Improved optical phenotyping of the grape berry surface using light-separation and automated RGB image analysis

T. Haucke[1], K. Herzog[2], P. Barré[1], R. Höfle[2], R. Töpfer[2] and V. Steinhage[1]

[1] Department of Computer Science IV, University of Bonn, Bonn, Germany
[2] Institute for Grapevine Breeding Geilweilerhof, Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Siebeldingen, Germany

Summary

Grape resilience towards *Botrytis cinerea* (*B. cinerea*) infections (*Botrytis* bunch rot) is an important concern of breeders and growers. Beside grape bunch architecture, berry surface characteristics like berry bloom (epicuticular wax) as well as thickness and permeability of the berry cuticle represent further promising physical barriers to increase resilience towards *Botrytis* bunch rot. In previous studies, two efficient sensor-based phenotyping methods were developed to evaluate both berry surface traits fast and objectively: (1) light-separated RGB (red-green-blue) image analysis to determine the distribution of epicuticular wax on the berry surface; and (2) electrical impedance characteristics of the grape berry cuticle based on point measurements.

The present proof-of-concept study aiming at the evaluation of light-separated RGB images for both phenotyping applications, phenotyping wax distribution pattern and berry cuticle impedance values. Within the selected grapevine varieties like 'Riesling', 'Sauvignon Blanc' or 'Calardis Blanc' five contributions were achieved: (1) Both phenotyping approaches were fused into one prototypic unified phenotyping method achieving a wax detection accuracy of 98.6 % and a prediction of electrical impedance with an accuracy of 95 %. (2) Both traits are derived using only light-separated images of the grapevine berries. (3) The improved method allows the detection and quantification of additional surface traits of the grape berry surface such as lenticels (punctual lignification) and the berry stem that are also known as being able to affect the grape susceptibility towards *Botrytis*. (4) The improved image analysis tools are further integrated into a comprehensive workbench allowing end-users, like breeders to combine phenotyping experiments with transparent data management offering valuable services like visualizations, indexing, etc. (5) Annotation work is supported by a sophisticated annotation tool of the image analysis workbench. The usage of light-separated images enables fast and non-invasive phenotyping of different optical berry surface characteristics, which saves time-consuming labor and additionally allows the reuse of the berry samples for subsequent investigations, e.g. *Botrytis* infection studies.

K e y   w o r d s : *Vitis vinifera* ssp. *vinifera*; *Botrytis* susceptibility; Convolutional Neural Network (CNN); computer-based phenotyping; semantic segmentation.

Introduction

In viticulture, one of the most relevant grape diseases is *Botrytis* bunch rot, which is caused by the necrotrophic fungus *Botrytis cinerea* (*B. cinerea*). Within the scope of grapevine breeding, physical barriers like loose grape bunch architecture and intact berry surface characteristics (*i.e.* cuticle and its intra- and epicuticular waxes) are the most promising traits to increase resilience of new grapevine varieties towards *Botrytis* bunch rot (Commenil *et al.* 1997, Herzog *et al.* 2015, Tello and Ibáñez 2018). The cuticle hereby represents the primary barrier to the atmosphere protecting berries against biotic stresses such as pathogens and pests (Hen-Avivi *et al.* 2014). The cuticle thickness, permeability as well as the distribution of epicuticular waxes are described as one of the most important indicators for grape susceptibility to the appearance of micro cracks and further towards *Botrytis* bunch rot (Blaich *et al.* 1984, Commenil *et al.* 1997, Mundy 2008, Becker and Knoche 2012, Herzog *et al.* 2015, Barré *et al.* 2019). In addition, the region around the pedicel (stem) and lenticels are further berry surface features that are correlated to an increased susceptibility towards *Botrytis* bunch rot (Brown and Considine 1982, Becker and Knoche 2012). For an objective and fast assessment of grape berry surfaces, two recently developed sensor-based approaches were important for the present proof-of-concept study: grape cuticle impedance (Herzog *et al.* 2015, Molitor *et al.* 2018, Barré *et al.* 2019) and structured light and light-separated RGB images (Barré *et al.* 2019). Herzog *et al.* (2015) proposed a simple-to-handle I-sensor that enables the fast and reliable point measurement of electrical impedance of the grape berry cuticles and its epicuticular waxes (CW). It is described as an indicator for the thickness and permeability of the grape berry surface (sum of cuticle and epicuticular waxes) and statistical experiments revealed significant correlations between relative impedance of CW and the resilience of grapes towards *Botrytis* bunch rot infection (Herzog *et al.* 2015). The I-sensor is a minimally invasive sensor technique using a Platin-Iridium wire, which has to be prick into intact

berries. Due to that, only one representative measurement per berry can be determined and the reuse of the same berry sample in further studies like induced *Botrytis* bunch rot infection tests is not feasible. The impedance-based method measures the sum of thickness and permeability and cannot distinguish between the cuticle thickness and wax quality, yet. To overcome these limitations, Barré *et al.* (2019) proposed a non-invasive approach to phenotype the distribution and intactness of epicuticular waxes (berry bloom) that uses illumination-separated images of grapevine berries as input. For image analysis, a convolutional neural network approach is used to derive the uniformity and intactness of waxes on berries. Method validation over six grapevine cultivars shows accuracies up to 97.3 %. In addition, electrical impedance of CW was correlated to the detected proportion of waxes with r = 0.76.

However, other traits of the surface of berry samples are also known as being able to affect the grape susceptibility towards *Botrytis* infestation, such as lenticels (punctual lignification) and the berry stem (Brown and Considine 1982, Evans and Emmett 2013). Therefore, the present study contributes enhancements to the method proposed by Barré *et al.* (2019) by expanding the range of identifiable traits and improving trait detection accuracy. This is achieved by employing a modified U-Net (Ronneberger *et al.* 2015) deep learning model.

The present study faces four additional challenges. (1) Small datasets: sample collection, preparation and image capture using the mobile Light-Separation-Lab (Fig. 2) is time consuming and expensive, resulting in small training datasets that had to be used in the present work. This chal-lenge is met by using dataset augmentation and transfer learning techniques. Based on that, grapevine varieties with economic importance, significant different impedance values and new PiWi varieties (PiWi – Pioneer Wines, fungal resistant varieties) were pre-selected. (2) Explainability: As neural networks used in deep learning are black box models, their results are difficult to interpret. This explainability is however crucial in grapevine breeding, as the information on how relevant different traits are for the model's prediction could be experimentally validated and used for guiding breeding decisions. This challenge is met by employing an attribution method (Sundararajan *et al.* 2017) to quantify the relevance of the identified traits. (3) Sustainability: Phenotyping experts should be supported to explore and validate evaluation results by integrating the management of data material, annotations and the exploration of evaluation results. This challenge is met by integrating the proposed improved phenotyping method into a comprehensive image analysis workbench offering a graphical user interface (GUI) for visualization services and procedures for efficient data management. (4) Annotation: The annotation of images for the detection of visual traits is tedious and time consuming. This challenge is met by reusing detection results starting points for a customized annotation approach.

