Identifying veraison process of colored wine grapes in field conditions combining deep learning and image analysis

Lei Shen[a, b, c], Shan Chen[a, b, c], Zhiwen Mi[a, b, c], Jinya Su[d], Rong Huang[e],
Yuyang Song[e], Yulin Fang[e], Baofeng Su[a, b, c*]

(a. *College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi, 712100, China*
b. *Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi, 712100, China*
c. *Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Services, Yangling, Shaanxi, 712100, China*
d. *Department of Computing Science, University of Aberdeen, Aberdeen AB24 3UE, U.K*
e. *College of Enology, Northwest A&F University, Yangling, Shaanxi, 712100, China*)

Abstract

Accurate identification of the veraison process is essential for improving wine quality, which is challenging due to the variability of veraison among berries of the same cluster in algorihtm design, and also the subjective and labor-intensive issues in mannual identification. Therefore, this study proposed a method combining deep learning and image analysis to identify veraison in colored wine grapes under natural field growing conditions. The removal of irrelevant background was first achieved by semantic segmentation model, and then Mask R-CNN instance segmentation pipeline was constructed with anchor parameters optimization. In particular, three kinds of backbone networks were analyzed and compared in Mask R-CNN, and the overall performance of ResNet50-FPN was the best, with the testset Average Precision reaching 81.53% and the inference time being only 45.70ms/frame. Then, a method for characterizing berry veraison by H component of HSV color space was proposed and the invariance of the H component of three colored wine grape berries under different light conditions was verified and discussed. An algorithm was developed to identify veraison progress by calculating the percentage of the number of berries of different grades in the total number of berries of the whole grape bunches. The test accuracy reached 92.50%, 91.25% and 91.88% for three wine grapes including Cabernet Sauvignon, Matheran and Syrah respectively. The proposed method is able to provide vital reference for automated monitoring and intelligent management decisions of grape growth.

Keywords: Grape veraison, Mask R-CNN, Segmentation, H component

1. Introduction

In the cultivation of wine grapes, veraison is the most critical period in the formation of wine grape, and the changes in the veraison stage play a crucial role in the quality of the grapes. Accurate identification of veraison process can provide intelligent decisions for vineyard cultivation management, which is important to improve the quality of veraison grapes and ensure the quality of wine (Costa et al., 2019; Santesteban, 2019).

Traditionally, the veraison of a single berry is judged by skilled experts through empirical methods such as color, gloss and taste, but this method is subjective and labor-intensive. Due to the asynchronous nature of the veraison, even the veraison of individual berries in the same cluster varies greatly, which makes it inaccurate and inefficient for viticulturists to identify the veraison process of the entire cluster (Parker et al., 2011). With the expansion of vineyard acreage, automated technologies can effectively reduce labor, save time expenses, and enable high-throughput analysis. Therefore, an automated analysis of veraison processes is necessary and valuable for viticulturists.

Kalt et al. (1995) investigated the relationship between surface color and other ripeness indicators (size, sugar, acid and anthocyanin content) in 72 blueberry samples. The results showed that sugar content was highly correlated with surface color, indicating that surface color can represent berry ripeness. Sadres and Petrie (2012) predicted the different maturity levels of grapes by measuring the soluble solids content within three wine grapes including Chardonnay, Shiraz and Cabernet Sauvignon. Grape veraison and ripening stage were found to be directly related. Extensive studies have shown that the process of veraison is accompanied by the accumulation of substances such as soluble solids and anthocyanins (Parker et al., 2013; Rienth et al., 2021). However, the accumulation of these compounds directly controls the degree of coloration of the grape berry skin, which in turn produces disturbances that affect the dynamics of color change in grapes (Llerena et al., 2019; Martins et al., 2012; Meng et al., 2015). This provides a theoretical basis for the identification of veraison processes by means of a pictorial approach.

Various traditional methods have been used for fruit identification. Early research on image segmentation mainly includes color thresholding, region growing, and edge detection. Wang and Zhang (2014) used the angle model thresholds established by the a and b components of the Lab color space and the segmentation thresholds established by the H and S components of the HSV color space to achieve melon fruit image segmentation in complex backgrounds. Region growth-based segmentation algorithm has been widely used to segment red tomato and apple images (Ji et al., 2012; Khoshroo et al., 2014). Rahman and Hellicar (2014) achieved the identification of white grape berries in field conditions based on the Hough transform method by setting an edge threshold of 0.9, an edge sensitivity of 0.05, a maximum radius of 35 and a minimum radius of 10, but there were a large number of false identifications. Although these methods can achieve high operating speeds, they suffer from crop variations, ambient light variations, shading and other problems which limit their practical applications (Xu et al., 2013).

