UNIVERSITÄT BONN

Dissertation
zur Erlangung des Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
der Landwirtschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn
Institut für Geodäsie und Geoinformation

Contributions to image-based high-throughput phenotyping in viticulture

von

Laura Zabawa

aus
Bergisch Gladbach, Germany

Bonn 2023

Referent:
Prof. Dr. Heiner Kuhlmann, University of Bonn, Germany

1. Korreferent:
Prof. Dr. Ribana Roscher, University of Bonn, Germany

2. Korreferent:
Prof. Dr. Reinhard Töpfer, Julius Kühn-Institut, Institut für Rebenzüchtung Geilweilerhof, Siebeldingen

Tag der mündlichen Prüfung: 24. August 2023

Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

Abstract

GRAPEVINE is a crop with significant economic importance. Unlike many other crops, grapevine is a quality crop with the focus on yield optimization instead of maximization. Regular monitoring of plant diseases and pests is crucial, as they have the potential to cause significant losses by the end of the season. Therefore phenotyping plays an important role not only for breeding purposes, but also for the monitoring of the plant performance during the growth season. To ensure an optimal end product, management decisions, such as leaf removal, berry reduction, or spraying, should be implemented based on carefully extracted information. These measures are crucial to mitigate the risk of losses and maximize the overall profitability of grapevine cultivation.

Due to the perennial nature of grapevine and the often challenging terrain situation, phenotyping procedures are still performed manually, which is labor-intensive, expensive and subjective. To address these challenges, there has been a growing interest in developing non-invasive, sensor-based methods. These offer fast, affordable and reliable solutions that are objective and non-invasive.

This thesis addresses three challenges in the field of image-based high-throughput phenotyping in viticulture. Firstly, we propose a novel instance segmentation method for the detection and counting of grapevine berries in images captured in the field. We evaluate our approach across two different training systems and varieties, and compare it with two state-of-the-art methods. Secondly, we investigate the potential and limitations of using the counted number of visible berries for yield estimation. We identify the variability of the leaf occlusion as the primary limiting factor. Finally, we present two different approaches for grapevine berry anomaly detection. The first is a supervised classification method that produces heatmaps with a sliding window approach. The second is a semi-supervised approach that utilizes a Variational Autoencoder (VAE) to learn the representation of a healthy phenotype and identifies anomalies as deviations from this healthy phenotype.

Overall, this thesis makes contributions to the field of image-based high-throughput phenotyping in viticulture, by proposing novel solutions for grapevine berry detection and counting, yield estimation, and anomaly detection. The ap-

proaches we present are evaluated thoroughly and compared to state-of-the-art methods, demonstrating their effectiveness in addressing these important challenges.

Zusammenfassung

Die Weinrebe ist eine Kulturpflanze mit großer wirtschaftlicher Bedeutung. Im Gegensatz zu vielen anderen Kulturpflanzen steht bei der Weinrebe die Qualität im Mittelpunkt, nicht der Ertrag. Letzterer soll optimiert und nicht maximiert werden, um ein möglichst hochwertiges Endprodukt zu gewährleisten. Auch die regelmäßige Überwachung von Pflanzenkrankheiten und von Auswirkungen von Schädlingen sind wichtig, da sie am Ende der Saison erhebliche Verluste verursachen können. Daher spielt die Phänotypisierung nicht nur für Züchtungszwecke eine wichtige Rolle, sondern auch für die Überwachung der Pflanzen während der Wachstumsphase. Auf der Grundlage der gewonnenen Informationen müssen Managemententscheidungen getroffen werden, z. B. die Entfernung von Blättern, die Reduzierung von Beeren oder die Applikation von Fungi- und Pestiziden. Diese Maßnahmen sind von entscheidender Bedeutung, um das Risiko von Verlusten zu mindern und die Gesamtrentabilität des Weinbaus zu maximieren.

Aufgrund des mehrjährigen Charakters der Weinrebe und der oft schwierigen Geländeverhältnisse werden Phänotypisierungsverfahren immer noch manuell durchgeführt, was arbeitsintensiv, teuer und subjektiv ist. Um diese Herausforderungen zu bewältigen, besteht ein wachsendes Interesse an der Entwicklung nicht-invasiver, sensorbasierter Methoden. Diese bieten schnelle, erschwingliche und zuverlässige Lösungen, die objektive und nicht-invasiv erhobene Daten liefern.

Diese Arbeit befasst sich mit drei Herausforderungen auf dem Gebiet der bildbasierten Hochdurchsatz-Phänotypisierung im Weinbau. Erstens schlagen wir eine neuartige Instanz-Segmentierungs-Methode für die Erkennung und Zählung von Weinbeeren in Bildern vor, die im Feld aufgenommen wurden. Wir evaluieren unseren Ansatz mit verschiedenen Anbau-Systemen und Sorten und vergleichen ihn mit zwei State-of-the-Art Methoden.

Zweitens untersuchen wir die Ertragsschätzung auf Basis der Anzahl der gezählten, sichtbaren Beeren. Dabei werden die Potentiale und Einschränkungen betrachtet. Wir stellen fest, dass die Variabilität der Blattverdeckung der wichtigste limitierende Faktor ist. Schließlich stellen wir zwei verschiedene Ansätze zur Erkennung von Anomalien an Weinstöcken vor. Der erste ist eine überwachte Klassifizierungsmethode, die Heatmaps mit einem Sliding-Window-

Ansatz erstellt. Der zweite ist ein halbüberwachter Ansatz, der einen Variational Autoencoder (VAE) verwendet, um die Repräsentierung eines gesunden Phänotyps zu erlernen, und Anomalien als Abweichungen von diesem zu identifiziert.

Insgesamt leistet diese Arbeit einen Beitrag zum Bereich der bildbasierten Hochdurchsatz-Phänotypisierung im Weinbau, indem sie neue Lösungen für die Erkennung und Zählung von Weinbeeren, die Ertragsschätzung und die Erkennung von Anomalien präsentiert. Die vorgestellten Ansätze werden gründlich evaluiert und mit den State-of-the-Art Methoden verglichen, um ihre Effektivität bei der Bewältigung dieser wichtigen Herausforderungen zu demonstrieren.

Acknowledgement

I would like to express my sincere gratitude and appreciation to the following individuals who have played significant roles in the completion of my thesis:

First and foremost, I would like to extend my deepest thanks to my supervisors Professor Heiner Kuhlmann and Professor Ribana Roscher for their guidance and support throughout this research journey. Their expertise and insightful feedback have been invaluable to me. I am especially deeply grateful to Ribana, who not only served as a supervisor but went beyond that role, providing me with valuable advice on publication strategies and sharing the experience of my first international conference with me. Our time together was filled with both fruitful discussions and enjoyable moments.

I would like to acknowledge the immense contribution of Lasse Klingbeil, whose assistance in navigating research projects and fostering my passion for interdisciplinary work has been instrumental in shaping my academic pursuits. I am truly grateful for the knowledge and inspiration he has shared with me. In this context, I also need to express my appreciation to my colleagues for creating a pleasant work environment and engaging in lively discussions. These interactions provided me with different perspectives. A special mention goes to Diana Pavlic, my favorite office buddy, whose presence and friendship made each day brighter. Her support has been a constant source of motivation for me. I am immensely thankful to Anna Kicherer, for imparting valuable knowledge about grapevine breeding and enhancing my understanding of this field.

To my parents, I want to express my deepest gratitude. Thank you for always believing in me, supporting me unconditionally, and being there for me every step of the way. Your faith in my abilities has been a driving force behind my achievements. I would like to thank my sister, Lina, for being my best friend. Our valuable discussions and shared experiences have shaped my perspective and helped me grow both personally and academically.

Last but not least, I want to acknowledge Dominik, whose steady support and grounding presence have been my rock during stressful times. Thank you for reminding me of my worth, providing me with comfort, and believing in my abilities.

To all those mentioned and to those whose names I may have inadvertently

omitted, please accept my heartfelt appreciation for your contributions, guidance, and unwavering support throughout this endeavor.

This work was partially funded by German Federal Ministry of Education and Research (BMBF, Bonn, Germany) in the framework of the project Novisys (FKZ 031A349).

This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2070 – 390732324.

And finally, the work was partially supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program in the framework PhytoMo (project number 2818718C19).

Contents

Chapter 1

Introduction

VITICULTURE describes the cultivation and harvesting of grapevine (*Vitis vinifera*). Although it is one of the oldest horticultural practices it is still one of the most intensive cultivation systems, which requires a substantial amount of manual labor. This is due to the often challenging terrain situations, the perennial nature of the plants and the high quality standards that need to be achieved.

In 1.1 we will state the motivation for this thesis, then we will provide a list of publications which contributed to this thesis in Section 1.2. In Section 1.3, we explicitly highlight the contributions to shared first-authorship publications and lastly, we summarize the main contributions in 1.4.

1.1 Motivation

In contrast to many other agricultural applications, vineyard management practices focus on yield optimization (Howell, 2001) instead of maximization (Ray et al., 2013). This means, that throughout the season, actions need to be taken based on phenotypic observations, including defoliation for the optimal sun exposure of the berries or the thinning of berries to achieve the desired yield at the end of the season. Currently, these decisions are based on the observation of a practitioner who goes to the field and samples single plants in the vineyard. These results are extrapolated to the whole field, without the possibility to account for any potential heterogeneity throughout the vineyard.

Consequently, objective sensor-based high-throughput phenotyping in viticulture became a focus for many research projects throughout the years (Matese and Di Gennaro, 2015). Different sensor platforms were developed, ranging from handheld solutions (Kicherer et al., 2014; Diago et al., 2012), over semi- or fully automatic ground vehicles (Nuske et al., 2014; Kicherer et al., 2017) to Unmanned

Aerial Vehicles (UAVs) (Di Gennaro et al., 2019) The platforms are equipped with different sensors, like scanners (Rist et al., 2018; Tagarakis et al., 2018; Hacking et al., 2019) or cameras, including, RGB (Roscher et al., 2014; Nuske et al., 2014; Millan et al., 2019), RGBD (Kurtser et al., 2020), multi- or hyper-spectral (Gutiérrez et al., 2018; Bendel et al., 2020) cameras. With this thesis we contribute to the field of image-based in-field phenotyping. We developed a novel method to detect single grapevine berries in images and a pipeline for the row-wise estimation of yield based on the count of visible berries in images. Furthermore, we contributed two approaches for anomaly detection.

Depending on the region, different grapevine training systems and varieties are popular. In Germany for example, the Vertical Shoot Positioned (VSP) system is widespread and common. Since 2011 (Intrieri et al., 2011), the Semi Minimal Pruned Hedge (SMPH) gained popularity, due to the mechanization potential. For more details on the relevant training systems, we refer the reader to Section 2.1.1). The different training systems and grown varieties all have different challenges, including the amount of leaves, the structure of the grape bunches or their color. Hence it is important to develop algorithms which work on in-field images, regardless of the variety or the training system. We contributed an image-based approach to tackle this problem.

Economically important phenotyping tasks include the early estimation of yield as well as the detection of anomalies. For the first task, the economic significance is apparent, since a detailed yield estimation for the whole vineyard can guide precise management decisions with regard to berry thinning. The higher the quality of the grown grapes, the higher the quality of the end product and the resulting profit. For the second task, anomaly detection, the implications are more complex. Grapevine is very vulnerable to pests and fungi, which leads to a high usage of fungicides and pesticides. In Germany, for example, only 1% of the arable land is used for viticulture, yet it accounts for nearly 30% of all fungicide applications. An anomaly detection could enable a better and more targeted management response. At the end of the season, the detected anomalies could also be used for a targeted harvesting, to ensure a higher quality harvest.

In this thesis, we tackled three different topics, which contribute to high-throughput phenotyping applications in viticulture. The first topic is presented in Section 3.1 and tackles the problem of segmenting and counting yield components in images, which are recorded with a field-phenotyping platform in the field. The second part in Section 3.2 uses the method from the first part to investigate an image-based yield estimation pipeline and identifies the challenges and limiting

factors. Lastly, in Section 3.3, two different methods for anomaly detection are presented.

1.2 Publications

Parts of this thesis have been published in the following peer-reviewed conference and journal articles:

- **Publication 1** (Peer-reviewed, Conference Workshop):
  L. Zabawa, A. Kicherer, L. Klingbeil, A. Milioto, R. Töpfer, H. Kuhlmann, and R. Roscher. Detection of single grapevine berries in images using fully convolutional neural networks. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2571–2579, 2019b. doi: 10.1109/CVPRW.2019.00313

- **Publication 2** (Peer-reviewed, Journal):
  L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, H. Kuhlmann, and R. Roscher. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 164:73–83, 2020. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2020.04.002. URL `https://www.sciencedirect.com/science/article/pii/S0924271620300939`

- **Publication 3** (Peer-reviewed, Conference Workshop):
  J. Bömer, L. Zabawa, P. Sieren, A. Kicherer, L. Klingbeil, U. Rascher, O. Muller, H. Kuhlmann, and R. Roscher. Automatic differentiation of damaged and unharmed grapes using rgb images and convolutional neural networks. In *Proc. of the Europ. Conf. on Computer Vision (ECCV) Workshops*, pages 347–359. Springer International Publishing, 2020. ISBN 978-3-030-65414-6

  Jonas Bömer and Laura Zabawa hold a shared first authorship

- **Publication 4** (Peer-reviewed, Journal):
  L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, R. Roscher, and H. Kuhlmann. Image-based analysis of yield parameters in viticulture. *Biosystems Engineering*, 218:94–109, 2022. ISSN 1537-5110. doi: https://doi.org/10.1016/j.biosystemseng.2022.04.009. URL `https://www.sciencedirect.com/science/article/pii/S1537511022000861`

- **Publication 5** (Peer-reviewed, Journal):
  M. Miranda, L. Zabawa, A. Kicherer, L. Strothmann, U. Rascher, and

R. Roscher. Detection of anomalous grapevine berries using variational autoencoders. *Frontiers in Plant Science*, 13, 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.729097. URL `https://www.frontiersin.org/articles/10.3389/fpls.2022.729097`

Miro Miranda and Laura Zabawa hold a shared first authorship

The content of each publication is summarized in Chapter 3.

1.3 Collaborations

Some of the work included in this thesis has been done in collaboration with other researchers and resulted in publications with a shared first authorship:

The publication *Automatic Differentiation of Damaged and Unharmed Grapes using RGB Images and Convolutional Neural Networks* (Bömer et al., 2020) has a shared first authorship with Jonas Bömer and was the result of a student project. Jonas helped with the implementation of the Convolutional Neural Network (CNN) and the training procedures. I supervised the student project and supported the implementation, helped to conceptualize the work, provided and prepared the data and wrote the majority of the manuscript.