Material and Methods

The overall workflow of our approach is depicted in Fig. 1. The structure of this contribution is derived in a coherent way from this overall workflow as follows: The
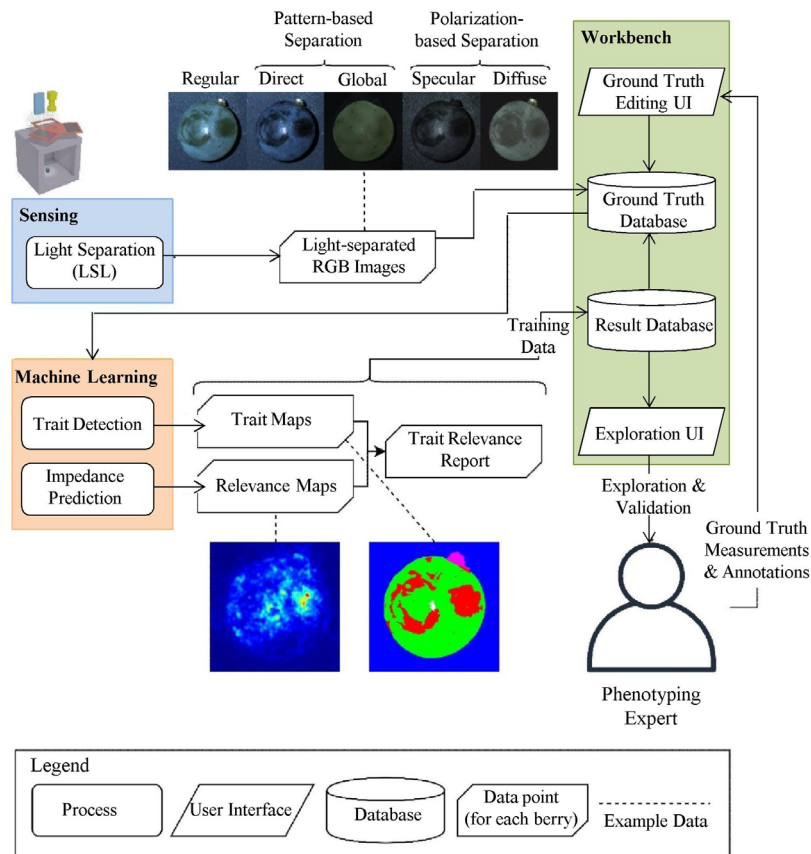

Fig. 1: Overall workflow of the phenotyping processes and the data management workbench (cf. text).

sensing step employs light separation, which is covered in the Methodology section. The results of the sensing step are then processed by two machine-learning steps: the detection step detects and localizes visual traits on berry surfaces like epicuticular waxes, lenticels, etc. and results in a map of the detected visual traits. This step is described in section Trait detection. The prediction step predicts the electrical impedance based on the extracted visual traits and results in a heat map that depicts the relevance of the surface parts with respect to the prediction result. This step is described in section Impedance prediction. The results of the two maps are combined in a ranked list of trait relevance, *i.e.* how much each class of traits contributes to the prediction result. The design and functionality of the Workbench and esp. the customized annotation tool are explained in section Workbench Implementation.

Plant material, image capture, ground truth data recording, and pre-processing steps: As plant material, grapevine varieties were selected showing differences in berry color, berry bloom and bunch density (Tab. 1). Representative bunches were sampled, and 45 visibly unharmed berries were cut off the bunches per variety and date.

Light-separation has been proven as a reliable tool to obtain detailed information about the berry surface and is conducted by separating different components of light reflected from a surface (Barré *et al.* 2019). We used a dataset comprised of light-separated grapevine berry images of different varieties, provided by the Julius Kühn-Institut (JKI), Institute for Grapevine Breeding Geilweilerhof, Siebeldingen in Rhineland-Palatinate, Germany. This dataset was acquired using the Light-Separation-Lab (LSL), developed by Barré *et al.* (2019) for this task (cf. Fig. 2). The LSL consists of an RGB camera for taking images of the specimen, a beamer acting as single source of illumination, a motorized, rotatable polarization filter wheel, a sealable enclosure in which the specimens are placed and a computer running the LSL's control software. In the capturing process,
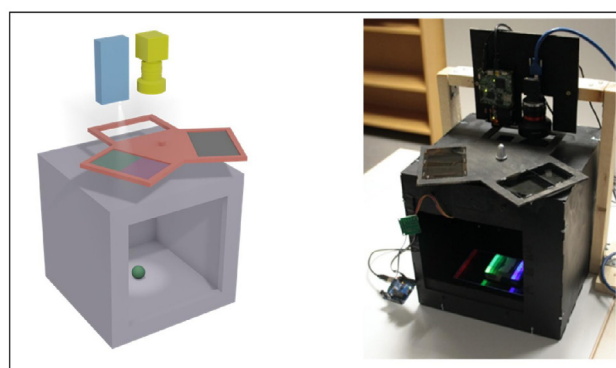

Fig. 2: Light separation lab (LSL).The schematic structure of the LSL can be seen in the left image. It consists of a camera (yellow), a beamer acting as light source (blue), a motorized, rotatable polarization filter wheel (red), and the specimen enclosure (grey) with a grapevine berry (green) inside. The camera, beamer and filter wheel are mounted on a structure which is not depicted for brevity. The right image shows the LSL in operation.

multiple light-separation methods are applied to the specimens. A regular image is captured using full illumination (without filters). Second, direct and global light components are separated using the pattern light separation method by Nayar *et al.* (2006). With this method, the beamer projects high frequency binary light patterns onto the specimens (again without filters). By capturing two images with alternating light patterns, each position on the specimen's surface is theoretically fully illuminated and shaded in one of the two images, respectively. This allows estimating the global and direct light components. In the last step, the specular and diffuse components are separated. The filter wheel first places two polarization filters, oriented in parallel (depicted as dark filters in Fig. 2) in front of both beamer and camera. An image using full illumination is then captured. This process is repeated using a set of two perpendicular polarization filters (depicted as red and green filters in Fig. 2), resulting in two images. These two images then allow estimating the specular and diffuse components. The output of a full LSL

Table 1

Grapevine varieties/accessions, assessed grape bunch/ berry characteristics, mean berry impedance, standard variation and coefficient of variation as well as maturity stage of sampled grapes (FTIR measurements). At both dates, 45 berries were sampled for each variety, resulting in a total of 495 samples. OIV – International Organization of Vine and Wine; FTIR – Fourier-transform infrared spectroscopy, B – Blanc = white varieties, N – Noir = red varieties. *V*IVC – *Vitis* International Variety Catalogue (www.vivc.de)

