry weight $W^b = 1.53g \pm 0.13g$. The standard deviation is 8.4% of the estimated weight. This agrees with Clingeleffer et al. (2001) who stated that the main factor for grapevine yield estimation is the number of berries, which account for 90% of the yield variation, while only 10% are influenced by the berry weight. We chose the highest number of visible berries in the VSP as a worst case scenario. For 2000 visible berries, the predicted yield without considering the leaf occlusion would be 3060 g. Since the weight factor is constant for the yield estimation. A reduction of 10% of the berry yield would lead to an estimate of 2754 g while an addition of 10% leads to 3.366 g. In comparison, taking into account the leaf occlusion with a visibility factor of 70% increases the yield by more than 2 kg, to 5202 g. Therefore the estimation of the leaf occlusion has a bigger impact then the variation of the berry weight.

For the VSP, we attempted an early yield estimation with the number of visible yield components and a factor which accounts for the invisible

berries. We tested three different approaches, correlating the number of visible berries and the reference yield directly and accounting for two different visibility factors , $P^{vis} = 1$ und $P^{vis} = 0.7$, for the number of invisible berries. In the end we achieved a MAE of 27% for $P^{vis} = 0.7$ in comparison to the reference yield. Our machine-driven approach showed the same accuracy compared to forecasts done by winegrowers, while being non-destructive, automatic and fast. Dunn (2010) compared five different yield forecasting methods at different points in time and showed that the mean absolute difference between winegrowers forecast and actual production was, with the best method, around 30% as well. After analysing many growers forecasts, Dunn and Martin (2003) suggested that although growers have a good understanding of average production over time they fail to adjust their forecast to the actual production of the year.

There are two ways to improve our yield forecast. One option is the improvement of the accuracy of berry detection. We assume that this would only lead to a small improvement, since the berry detection itself already achieves very good results. We identify the occlusion of yield parameters through leaves as the main limiting factor. The improvement of this could be the other option, e.g. by adaptions in the VSP system using a more defoliated management system. Approaches which were developed for view point planning in horticulture, especially applications for sweet pepper harvesting (Kurtser and Edan, 2018b, Harel, van Essen, Parmet and Edan, 2020) would be difficult to extent to viticulture. Furthermore the goal is not interaction with the crop, e.g. harvesting, but the observation.

5. Conclusions

We presented a yield estimation pipeline, which is an extension of an image based berry detection method from an earlier paper (Zabawa et al., 2020). We extended the method to work on horizontally and vertically overlapping images taken with a multi-camera system on a moving platform, the Phenoliner. We collected an extensive dataset of 767 plants (including 3 genotypes) planted in 15 rows, featuring two different training systems. All of them were observed over three years and destructive yield measurement of selected plants were performed. We evaluate the system on plant level with manual annotations and showed, that our systems achieves a MAE of the berry counting on plant level between 6 - 8% compared to the reference. Furthermore we showed that the results are reproducible and stable over multiple data collection processes.

We also analysed the influence of the leaf occlusion by dedicated experiments. We were able to show that the leaf occlusion in SMPH training systems has such a high variability that a relation between the number of visible berries and the yield is not possible. In VSP systems however, we were able to achieve an MAE of 26% for the yield estimation.

This is to be expected, since the estimation of yield is highly dependent of the amount of visible yield components. This visibility factor is the limiting factor for the estimation while the berry weight does not play a major role in comparison. The actual detection of the visible berries is the most uncritical part.

A way to improve our yield estimation would be to enhance the visibility of these components through management decisions regarding the leaf and

grape occlusion.

Acknowledgment

This work was funded by German Federal Ministry of Education and Research (BMBF, Bonn, Germany) in the framework of the project novisys (FKZ 031A349) and partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germanys Excellence Strategy - EXC 2070-390732324.

References

Alercia, A., Becher, R., Boursiquot, J.M., Carara, R., amd A. Costacurta, P.C., et al., 2009. 2nd edition of the oiv descriptor list for grape varieties and vitis species.

Anderson, N.T., Walsh, K., Wulfsohn, D., 2021. Technologies for forecasting tree fruit load and harvest timing—from ground, sky and time. Agronomy 11.

Aquino, A., Diago, M.P., Millan, B., Tardaguila, J., 2016. A new methodology for estimating the grapevine-berry number per cluster using image analysis. Biosystem Engineering 159, 80 – 95.

Aquino, A., Millan, B., Diago, M.P., Tardaguila, J., 2018. Automated early yield prediction in vineyards from on-the-go image acquisition. Computers and Electronics in Agriculture 144, 26 – 36.