The publication *Detection of Anomalous Grapevine Berries Using Variational Autoencoders* (Miranda et al., 2022) was mainly done in collaboration with Miro Miranda. Miro helped with the implementation of the Variational Autoencoder (VAE), and conducted the network training. Together we designed and conducted the analyses. I provided and prepared the data and wrote the majority of the manuscript. The detailed contributions are also listed in the publication.

1.4 Main Contributions

In the following we will state the main contributions of each part of the thesis:

Detection and Counting

Detection and counting often pose as a preliminary task for yield estimation or anomaly detection. Depending on the target application, other information beside the object number itself or the position can be of interest, for example the object size and shape. This is particularly relevant for phenotyping applications, where these additional information can be very valuable. To translate these requirements into Machine Learning (ML) or Deep Learning (DL) terms, counting

can be realized with many different approaches, e.g. counting with regression, object detection or instance segmentation. Each of these problem formulations have their own advantages and disadvantages. For example, many instance segmentation approaches, like Mask-RCNN (He et al., 2017) deliver object positions, as well as shape and size of the object. It features a complex network structure with many parameters, which need a large amount of training data to optimize the model. Challenging situations include the detection of small objects or large quantities of items in images. Regression approaches on the other hand are lightweight, but give only the object number and the position. Therefore, we developed our own instance segmentation approach, which is based on a semantic segmentation. Besides the classes *berry* and *background*, we introduced a third class *edge* to separate between single instances of the same class. The network can be a lightweight CNN, without the need for an additional detection head or a large number of parameters. Therefore this approach is especially suited for small data sets. We first introduced this idea in Publication 1 (Zabawa et al., 2019b) and extended the evaluation in Publication 2 (Zabawa et al., 2020). We compared our method to two other approaches, including an instance segmentation and a regression approach and showed that our results outperformed both for our particular problem (Zabawa et al., 2020).

Yield Estimation

In Germany, different wine quality levels exist, which are determined by the yield per hectare. Therefore wine growers are interested in yield estimation methods, to optimize the yield to the respective desired quality level. Early yield estimations especially, can assist in making informed management decisions, like berry thinning. Traditionally, these estimations are performed by skilled experts, who sample single plants destructively in the vineyard, and estimate a yield based on the number and weight of certain yield components, for example grape bunches. The average achieved accuracy of these procedures is around 30% (Dunn and Martin, 2003), meaning that the actual yield in kg deviates by 30% from the predicted one. Hence, sensor-based approaches which can monitor the whole vineyard came into focus. Many of the works done in this direction are in highly defoliated vineyards, where the yield components are well visible. In Germany on the other hand, mild defoliation or even hedge like canopies are prominent. These types of training system or management styles highly influence the visibility of the grape bunches. In Publication 4 (Zabawa et al., 2022), we present a pipeline which automatically evaluates overlapping images from a whole grapevine row and perform a yield mapping for the medium defoliated rows trained in the VSP system. The achieved results of 27% yield variation are slightly better than the

average industrial standard, and show potential for improvement, if the defoliation is increased leading to a lower occlusion of the yield components. To quantify this, we performed a leaf-removal experiment to investigate the leaf-occlusion in detail. The result showed, that the leaf occlusion is the main limiting factor for our method, especially for more bush-like canopies in the SMPH training system.

Anomaly detection

Anomaly detection is one of the key tasks in phenotyping. Anomalies can be caused either by abiotic stress, for example drought or sunburn, or by biotic stress like fungi and insects. For some applications it is important to identify the exact damage cause to decide on a treatment. This can be very challenging, since sometimes different diseases look very similar and can only be distinguished on the molecular level (e.g. *Bois noir* or *Flavenscence Dorée*). Furthermore it is hard to collect sufficient data for each kind of anomaly, since most of the time the majority of the investigated plants are healthy. In other cases, the exact cause of damage is of subordinate interest, for example if the quality of the product is in the focus, or if a preliminary screening is enough to identify affected plants. We contributed with two works to the latter problem. In Bömer et al. (2020) we only defined the classes *healthy* and *damaged* and showed, that a shallow neural network was enough to classify in-field berry patches correctly. In combination with a sliding window approach we produced heatmaps highlighting anomalous areas. As an extension to this work we developed a new approach which trains only with non-anomalous plant material (Miranda et al., 2022). The network is a VAE which learned a healthy phenotype and identifies anomalies as variations from these. Important to note, is that we used a Feature Perceptual Loss (FPL), which yielded superior and sharper results, compared to a pixel wise loss (Miranda et al., 2022).

Chapter 2

Basic Techniques

This thesis describes contributions to image-based high-throughput phenotyping with special focus on viticulture. We will start with introducing some important viticulture terminology in Section 2.1. In Section 2.2 the important Deep Learning (DL) techniques will be introduced, including the general idea in Section 2.2.1 and special network structures which are used in the publications in Section 2.2.2 and Section 2.2.3. Finally the most common evaluation metrics will be explained in Section 2.2.4.

2.1 Viticulture

In order to understand the contributions of the presented publications, it's important to be familiar with some of the terminology frequently used in viticulture. Therefore, we will begin by providing a concise explanation of the two training systems discussed in the Publication 1 (Zabawa et al., 2019b), Publication 2 (Zabawa et al., 2020) and Publication 4 (Zabawa et al., 2022) in Section 2.1.1. Following that, we will briefly introduce the essential growth stages in Section 2.1.2.

2.1.1 Grapevine Training Systems

Training grapevines (*Vitis Vinifera*) involves the manipulation of the vine form (Reynolds and Vanden Heuvel, 2009) and depends on the terrain, regional climate and technical requirements of the vineyard. The training system influences for example the photosynthetic capabilities of the vine, the grape bunch structure as well as the vine microclimate. It also determines the degree of mechanization possible within vineyard management.

The Vertical Shoot Positioned (VSP) system is one of the most common training systems in Germany, it is typically used in cooler climates (Jackson, 1997). It

(a) VSP (b) SMPH

Figure 2.1: Schematic overview of two different grapevine training systems. Figure 2.1a shows the VSP system. Figure 2.1b shows the SMPH.(Fig. adapted from Zabawa et al. (2020)).

features one or two perennial canes with several thinner annual shoots branching off each season (see Figure 2.1a). After each season most canes and shoots are removed, annual shoots become canes in the following year. During the season, defoliation of the grape zone takes place to ensure optimal light conditions for the development of the berries as well as fast drying after rainfall or morning dew. This leads to a small and airy leaf wall with few but large leaves. Although this is an advantage, it can also lead to sunburn during summer. Grape bunches mainly occur in the bottom part of the canopy, are rarely covered by leaves and feature a compact and homogeneous berry structure. The compactness of the grape bunches makes them vulnerable to certain fungal diseases, e.g. Gray Mold (*Botrytis cinerea*). To handle this and other risks, an extensive spraying of the grapes is needed throughout the season.

The Semi Minimal Pruned Hedge (SMPH) is an extension to the Minimal Pruning (MP), which became popular in the 1979s in America. The MP did not gain the same popularity in Europe as in America and Australia, since the farmers were not satisfied with the quality of the grown grapes (Martinez de Toda and Sancha, 1998; Intrieri et al., 2011). The SMPH features a mechanical, rough winter pruning of former VSP rows, resulting in a hedge-like leaf wall with several perennial canes, in contrast to the VSP trained plants which only have one. It has a thick canopy, with many smaller leaves which impedes the drying of the leaf wall. The grape bunches are spread through the whole canopy although they primarily occur in the upper part and are mainly occluded by leaves (see Figure 2.1b). The bunches themselves have a loose structure and

(a) BBCH75 (b) BBCH77 (c) BBCH89

(d) BBCH75 (e) BBCH77 (f) BBCH89

Figure 2.2: Riesling grape bunch at three different phenological stages. The upper row shows a bunch from a VSP trained row, while the lower row shows one from a fine trained in the SMPH. The images for each stage were taken at the same day for both training systems.

the berries feature an in-homogeneous size. The loose berry structure makes the grape bunches more resilient to bunch rot diseases like *Botrytis* and facilitates a mainly mechanized thinning and pruning (Molitor et al., 2019). A comparison of the microclimate and the canopy architecture with respect to the training system can be found in Kraus et al. (2018).

2.1.2 Relevant Growth Stages

Traditional phenotyping methods in viticulture consist of visual screening by skilled experts in the field. The desired traits are estimated using different descriptors, either the Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) scale (Lorenz et al., 1995) or the International Organisation of Vine and Wine (OIV) descriptors (Alercia et al., 2009).

The OIV descriptors are a viticulture specific measure and serves to provide a standardized and objective description of grape varieties and species (Alercia et al., 2009). Described traits include for example shoot, leaf or berry characteristics. For berries these include the shape, size, color or firmness of the flesh.

The BBCH scale exists for many different crops and describes the respective growth stages of the plants. For many of our presented works, four distinct BBCH are of great interest. The first is the BBCH75 stage. It features pea-sized berries (see Figure 2.2a and 2.2d) and is often the stage, where decisions about berry thinning procedures are made. BBCH77 is when the majority of the berries

are touching each other, which is often shortly after the thinning procedures (see Figure 2.2b and 2.2e). BBCH83 corresponds to the véraison, where the berries develop color and get soft. This is important, since most of our developed algorithms focus on green berries, since all grapevine varieties have green berries before the véraison. The last stage is the BBCH89, when the grapes are ready for harvest (see Figure 2.2c and 2.2f). At this point in time the reference measures for the yield are taken.

Although the images for the different BBCH stages are taken at the same time in 2.2, we can see that the berry structure is very different between the two training systems. For the SMPH, the berries are smaller and appear to develop slower than the ones in the VSP. Even close to harvest, the majority of the berries in the SMPH are not touching.

2.2 Deep Learning

Deep Learning (DL) is a subdomain of Machine Learning (ML) and plays an important role in the publications presented in this thesis. ML in general describes algorithms which are supposed to mimic human decision making by using so called models. These models are mathematical representations of real world problems resulting in functions f, which turn a given input $\mathcal{X}$ into a desired output $\mathcal{Y}$, given internal parameters θ:

$$\mathcal{Y} = f(\mathcal{X}, \theta) \tag{2.1}$$

Some of the tackled tasks include classification (the assignment of a single label to a piece of data), regression (a continuous output from a given input) or clustering (the grouping of similar data). To achieve this, three main components are required: data as an input, a model for the decision making and some sort of learning of the model. Especially in the field of Computer Vision (CV), it is very challenging to create a model for solving tasks which are naturally easy for humans. Viewpoint variations, object deformations, occlusions, changing illumination, background or intra-class variations are easy to handle for a human but not for a model. Classical CV algorithms rely on hand-crafted features to break down the high-dimensional information presented in images. In contrast to this, DL methods, which emerged mainly after 2012, extract features directly from the data.

Since this work mainly relies on DL methods, strictly speaking Neural Networks (NNs), the following section will solemnly focus on this. Classical methods which were used in similar contexts are cited in the respective places in Chapter 3.

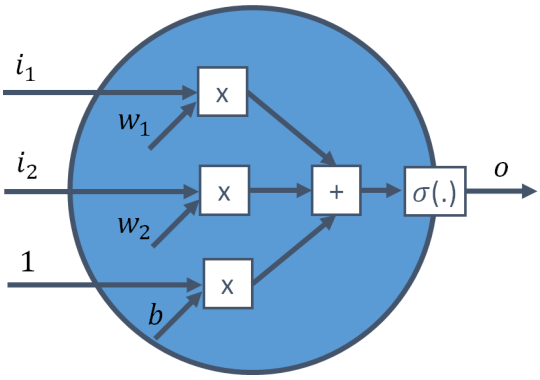


Figure 2.3: Model of a neuron with two inputs i_1 and i_2, one output o, one non-linearity $\sigma(.)$. The parameters are the weights w_1 and w_2 and the bias b.

2.2.1 Neural Networks

NNs have proven to be very successful in many ML tasks, which is the reason why we chose to use them in our publications. NNs are the base model for DL techniques, but special models exist for different types of applications, including CNNs, recurrent NNs or Generative Adversarial Networks (GANs). But the main idea behind all of them is the assembly of simple processing units into layers to approximate high dimensional functions. The smallest unit of a neural network is called a neuron. Figure 2.3 shows a simple neuron with two inputs i. The inputs are multiplied with weights w and a bias b is added. The last operation is the application of a non-linearity $\sigma(.)$. The output o is computed as following:

$$o = \sigma(a) = \sigma(i_1 w_1 + i_2 w_2 + b) \tag{2.2}$$

Corresponding to Equation 2.1, this means that i_1 and i_2 correspond to the inputs $\mathcal{X}$, the trainable parameters w_1, w_2 and b to θ and the output o to $\mathcal{Y}$. The non-linearities σ, also called activation functions, influence the value range of the output and can be chosen based on the respective problem. But in general, activation functions are used to counteract vanishing gradients in the training process, and to ensure a better convergence behaviour. In our works we use either the Rectified Linear Unit (ReLU), which are computationally efficient, counteract the vanishing gradient problem and show a better convergence behaviour compared to other functions like the sigmoid activation or the LeakyReLU, which is an advancement to the ReLU. It prevents neurons to die when not activated due to the factor α which is usually chosen to be a small value:

$$\text{ReLU:} \qquad \sigma(a) = \max(0, a) \tag{2.3}$$

$$\text{LeakyReLU:} \qquad \sigma(a) = \max(\alpha a, a) \tag{2.4}$$

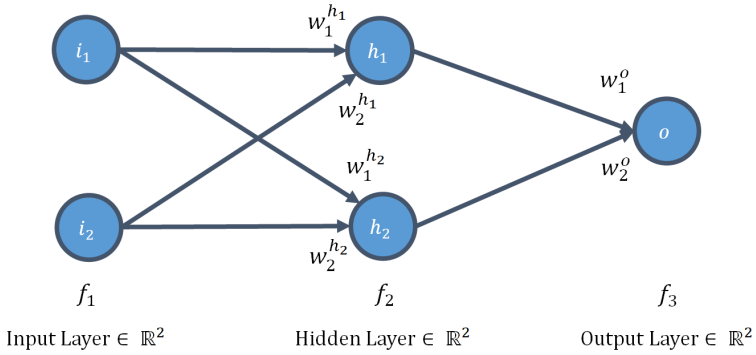


Figure 2.4: A simple FCNN with two inputs i and one hidden layer with two hidden neuron h and one output neuron o.