Date of sampling and data acquisition	Grapevine variety	*V*IVC no.	Berry color	Berry bloom (OIV 227)	Bunch density (OIV 204)	Mean berry impedance Z_{rel} CW	Standard variation of Z_{rel} CW	Coeff. of var.	FTIR sugar content (°Brix)	FTIR total acidity (g·L⁻¹)
11th September 2018	Cabernet Blanc	22355	B	7	1	820	94.2	0.11	22.4	8.1
	Riesling	10077	B	5	7	589	73.3	0.12	20.1	8.7
	Sauvignon Blanc	10790	B	5	7	567	99.4	0.18	22.8	7.6
18th September 2018	Cabernet Blanc	22355	B	7	3	862	83.9	0.10	25.6	7.4
	Cabernet Sauvignon	1929	N	5-7	3	757	88.2	0.12	22.4	9.6
	Calardis Blanc	22828	B	5-7	5	743	123.7	0.17	21.9	6.2
	Dakapo	14728	N	5-7	5	563	127.0	0.23	23.5	9.2
	Norton	3304	N	7	5	1300	143.3	0.11	23.5	25.8
	Riesling	10077	B	3	7	622	68.1	0.11	21.2	9.1
	Sauvignon Blanc	10790	B	3	7	628	73.5	0.12	24.2	6.4
	Seibel 7511*	11249	B	5-7	5	871	128.9	0.15	27.3	10.4
	* no variety protection, it is a grapevine accession									

Z_{rel} CW where the superscript notation °Brix is (°Brix) and total acidity (g·L^{-1}).

capturing process is then five RGB images, the regular, direct, global, specular and diffuse components, taking a total of 12.5 s to complete.

For ground truth data recording, the LSL capture of each berry was then followed by a measurement of the impedance of the berry cuticle at 2 KHz and 30 KHz, $Z_{rel}\,CW$ using the I-Sensor and calculation of relative berry impedance $Z_{rel}\,CW$ (Herzog *et al.* 2015). The mean $Z_{rel}\,CW$ per variety is given in Tab. 1. In addition to the ground-truth $Z_{rel}\,CW$ measurements, a subset of images (three samples of each of the eight varieties surveyed for a total of 24) was manually partitioned (annotated) into the classes 'wax', 'no-wax', 'stem', 'lenticel', 'background' and 'overexposure'.

M e t h o d o l o g y : According to the overall workflow depicted in Fig. 1, this section will present the algorithms that implement all steps of the workflow in detail.

T r a i t d e t e c t i o n : In this proof-of-concept study, we treat the task of detecting traits of the grapevine berry surface as an image segmentation task. Image segmentation within this study is applied by using a supervised deep learning approach and light-separated RGB images as input. Our method allows finer-grained trait detection results and stronger involvement of experts than the approach by Barré *et al.* (2019). In contrast to the approach of Barré *et al.* (2019), we use a single fully convolutional neural network (FCN) applied to pixel-wise annotated ground truth.

We introduce three new classes in addition to the 'wax', 'no-wax' and 'background' classes targeted by the original method (Fig. 3). The 'stem' and 'lenticel' classes represent traits of the berry surface, which are described as relevant towards berry susceptibility towards *Botrytis* bunch rot. The 'overexposure' class represents areas which were overexposed during the light-separation process. These regions contain no useful information about the berry surface and are therefore annotated as their own trait instead of guessing which real surface trait this area contains. The annotation process is described in section "Annotation process".

FCNs are a subset of convolutional neural networks (CNNs), only convolutional and pooling layers in addition to activation functions. The resulting networks can therefore take inputs of arbitrary spatial dimension (width and height) with the spatial dimensions of the output being a function of both input size and network architecture. This is potentially a
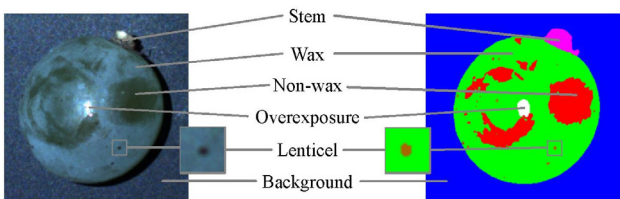

Fig. 3: Exemplary annotation of a 'Sauvignon Blanc' berry. The annotation shows the different surface traits our method detects, the stem (magenta), wax (green), non-wax (red) and lenticel (brown), in addition to two other image features: overexposure (white) and background (blue).

desirable property because it allows the model to be trained and evaluated using differently sized inputs. Due to the increased granularity of the required ground truth annotation and the increase from three to six target classes - compared to

the approach from Barré *et al.* (2019) - the training dataset has to be labeled from scratch. Therefore, the chosen FCN architecture should pose as few requirements on training dataset size as possible.

U-Net is an FCN architecture which was developed for biomedical image segmentation with the explicit goal of requiring only small training datasets, compared to other architectures like those proposed by Long *et al.* (2015). The original U-Net architecture (Ronneberger *et al.* 2015) outputs segmentation maps which are smaller than the input images. This is because the biomedical images that have to be segmented are potentially too large to fit into memory in a single forward pass and are therefore segmented in patches (*i.e.* one image region at a time). To achieve seamless segmentation in this scenario, the input has to contain additional padding proportional to the receptive field size. Otherwise, border regions would have less spatial context. In this work, the input images are of size 512 × 512 pixel, allowing segmentation without patches.

In this study, the U-Net architecture is modified to output segmentation maps of same size as the input (full-image U-Net). As the berries are located at the center of the images, the border regions contain only uniform background which needs less spatial context to be classified correctly. Therefore, the input feature maps are padded using the nearest neighbor strategy if this is required for preserving the size during convolutional operations. This simplifies the implementation of the architecture in two steps. (1) It allows to segment an entire berry image in a single forward propagation yielding the "full-image U-Net" as first efficient simplification). Additionally, the channel count is reduced by a factor of 2 in each layer to limit the potential of overfitting yielding the "narrow full-image U-Net" as second simplification. The architecture resulting from both simplifications is depicted in Fig. 4. The last convolutional layer is followed by a Softmax activation function, allowing the output segmentation map (or trait map) of size 512 × 512 × 6 to be interpreted as a probability distribution with respect to the six classes at each pixel.