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science 19, 52 – 61.

Bargoti, S., Underwood, J., 2017. Image segmentation for fruit detection and yield estimation in apple orchards. Journal of Field Robotics 34, 1039–1060.

Bendel, N., Kicherer, A., Klück, A.B.H.C., Seiffert, U., Fischer, M., Voegele, R., Töpfer, R., 2020. Evaluating the suitability of hyper- and multispectral imaging to detect foliar symptoms of the grapevine trunk disease esca in vineyards. Plant methods 16.

Bloesch, B., Viret, O., 2008. Stades phnologiques repres de la vigne. Revue Suisse de Viticult. Arboricult. Horticult 40, 1–4.

Cecotti, H., Rivera, A., Farhadloo, M., Pedroza, M.A., 2020. Grape detection with convolutional neural networks. Expert Systems with Applications 159, 113588. URL: https://www.sciencedirect.com/science/article/pii/S0957417420304127, doi:https://doi.org/10.1016/j.eswa.2020.113588.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. CoRR abs/1802.02611.

Clingeleffer, P., Dunn, G., Krstic, M., Martin, S., 2001. Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties: A national approach. Technical report, Grape and Wine Research and Development Corporation .

Coviello, L., Cristoforetti, M., Jurman, G., Furlanello, C., 2020. Gbcnet: In-

field grape berries counting for yield estimation by dilated cnns. Applied Sciences .

Darwin, B., Dharmaraj, P., Prince, S., Popescu, D.E., Hemanth, D.J., 2021. Recognition of bloom/yield in crop images using deep learning models for smart agriculture: A review. Agronomy 11. URL: https://www.mdpi.com/2073-4395/11/4/646, doi:10.3390/agronomy11040646.

Diago, M.P., Correa, C., Millán, B., Barreiro, P., Valero, C., Tardaguila, J., 2012. Grapevine yield and leaf area estimation using supervised classification methodology on rgb images taken under field conditions. Sensors 12, 16988 – 17006.

Dorj, U.O., Lee, M., Yun, S.S., 2017. An yield estimation in citrus orchards via fruit detection and counting using image processing. Computers and Electronics in Agriculture 140, 103–112. URL: https://www.sciencedirect.com/science/article/pii/S0168169916312455, doi:https://doi.org/10.1016/j.compag.2017.05.019.

Dunn, G.M., 2010. Yield forecasting, in: Fact Sheet June 2010, Australian Government, Grape and Wine Research and Development Corporation.

Dunn, G.M., Martin, S.R., 2003. The current status of crop forecasting in theaustralian wine industry, in: Proceedings of the ASVO Seminar Series: Grape-growing at the Edge, Tanunda, Barossa Valley, South Australia, pp. 4–8.

de la Fuente, M., Linares, R., Baeza, P., Miranda, C., Lissarrague, J.R., 2015. Comparison of different methods of grapevine yield prediction in the time

window between fruitset and veraison. OENO One 49, 27–35. URL: https://oeno-one.eu/article/view/96, doi:10.20870/oeno-one.2015.49.1.96.

Gan, H., Lee, W.S., Alchanatis, V., Abd-Elrahman, A., 2020. Active thermal imaging for immature citrus fruit detection. Biosystems Engineering 198, 291–303. URL: https://www.sciencedirect.com/science/article/pii/S1537511020302348, doi:https://doi.org/10.1016/j.biosystemseng.2020.08.015.

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M., Zhang, Q., 2020. Multi-class fruit-on-plant detection for apple in snap system using faster r-cnn. Computers and Electronics in Agriculture 176, 105634. URL: https://www.sciencedirect.com/science/article/pii/S0168169920314009, doi:https://doi.org/10.1016/j.compag.2020.105634.

Gennaro, S.F.D., Toscano, P., Cinat, P., Berton, A., Matese, A., 2019. Low-cost and unsupervised image recognition methodology for yield estimation in a vineyard. Frontiers in plant science 10, 559. doi:10.3389/fpls.2019.00559.

Gené-Mola, J., Gregorio, E., Auat Cheein, F., Guevara, J., Llorens, J., Sanz-Cortiella, R., Escolà, A., Rosell-Polo, J.R., 2020. Fruit detection, yield prediction and canopy geometric characterization using lidar with forced air flow. Computers and Electronics in Agriculture 168, 105121. URL: https://www.sciencedirect.com/

science/article/pii/S0168169919313390, doi:https://doi.org/10.1016/j.compag.2019.105121.