The parameters, the weights w_i and bias b, can be adjusted to approximate a simple non-linear function. Nonetheless the representational capability of one neuron by itself is limited, but the accumulation and connection of many neurons can approximate arbitrary functions. The most simple way to do this, is to arrange the neurons in a feed-forward fashion in a so called Fully Connected Neural Network (FCNN). Here, each neuron is connected with every input and every output. Figure 2.4 shows a very simple FCNN with two inputs i and two hidden neurons h and one output neuron o. Each neuron is modelled after the one presented in Figure 2.3. The output can be computed as followed:

$$o(h_1, h_2) = f(w_1^o h_1 + w_2^o h_2 + b^o) \tag{2.5}$$

$$h_1(i_1, i_2) = f(w_1^{h_1} 1_1 + w_2^{h_1} i_2 + b^{h_1}) \tag{2.6}$$

$$h_2(i_1, i_2) = f(w_1^{h_2} 1_1 + w_2^{h_2} i_2 + b^{h_2}) \tag{2.7}$$

In short we can stack and encapsulate the layers to the following representation:

$$o = f^{(3)}\left(f^{(2)}\left(f^{(1)}\left(x, \theta^1\right), \theta^2\right), \theta^3\right) \tag{2.8}$$

Networks can be very large and complex with many layers and countless neurons. Therefore larger networks are characterized by the number of layers, which is called the network *depth* and the number of neurons in each layer, which is called the network *width*.

Training of Neural Networks

To fully facilitate the potential of NNs, it is important to optimize the networks, which is called training. The training of neural networks involves the minimization of a training objective. This objective is called a loss function $\mathcal{L}$ and compares

the network output o with the known, desired output $\hat{o}$. The choice of the loss function depends on the problem which needs to be solved. Popular losses include:

$$\text{Mean Absolute Error:} \qquad \mathcal{L}_{l1} = \frac{1}{n}\sum_{i=1}^{n}|\hat{o} - o| \tag{2.9}$$

$$\text{Mean Squared Error:} \qquad \mathcal{L}_{\text{MSE}}(w, b, i) = \frac{1}{n}\sum_{i=1}^{n}(\hat{o} - o)^2 \tag{2.10}$$

$$\text{Cross Entropy:} \qquad \mathcal{L}_{\text{CE}} = \frac{1}{n}\sum_{i=1}^{n} o * \log(\hat{o}) \tag{2.11}$$

In our Publication 1 (Zabawa et al., 2019b) and Publication 2 (Zabawa et al., 2020), we optimize the network using the Intersection over Union (IoU), which is a suitable loss function for a segmentation problem. The IoU will be explained in detail in Section 2.2.4 and further details can be found in Publication 2 (Zabawa et al., 2020). In Publication 5 (Miranda et al., 2022) we used a feature perception loss which does not compare the output o directly with the desired output $\hat{o}$, but their respective representations in a latent space.

NNs are fully differentiable functions, since they are composed of a multitude of small, differentiable functions. The loss function is minimized using an optimizer, for example Adaptive Moment Estimation (Adam) (Kingma and Ba, 2015), Adaptive Gradient Algorithm (AdaGrad) (Duchi et al., 2011) or Gradient Descent (GD). GD is used to find the steepest descent in the parameter space ∇_θ. The loss function is then iteratively minimized as followed:

$$\theta' = \theta - \epsilon \nabla_\theta \mathcal{L}(\theta) \tag{2.12}$$

The parameter ϵ is the learning rate, which specifies the steps length in the direction of the steepest descent. The learning rate is a parameter which needs to be carefully chosen. If the learning rate is to large, the convergence behaviour is negatively affected, if it is too small, the learning process is very slow. Therefore, the learning rate is often adjusted during the training process, starting with a large learning rate which is decreased after some time. In Publication 3 (Bömer et al., 2020) and Publication 5 (Miranda et al., 2022) we used constant learning rates during the training process. In contrast to this, we used adaptive learning rate in Publication 1 (Zabawa et al., 2019b) and Publication 2 (Zabawa et al., 2020).

The optimization and computation of the gradients are computationally intensive. As an alternative, the parameter updates are computed using randomly sampled batches of the data. Each sample is called a mini-batch with a batch

size $\mathcal{B}$. An Epoch is processed when all mini-batches, the whole data set, was fed through the network. The problem with this batch-wise computation is a challenging converging behaviour, since the inputs to the next layer change between the single batches. To prevent these problems, batch-normalization is applied. It means that the mean of the inputs is set to zero and the standard deviation is forced to be one. A network is always trained for multiple epochs, with randomly sampled mini-batches each epoch. Therefore, the optimization process is called Stochastic Gradient Descent (SGD). We used batch normalization in Publication 5 (Miranda et al., 2022), to ensure a faster training time and introduce a regularization.

2.2.2 Convolutional Neural Networks

We used CNNs in all of our publications, therefore it is important to introduce the concepts in this section. Images have a special data structure, pixels are organized in a regular grid with a fixed size, and local image regions are stronger correlated with each other than far away image regions (Schmidhuber, 2015). These characteristics can be exploited using Convolutional Neural Network (CNN). These networks share similarities with FCNNs, like the layer-wise architecture, but are highly optimized for the usage of image data. The before-mentioned neurons are replaced by convolutions, namely image-filters.

Convolutional Layer

The basic building block of a CNN is a discrete convolution with a filter. A filter with a predetermined size is slid across an input image with a fixed stride. The result of the convolutional operation is the dot-product between the filter and input values, resulting in a single value for each filter position (see Figure 2.5). The filter values themselves are the learnable parameters of each layer. Figure 2.5 shows an example of a 3×3 kernel applied on a 5×5 input with a stride of one, leading to a 3×3 output.

Similar to FCNNs, convolutional layers are followed by an activation function and a channel-wise batch-normalization, resulting in a single feature map. In each layer, multiple kernels are applied to the same input producing multiple feature maps. The aggregation of these is called a feature volume, which is the input to the next layer. For example if we have 5 kernels in one layer, the output feature volume would have 5 channels.

As we can see in Figure 2.5, the output has a smaller spatial dimension than the input. If we want to keep the same spatial resolution, we need to surround the input with a boundary, this is called zero padding. In some cases the reduction of the spatial dimensions is desired, either for memory efficiency reasons or to

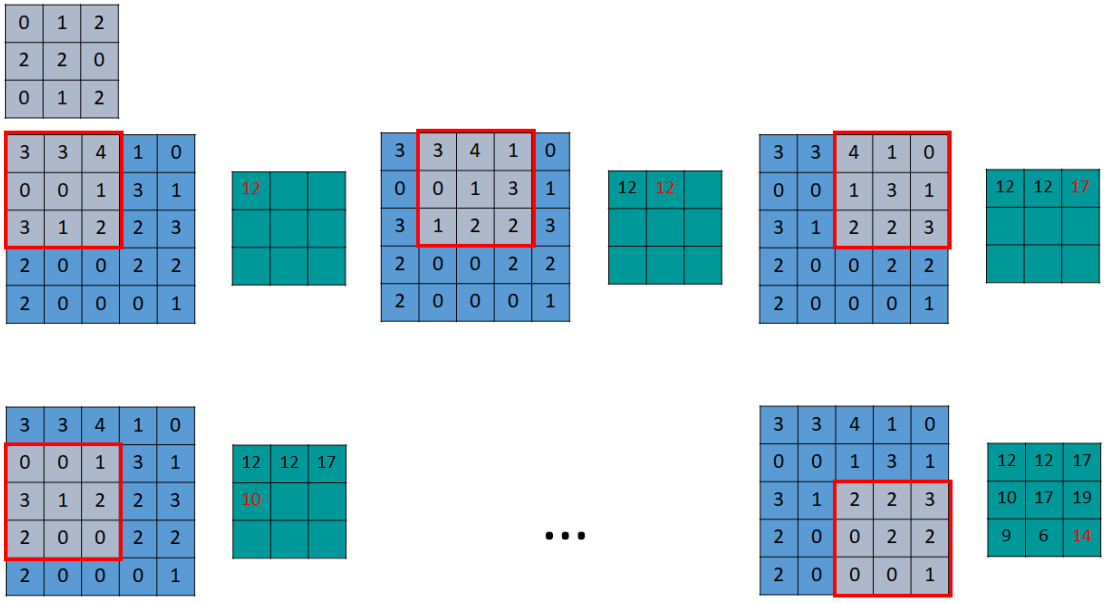
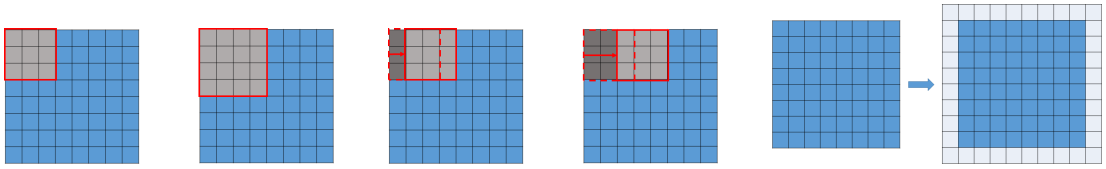


Figure 2.5: Detailed example of a convolution. The grey 3×3 kernel is slid over the blue 5×5 input. The result is the green output, which is a linear transformation of the input. Image inspired by Dumoulin and Visin (2016).



(a) Filter size: 3×3 vs. 4×4 (b) Stride: 1 vs. 2 (c) Zero padding: none vs. 1

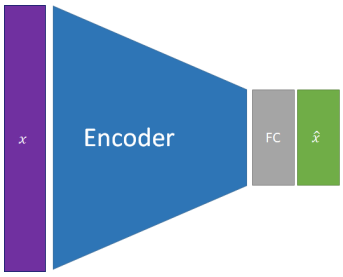
Figure 2.6: Configuration options for convolutions in a CNN. Figure 2.6a shows two different filter sizes, Figure 2.6b shows two different strides and Figure 2.6c the difference between a non zero padded input and an input with a zero padding of 1.

enhance the information density for the following layers. This can be realized either with a larger stride or so called pooling operations. Different pooling operations exist, including max-pooling or mean-pooling, where the max or mean value is copied into the down-sampled feature map. Settings which can be chosen for each layer include the filter size, the stride or the zero padding, examples are shown in Figure 2.6.

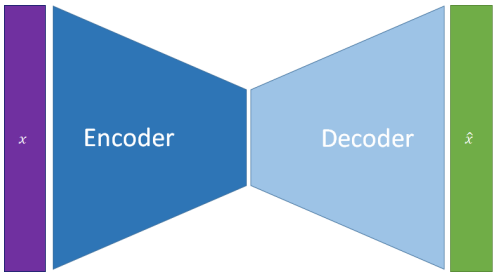
In general, CNNs have fewer parameters than FCNNs, since the same filters are applied to the whole input. This leads to less complex networks and a better generalization of the networks, reducing the risk of over-fitting.

Encoder-Decoder

Depending on the task which needs to be solved by our model, different network architectures can be used. Classification tasks need at least one fully connected layer at the end, since the expected output is a distinct class label (see Fig-



(a) CNN with fully connected layers at the end for classification. The input is an image and the output a label.



(b) Fully convolutional NN with an encoder-decoder structure. The input is an image and the output a mask with the same dimension as the image.

Figure 2.7: Different types of CNNs. Figure 2.7a is an example of a CNN with fully connected layers at the end, which produces a class label $\hat{x}$ given a input x. Figure 2.7b on the other hand is a fully convolutional CNN, solving a segmentation task, where the output $\hat{x}$ has the same dimension as the input x.

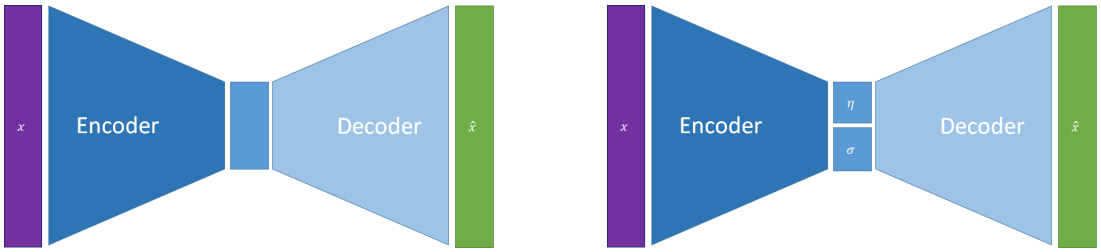
ure 2.7a). Segmentation tasks on the other hand should deliver an output mask of the same dimension as the input image (see Figure 2.7b). A so called encoder reduces the feature volume using pooling operations and a decoder up-samples this feature volume back to the original image dimensions. The choices for the encoder and decoder design depend on the respective task at hand. For example in Publication 1 (Zabawa et al., 2019b) and Publication 2 (Zabawa et al., 2020), we use the DeepLabV3+ decoder, which was developed by Chen et al. (2018) to refine the segmentation results with special focus on object boundaries.

2.2.3 Variational Autoencoder

In Publication 5 (Miranda et al., 2022) we developed a VAE for anomaly detection, therefore we explain the basic concept behind an Autoencoder (AE) and the special case VAE in the following section. AEs are a special kind of NN, with the main task to learn a data encoding in an unsupervised fashion. Tasks which are often solved using AEs include dimensionality reduction, image denoising or anomaly detection.

The general structure of a AE is based on the before-mentioned encoder-decoder structure. But in this case, the encoder is used to find a very dense and highly expressive representation of the data in the latent space (see Figure 2.8a). In dimensionality reduction scenarios, this dense latent representation is already the desired output, but in other cases a reconstruction of an image from the latent representation is needed. In this case the decoder tries to revert the compression. VAEs are a special case of a AE, where the latent representation is forced to represent the mean and a standard deviation of the input data (see Figure 2.8b).

The training objective of AE can either be to make the output as similar to the



(a) Convolutional Autoencoder (CAE). (b) Variational Autoencoder (VAE)

Figure 2.8: Two different kinds of AEs. The difference is the representation of the latent space. in both cases the input x has the same dimension as the output $\hat{x}$.

input as possible. This corresponds to a pixel-wise loss, which directly compares the desired output $\hat{o}$ to the predicted output o. Alternatively, in Publication 5 (Miranda et al., 2022) we used a subsequent CNN that can be used to fit a neural representation of the output to the neural representation of the input. This creates a loss based on the latent space, called Feature Perceptual Loss (FPL).

2.2.4 Evaluation Metric

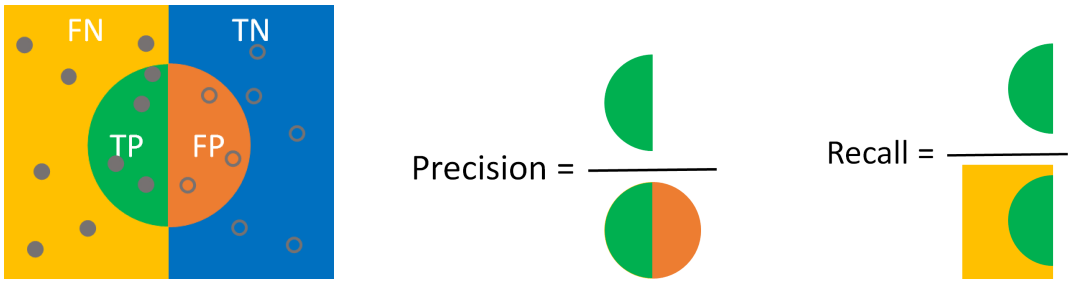
We will describe the evaluation metrics which are used throughout this thesis. This includes metrics to evaluate classification as well as segmentation problems.