The resulting architecture is implemented using the PyTorch framework (Paszke *et al.* 2017). The weighted Cross-Entropy loss function, defined as
with respect to the six classes C, the class weight w_c, model prediction y_c and true probability $\hat{y}_c$ for some pixel, is used for training. As the last layer of the described U-Net architecture consists of a Softmax activation function, the model's output at each pixel can be interpreted as a probability distribution with respect to the six classes. The weighting is

$$L(y) = -\sum_{c \in C} w_c \hat{y}_c \log y_c$$

necessary because of class imbalance. The background, wax and non-wax classes occur much more frequently than the other classes, leading the model being biased towards the frequent classes. The weight of each class is calculated by taking the inverse number of occurrences of each class in the ground truth set and normalizing the resulting weight vector to length one. The model is then trained by minimizing the loss function, averaged over each pixel, using the Adam
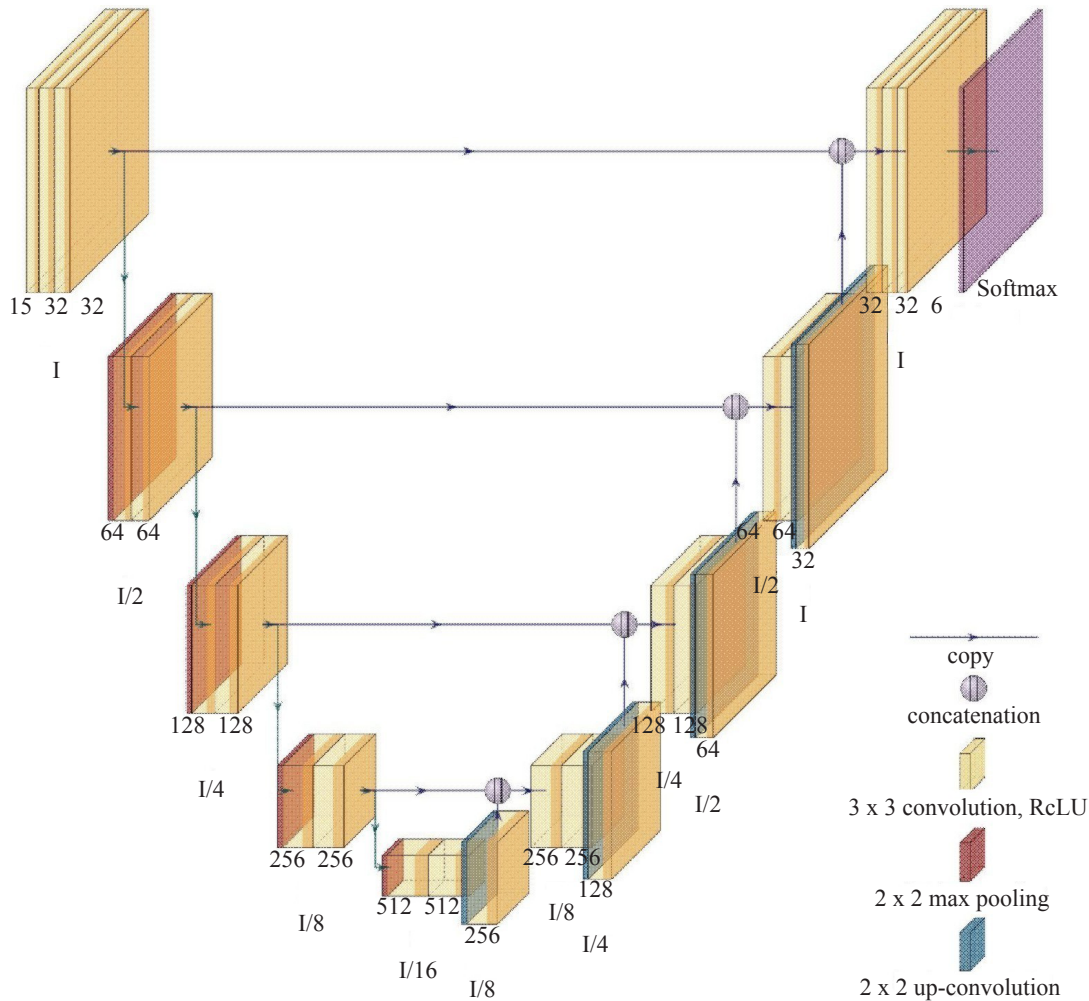
Fig. 4: Narrow full-image U-Net variant designed for this study.

optimization algorithm (KINGMA and BA 2014), which is shown to yield similar convergence to stochastic gradient descent while allowing faster training. The recommended parameters (KINGMA and BA 2014) of learning rate $\alpha = 10^{-3}$, exponential moment estimate decay rates $\beta = (0.9, 0.999)$ and numerical stability constant $\varepsilon = 10^{-8}$ are used. This model is trained for 500 epochs, as it shows overfitting (indicated by decreasing loss for the test set and strongly increasing loss for the validation set) in every configuration by epoch 450.

We employ training dataset augmentation on the fly during training. First, we apply color shifting in HSV (hue, saturation, value) space by a uniform amount from [-10 %, 10 %] to make the training less dependent on exact color values. Additionally, we apply Gaussian noise with mean $\mu = 0$ and standard deviation of $\sigma = \frac{1}{20}$ to the input image. The last augmentation method involves randomly rotating both the input and target trait map by some angle $\alpha \in \{0°, 90°, 180°, 270°\}$ and mirroring around the horizontal and/or vertical axis. As the berry orientation is arbitrary, this was done to improve generalization by helping the network to learn all traits in different orientations. For evaluation we use the Jaccard Index, or Intersection over Union (IoU), defined as: Let Ac be the set of points for which the model's prediction for class is maximal and the set of points which true class is c. $J(A_c, B_c)$ is then a measure of the overlap of prediction and ground truth for some class. Each class can therefore

be weighed independent of its spatial frequency. Integrated gradients (SUNDARARAJAN *et al.* 2017) are employed to determine the attribution of the output trait map with respect to the

$$J(A_c, B_c) = \frac{|A_c \cap B_c|}{|A_c \cup B_c|}$$

input (*relevance*). As the output at each pixel depends only on the local neighborhood in the input, the spatial relevance is highly uniform. Much more interesting is the contribution of the different color channels and therefore light-separation methods. The relevance with respect to the entire input is hence averaged over the spatial image dimensions to yield the attribution per channel. An image with the original input's average color at each channel is used as baseline. As the different light-separated images have varying levels of average intensity, this prevents high-intensity light-separated images to have a disproportionate relevance.

Impedance prediction: In this proof-of-concept study, the prediction of relative impedance of the berry cuticle ($Z_{rel}CW$, Tab. 1) from light-separated images of berry samples is implemented as a deep learning-based approach to regression. Commonly, in deep learning, regression models (as well as classification models) are implemented in two stages. First is the feature extraction, which is responsible for transforming the input into a form which contains only the relevant features. The second part is the regression itself,

which takes these extracted features as input and produces the final regression result as output. The VGG16 CNN architecture (Simonyan and Zisserman 2014) is used for feature extraction due to its simplicity in architecture and implementation. As described, the VGG16 architecture is split into feature extraction and classification stages. Analogous to the segmentation model, the input to the feature extraction stage is a single berry capture, consisting of five RGB images resulting from the regular image and the two light-separation methods. The input is therefore of shape *width × height ×* 15. However, as the VGG16 architecture is designed to classify regular RGB images, it must be modified to take input images with 12 additional channels as input. This can be done exclusively in the first layer, which comprises 64 convolutional kernels of size 3 × 3 per input channel, resulting in weights of shape $3 × 64 × 3^2$. To avoid losing the weights of the pretrained model in the first layer, the original weights are duplicated along the first axis five times, preserving the meaning of the original weights that correspond to each of the RGB channels. The resulting weights are then divided by five to achieve the same average activation compared to the original architecture when given a regular RGB image. This allows to use weights of a model pretrained on the ImageNet dataset as initialization.