Gongal, A., Amatya, A., Karkee, M., Zhang, Q., Lewis, K., 2015. Sensors and systems for fruit detection and localization: A review. Computers and Electronics in Agriculture 116, 8–19.

Hacking, C., Poona, N., Manzan, N., Poblete-Echeverría, C., 2020. Investigating 2-d and 3-d proximal remote sensing techniques for vineyard yield estimation. Sensors 10.

Harel, B., van Essen, R., Parmet, Y., Edan, Y., 2020. Viewpoint analysis for maturity classification of sweet peppers. Sensors 20, 3783. doi:https://doi.org/10.3390/s20133783.

Hemming, J., Ruizendaal, J., Hofstee, J.W., Henten, E.J.V., 2014. Fruit detectability analysis for different camera positions in sweet-pepper. Sensors 14, 6032–6044. URL: https://www.mdpi.com/1424-8220/14/4/6032, doi:10.3390/s140406032.

Howell, G., 2001. Sustainable grape productivity and the growth-yield relationship: A review. American Journal of Enology and Viticulture 52, 165–174.

Ivorra, E., Sánchez, A., Camarasa, J., Diago, M.P., Tardáguila, J., 2015. Assessment of grape cluster yield components based on 3d descriptors using stereo vision. Food Control 50, 273 – 282.

Kalantar, A., Edan, Y., Gur, A., Klapp, I., 2020. A deep learning system for single and overall weight estimation of melons us-

ing unmanned aerial vehicle images. Computers and Electronics in Agriculture 178, 105748. URL: https://www.sciencedirect.com/science/article/pii/S0168169920304804, doi:https://doi.org/10.1016/j.compag.2020.105748.

Kicherer, A., Herzog, K., Bendel, N., Klück, H.C., Backhaus, A., Wieland, M., Rose, J.C., Klingbeil, L., Läbe, T., Hohl, C., Petry, W., Kuhlmann, H., Seiffert, U., Töpfer, R., 2017. Phenoliner: A new field phenotyping platform for grapevine research. Sensors .

Kicherer, A., Herzog, K., Pflanz, M., Wieland, M., Rüger, P., Kecke, S., Kuhlmann, H., Töpfer, R., 2015. An automated field phenotyping pipeline for application in grapevine research. Sensors 15, 4823–4836. doi:10.3390/s150304823.

Kicherer, A., Roscher, R., Herzog, K., Förstner, W., Töpfer, R., 2014. Image based evaluation for the detection of cluster parameters in grapevine, in: XI International Conference on Grapevine Breeding and Genetics 1082, pp. 335–340.

Kliewer, M., Dokoozlian, N., 2005. Leaf area/crop weight ratios of grapevines: Influence on fruit composition and wine quality. American Journal of Enology and Viticulture 56, 170–181.

Koirala, A., Walsh, K., Wang, Z., 2021. Attempting to estimate the unseen—correction for occluded fruit in tree fruit load estimation by machine vision with deep learning. Agronomy 11.

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., . Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'mangoyolo'. Precision Agriculture 20, 1107–1135.

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning – method overview and review of use for fruit detection and yield estimation. Computers and Electronics in Agriculture 162, 219–234. URL: https://www.sciencedirect.com/science/article/pii/S0168169919301164, doi:https://doi.org/10.1016/j.compag.2019.04.017.

Kurtser, P., Edan, Y., 2018a. Statistical models for fruit detectability: spatial and temporal analyses of sweet peppers. Biosystems Engineering 171, 272–289. doi:10.1016/j.biosystemseng.2018.04.017.

Kurtser, P., Edan, Y., 2018b. The use of dynamic sensing strategies to improve detection for a pepper harvesting robot, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8286–8293. doi:10.1109/IROS.2018.8593746.

Kurtser, P., Ringdahl, O., Rotstein, N., Berenstein, R., Edan, Y., 2020. In-field grape cluster size assessment for vine yield estimation using a mobile robot and a consumer level rgb-d camera. IEEE Robotics and Automation Letters 5, 2031–2038. doi:10.1109/LRA.2020.2970654.

Kurtulmus, F., Lee, W.S., Vardar, A., 2011. Green citrus detection using 'eigenfruit', color and circular gabor texture features under natural outdoor conditions. Computers and Electronics in

Agriculture 78, 140–149. URL: https://www.sciencedirect.com/science/article/pii/S0168169911001475, doi:https://doi.org/10.1016/j.compag.2011.07.001.

Linker, R., Cohen, O., Naor, A., 2012. Determination of the number of green apples in rgb images recorded in orchards. Computers and Electronics in Agriculture 81, 45–57. URL: https://www.sciencedirect.com/science/article/pii/S0168169911002638, doi:https://doi.org/10.1016/j.compag.2011.11.007.