Precision and Recall

The evaluation of classification tasks is often done using precision and recall. For a binary classification problem this can be achieved by reasoning over the classification results. In our case we will look into the classification of *berry* and *non-berry*. True positive (TP) describes the data points which belong to the class *berry* and are correctly classified as *berry*. True negative (TN) does the same for the class *non-berry*. Data points which belong to the class *berry*, but are classified as *non-berry* are called False negative (FN), while *non-berry* points which are classified as *berry* are called False positive (FP).

The precision and recall are metrics which are ratios of these sets (see Figure 2.9). The precision measures how likely it is that a classifier's prediction is correct. It is computed as the ratio between the data points which are correctly classified as berries (TP) and all data points which were classified as berries (sum of TP and FP), see Equation (2.13). The precision takes values between 0 and 1, values close to 1 indicate a better classifier performance. That means if the precision is low, the classifier predicts too many berries which are in truth non-berries.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.13}$$



(a) Set (b) Precision (c) Recall

Figure 2.9: Computation of precision and recall for a binary classification problem. For our example, the empty circles in Figure 2.9a show samples belonging to the class *non-berry*, while the filled circles represent *berries*. The filled circles in the green area show correctly classified berries (TP), and the circles in the blue area show correctly classified non-berries (TN). The circles in the red and orange area show wrongly classified samples (FP and FN)

The recall is also called sensitivity and measures the number of actual berries which are found by the classifier. It is the ratio between the correctly classified berries (TP) and all the actual berries, even the not classified ones (sum of TP and FN). The recall takes values between 0 and 1, higher values indicate a better classifier performance. It is computed as followed:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.14}$$

Since there is always a trade-off between the precision and recall, it is good to have an additional measure which makes the comparison of classifiers easier. The F1-score is a single metric that combines the two above mentioned metrics using the harmonic mean and is computed as followed:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.15}$$

Intersection over Union

First, the IoU was used to evaluate detections by computing the overlap between Bounding Boxs (BBs) (see Figure 2.10). Later, it was used to evaluate segmentation on pixel-level. But nonetheless, the IoU describes the ratio between the area of overlap (the TP) and the union areas (the sum of TP, FP and FN).

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{2.16}$$

The principle is the same compared to the bounding box, but the area can have an arbitrary shape, as can be seen in Figure 2.11.

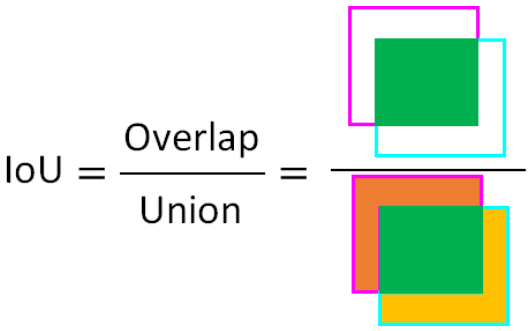
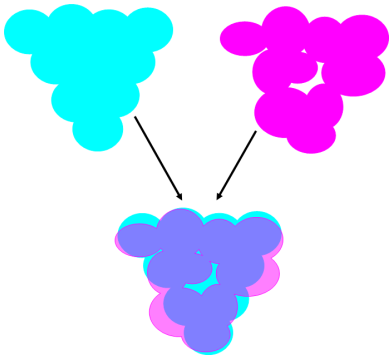
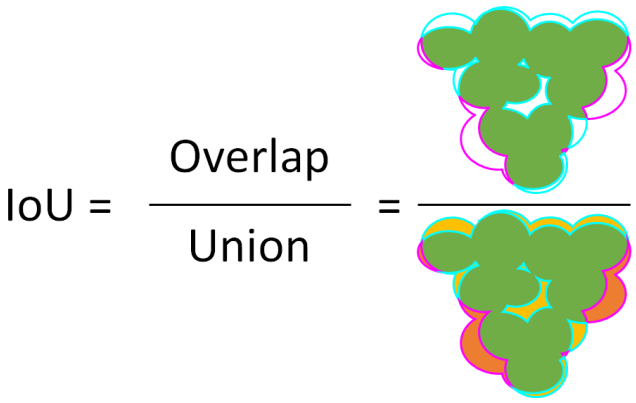


Figure 2.10: Explanation of the general concept of IoU using BBs. The turquoise rectangle is the ground-truth BB, the pink one the prediction. The overlap between both boxes is shown in green and represents the TP. The yellow area is not recognized by the classifier, therefore it represents the FN, while the orange area is recognized although it is not part of the ground-truth, representing the FP.



(a) overlap between the ground truth mask (turquoise) and the predicted mask (pink).



(b) IoU for an arbitrary object shape.

Figure 2.11: Example for the IoU for arbitrarily shaped objects. The turquoise object depicts the ground truth mask and the pink one the predicted one. The TP area is shown in green, the FP in orange and the FN in yellow.

Chapter 3

Summary of relevant publications

The thesis can be structured into three main topics. An allocation of the relevant publications into these topics can be seen in Figure 3.1.

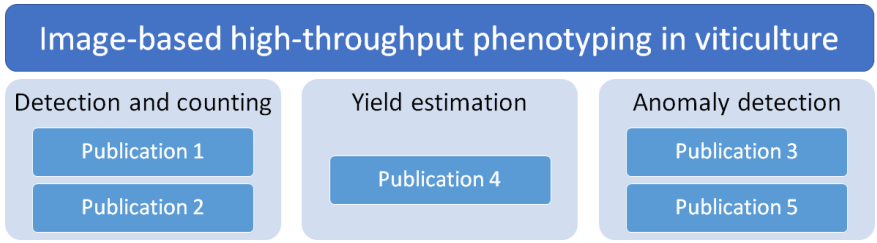


Figure 3.1: Content assignment of the relevant publications to clarify their respective contribution to the dissertation.

The two publications which contributed to the topic of detection and counting are summarized in Section 3.1. The second main topic deals with the estimation of yield and is presented in Section 3.2. The last part highlights the contribution to anomaly detection in Section 3.3.

3.1 Detection and Counting

Object detection and counting is often an important preliminary task for other phenotyping applications. Especially in agriculture, the objects can be very small, appear in large quantities or close clusters. Many of these characteristics pose challenges to state-of-the-art approaches which are very successful for other applications. Therefore an application-specific instance segmentation for large numbers of small objects was proposed in Publication 1 (Zabawa et al., 2019b) and further developed in Publication 2 (Zabawa et al., 2020).

Scientific Context

Sensor-based high-throughput phenotyping gained a lot of attention with the development of more affordable sensors, for example RGB-, RGBD- or multi-spectral cameras and laser scanners (Gongal et al., 2015; Tang et al., 2020).

Due to the perennial nature of grapevine, phenotypic data have to be acquired in the field. Early approaches started to use handheld cameras for image acquisition and extract geometric structures from the images. Roscher et al. (2014) and Nuske et al. (2011) both use a circular Hough-transform to detect round objects (berries) in images, other works detect convex surfaces (Nyarko et al., 2018). Liu et al. (2020) perform a detailed bunch architecture estimation from images taken with an artificial background. Later on, field phenotyping platforms emerged for the (semi)automatic acquisition of images. Nuske et al. (2014) developed a moving platform with a semi-automatic image acquisition and an artificial lighting setup, which was used in a large scale phenotyping experiment. Kicherer et al. (2014) remodelled a grapevine harvester into a field phenotyping platform with a multi-camera system covering a large part of the vertical canopy and artificial lighting. By utilizing overlapping images, Rose et al. (2016) reconstruct 3D point-clouds and detect berries based on their distinct color and shape features.

Especially the detection of green fruit in front of green canopy is very challenging and sometimes unavoidable. In some cases, the fruit itself is green and stays green throughout the growth period, for example limes or green apples. In other cases, an observation is necessary at certain growth stages, where the fruit is still green and changes color later, for example in grapevine where the color changes at véraison. For the detection of green apples, Wachs et al. (2010) rely on a combination of RGB and thermal images. They define high level features as global attributes and local features for the use of primitive parts-based filters. Gan et al. (2020) detect green citrus fruits using thermal images in combination with a water spraying system, taking the different temperature change rates into account. Other works rely on a combination of color and texture or shape features for either the detection of green citrus fruit (Kurtulmus et al., 2011) or green apples (Linker et al., 2012).

In 2012, AlexNet (Krizhevsky et al., 2012) started the road to success for DL methods and NNs became the state of the art for CV problems, including classification, segmentation and image generation. Especially the development of CNNs (Long et al., 2015) boosted this research direction. A review on deep learning applications in agriculture was done by Kamilaris and Prenafeta-Boldu (2018), and by Tardaguila et al. (2021) and Mohimont et al. (2022) for viticulture, but in the following we will mainly focus on the detection and counting aspects.

Different problem formulations can be used to count objects in images. Regression networks for example (Lempitsky and Zisserman, 2010; Cohen et al.,

2017) are straightforward, do not require detailed semantic annotations and can be trained with small data sets due to the comparably small networks used. They showed promising results in other domains, for example the counting of penguins (Arteta et al., 2016), cells in microscopy images (Xie et al., 2016; Guo et al., 2019) or buildings (Lobry and Tuia, 2019). Lu et al. (2018) presented a class agnostic approach which can be applied to different tasks. Most applications only desire a number of objects as the output of the pipeline. Coviello et al. (2020) adapt a crowd counting algorithms using a dilated CNN for the counting of grapevine berries in smartphone images.

Particular instance segmentation approaches are another option for the counting of objects in images. Until today, the most commonly used one is the Mask-RCNN by He et al. (2017). The approach is comprised of two stages: the first stage involves segmentation and region proposals, while the second stage focuses on classification and bounding box estimation. Networks with region proposal stages require the specification of a certain number of proposals, which makes them hard to use if many objects are in an image. Furthermore many instance segmentation approaches struggle with small objects. Both of the described problems are often relevant for phenotyping applications. Gené-Mola et al. (2020b) use the Mask-RCNN on images and simultaneously compute a point-cloud using Structure from Motion (SfM) to detect apples in the orchard. Nellithimaru and Kantor (2019) and Yin et al. (2021) do the same for the detection of grapevine berries followed by a sphere fitting. High defoliation was observed in both cases and the study focused solely on detecting red grapes.

Many works rely on the detection and semantic segmentation methods to identify grapevine berries. Aquino et al. (2016) detect circular light reflections in images with an artificial background as berry candidates, while they later introduce a NN (Aquino et al., 2018) and discard the artificial background. Bargoti and Underwood (2017) use monocular images taken with a ground-based vehicle for the detection of apples. They perform a semantic segmentation using a CNN followed by a Hough-transform for the detection of single instances. Cecotti et al. (2020) detect a grape cluster with green and red grapes. They investigate the influence of the feature space (color or gray scale or color histograms), different network parameters and augmentation strategies as well as the impact of pre-training. Kurtser et al. (2020) use a RGBD camera mounted on a mobile robot platform to estimate the cluster number, size, volume, length and width of grape bunches based on colors.

In Publication 1 (Zabawa et al., 2019b) and Publication 2 (Zabawa et al., 2020), we developed a novel approach to extract detailed information about grapevine berries from images using CNNs. We chose to use CNNs because they have proven to deliver superior results in the image domain compared to classical

ML methods. Classical methods are effective if the fruit has a distinct color, but they do not work well for green fruit. Additionally, the detection of geometric objects in images can require assumptions, such as the circle radius or the Hough-transform (Roscher et al., 2014; Nuske et al., 2014), which is highly dependent on the image collection setup. We circumvent these limitations by adopting a NN based approach and achieved convincing results without the need to specify the distinct object color or the approximate object radius. Regression methods can provide object counts (Lu et al., 2018; Coviello et al., 2020) but they are unable to provide details about the object's geometry, such as its shape. These are information which we can provide with our method, without the computational overhead required by instance segmentation methods. Instance segmentation networks are generally more complex than lightweight CNNs (He et al., 2017), making them harder to train, especially for small data sets. In contrast, we used a very lightweight network with only around 300k parameters, which made it very suitable for our limited data set, while outperforming the segmentation results of the Mask-RCNN.

Publication 1 (Peer-reviewed, Conference)

L. Zabawa, A. Kicherer, L. Klingbeil, A. Milioto, R. Töpfer, H. Kuhlmann, and R. Roscher. Detection of single grapevine berries in images using fully convolutional neural networks. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2571–2579, 2019b. doi: 10.1109/CVPRW.2019.00313

This publication focuses on the detection and counting of single grapevine berries in images, independent from the observation time, the variety or the training system. We chose to work on images taken before the véraison, the onset of the ripening and color change of the berries. This has the advantage that we can handle different varieties since all varieties have green berries before this point in time (Section 2.1.2), but makes the overall detection more challenging, since green berries are harder to distinguish from the green canopy in the background. Furthermore grapevine berries look very similar to one another and appear in large quantities.

To tackle these challenges, we propose a new instance segmentation based on a semantic segmentation. The main idea is the introduction of a new class called *edge* which helps to separate two instances of the same class. This enables the usage of a lightweight semantic segmentation network (Milioto and Stachniss, 2019) with a classical encoder-decoder structure (Section 2.2.2), which can be trained with a relatively small data set. We evaluate the influence of the chosen edge-thickness and different geometric filter strategies, which take the shape and

area of the berries into account. We can achieve convincing results for the two different training systems VSP and SMPH (Section 2.1.1), with 92% and 87% correctly detected berries respectively.

Publication 2 (Peer-reviewed, Journal)

L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, H. Kuhlmann, and R. Roscher. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 164:73–83, 2020. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2020.04.002. URL `https://www.sciencedirect.com/science/article/pii/S0924271620300939`

The second publication builds upon the first, by extending the investigation with an expanded data set and improved evaluation metrics (Section 2.2.4) using classical computer vision methods. To compare the performance of the approach, a systematic comparison was conducted with two state-of-the-art methods: Mask-RCNN (He et al., 2017) and a regression approach using U-Net (Ronneberger et al., 2015).

Mask-RCNN, with 20 times more parameters compared to the Bonnet (Milioto and Stachniss, 2019) used in this study, faced challenged in training with the limited data set. The inference was much slower and the network struggled with the large amount of small objects in the images. On the other hand, the regression approach had only twice as many parameters as our network, making it easier to train and faster at inference time. However, due to the dot detection no further information about the berry size or shape can be extracted.