We then approach the regression task by repurposing the classification stage of the VGG architecture. The original classification stage of the VGG architecture consists of two fully connected layers of size 4096 (first layer) and 1000 (second layer). The first layer takes as input the output of the last feature extraction layer. The second layer then has a Softmax instead of a ReLU activation function, with each of the 1000 items in the output corresponding to a class in the ImageNet dataset. However, as the regression only needs a scalar output, the last layer is replaced with a layer of size 1 and a logistic activation function

$$f(x) = \frac{y_{max}}{1 + e^{-x}}$$

where is the maximum target value of the given regression problem. This *VGG regression* is then combined with the *VGG feature extractor* into a single model and fine-tuned to one of the described datasets (*fine-tuned VGG regression*). This is achieved by minimizing the mean squared error using stochastic gradient descent with a learning rate of 0.005, momentum of 0.9 and mini-batches of size 32 *via* the PyTorch framework (Paszke *et al.* 2017). The small learning rate is due to fine-tuning, which benefits from smaller weight updates. The model is fine-tuned for 100 epochs.

The augmentation methods used for the segmentation model were adopted for the regression. Additionally, a continuous rotation augmentation, instead of only multiples of 90°, is implemented to further help improving generalization. As the VGG architecture expects an input image of *width = height =* 224, the original input images of size 512 have to be down sampled. To reduce loss in quality due to multiple resamples required by continuous rotations and resizing, these operations are combined into a single affine transformation of the input. During this affine transformation, the berry is also centered in the image, using the geometric center of the inverse of the background trait obtained

from the trait map. During the LSL capture process, some regions of the berry surface may get overexposed by the camera depending on berry position and surface orientation. These regions contain no information about the underlying surface morphology but tend to benefit overfitting. They are therefore replaced by their neighborhood, which is assumed to be the most accurate approximation, using the inpainting method by (Bertalmio *et al.* 2001) implemented via the OpenCV library (Bradski 2000). This is done for all light-separated images, except for the global image captured using the pattern light-separation method because it consists of only the minimum intensity obtained during all pattern steps and is therefore not affected by overexposure. Each resulting input is also normalized by subtracting the dataset mean and dividing by the average dataset standard deviation for each channel. The dataset is split into training and validation sets using a ratio of $\frac{11}{15}$ and $\frac{4}{15}$, respectively.

The integrated gradients attribution method is applied to the regression model and the resulting attribution is averaged over the 15 input channels to create a *relevance map*. The manual interpretation of a large number of relevance maps is infeasible.

The relevance map is therefore multiplied by the corresponding trait map and then averaged over the spatial image dimensions. The result is the *absolute surface trait relevance*, which describes how much each surface trait contributed to the model's prediction. Each trait's absolute relevance is then divided by the relative frequency of the trait in the corresponding trait map, yielding the trait relevance per surface area (*relative surface trait relevance*). As the resulting absolute and relative surface trait relevance values are scalar values, their interpretation is less complex, and they can be aggregated by categories such as cultivar.

Apart from the already described traits, the border regions between some traits could be even more relevant than the traits themselves, e.g. the region around lenticels and the transition between the berry and stem where *Botrytis* spores can more easily clip to the berry surface (Evans and Emmett 2013). Both the region around lenticels and the berry to stem transition do not need to be annotated manually. These border traits are instead inferred from the existing trait map using dilation and erosion operations on binary masks. The predictive accuracy of the regression models is measured by the maximum-normalized mean absolute error with prediction y_i, true value $\hat{y}_i$:

$$MAE = \frac{1}{\max(\hat{y}) \cdot n} \cdot \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

The prediction therefore improves as the mean absolute error goes against 0.

Workbench implementation: The purpose of the Workbench is to make the proposed methods accessible to end-users, e.g. breeders. Three main goals were set to fulfill this purpose. First, it should act as an expert-editable database for the described dataset consisting of light-separated grapevine berry images and corresponding impedance measurements. Second, the Workbench should allow experts to start training processes which use the current state of the datasets and set custom hyperparameters. Third, it should allow to visually explore and aggregate the training results.

Fourth, the trait annotation process should be simplified and accelerated to allow experts to focus more on results than on tedious work. We propose a graphical user interface to provide the required functionality without the need of extensive training or programming knowledge.

The Workbench is implemented using a client-server-architecture. This allows the state of the database to be centralized on the server, while multiple clients can access the data simultaneously. The client is implemented as a JavaScript web application compatible with modern browsers. This reduces initial hurdles such as software distribution and supporting multiple operating systems.

B o o t s t r a p p i n g a n n o t a t i o n : To simplify and accelerate the annotation process, we propose a bootstrapping method which uses existing trait maps to initialize a custom annotation. The currently edited annotation is overlaid on the regular berry image and a continuous scale allows to arbitrarily blend the two images to be able to assess the berry morphology while annotating the corresponding traits.

A n n o t a t i o n p r o c e s s : The goal of the manual annotation process is to label different regions of light-separated grapevine berry images as one of the six classes: wax, non-wax, background, overexposure, lenticel and stem. Overexposures are labeled due to the underlying traits not being visible. The differentiation between the different traits is however not unambiguous and subjective, especially for smooth transitions between wax and non-wax areas. The resulting annotation is an image of the same resolution as, and congruent to the source image. Six distinct colors with 8 bits per RGB channel are defined to represent the different classes. The wax class is represented by rgb(255, 0, 0), non-wax by rgb(0, 255, 0), background by rgb(0, 0, 255), overexposure by rgb(255, 255, 255), lenticel by rgb(171, 102, 0) and stem by rgb(255, 0, 255). PNG is used as file format, as it is lossless and ubiquitous. The annotations for this work were created by enclosing the different areas using the pencil tool and then filling these areas by their respective color.

H a r d w a r e : As deep learning is infeasible on low-powered hardware, the training is designed to be executed by a remote host with graphics processing units (GPUs). In this work we run training tasks on a Google Cloud n1-highmem-2 instance (2 cores of an Intel® Xeon CPU clocked at 2.2 GHz, 13GB RAM) with a single GK120 GPU of a NVIDIA® Tesla® K80.

Results and Discussion

Unless otherwise noted, all results are obtained by averaging over three independent training processes, running on the environment described in section above using three-fold cross-validation.