Long, J., Shelhamer, E., T.Darell, 2015. Fully convolutional networks for semantic segmentation. Proceedings in the IEEE Conference on Computer Vision and Pattern Recognition , 3431 – 3440.

Lorenz, D., Eichhorn, L., Bleiholder, H., Klose, R., Meier, U., Weber, E., 1995. Growth stages of the grapevine: Phenological growth stages of grapevine (vitis vinifera l. ssp. vinifera) - codes and descriptions according to the extended bbch scale. Australian Journal of Grape and Wine Research 1, 133–103.

Matese, A., Gennaro, S.D., 2015. Technology in precision viticulture: a state of the art review. International Journal of Wine Research 156, 69 – 81. doi:/10.2147/IJWR.S69405.

Nellithimaru, A.K., Kantor, G.A., 2019. Rols : Robust object-level slam for grape counting, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2648–2656. doi:10.1109/CVPRW.2019.00321.

Nuske, S., Achar, S., Bates, T., Narasimhan, S., Singh, S., 2011. Yield estimation in vineyards by visual grape detection. IEEE/RSJ International Conference on Intelligent Robots and Systems , 2352–2358.

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., Singh, S., 2014. Automated visual yield estimation in vineyards. Journal of Field Robotics 31, 837 – 860.

Nyarko, E.K., Vidović, I., Radočaj, K., Cupec, R., 2018. A nearest neighbor approach for fruit recognition in rgb-d images based on detection of convex surfaces. Expert Systems with Applications 114, 454–466.

Pereira, C., Morais, R., Reis, M., 2018. Pixel-based leaf segmentation from natural vineyard images using color model and threshold techniques, in: ICIAR 2018: Image Analysis and Recognition, Springer International Publishing. pp. 96–106.

Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. PLoS One 8.

Rist, F., Herzog, K., Mack, J., Richter, R., Steinhage, V., Töpfer, R., 2018. High-precision phenotyping of grape bunch architecture using fast 3d sensor and automation. Sensors 18, 763. doi:10.3390/s18030763.

Roscher, R., Herzog, K., Kunkel, A., Kicherer, A., Töpfer, R., Förstner, W., 2014. Automated image analysis framework for high throughput determination of grapevine berry size using conditional random fields. Computers and Electronics in Agriculture 100, 148–158.

Rose, J., Kicherer, A., Wieland, M., Klingbeil, L., Töpfer, R., Kuhlmann, H., 2016. Towards automated large-scale 3d phenotyping of vineyards under field conditions. Sensors 16, 2136. doi:10.3390/s16122136.

Rudolph, R., Herzog, K., Töpfer, R., Steinhage, V., 2018. Efficient identification, localization and quantification of grapevine inflorescences in unprepared field images using fully convolutional networks. CoRR .

Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., Chen, L., 2018. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. CoRR abs/1801.04381.

Santos, T.T., de Souza, L.L., dos Santos, A.A., Avila, S., 2020. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. Computers and Electronics in Agriculture 170, 105247. URL: https://www.sciencedirect.com/science/article/pii/S0168169919315765, doi:https://doi.org/10.1016/j.compag.2020.105247.

Silver, D.L., Monga, T., 2019. In vino veritas: Estimating vineyard grape yield from images using deep learning, in: Advances in Artificial Intelligence, Springer International Publishing. pp. 212–224.

Stein, M., Bargoti, S., Underwood, J., 2016. Image based mango fruit detection, localisation and yield estimation using multiple view geometry. Sensors 16. doi:10.3390/s16111915.

Tagarakis, A., Koundouras, S., Fountas, S., Gemtos, T., 2018. Evaluation of the use of lidar laser scanner to map pruning wood in vineyards and

its potential for management zones delineation. Precision Agriculture 19, 334–347. doi:10.1007/s11119-017-9519-4.

Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., Zou, X., 2020. Recognition and localization methods for vision-based fruit picking robots: A review. Frontiers in Plant Science 11.

Wachs, J.P., Stern, H., Burks, T., Alchanatis, V., 2010. Low and high-level visual feature-based apple detection from multi-modal images. Precision Agriculture 11, 717–735. doi:https://doi.org/10.1007/s11119-010-9198-x.

Wang, Z., Walsh, K., Koirala, A., 2019. Mango fruit load estimation using a video based mangoyolo-kalman filter-hungarian algorithm method. Sensors 19.

Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Kuhlmann, H., Roscher, R., 2020. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing 164, 73–83.