Summarized, our semantic segmentation approach outperformed both other methods in predicting berry numbers in an image, showing a higher correlation between the manual counting and predicted values. The Mask-RCNN performed significantly worse, while the regression approach performed similarly.

3.2 Image-based Analysis of Yield Parameters in Viticulture

Yield estimation is a major research area for agriculture. Many applications have the goal to maximize the yield under given circumstances and therefore need to monitor the plants throughout a whole season to achieve this. However, in viticulture yield optimization is the key challenge, not maximization (Howell, 2001). Winegrowers try to achieve an optimal yield per hectare to ensure a high quality end product. Therefore, an early yield estimation is of great interest, to guide

informed decisions on berry thinning procedures. The current industrial standard involves sampling of single vines and extrapolation to the whole vineyard, which is labor-intensive and at the time error-prone due to the limited sample size. Hence sensor-based, especially image-based systems are in the focus of many research applications.

Scientific Context

Yield optimization is a crucial task for grapevine growers, as it directly impacts the quality and quantity of grapes produced per hectare. In the context of wine production, yield optimization is especially important, as regulations exist to ensure that only quality wine is produced. In the German state of Rheinland-Pfalz, for instance, regulations mandate a yield range of 105 - 125 hectoliters per hectare (hl/ha) for quality wine production (Landwirtschaftskammer Rheinland-Pfalz, 2012). While it may be tempting for growers to maximize yields in order to increase profits, there are potential downsides to doing so. Overly high yields can result in lower grape quality, as the vines are forced to produce more grapes than they can support. This can lead to diluted flavors and lower sugar levels, which can negatively impact wine quality. In addition, high yields can put stress on the vines and make them more susceptible to disease and pests. Overall, yield optimization is a delicate balancing act for grapevine growers. They must strive to achieve optimal yields that meet regulatory requirements for quality wine while also ensuring that grape quality is not compromised. This requires careful attention to a wide range of factors, from grape variety selection to vineyard management practices, in order to achieve the best possible results.

Traditional forecasting methods in viticulture include the counting and weighing of sampled yield components in the field (Clingeleffer et al., 2001), as well as the usage of historic data relating to weather conditions. The estimation can be performed at different points in time, taking different yield components into account. For example, de la Fuente et al. (2015) compare different methods for grapevine yield prediction between fruit-set and véraison and found that estimation methods at the véraison yielded the most promising results.

Using CV, counting or detection is often a preliminary step for the yield estimation, and often the count itself is used as the yield estimate. For example, Dorj et al. (2017) detect yellow citrus fruits in images using a color threshold, and use the fruit count as a proxy for the yield. Similarly, Bargoti and Underwood (2017) segment apples in images using a CNN, followed by a Hough-transform to distinguish and count the single instances. Wang et al. (2019) developed a video based tracking approach for the counting of mango adapting a deep learning detection network called YOLO and a Kalman-Filter. They as well only count

the fruits and do not estimate a yield in kg. This seem reasonable for crops which are sold per unit, like apples or mango, but is not sufficient for grapevine. In addition to using YOLO, Shen et al. (2023) incorporate SORT to enable multi-object tracking in video sequences. In most cases, the argumentation is that a better detection enables a better yield estimation (Kurtser et al., 2020). Kalantar et al. (2020) take the step from fruit detection to the estimation of yield in kg. They detect melons in drone images using a CNN, estimate geometric features and directly regress a weight for each melon. Sayago and Bocco (2018) go even further and directly estimate corn and soybean yield from satellite images without the detection of single fruits. For a more detailed review of deep learning based yield estimation methods for different fruits, we refer the reader to Koirala et al. (2019a), Anderson et al. (2021) and Darwin et al. (2021).

For viticulture, observed yield components can include the cluster number and size, as well as the volume, length or width (Kicherer et al., 2014; Di Gennaro et al., 2019; Kurtser et al., 2020). Under laboratory conditions, a detailed investigation of single grape bunches was performed by Ivorra et al. (2015) using a stereo camera and Hacking et al. (2019) using a RGBD camera respectively. Obtaining these results in the field can be challenging due to various environmental factors, such as occluded bunches. To overcome these challenges, researchers have developed a handheld laser scanner, as demonstrated by Rist et al. (2018), which can be used both in the laboratory and in the field.

Instead of focusing on the counting of yield components, Diago et al. (2012) perform a supervised pixel-wise classification of red grapes in a highly defoliated system using the Mahalanobis-distance and correlate the number of berry pixels with the yield. Silver and Monga (2019) estimate grape yield from smartphone images directly, comparing 5 different CNNs. Their approach also focuses on red grapes in front of green canopy. Di Gennaro et al. (2019) developed an unsupervised approach for the estimation of cluster number and size from high-resolution RGB images collected with a low-cost UAV. The approach works for red grapes with a high degree of defoliation. Palacios López et al. (2022) perform a yield estimation, based on a count of visible berries, for 6 different grapevine varieties achieving yield estimations, or rather a counting accuracy, between 16% and 40% depending on the variety.

Fruit occlusion is one of the biggest challenges of yield estimation methods, especially for horticultural environments. Different strategies were developed to handle these problems. Especially in greenhouse environments, the usage of robotic arms with real-time viewpoint planning strategies were investigated (Kurtser and Edan, 2018a,b; Harel et al., 2020). In other cases, where the application of robotic arms is difficult, multi-camera systems or multi-viewpoint

approaches were developed (Hemming et al., 2014; Koirala et al., 2019b). Gené-Mola et al. (2020a) and Nellithimaru and Kantor (2019) proposed a forced airflow-system to move occluding canopy out of the way. Gao et al. (2020) on the other hand estimate occlusion classes for detected fruits for the planning of different picking strategies.

Another strategy involves the estimation of occlusion factors, to correct for invisible or only partly visible fruits. Koirala et al. (2019b) estimate an occlusion correction factor, which they determine beforehand by investigating manual counts on the tree. In a following work Koirala et al. (2021) compare 5 different methods for the estimation of mango fruit yield and show that the estimation of a yield occlusion factor is only possible per tree, since the variability of the occlusion is too large. Also their approach works only for one season but is not expendable to a new one. Kierdorf et al. (2022) use the same data as our occlusion experiment to estimate the unseen berries behind the leaves, using a GAN, showing promising results to bypass the need for a distinct occlusion factor.

Yield estimation for grapevines traditionally involves manual sampling, reliance on historical knowledge, and scaling up to the whole vineyard. However, this process heavily relies on the experience and expertise of the person performing the estimation, making it challenging to achieve accurate estimates at the optimal time. On average, traditional methods achieve an accuracy of only around 30% (Dunn and Martin, 2003). In contrast to these traditional methods, we offer a sensor-based solution, which does not depend on historic weather data or personal experience, only a berry weight factor for the variety. This also sets us apart from methods, which only provide fruit counts, like Bargoti and Underwood (2017), Wang et al. (2019) or Diago et al. (2012). Many works in the field of grapevine yield estimation also focus on red grapes (Di Gennaro et al., 2019; Silver and Monga, 2019), which makes the detection easier. While other works have achieved even better yield estimations compared to our method, they have typically observed highly defoliated canopies with excellent visibility of the yield components (Aquino et al., 2016; Nuske et al., 2011). In contrast, our experiments were conducted under realistic conditions in German vineyards, where moderate defoliation is more common. We also conducted leaf removal experiments, which showed that the variability of leaf occlusion was the limiting factor for our method.

Publication 4 (Peer-reviewed, Journal)

L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, R. Roscher, and H. Kuhlmann. Image-based analysis of yield parameters in viticulture. *Biosystems Engineering*, 218:94–109, 2022. ISSN 1537-5110. doi: https://doi.org/10.1016/j.biosystemseng.2022.04.009. URL `https://www.sciencedirect.com/science/article/pii/`

S1537511022000861

The publication describes an automated framework for estimating grapevine yield from geo-referenced image sequences captured using a semi-automatic platform equipped with a multi-camera system. The system captures three vertically overlapping images at each point in time to observe the whole grapevine canopy.

To account for the appearance of objects in multiple images, the pipeline selects horizontally, minimally overlapping images along the driving direction by taking the sensor position and the distance to the canopy into account. The vertical overlap is taken into account by matching image patches showing the same grape bunch in both images.

The most significant challenge in yield estimation is the variable leaf occlusion, which also varies between different training systems (Section 2.1.1). To quantify this, a large leaf occlusion experiment was conducted over two years. However, the correlation between the number of visible berries and yield could not be established for the SMPH system, since the leaf occlusion varied extremely. As a result, the yield estimation was only performed on plants trained in the VSP system, yielding results comparable to the industrial standard (Dunn, 2010).

3.3 Grapevine Anomaly Detection

The monitoring of plants throughout the growth season is one of the main aspects in precision agriculture. The identification of anomalies, e.g. diseases, nutrient deficiencies or water stress, can help farmers to take action at the appropriate moments in time. In some cases, the main goal is an early detection of a certain disease to take action, for example the application of fungicides or pesticides. For other applications the exact kind of damage is of subordinate importance. Furthermore, it is hard to design classifiers which handle multiple kinds of diseases or crop damages at the same time, since it's often hard to acquire sufficient data for all cases. Publication 3 (Bömer et al., 2020) and Publication 5 (Miranda et al., 2022) are works which contributed anomaly independent classifiers for application in vineyards.

Scientific Context

Regular monitoring of the plant performance is crucial to ensure the optimal growth of fruits and crops. This is especially important for fruits, since the climate change causes more extreme and on average higher temperatures, increased water and drought stress, higher CO_2 concentrations in the atmosphere, and changing abundance of pests (Jones, 2007). In some cases it is important

to know and react to certain diseases or damages, but in some cases it is only important to remove damaged or diseased plant material before the harvest, to produce a fine wine (Charters and Pettigrew, 2007).

Imaging sensors are very suitable for this task (Khirade and Patil, 2015; Ma et al., 2019). Especially hyper-spectral cameras showed a great potential for disease classification in laboratory environments (Behmann et al., 2015; Foerster et al., 2019). These cameras are very expensive and hard to operate under field conditions, therefore most research focuses on multi-spectral or RGB cameras instead. For example, Bendel et al. (2020) investigate the detection of Esca, first with a hyper-spectral camera mounted on a ground vehicle. They identify relevant channels to transfer the approach to a multi-spectral camera on a UAV. Similarly, Di Gennaro et al. (2016) use multi-spectral images collected with a UAV to estimate Esca symptoms. They found a high correlation between the Normalized Difference Vegetation Index (NDVI) and the expression of the symptoms. Another anomaly, the infestation with *Phylloxera* and the respective symptoms was detected by Vanegas et al. (2018) in UAV multi-spectral images.

The identification of diseases can be achieved with different problem formulations, for example detection, classification or segmentation. Many of these approaches are trained in a supervised manner (Kaur et al., 2019), which require many costly and labor-intensive annotations. Yadhav Yegneshwar et al. (2020) for example use a shallow CNN to detect diseases by performing a multi-class classification, while Amara et al. (2017) use a LeNet to detect diseases on banana leaves. Foerster et al. (2019) forecast symptoms of powdery mildew with the use of multi-spectral images acquired in a laboratory using a cycle consistent GAN. Many of these works were trained on the same data set, the PlantVillage data set (Hughes and Salathé, 2015), which shows single leaves in front of an artificial background.

Since it is very hard to acquire enough data samples showing diseased plants, unsupervised approaches came into focus. Un- or semi-supervised approaches are used to train on non-anomalous data only, forcing a network to learn a representation of the normal state. Anomalies are then found as deviations from this normal state (Pang et al., 2021). This bypasses the need for expensive annotations of anomalous examples and the need to fully capture the variability of anomalies. For example Akçay et al. (2018) use a Convolutional Autoencoder (CAE) for the detection of anomalies in flight luggage, and Baur et al. (2019) use a deep auto-encoding models for anomaly segmentation in magnetic resonance images (MRI) of brains. In the agricultural domain, Pardede et al. (2018) use a CAE as a feature extractor and detected plant diseases with a Support Vector Machine (SVM). Strothmann et al. (2019) did a reconstruction based anomaly detection

using a CAE and compared the reconstructed image patches with the original ones to find deviations from the learned healthy phenotype. For a more in depth literature review, including research in the medical field (Shvetsova et al., 2021) or landmine detection (Picetti et al., 2018), we refer the reader to the respective publications.

Agricultural disease and anomaly detection is often studied in laboratory conditions using hyper- or multi-spectral sensors (Behmann et al., 2015; Foerster et al., 2019) or limited data sets like the PlantVillage data set (Hughes and Salathé, 2015), where single leaves are presented in front of an artificial background. However, only a few studies have investigated the use of RGB cameras in the field, such as in the works of Amara et al. (2017), Publication 3 (Bömer et al., 2020), and Publication 5 (Miranda et al., 2022). In order to restrict the problem to regions of interest, we used the berry detection from Publication 2 (Zabawa et al., 2020), allowing for in-field applications.

In Publication 3 (Bömer et al., 2020), we demonstrated that a lightweight classification CNN and a broad anomaly definition in the data set can achieve convincing results. We defined a single class for damaged grapevine berries and were able to detect a variety of damages, ranging from sunburn to fungus infections. In contrast to defining a class for each disease type (Kaur et al., 2019), which requires more complex models, our approach simplified the process. Furthermore, comprehensive anomaly description and labeling is often challenging for methods that define a class for each disease type.

In Publication 5 (Miranda et al., 2022), we took the step further and eliminated the need for labeled anomaly data altogether by using a VAE trained on non-anomalous data only. Unlike other studies that used a pixel-wise loss (Akçay et al., 2018; Baur et al., 2019; Strothmann et al., 2019), we used a feature-wise perceptual loss, which resulted in qualitatively and quantitatively better results. Other studies, for example Pardede et al. (2018) used the extracted features from the CAE and classified using an SVM, bypassing the reconstruction step and yielding only classification results. In contrast, we compared the reconstructed image patches directly with the original ones and used the pixel-wise differences for the classification.

Publication 3 (Peer-reviewed, Conference)

J. Bömer, L. Zabawa, P. Sieren, A. Kicherer, L. Klingbeil, U. Rascher, O. Muller, H. Kuhlmann, and R. Roscher. Automatic differentiation of damaged and unharmed grapes using rgb images and convolutional neural networks. In *Proc. of the Europ. Conf. on Computer Vision (ECCV) Workshops*, pages 347–359. Springer International Publishing, 2020. ISBN 978-3-030-65414-6

The detection of damaged grapevine berries is crucial for winegrowers, as there are various factor that can cause damage. These factors include diseases, pests and abiotic stress like sun damage or damages caused by mechanical machines. However, identifying the underlying cause of the damage can be challenging, as the visual symptoms can be similar and may only be distinguishable by a trained experts. Furthermore, the same disease can manifest differently, making it challenging to develop a model that can account for all variations. To address this problem in the field of grapevine berry anomaly detection, we decided to only label two classes, *healthy* and *damaged* berries, using data from different grapevine varieties.