T r a i t d e t e c t i o n : Both implemented variations of the U-Net architecture were evaluated by combining all proposed augmentation methods. While the full-image U-Net achieves a validation IoU of 89 %, the narrow full-image U-Net achieves a validation IoU of 92 %. The training of the full-image U-Net on the described dataset takes 43 s per Epoch, while the narrow full-image U-Net takes 23 s per

Epoch, resulting in training times of 6 h and 3.25 h respectively. Overfitting is less present using the narrow full-image U-Net architecture, as can be seen in Fig. 5.
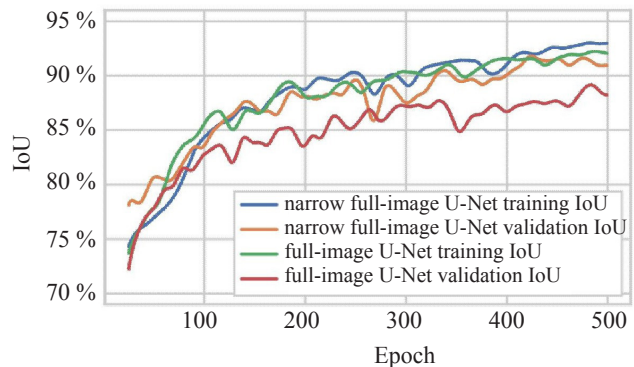

Fig. 5: Segmentation IoU during training.

As described in section *Annotation Process*, the discrimination of different classes during the manual annotation is not always strictly defined and subjective. To be able to exemplarily quantify the consistency in annotation, four different berries were annotated twice within an interval of seven days, achieving an IoU of 96 %. This suggests that the segmentation model's IoU of 92 % is approaching the practical optimum given the inconsistent annotation.

Compared to the results of the approach of BARRÉ *et al.* (2019) that focuses only on the detection of epicuticular waxes, our narrow full-image U-Net approach with full augmentation achieves a 99.4 % background detection accuracy, compared to 97.9 % and a wax detection accuracy that is improved by 1.3 percentage points from 97.3 % to 98.6 %.

We infer the relevance of the different light-separation results and their respective color channels using the integrated gradients method (SUNDARARAJAN *et al.* 2017). As seen in Fig. 6, the pattern light-separation method dominates the model's relevance by 49 %. Second is the regular method, which accounts for a relevance of 38 %, even though it consists of only a single capture and therefore provides only a fifth of the total network input. The polarization method accounts only for 13 % of the total relevance. Generally, the red (38 %) and blue (43 %) channels contribute more to the total relevance than the green (19 %) channel. An exception to this is the global light-separation result produced by the pattern method, where the green channel contributes the most. This is expected, as the global capture primarily images subsurface structures, in this case the green berry flesh. The method relevance generally confirms the experimental results from BARRÉ *et al.* (2019), which show that the pattern-based separation alone yields better wax detection accuracy than the regular capture or polarization-based separation in isolation.

I m p e d a n c e p r e d i c t i o n : The fine-tuning of the VGG regression model requires 7.15 h for training until convergence. It achieves a mean absolute error of 7.74 % for the variety 'Dakapo', 4.19 % for 'Calardis Blanc' and 4.23 % for 'Sauvignon Blanc', as can be seen in Fig. 7.

Tab. 2 shows the absolute and relative trait relevance values obtained by the combination of trait and relevance maps. Combined, the wax and non-wax distribution contrib-
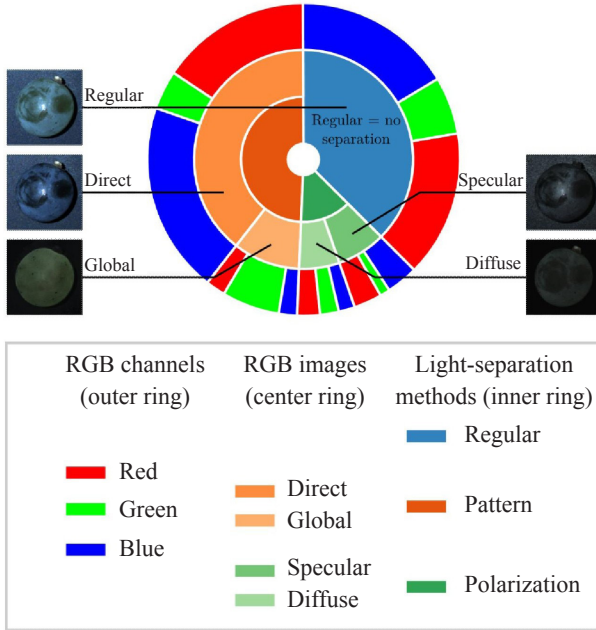
Trait relevance, computed using the integrated gradients method. Absolute relevance is a measure of how much the regression model is influenced by a specific trait. Relative relevance is the absolute relevance divided by the spatial frequency of the respective trait. Both measures are therefore not necessarily representative of true relevance, but only of how relevant they are to the regression model. The two border traits are automatically inferred from the segmentation maps (cf. Section Impedance Prediction)

Trait	Trait Relevance	
	Absolute	Relative
Wax	71.67%	17.68%
Non-wax	19.27%	20.97%
Lenticels	0.22%	11.39%
Stem	3.54%	18.81%
Stem-berry-border	2.05%	16.83%
Lenticel-berry-border	3.24%	14.32%

Fig. 6: Relevance of different light-separation results by the segmentation model. We infer the relevance of the different light-separation results (cf. Section Plant Material, Image capture, ground truth data recording, and pre-processing steps) for the segmentation of the four surface traits (stem, wax, non-wax, lenticel [cf. Section Trait detection]) provided by the LSL, computed using integrated gradients. The "outermost ring" shows the relevance of the red, green and blue color channels of the respective light components, which are shown in the "center ring". The "innermost ring" shows the relevance of the different light-separation methods by which the light components are obtained.

utes 91 % to the total relevance, which reflects the factors relevant to $Z_{rel}CW$.

The image analysis workbench: The graphical user interface (GUI) is implemented using five main screens. First is the ground truth overview screen, which is a table containing the ground truth data of each berry in the dataset. Each row links to its corresponding berry detail screen which allows to add or edit both impedance measure-

ments using a graphical editor. The berry detail screen also allows to view and manually update light-separated images, as well as view the trait maps and regression relevance maps.

Bootstrapping-based annotation: The annotation screen is invoked via the berry detail screen and provides an annotation tool tailored to light-separated berry images (Fig. 8). The reduction in annotation effort facilitated by the annotation bootstrapping was exemplarily examined by comparing the required mouse click amount, physical mouse traveling distance and time required with and without bootstrapping for annotations of four different berries, as seen in Tab. 3. The IoU of the two resulting annotations per berry is 95 % averaged. The bootstrapping especially reduced the annotation effort for larger areas, such as the background and large wax areas, which results in the large mouse traveling distance reduction. The aggregation tool provides pivot table functionality and allows experts to aggregate and examine the data from light-separation, segmentation and regression. Dimensions can be arranged along two axes using a drag-and-drop interface to create different high-quality visualizations (Fig. 9).
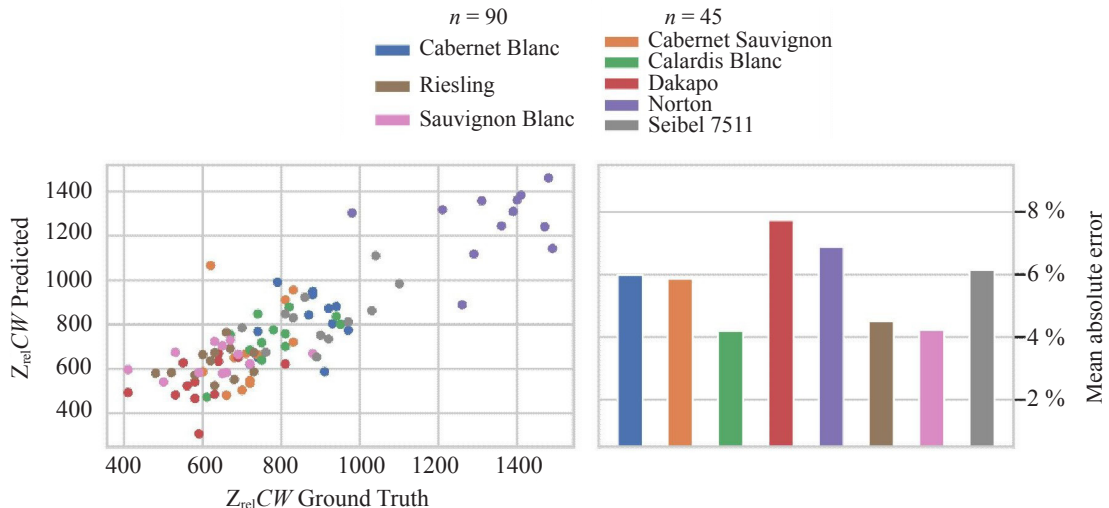

Fig. 7: Predicted impedance against ground truth per sample (left) and the mean absolute error of the impedance regression averaged by grapevine variety (right, lower is better), both on the validation set.
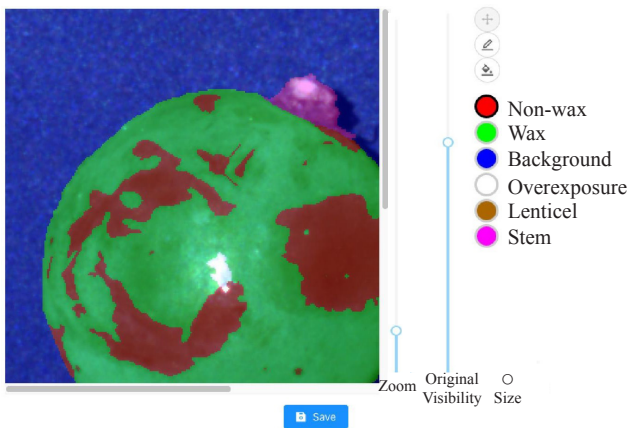
Fig. 8: Screenshot of the berry annotation tool GUI (cf. text). Depicted here is a 'Sauvignon Blanc' berry, the same specimen as in Fig. 3.

T a b l e   3

Annotation effort with and without bootstrapping

	Computer mouse clicks	Total distance of computer mouse movement	Time
w/o Bootstrapping	317	8.2 m	33 min
w/ Bootstrapping	102	1.6 m	10 min
Speedup	3.1	5.1	3.3

Conclusions

C o n t r i b u t i o n s   a n d   b e n e f i t s : In the present proof-of-concept study, five contributions could be delivered: (1) A unified method is introduced for the detection of grapevine berry surface traits and the prediction of grape berry cuticle impedance. This method achieves a wax detection accuracy of 98.6 %, which significantly exceeds existing methods (BARRÉ *et al*. 2019). Within the investigated data set, the impedance prediction based on image analysis has a mean absolute error of 5 %. This is an important improvement for the light-separation method, which should be validated as a suitable and non-invasive alternative to the physical impedance measurement for future applications. (2) Therefore, all traits are derived using only light-separated images of the grapevine berries. Light-separation provides valuable information for trait detection and impedance prediction. (3) While the approach of BARRÉ *et al.* (2019) allows only to estimate the distribution of epicuticular wax on berry surfaces, our improved method allows the detection and quantification of additional surface traits of the berry surface such as lenticels (punctual lignification) and the berry stem that are also known as being able to affect the grape susceptibility towards *Botrytis*. (4) The workbench unifies the proposed methods and provides a cohesive interface for breeding experts. This interface facilitates the full lifecycle from data acquisition and management, over the conduction of experiments with custom parameters, to the visualization and evaluation of experiment results. To foster explainability, the integrated gradients method is applied to both trait detection and impedance prediction, which allows breeding experts to gain insight into how model decisions and parts of the input are related instead of treating the models as black-boxes. Combining the relevance maps of the impedance prediction with the trait maps provided by the trait detection allows quantification of the relevance of different surface traits. (5) Additionally, the manual surface trait annotation process is supported by a sophisticated annotation tool which employs bootstrapping to reduce the annotation effort, leading to a time reduction of 3.3x in our exemplary study.

F u t u r e   w o r k : Sensor-based recognition of important fleshy fruit surface traits including impedance of the berry cuticle provides certain benefits with regard to objective screenings, comparable investigation studies and the development of molecular markers with application in marker-assisted-selection (MAS). For applications in grapevine breeding and research, the image-based prediction of berry impedance must be validated in the near future with regard to data replicability, result reliability and accuracy. Therefore, screening of highly variable breeding material or grapevine varieties within genetic repositories should be conducted over several years. However, from the breeder point of view, one further important goal is to employ the same machine-learning methods for end-to-end prediction
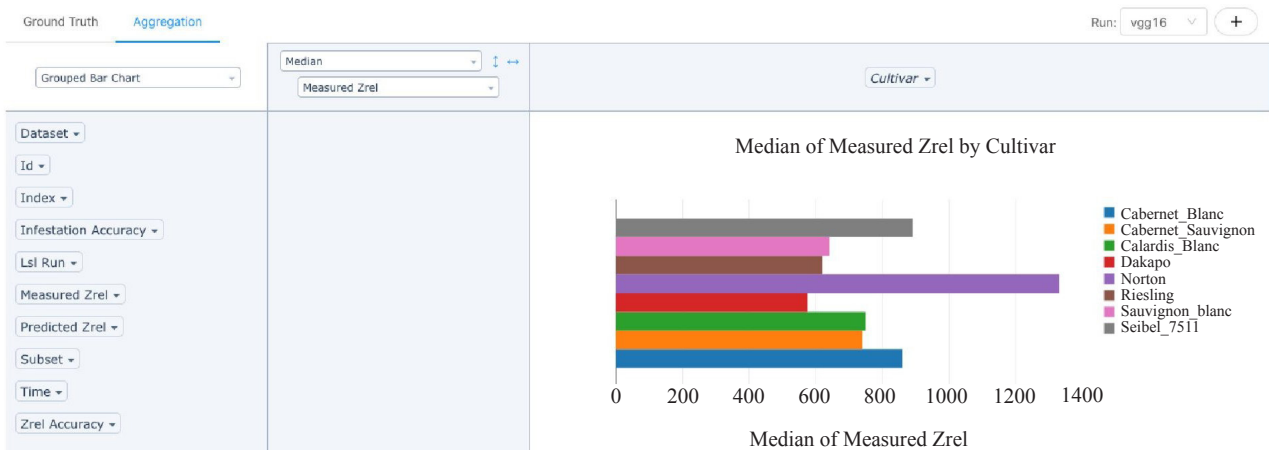