Our goal is to detect anomalies and guide farmers to damaged plant regions for further investigation, rather than to identify every type of damage. We realize this, by creating heatmaps indicating damaged berries in image. First we identify image regions containing berries and extract image patches containing these using a sliding window approach. We trained a shallow Convolutional Neural Network (CNN) (Section 2.2.2) to classify each patch and produce heatmaps showing image regions containing damaged berries.

Our approach was successful in identifying various types of damage, including cracked berry skin, withered berries, and color variations. We found that a shallow, non-pretrained NN outperformed a deeper, ImageNet pre-trained network. Our model was able to detect damaged plant material in images in the field under natural illumination, regardless of the grapevine variety. Overall, our approach provides a practical solution for detecting damaged grapevine berries and can help guide farmers to investigate potentially problematic areas.

Publication 5 (Peer-reviewed, Journal)

M. Miranda, L. Zabawa, A. Kicherer, L. Strothmann, U. Rascher, and R. Roscher. Detection of anomalous grapevine berries using variational autoencoders. *Frontiers in Plant Science*, 13, 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.729097. URL `https://www.frontiersin.org/articles/10.3389/fpls.2022.729097`

In this publication, we went beyond the previous work by elimination the need for labelled anomalous data altogether. Instead, we trained a Variational Autoencoder (VAE) (Section 2.2.3) using a Feature Perceptual Loss (FPL) exclusively on healthy data acquired in the field using a modified grapevine harvester. The measurements were collected in a closed chamber with artificial lighting. We used images from different varieties and growth stages and employed a method from Publication 2 (Zabawa et al., 2020) to identify regions containing berries to limit complexity.

To evaluate our approach, we compared it with an Autoencoder (AE) using a

pixel-wise loss, Structural Similarity Index Measure (SSIM), and the same FPL used in the VAE. Our results showed that the performance of the FPL improved the performance of the AE, but the combination of VAE and FPL produced the best results.

The anomaly detection itself is realized by comparing the reconstructed image patch with the original image patch using different losses, the Least Absolute Deviations (L1), Binary Cross Entropy (BCE) and the Mean Squared Error (MSE). Using a iterative optimization strategy for the loss histograms, we determined a threshold to distinguish healthy and anomalous image patches.

Overall, our framework generated heatmaps that indicate areas with diseased or damaged berries, providing an effective solution for anomaly detection without the need for labelled anomalous data.

Chapter 4

Conclusion and Outlook

High-throughput phenotyping plays a crucial role in enhancing the profitability and sustainability of vineyards. However, the current practice of relying on skilled experts for screening often yields subjective outcomes, heavily influenced by historical data and personal experiences. Additionally, conventional sampling strategies fail to consider the spatial variability within vineyards, leading to potential inaccuracies. To address these challenges, image-based methods offer a more comprehensive and objective approach, empowering winegrowers and breeders to make informed decisions regarding vineyard management.

This thesis presents contributions to three key challenges in viticulture phenotyping: Detection and counting, yield estimation and anomaly detection. In the first part of the thesis we presented a novel instance segmentation approach for the detection and counting of single grapevine berries in images. This method was used in the subsequent works, either to provide the number of visible yield components, in this case grapevine berries, for a yield estimation, or to identify regions of interests for berry anomaly detection.

The following sections provide a concise summary of our contributions to each of the three main phenotyping tasks. Furthermore, an outlook on future research directions will be presented.

- **Detection and Counting**

  In Section 3.1 we presented a lightweight and novel instance segmentation approach which is highly suited for the detection of grapevine berries. The main concept is to enhance the semantic segmentation by introducing an additional class for edges, which facilitates the differentiation of individual instances within a particular class. The data annotation strategy is compatible with various semantic segmentation networks, however, we found that utilizing a lightweight network architecture like the MobileNetV2 can

deliver both rapid and reliable results. Our method outperforms a state-of-the-art instance segmentation network, the Mask-RCNN and achieves a higher correlation between the actual and predicted count of grapevine berries in images. Furthermore, it offers significant advantages over a regression-based approach, as it can extract more comprehensive traits like the berry size and shape. Our studies have shown that the method delivers robust performance across two different training systems and three different grapevine varieties, even under realistic conditions in German vineyards.

One promising area for future research is to enhance the robustness of fruit detection in varying environments and with different sensors. Currently, the method has been developed primarily for a limited dataset collected with a field phenotyping platform with an artificial background and lighting. Although we showed, that the algorithm is able to detect fruits in images taken with a handheld camera and natural background, the performance is significantly worse compared to the original data set. Other studies have explored domain adaptation techniques for agricultural applications, specifically for crop-weed segmentation in images. These approaches utilized methods such as GANs (Gogoll et al., 2020) and Fourier transforms (Vasconcelos et al., 2021) to investigate transferability across different cameras, study sites and growth stages. Results showed, that transfer across different cameras and platforms was possible, but difficulties were encountered when transferring across growth stages. Bertoglio et al. (2023) directly compared the approaches and came to similar results. It is worth exploring how those techniques perform when transferring from an artificial background to a natural one, as well as simulating various lighting conditions.

- **Yield estimation**

In Section 3.2 we presented an image-based analysis of yield parameters for viticultural applications. Our experiments were conducted under realistic conditions in German vineyards, where a moderate defoliation is customary. This sets them apart from many other works which operate on highly defoliated vines, enabling a very good visibility of the yield components. Our work identified leaf occlusion variability as the most critical limiting factor, especially for the Semi Minimal Pruned Hedge (SMPH) system. Despite this challenge, we achieved an image-based yield estimation for vines trained in the Vertical Shoot Positioned (VSP) that is slightly better (27% yield variation) than the average industrial standard (30 %), without the need of a highly specialized expert with years of experience and the incorporation of historic data.

To address this limitation and improve our yield estimation accuracy, one potential future direction is to build upon the work of Kierdorf et al. (2022), who successfully generated images of berries hidden behind leaves. Their approach demonstrated better performance compared to using a pre-set occlusion factor, which is unable to account for the high variability of occlusion in the field. By using machine learning techniques to generate images of occluded yield components, we may be able to improve our yield estimation accuracy. In addition to improving our methodology, we also need to conduct a large-scale experiment to demonstrate the superiority of our machine-learning-based approach compared to traditional sample-based methods. This requires collecting more data and reference measurements, as well as conducting row-wise measurements to better capture variability across the vineyard. Another critical aspect of our work is the transferability of our approach across different sites and varieties. To address this, we could draw inspiration from the work of Ma et al. (2021), who proposed an adaptive adversarial domain adaptation method for corn yield estimation using satellite data. By developing a general yield estimation method, that can work across different varieties and locations, we can significantly enhance our ability to estimate yield in different vineyard settings.

- **Anomaly detection**

  In Section 3.3 we presented two different approaches for the detection of anomalous grapevine berries in images. Both rely on the assumption that it is highly improbable for all possible expressions of different diseases or damages to be fully defined and accounted for in a data set. In Publication 3 (Bömer et al., 2020) we established two classes: *healthy* and *damaged*, and demonstrated, that a shallow Neural Network (NN) can accurately classify image patches into these two categories. By utilizing a sliding window approach, we were able to generate heatmaps that identify the locations of damages across entire grape bunches. We made a deliberate decision not to specify distinct classes of damage or disease, as it would have been impossible to collect data representing all potential types of damage. By taking this approach, we aimed to avoid the potential for incomplete or biased data, and instead focused on identifying common patterns and characteristics across a broad range of damage scenarios. In Publication 5 (Miranda et al., 2022), we build upon these ideas and take them further by eliminating the need for anomalous data in our network training process. We proposed the usage of a Variational Autoencoder (VAE) in combination with a Feature Perceptual Loss (FPL). The network is designed to learn the characteristics

of a healthy phenotype and use this knowledge to identify any deviations from the norm as anomalies.

One potential research direction to improve the current approach is to develop an AE that can jointly handle the region of interest and complex backgrounds. This could lead to more accurate and robust anomaly detection on whole images, including leaves or other plant material. In addition to the need for improved anomaly detection models, there is also a growing demand for uncertainty measures to facilitate practical applications. While research in this area is primarily focused on life-safety applications like autonomous driving or medical image processing (Dolezal et al., 2022), uncertainty estimation is also critical in the field of agriculture. Several methods can be used to estimate uncertainty, including dropout, deep ensembles, and test time augmentations. A comprehensive review of uncertainty estimation research for deep learning techniques can be found in Mena et al. (2021) and Abdar et al. (2021). By incorporating uncertainty measures into our anomaly detection models, we can provide more transparent and reliable decision-making for practical applications in agriculture.

Although we have made significant contributions to the field of high-throughput phenotyping in viticulture, practical applications that can be used by winemakers and vineyard managers still pose significant challenges. Robust, reliable, and transferable algorithms are necessary, but scalability, standardization, and cost-effectiveness must also be addressed. Developing smaller, more agile field phenotyping platforms is essential. Addressing these challenges will require further research and development to refine and optimize high-throughput phenotyping methods for use in the wine industry. Collaboration between researchers, winemakers, and other stakeholders can identify specific needs and opportunities for improvement. Ongoing validation and testing of new methods in real-world settings will be crucial to this endeavor.

Chapter 5

List of further publications

This chapter gives a chronological overview of further publications in which the author of this dissertation was involved. The publications listed here are excluded from the main contributions as they are not directly related to this thesis or the author of the thesis participated only as a coauthor.

Peer-reviewed publications:

- L. Zabawa, A. Kicherer, L. Klingbeil, A. Milioto, R. Töpfer, and H. Kuhlmann. Detektion von Weintrauben in Bildern mit Hilfe von Fully Convolutional Neural Nets. In Leibniz Institut für Agrartechnik und Bioökonomie e. V. (ATB) Potsdam, Michael Pflanz, Michael Schirrmann, Marius Hobart, Lasse Klingbeil, and Jan Behmann, editors, *25. Workshop Computer und Bildanalyse in der Landwirtschaft*, 17. April 2019, Bonn, pages 15–20, 2019a

- J. Kierdorf, L. Zabawa, L. Lucks, L. Klingbeil, and H. Kuhlmann. Erkennung und Zählung von Weizenähren mit Hilfe bodengestützten Bildaufnahmen. In Leibniz Institut für Agrartechnik und Bioökonomie e. V. (ATB) Potsdam, Michael Pflanz, Michael Schirrmann, Marius Hobart, Lasse Klingbeil, and Jan Behmann, editors, *25. Workshop Computer und Bildanalyse in der Landwirtschaft*, 17. April 2019, Bonn, pages 158–167, 2019

- S. Yang, L. Zheng, X. Chen, L. Zabawa, M. Zhang, and M. Wang. Transfer learning from synthetic in-vitro soybean pods dataset for in-situ segmentation of on-branch soybean pods. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1665–1674, 2022. doi: 10.1109/CVPRW56347.2022.00173

- J. Kierdorf, I. Weber, A. Kicherer, L. Zabawa, L. Drees, and R. Roscher. Behind the leaves: Estimation of occluded grapevine berries with conditional generative adversarial networks. *Frontiers in Artificial Intelligence*, 5, 2022. ISSN 2624-8212. doi: 10.3389/frai.2022.830026. URL `https://www.frontiersin.org/articles/10.3389/frai.2022.830026`

Open Source Contributions

Besides the peer-reviewed papers, the data set used in Publication 1 (Zabawa et al., 2019b) and Publication 2 (Zabawa et al., 2020) was published:

- L. Zabawa and A. Kicherer. Segmentation of wine berries, 2021. URL `https://www.openagrar.de/receive/openagrar_mods_00067631` The data set is available online at: `https://www.openagrar.de/receive/openagrar_mods_00067631`

Acronyms

AdaGrad Adaptive Gradient Algorithm.

Adam Adaptive Moment Estimation.

AE Autoencoder.

BB Bounding Box.

BBCH Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie.

BCE Binary Cross Entropy.

CAE Convolutional Autoencoder.

CNN Convolutional Neural Network.

CV Computer Vision.

DL Deep Learning.

FCNN Fully Connected Neural Network.

FN False negative.

FP False positive.

FPL Feature Perceptual Loss.

GAN Generative Adversarial Network.

GD Gradient Descent.

IoU Intersection over Union.

L1 Least Absolute Deviations.

ML Machine Learning.

MP Minimal Pruning.

MSE Mean Squared Error.

NDVI Normalized Difference Vegetation Index.

NN Neural Network.

OIV International Organisation of Vine and Wine.

ReLU Rectified Linear Unit.

SfM Structure from Motion.

SGD Stochastic Gradient Descent.

SMPH Semi Minimal Pruned Hedge.

SSIM Structural Similarity Index Measure.

SVM Support Vector Machine.

TN True negative.

TP True positive.

UAV Unmanned Aerial Vehicle.

VAE Variational Autoencoder.

VSP Vertical Shoot Positioned.

Bibliography

M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, 2021. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2021.05.008.

A. Akçay, A. Atapour-Abarghouei, and T. P. Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Proc. of the Asian Conf. on Computer Vision (ACCV)*, pages 622–637. Springer, 2018.

A. Alercia, R. Becher, J.-M. Boursiquot, R. Carara, P. Chomé amd A. Costacurta, and et al. 2nd edition of the oiv descriptor list for grape varieties and vitis species, 2009. URL `http://www.oiv.int/en/`.

J. Amara, B. Bouaziz, and A. Algergawy. A deep learning-based approach for banana leaf diseases classification. *BTW Workshop*, pages 79–88, 2017.

N. T. Anderson, K. Walsh, and D. Wulfsohn. Technologies for forecasting tree fruit load and harvest timing—from ground, sky and time. *Agronomy*, 11(7), 2021. doi: https://doi.org/10.3390/agronomy11071409.

A. Aquino, M. P. Diago, B. Millan, and J. Tardaguila. A new methodology for estimating the grapevine-berry number per cluster using image analysis. *Biosystems Engineering*, 159:80 – 95, 2016. doi: https://doi.org/10.1016/j.biosystemseng.2016.12.011.

A. Aquino, B. Millan, M.-P. Diago, and J. Tardaguila. Automated early yield prediction in vineyards from on-the-go image acquisition. *Computers and Electronics in Agriculture*, 144:26 – 36, 2018. doi: https://doi.org/10.1016/j.compag.2017.11.026.

C. Arteta, V. Lempitsky, and A. Zisserman. Counting in the wild. In *Proc. of the Europ. Conf. on Computer Vision (ECCV)*, 2016.

S. Bargoti and J. Underwood. Image segmentation for fruit detection and yield estimation in apple orchards. *Journal of Field Robotics (JFR)*, 34:1039–1060, 2017. doi: [10.1002/rob.21699](http://dx.doi.org/10.1002/rob.21699).

C. Baur, B. Wiestler, S. Albarqouni, and N. Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 161–169. Springer International Publishing, 2019. ISBN 978-3-030-11723-8.

J. Behmann, A.-K. Mahlein, T. Rumpf, C. Römer, and L. Plümer. A review of advanced machine learning methods for the detection of biotic stress in crop protection. *Agriculture*, 16:239–260, 06 2015. doi: 10.1007/s11119-014-9372-7.

N. Bendel, A. Kicherer, A. Backhaus H.-C. Klück, U. Seiffert, M. Fischer, R.T. Voegele, and R. Töpfer. Evaluating the suitability of hyper- and multispectral imaging to detect foliar symptoms of the grapevine trunk disease esca in vineyards. *Plant methods*, 16, 2020. doi: https://doi.org/10.1186/s13007-020-00685-3.

R. Bertoglio, A. Mazzucchelli, N. Catalano, and M. Matteucci. A comparative study of fourier transform and cyclegan as domain adaptation techniques for weed segmentation. *Smart Agricultural Technology*, 4:100188, 2023. ISSN 2772-3755. doi: https://doi.org/10.1016/j.atech.2023.100188.

J. Bömer, L. Zabawa, P. Sieren, A. Kicherer, L. Klingbeil, U. Rascher, O. Muller, H. Kuhlmann, and R. Roscher. Automatic differentiation of damaged and unharmed grapes using rgb images and convolutional neural networks. In *Proc. of the Europ. Conf. on Computer Vision (ECCV) Workshops*, pages 347–359. Springer International Publishing, 2020. ISBN 978-3-030-65414-6.

H. Cecotti, A. Rivera, M. Farhadloo, and M. A. Pedroza. Grape detection with convolutional neural networks. *Expert Systems with Applications*, 159:113588, 2020. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2020.113588. URL `https://www.sciencedirect.com/science/article/pii/S0957417420304127`.

S. Charters and S. Pettigrew. The dimensions of wine quality. *Food Quality and Preference*, 18(7):997–1007, 2007.

L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proc. of the Europ. Conf. on Computer Vision (ECCV)*, abs/1802.02611:833 – 851, 2018. doi: https://doi.org/10.1007/978-3-030-01234-2_49.

P. Clingeleffer, G. Dunn, M. Krstic, and S. Martin. 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*, 2001.

J. P. Cohen, G. Boucher, C. A. Glastonbury, H. Z. Lo, and Y. Bengio. Count-ception: Counting by fully convolutional redundant counting. *Proc. of the Int. Conf. on Computer Vision (ICCV) Workshops*, abs/1703.08710:18 – 26, 2017. doi: 10.1109/ICCVW.2017.9.

L. Coviello, M. Cristoforetti, G. Jurman, and C. Furlanello. Gbcnet: In-field grape berries counting for yield estimation by dilated cnns. *Applied Sciences*, 10(14), 2020. doi: https://doi.org/10.3390/app10144870.

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

M. de la Fuente, R. Linares, P. Baeza, C. Miranda, and J. R. Lissarrague. Comparison of different methods of grapevine yield prediction in the time window between fruitset and veraison. *OENO One*, 49(1):27–35, March 2015. doi: 10.20870/oeno-one.2015.49.1.96. URL `[https://oeno-one.eu/article/view/96](https://oeno-one.eu/article/view/96)`.

S. Di Gennaro, E. Battiston, S. Di Marco, O. Facini, A. Matese, M. Nocentini, A. Palliotti, and L. Mugnai. Unmanned aerial vehicle (uav)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex. *Phytopathologia Mediterranea*, 55:262−275, 08 2016. doi: 10.14601/Phytopathol_Mediterr-18312.

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

M.-P. Diago, C. Correa, B. Millán, P. Barreiro, C. Valero, and J. Tardaguila. Grapevine yield and leaf area estimation using supervised classification methodology on rgb images taken under field conditions. *IEEE Sensors Journal*, 12:16988 – 17006, 2012. doi: https://doi.org/10.3390/s121216988.

J. M. Dolezal, A. Srisuwananukorn, D. Karpeyev, S. Ramesh, S. Kochanny, B. Cody, A. S. Mansfield, S. Rakshit, R. Bansal, M. C. Bois, A. O. Bungum,

J. J. Schulte, E. E. Vokes, M. C. Garassino, A. N. Husain, and A. T. Pearson. Uncertainty-informed deep learning models enable high-confidence predictions for digital histopathology. *Nature Communications*, 13, 2022. doi: 10.1038/s41467-022-34025-x.

U.-O. Dorj, M. Lee, and S.-S. Yun. An yield estimation in citrus orchards via fruit detection and counting using image processing. *Computers and Electronics in Agriculture*, 140:103–112, 2017. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2017.05.019. URL https://www.sciencedirect.com/science/article/pii/S0168169916312455.

J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12 (61):2121–2159, 2011. URL http://jmlr.org/papers/v12/duchi11a.html.

V. Dumoulin and F. Visin. A guide to convolution arithmetic for deep learning, 2016. URL https://arxiv.org/abs/1603.07285.

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

G. M. Dunn and S. R. Martin. The current status of crop forecasting in the australian wine industry. In *Proceedings of the ASVO Seminar Series: Grapegrowing at the Edge, Tanunda, Barossa Valley, South Australia*, pages 4–8, 2003.

A. Foerster, J. Behley, J. Behmann, and R. Roscher. Hyperspectral plant disease forecasting using generative adversarial networks. In *International Geoscience and Remote Sensing Symposium*, 2019. doi: https://doi.org/10.1109/IGARSS.2019.8898749.

H. Gan, W. S. Lee, V. Alchanatis, and A. Abd-Elrahman. Active thermal imaging for immature citrus fruit detection. *Biosystems Engineering*, 198:291–303, 2020. ISSN 1537-5110. doi: https://doi.org/10.1016/j.biosystemseng.2020.08.015. URL https://www.sciencedirect.com/science/article/pii/S1537511020302348.

F. Gao, L. Fu, X. Zhang, Y. Majeed, R. Li, M. Karkee, and Q. Zhang. Multi-class fruit-on-plant detection for apple in snap system using faster r-cnn. *Computers and Electronics in Agriculture*, 176:105634, 2020. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2020.105634. URL https://www.sciencedirect.com/science/article/pii/S0168169920314009.

J. Gené-Mola, E. Gregorio, F. Auat Cheein, J. Guevara, J. Llorens, R. Sanz-Cortiella, A. Escolà, and J. R. Rosell-Polo. Fruit detection, yield prediction and canopy geometric characterization using lidar with forced air flow. *Computers and Electronics in Agriculture*, 168:105121, 2020a. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2019.105121. URL `https://www.sciencedirect.com/science/article/pii/S0168169919313390`.

J. Gené-Mola, R. Sanz-Cortiella, J. R. Rosell-Polo, J.-R. Morros, J. Ruiz-Hidalgo, V. Vilaplana, and E. Gregorio. Fruit detection and 3d location using instance segmentation neural networks and structure-from-motion photogrammetry. *Computers and Electronics in Agriculture*, 169:105165, 2020b. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2019.105165. URL `https://www.sciencedirect.com/science/article/pii/S0168169919321507`.

D. Gogoll, P. Lottes, J. Weyler, N. Petrinic, and C. Stachniss. Unsupervised domain adaptation for transferring plant classification systems to new field environments, crops, and robots. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 2636–2642, 2020. doi: 10.1109/IROS45743.2020.9341277.

A. Gongal, A. Amatya, M. Karkee, Q. Zhang, and K. Lewis. Sensors and systems for fruit detection and localization: A review. *Computers and Electronics in Agriculture*, 116:8–19, 2015. doi: https://doi.org/10.1016/j.compag.2015.05.021).

Y. Guo, J. Stein, G. Wu, and A. Krishnamurthy. Sau-net: A universal deep network for cell counting. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '19, pages 299–306, 2019. doi: https://doi.org/10.1145/3307339.3342153.

S. Gutiérrez, J. Fernández-Novales, M. P. Diago, and J. Tardaguila. On-the-go hyperspectral imaging under field conditions and machine learning for the classification of grapevine varieties. *Frontiers in Plant Science*, 9, 2018. ISSN 1664-462X. doi: 10.3389/fpls.2018.01102. URL `https://www.frontiersin.org/articles/10.3389/fpls.2018.01102`.

C. Hacking, N. Poona, N. Manzan, and C. Poblete-Echeverría. Investigating 2-d and 3-d proximal remote sensing techniques for vineyard yield estimation. *IEEE Sensors Journal*, 10(17), 2019. doi: https://doi.org/10.3390/s19173652.

B. Harel, R. van Essen, Y. Parmet, and Y. Edan. Viewpoint analysis for maturity classification of sweet peppers. *IEEE Sensors Journal*, 20(13):3783, 2020. doi: https://doi.org/10.3390/s20133783.

K. He, G. Gkioxari, P. Dollár, and R. B. Girshick. Mask r-cnn. *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*, pages 2980–2988, 2017. doi: https://doi.org/10.1109/ICCV.2017.322.

J. Hemming, J. Ruizendaal, J. W. Hofstee, and E. J. Van Henten. Fruit detectability analysis for different camera positions in sweet-pepper. *IEEE Sensors Journal*, 14(4):6032–6044, 2014. ISSN 1424-8220. doi: 10.3390/s140406032. URL `https://www.mdpi.com/1424-8220/14/4/6032`.

G. Howell. Sustainable grape productivity and the growth-yield relationship: A review. *American Journal of Enology and Viticulture*, 52:165–174, 01 2001. doi: 10.5344/ajev.2001.52.3.165.

D. P. Hughes and M. Salathé. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. *CoRR*, abs/1511.08060, 2015. URL `http://arxiv.org/abs/1511.08060`.

C. Intrieri, I. Filippetti, G. Allegro, G. Valentini, C. Pastore, and E. Colucci. The semi-minimal-pruned hedge: A novel mechanized grapevine training system. *American Journal of Enology and Viticulture*, 62(3):312–318, 2011. ISSN 0002-9254. doi: 10.5344/ajev.2011.10083. URL `https://www.ajevonline.org/content/62/3/312`.

E. Ivorra, A. Sánchez, J. Camarasa, M. P. Diago, and J. Tardaguila. Assessment of grape cluster yield components based on 3d descriptors using stereo vision. *Food Control*, 50:273 – 282, 2015. doi: https://doi.org/10.1016/j.foodcont.2014.09.004.

D. Jackson. *Pruning and Training. Monographs in Cool Climate Viticulture*. Dunmore Publishing Limited, 1997.

G. V. Jones. Climate change: observations, projections, and general implications for viticulture and wine production. *Whitman College Economics Department working paper*, 7:14, 2007.

A. Kalantar, Y. Edan, A. Gur, and I. Klapp. A deep learning system for single and overall weight estimation of melons using unmanned aerial vehicle images. *Computers and Electronics in Agriculture*, 178:105748, 2020. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2020.105748. URL `https://www.sciencedirect.com/science/article/pii/S0168169920304804`.

A. Kamilaris and F. X. Prenafeta-Boldu. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90, 2018. doi: https://doi.org/10.1016/j.compag.2018.02.016.

S. Kaur, S. Pandey, and S. Goel. Plants disease identification and classification through leaf images: A survey. *Archives of Computational Methods in Engineering*, 26(2):507–530, 2019.

S. D. Khirade and A.B. Patil. Plant disease detection using image processing. In *International Conference on Computing Communication Control and Automation*, pages 768–771, 2015.

A. Kicherer, R. Roscher, K. Herzog, W. Förstner, and R. Töpfer. Image based evaluation for the detection of cluster parameters in grapevine. In *Proc. of the XI International Conference on Grapevine Breeding and Genetics*, pages 335–340, 2014. doi: https://doi.org/10.17660/ActaHortic.2015.1082.46.

A. Kicherer, K. Herzog, N. Bendel, H.-C. Klück, A. Backhaus, M. Wieland, J. C. Rose, L. Klingbeil, T. Läbe, C. Hohl, W. Petry, H. Kuhlmann, U. Seiffert, and R. Töpfer. Phenoliner: A new field phenotyping platform for grapevine research. *IEEE Sensors Journal*, 2017. doi: https://doi.org/10.3390/s17071625.

J. Kierdorf, L. Zabawa, L. Lucks, L. Klingbeil, and H. Kuhlmann. Erkennung und Zählung von Weizenähren mit Hilfe bodengestützten Bildaufnahmen. In Leibniz Institut für Agrartechnik und Bioökonomie e. V. (ATB) Potsdam, Michael Pflanz, Michael Schirrmann, Marius Hobart, Lasse Klingbeil, and Jan Behmann, editors, *25. Workshop Computer und Bildanalyse in der Landwirtschaft*, 17. April 2019, Bonn, pages 158–167, 2019.

J. Kierdorf, I. Weber, A. Kicherer, L. Zabawa, L. Drees, and R. Roscher. Behind the leaves: Estimation of occluded grapevine berries with conditional generative adversarial networks. *Frontiers in Artificial Intelligence*, 5, 2022. ISSN 2624-8212. doi: 10.3389/frai.2022.830026. URL `https://www.frontiersin.org/articles/10.3389/frai.2022.830026`.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proc. of the Int. Conf. on Learning Representations (ICLR*, 2015. URL `http://arxiv.org/abs/1412.6980`.

A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy. Deep learning – method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, 162:219–234, 2019a. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2019.04.017. URL `https://www.sciencedirect.com/science/article/pii/S0168169919301164`.

A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'man-

goyolo'. *Agriculture*, 20:1107–1135, 2019b. doi: https://doi.org/10.1007/s11119-019-09642-0.

A. Koirala, K. Walsh, and Z. Wang. Attempting to estimate the unseen—correction for occluded fruit in tree fruit load estimation by machine vision with deep learning. *Agronomy*, 11(2), 2021. doi: https://doi.org/10.3390/agronomy11020347.

C. Kraus, T. Pennington, K. Herzog, A. Hecht, M. Fischer, R. T. Voegele, C. Hoffmann, R. Töpfer, and A. Kicherer. Effects of canopy architecture and microclimate on grapevine health in two training systems. *Vitis*, 57:53–60, 2018.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Proc. of the Advances in Neural Information Processing Systems (NIPS)*, volume 25. Curran Associates, Inc., 2012. URL `https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf`.