Fig. 9: Screenshot of the Aggregation tool GUI. The Aggregation tool provides a pivot table functionality for aggregating over various dimensions and plotting the results.

of the susceptibility towards *Botrytis* directly, e.g. without predicting intermediate traits like impedance or wax distribution. In a preliminary test, we implemented this approach using the same model architecture as in the impedance prediction. It consists of pairs of light-separated berry images acquired using the LSL and corresponding *Botrytis* infestation assessments. Therefore, intact berries were inoculated with *Botrytis* spores followed by an infestation assessment seven days post inoculation. We achieve promising accuracy for some cultivars such as Norton. But, traits of the berry surface are highly variable between varieties, location, and years and thus, *Botrytis* bunch rot infestation rate differs between years. This means, that this kind of investigations have to be done over at least three years and within highly variable breeding material like genetic repositories or breeding populations. Therefore, future work should aim at collecting an extensive dataset of light-separated images and *Botrytis* infestation measurements. Such an extensive dataset can be used to train a reliable model to predict the susceptibility to *Botrytis* directly, instead of key indicators such as specific surface features or the impedance of the berry cuticle. The relevance maps could then be used to identify regions which the model uses for its prediction, therefore highlighting features which are potentially worth evaluating for breeding decisions.

References

Barré, P.; Herzog, K.; Höfle, R.; Hullin, M. B.; Töpfer, R.; Steinhage, V.; 2019: Automated phenotyping of epicuticular waxes of grapevine berries using light separation and convolutional neural networks. Comput. Electron. Aric. **156**, 263-274. DOI: https://doi.org/10.1016/j.compag.2018.11.012

Becker, T.; Knoche, M.; 2012: Deposition, strain, and microcracking of the cuticle in developing "Riesling" grape berries. Vitis **51**, 1-6. DOI: https://doi.org/10.5073/vitis.2012.51.1-6

Bertalmio, M.; Bertozzi, A. L.; Sapiro, G.; 2001: Navier-stokes, fluid dynamics, and image and video inpainting. In: Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. IEEE. DOI: https://doi.org/10.1109/CVPR.2001.990497

Blaich, R.; Stein, U.; Wind, R.; 1984: Perforationen in der cuticula von Weinbeeren als morphologischer Faktor der Botritisresistenz. Vitis **23**, 242-256. DOI: https://doi.org/10.5073/vitis.1984.23.242-256

Bradski, G.; 2000: The OpenCV Library. Dr. Dobb's Journal of Software Tools (https://www.drdobbs.com/open-source/the-opencv-library/184404319).

Brown, K.; Considine, J.; 1982: Physical aspects of fruit growth: Stress distribution around lenticels. Plant Physiol. **69**, 585-590. DOI: https://doi.org/10.1104/pp.69.3.585

Commenil, P.; Brunet, L.; Audran, J. C.; 1997: The development of the grape berry cuticle in relation to susceptibility to bunch rot disease. J. Exp. Bot. **48**, 1599-1607. DOI: https://doi.org/10.1093/jxb/48.8.1599

Evans, K.; Emmett, R.; 2013: Botrytis: Questions & Answers. Wine Australia Factsheet (www.wineaustralia.com).

Hen-Avivi, S.; Lashbrooke, J.; Costa, F.; Aharoni, A.; 2014: Scratching the surface: genetic regulation of cuticle assembly in fleshy fruit. J. Exp. Bot. **65**, 4653-4664. DOI: https://doi.org/10.1093/jxb/eru225

Herzog, K.; Wind, R.; Töpfer, R.; 2015: Impedance of the grape berry cuticle as a novel phenotypic trait to estimate resistance to *Botrytis cinerea*. Sensors **15**, 12498-12512. DOI: https://doi.org/10.3390/s150612498

Kingma, D. P.; Ba, J.; 2014: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (https://arxiv.org/abs/1412.6980).

Molitor, D.; Biewers, B.; Junglen, M.; Schultz, M.; Clementi, P.; Permesang, G.; Regnery, D.; Porten, M.; Herzog, K.; Hoffmann, L.; Beyer, M.; Berkelmann-Löhnertz, B.; 2018: Multi-annual comparisons demonstrate differences in the bunch rot susceptibility of nine *Vitis vinifera* L. "Riesling" clones. Vitis **57**, 17-25. DOI: https://doi.org/10.5073/vitis.2018.57.17-25

Mundy, D. C.; 2008: A review of the direct and indirect effects of nitrogen on botrytis bunch rot in wine grapes. New Zealand Plant Protect. **61**, 306-310. DOI: https://doi.org/10.30843/nzpp.2008.61.6841

Nayar, S. K.; Krishnan, G.; Grossberg, M. D.; Raskar, R.; 2006: Fast separation of direct and global components of a scene using high frequency illumination. In: ACM Transactions on Graphics (TOG). ACM 935-944 (https://www.cs.columbia.edu/CAVE/projects/separation/).

Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A.; 2017: Automatic differentiation in PyTorch. 31st Conf. Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA (https://openreview.net/forum?id=BJJsrmfCZ).

Ronneberger, O.; Fischer, P.; Brox, T.; 2015: U-net: Convolutional networks for biomedical image segmentation, 234-241. In: Int. Conf. Medical Image Computing and Computer-Assisted Intervention. Springer (https://arxiv.org/abs/1505.04597).

Simonyan, K.; Zisserman, A.; 2014: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556[cs.CV] (https://arxiv.org/abs/1409.1556).

Sundararajan, M.; Taly, A.; Yan, Q.; 2017: Axiomatic Attribution for Deep Networks. arXiv:1703.1365[cs.LG] (https://arxiv.org/abs/1703.01365).

Tello, J.; Ibáñez, J.; 2018: What do we know about grapevine bunch compactness? A state-of-the-art review. Aust. J. Grape Wine Res. **24**, 6-23. DOI: https://doi.org/10.1111/ajgw.12310