P. Kurtser and Y. Edan. Statistical models for fruit detectability: spatial and temporal analyses of sweet peppers. *Biosystems Engineering*, 171:272–289, 07 2018a. doi: 10.1016/j.biosystemseng.2018.04.017.

P. Kurtser and Y. Edan. The use of dynamic sensing strategies to improve detection for a pepper harvesting robot. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 8286–8293, 2018b. doi: 10.1109/IROS.2018.8593746.

P. Kurtser, O. Ringdahl, N. Rotstein, R. Berenstein, and Y. Edan. 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(2): 2031–2038, 2020. doi: 10.1109/LRA.2020.2970654.

F. Kurtulmus, W. S. Lee, and A. Vardar. Green citrus detection using 'eigenfruit', color and circular gabor texture features under natural outdoor conditions. *Computers and Electronics in Agriculture*, 78(2):140–149, 2011. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2011.07.001. URL `https://www.sciencedirect.com/science/article/pii/S0168169911001475`.

Landwirtschaftskammer Rheinland-Pfalz. Grundsätze für die Durchführung der Hektarertragsregelung - Erzeugerstufe in Rheinland-Pfalz. `https://www.lwk-rlp.de/fileadmin/lwk-rlp.de/Weinbau/PDF/Infoblatt_GHE_eigene_Erzeugnisse.pdf`, 2012. Accessed: 2023-05-05.

V. Lempitsky and A. Zisserman. Learning to count objects in images. *Proc. of the Advances in Neural Information Processing Systems (NIPS)*, 23:1324–1332, 2010. URL https://proceedings.neurips.cc/paper/2010/file/fe73f687e5bc5280214e0486b273a5f9-Paper.pdf.

R. Linker, O. Cohen, and A. Naor. Determination of the number of green apples in rgb images recorded in orchards. *Computers and Electronics in Agriculture*, 81:45–57, 2012. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2011.11.007. URL https://www.sciencedirect.com/science/article/pii/S0168169911002638.

S. Liu, X. Zeng, and M. Whitty. A vision-based robust grape berry counting algorithm for fast calibration-free bunch weight estimation in the field. *Computers and Electronics in Agriculture*, 173:105360, 2020. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2020.105360. URL https://www.sciencedirect.com/science/article/pii/S0168169919326432.

S. Lobry and D. Tuia. Deep learning models to count buildings in high-resolution overhead images. In *2019 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4, May 2019. doi: 10.1109/JURSE.2019.8809058.

J. Long, E. Shelhamer, and T.Darell. Fully convolutional networks for semantic segmentation. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 3431 – 3440, 2015. doi: https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298965.

D. Lorenz, L. Eichhorn, H. Bleiholder, R. Klose, U. Meier, and E. Weber. 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*, 1:100–103, 1995. doi: https://doi.org/10.1111/j.1755-0238.1995.tb00085.x.

E. Lu, W. Xie, and A. Zisserman. Class-agnostic counting. In *Proc. of the Asian Conf. on Computer Vision (ACCV)*, 2018.

L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 152:166–177, 2019.

Y. Ma, Z. Zhang, H. L. Yang, and Z. Yang. An adaptive adversarial domain adaptation approach for corn yield prediction. *Computers and Electronics in Agriculture*, 187:106314, 2021. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2021.106314. URL https://www.sciencedirect.com/science/article/pii/S0168169921003318.

F. Martinez de Toda and J.C. Sancha. Long-term effects of zero pruning on grenache vines under drought conditions. *Vitis*, 37:155–157, 12 1998.

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

J. Mena, O. Pujol, and J. Vitrià. A survey on uncertainty estimation in deep learning classification systems from a bayesian perspective. *ACM Comput. Surv.*, 54(9), 2021. ISSN 0360-0300. doi: 10.1145/3477140. URL `https://doi.org/10.1145/3477140`.

A. Milioto and C. Stachniss. Bonnet: An open-source training and deployment framework for semantic segmentation in robotics using cnns. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7094–7100, 2019. doi: 10.1109/ICRA.2019.8793510.

B. Millan, M.P. Diago, A. Aquino, F. Palacios, and J. Tardaguila. Vineyard pruning weight assessment by machine vision: Towards an on-the-go measurement system. *OENO One*, 53(2):334–347, 2019. doi: 10.20870/oeno-one.2019.53.2.2416.

M. Miranda, L. Zabawa, A. Kicherer, L. Strothmann, U. Rascher, and R. Roscher. Detection of anomalous grapevine berries using variational autoencoders. *Frontiers in Plant Science*, 13, 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.729097. URL `https://www.frontiersin.org/articles/10.3389/fpls.2022.729097`.

L. Mohimont, F. Alin, M. Rondeau, N. Gaveau, and L. A. Steffenel. Computer vision and deep learning for viticulture. *Agronomy*, 12, 10 2022. doi: 10.3390/agronomy12102463.

D. Molitor, M. Schultz, R. Mannes, M. Pallez-Barthel, L. Hoffmann, and M. Beyer. Semi-minimal pruned hedge: A potential climate change adaptation strategy in viticulture. *Agronomy*, 9(4), 2019. ISSN 2073-4395. doi: 10.3390/agronomy9040173. URL `https://www.mdpi.com/2073-4395/9/4/173`.

A. K. Nellithimaru and G. A. Kantor. Rols : Robust object-level slam for grape counting. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2648–2656, June 2019. doi: 10.1109/CVPRW.2019.00321.

S. Nuske, S. Achar, T. Bates, S. Narasimhan, and S. Singh. Yield estimation in vineyards by visual grape detection. *IEEE/RSJ International Conference on*

Intelligent Robots and Systems, pages 2352–2358, 2011. doi: https://doi.org/10.1109/IROS.2011.6095069.

S. Nuske, K. Wilshusen, S. Achar, L. Yoder, S. Narasimhan, and S. Singh. Automated visual yield estimation in vineyards. *Journal of Field Robotics (JFR)*, 31:837 – 860, 2014. doi: https://doi.org/10.1109/IROS.2011.6095069.

E. K. Nyarko, I. Vidović, K. Radočaj, and R. Cupec. A nearest neighbor approach for fruit recognition in rgb-d images based on detection of convex surfaces. *Expert Systems with Applications*, 114:454–466, 2018. doi: https://doi.org/10.1016/j.eswa.2018.07.048.

F. Palacios López, M.-P. Diago, P. Melo-Pinto, and J. Tardaguila. Early yield prediction in different grapevine varieties using computer vision and machine learning. *Agriculture*, 24:1–29, 08 2022. doi: 10.1007/s11119-022-09950-y.

G. Pang, C. Shen, L. Cao, and A. Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2):1–38, 2021.

H. F. Pardede, E. Suryawati, R. Sustika, and V. Zilvan. Unsupervised convolutional autoencoder-based feature learning for automatic detection of plant diseases. In *International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pages 158–162, 2018.

F. Picetti, G. Testa, F. Lombardi, P. Bestagini, M. Lualdi, and S. Tubaro. Convolutional autoencoder for landmine detection on gpr scans. In *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, pages 1–4, 2018.

D. K. Ray, N. D. Mueller, P. C. West, and J. A. Foley. Yield trends are insufficient to double global crop production by 2050. *PLoS One*, 8, 2013. doi: https://doi.org/10.1371/journal.pone.0066428.

A. G. Reynolds and J. E. Vanden Heuvel. Influence of grapevine training systems on vine growth and fruit composition: A review. *American Journal of Enology and Viticulture*, 60(3):251–268, 2009. ISSN 0002-9254. doi: 10.5344/ajev.2009.60.3.251. URL `https://www.ajevonline.org/content/60/3/251`.

F. Rist, K. Herzog, J. Mack, R. Richter, V. Steinhage, and R. Töpfer. High-phenotyping of grape bunch architecture using fast 3d sensor and automation. *IEEE Sensors Journal*, 18:763, 03 2018. doi: 10.3390/s18030763.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-*

Assisted Intervention (MICCAI), 9351:234–241, 2015. doi: https://doi.org/10.1007/978-3-319-24574-4_28.

R. Roscher, K. Herzog, A. Kunkel, A. Kicherer, R. Töpfer, and W. Förstner. Automated image analysis framework for high throughput determination of grapevine berry size using conditional random fields. *Computers and Electronics in Agriculture*, 100:148–158, 2014. doi: https://doi.org/10.1016/j.compag.2013.11.008.

J.C. Rose, A. Kicherer, M. Wieland, L. Klingbeil, R. Töpfer, and H. Kuhlmann. Towards automated large-scale 3d phenotyping of vineyards under field conditions. *IEEE Sensors Journal*, 16, 2016. doi: https://doi.org/10.3390/s16122136).

S. Sayago and Monica Bocco. Crop yield estimation using satellite images: Comparison of linear and non-linear models. *AgriScientia*, 1:1, 06 2018. doi: 10.31047/1668.298x.v1.n35.20447.

J. Schmidhuber. Deep learning in neural networks: an overview. *Neural networks : the official journal of the International Neural Network Society*, 61:85–117, 2015. doi: 10.1016/j.neunet.2014.09.003.

L. Shen, J. su, R. He, L. Song, R. Huang, Y Fang, Y. Song, and B. Su. Real-time tracking and counting of grape clusters in the field based on channel pruning with yolov5s. *Computers and Electronics in Agriculture*, 206:107662, 03 2023. doi: 10.1016/j.compag.2023.107662.

N. Shvetsova, B. Bakker, I. Fedulova, H. Schulz, and D. V. Dylov. Anomaly detection in medical imaging with deep perceptual autoencoders. *IEEE Access*, 9:118571–118583, 2021. doi: 10.1109/ACCESS.2021.3107163.

D. L. Silver and T. Monga. In vino veritas: Estimating vineyard grape yield from images using deep learning. In *Advances in Artificial Intelligence*, pages 212–224. Springer International Publishing, 2019. ISBN 978-3-030-18305-9. doi: https://doi.org/10.1007/978-3-030-18305-9_17.

L. Strothmann, U. Rascher, and R. Roscher. Detection of anomalous grapevine berries using all-convolutional autoencoders. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 3701–3704, 2019. doi: 10.1109/IGARSS.2019.8898366.

A.C. Tagarakis, S. Koundouras, S. Fountas, and T. Gemtos. Evaluation of the use of lidar laser scanner to map pruning wood in vineyards and its potential for management zones delineation. *Agriculture*, 19:334–347, 2018. doi: 10.1007/s11119-017-9519-4.

Y. Tang, M. Chen, C. Wang, L. Luo, J. Li, G. Lian, and X. Zou. Recognition and localization methods for vision-based fruit picking robots: A review. *Frontiers in Plant Science*, 11, 2020. doi: https://doi.org/10.3389/fpls.2020.00510.

J. Tardaguila, M. Stoll, S. Gutiérrez, T. Proffitt, and M. P. Diago. Smart applications and digital technologies in viticulture: A review. *Smart Agricultural Technology*, 1:100005, 2021. ISSN 2772-3755. doi: https://doi.org/10.1016/j.atech.2021.100005. URL `https://www.sciencedirect.com/science/article/pii/S2772375521000058`.

F. Vanegas, D. Bratanov, J. Weiss, K. Powell, and F. Gonzalez. Multi and hyperspectral uav remote sensing: Grapevine phylloxera detection in vineyards. In *IEEE Aerospace Conference*, pages 1–9, 2018. doi: 10.1109/AERO.2018.8396450.

G. J. Q. Vasconcelos, T. V. Spina, and H. Pedrini. Low-cost domain adaptation for crop and weed segmentation. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 141–150. Springer International Publishing, 2021. ISBN 978-3-030-93420-0.

J. P. Wachs, H.I. Stern, T. Burks, and V. Alchanatis. Low and high-level visual feature-based apple detection from multi-modal images. *Agriculture*, 11(6): 717–735, 2010. doi: https://doi.org/10.1007/s11119-010-9198-x.

Z. Wang, K. Walsh, and A. Koirala. Mango fruit load estimation using a video based mangoyolo-kalman filter-hungarian algorithm method. *IEEE Sensors Journal*, 19, 2019. doi: 10.3390/s19122742.

W. Xie, J. A. Noble, and A. Zisserman. Microscopy cell counting with fully convolutional regression networks. *Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2016. doi: http://dx.doi.org/10.1080/21681163.2016.1149104.

S. Yadhav Yegneshwar, T. Senthilkumar, S. Jayanthy, and J. Kovilpillai Judeson Antony. Plant disease detection and classification using cnn model with optimized activation function. In *International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pages 564–569, 2020. doi: 10.1109/ICESC48915.2020.9155815.

S. Yang, L. Zheng, X. Chen, L. Zabawa, M. Zhang, and M. Wang. Transfer learning from synthetic in-vitro soybean pods dataset for in-situ segmentation of on-branch soybean pods. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1665–1674, 2022. doi: 10.1109/CVPRW56347.2022.00173.

W. Yin, H. Wen, Z. Ning, J. Ye, Z. Dong, and L. Luo. Fruit detection and pose estimation for grape cluster–harvesting robot using binocular imagery based on deep neural networks. *Frontiers in Robotics and AI*, 8, 2021. ISSN 2296-9144. doi: 10.3389/frobt.2021.626989. URL https://www.frontiersin.org/articles/10.3389/frobt.2021.626989.

L. Zabawa and A. Kicherer. Segmentation of wine berries, 2021. URL https://www.openagrar.de/receive/openagrar_mods_00067631.

L. Zabawa, A. Kicherer, L. Klingbeil, A. Milioto, R. Töpfer, and H. Kuhlmann. Detektion von Weintrauben in Bildern mit Hilfe von Fully Convolutional Neural Nets. In Leibniz Institut für Agrartechnik und Bioökonomie e. V. (ATB) Potsdam, Michael Pflanz, Michael Schirrmann, Marius Hobart, Lasse Klingbeil, and Jan Behmann, editors, *25. Workshop Computer und Bildanalyse in der Landwirtschaft*, 17. April 2019, Bonn, pages 15–20, 2019a.

L. Zabawa, A. Kicherer, L. Klingbeil, A. Milioto, R. Töpfer, H. Kuhlmann, and R. Roscher. Detection of single grapevine berries in images using fully convolutional neural networks. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2571–2579, 2019b. doi: 10.1109/CVPRW.2019.00313.

L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, H. Kuhlmann, and R. Roscher. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 164:73–83, 2020. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2020.04.002. URL https://www.sciencedirect.com/science/article/pii/S0924271620300939.

L. Zabawa, A. Kicherer, L. Klingbeil, R. Töpfer, R. Roscher, and H. Kuhlmann. Image-based analysis of yield parameters in viticulture. *Biosystems Engineering*, 218:94–109, 2022. ISSN 1537-5110. doi: https://doi.org/10.1016/j.biosystemseng.2022.04.009. URL https://www.sciencedirect.com/science/article/pii/S1537511022000861.