In recent years, the development of inexpensive sensors and electronic systems has driven the acquisition of field phenotypes, and emerging image analysis techniques have provided the necessary conditions for efficient and automated analysis and extraction of the necessary agronomic traits and phenotypic characteristics. With the advancement of deep learning methods, especially convolutional neural network (CNNs), the adaptability and robustness of image recognition methods have improved tremendously and many successes have been achieved (Wang and He, 2022). Earlier studies also show that CNNs show great promise for image classification, object detection and segmentation. Lin et al. (2019) used fully convolutional network (FCN), a fully convolutional segmentation network, to segment pomegranate images in natural environments, and the results showed that the algorithm achieved an accuracy of 0.893 and an IoU of 0.806 for pomegranate segmentation. Liang et al. (2020) used YOLOv3 to detect litchi fruits in natural environments at night, and then determined the region of interest (RoI) of fruit stems based on the bounding boxes of litchi fruits. Finally, the fruit stems were segmented one by one based on U-Net to achieve the detection of litchi fruits and fruit stems at night. Kang and Chen (2019) used a deep convolutional neural network for real-time detection and semantic segmentation of apples in an apple orchard, and finally obtained a segmentation accuracy of 86.5%. Kestur et al. (2019) proposed a new MangoNet semantic segmentation network with better robustness in terms of scale, illumination, contrast and occlusion to accurately segment mangoes in an orchard environment. Despite some success and progress in the application of CNN and artificial vision systems in agriculture, a comprehensive analysis of the usability of these methods in real field conditions is still lacking. It can be seen that there are still much room to explore in using different CNN architectures for different agricultural application scenarios. Especially for the more complex tasks in agriculture, combining multiple CNN architectures helps to fully utilize the advantages of each.

Some semantic segmentation models have also been used in cluster fruit recognition. Santos et al. (2020) evaluated the performance of three models, Mask R-CNN, YOLOv2 and YOLOv3, in order to detect and segment grape clusters in the field, and achieved the detection and counting of clusters. Similarly, Marani et al. (2021) used consumer-grade RGB-D cameras for automatic segmentation of grape bunches in color images. However, the segmentation of individual berries was not achieved. To identify individual berries, Grimm et al. (2019) proposed a deep semantic segmentation method by using VGG16 as an encoder to identify grape berries, and although the recognition accuracy was high, the method labeled berries with constant radius circles, which made it difficult to segment complete berry individuals. Zabawa et al. (2020) used DeepLabV3+ to segment grape berries in the field by adding "edge" labels and achieved surprisingly good segmentation results. Several studies have also been conducted by using deep learning for the detection and segmentation of individual berries (Buayai et al., 2020;Ni et al., 2020), but the method for identifying the veraison of clusters by the veraison of the berries is not yet clear.

This study proposes a method that combines deep learning and image analysis to identify colored wine grape veraison in field environments. The method can be used as a reference for automated monitoring and intelligent management decisions of wine grapes during their growth. The main contributions are summarized as below:

（1） A pipeline for extracting individual berries in field conditions was developed by combining semantic segmentation and instance segmentation.

（2） A method for characterizing berry veraison by H component of HSV color space was proposed.

（3） The invariance of the H component of three colored wine grape berries under different light conditions was verified and discussed.

（4） An algorithm was developed to identify the veraison process of grapes, and the accuracy of the test on three varieties was able to reach more than 91.25%.

2. Materials and methods

2.1 Image preparation

2.1.1 Image acquisition

The experiment was conducted in a wine grape cultivation site in Yangling, Shaanxi Province (34°18′7″N, 108°05′10″E) with a continental monsoon climate. The wine grapes were cultivated in single hedge frame with north-south rows, with rain shelters and a spacing of about 3m between rows and 1.5m between vines.

The wine grape image data collection took place in July-August 2021 and covered all stages of the veraison. The image acquisition equipment was a SONY ILCE-5100L digital camera manufactured by Sony. The camera resolution was 3008 × 1668 pixels, the aperture value was f/3.2, the exposure time was 1/60 s, and all images were saved in .JPG format. A total of three wine grape varieties were collected including Cabernet Sauvignon, Matheran and Syrah. To ensure the diversity of the samples, 20 images of clusters were collected at each time for each variety under different weather conditions, such as sunny and cloudy days, and different lighting conditions, such as normal, direct sunlight and backlight. A total of 45 acquisitions were performed during this period, with a total of 2700 images. Some sample wine grape images under various imaging conditions are shown in Fig 1. The number of images for each grape variety under different environmental conditions is shown in the Table 1.

Table 1 Number of images of different wine grape varieties.

Parameters	Sunny			Cloudy day	Total /images
	Direct sunlight	Backlight	Normal		
Cabernet Sauvignon	147	108	225	420	900
Matheran	138	96	246	420	900
Syrah	145	110	225	420	900
Total/images	430	314	696	1260	2700

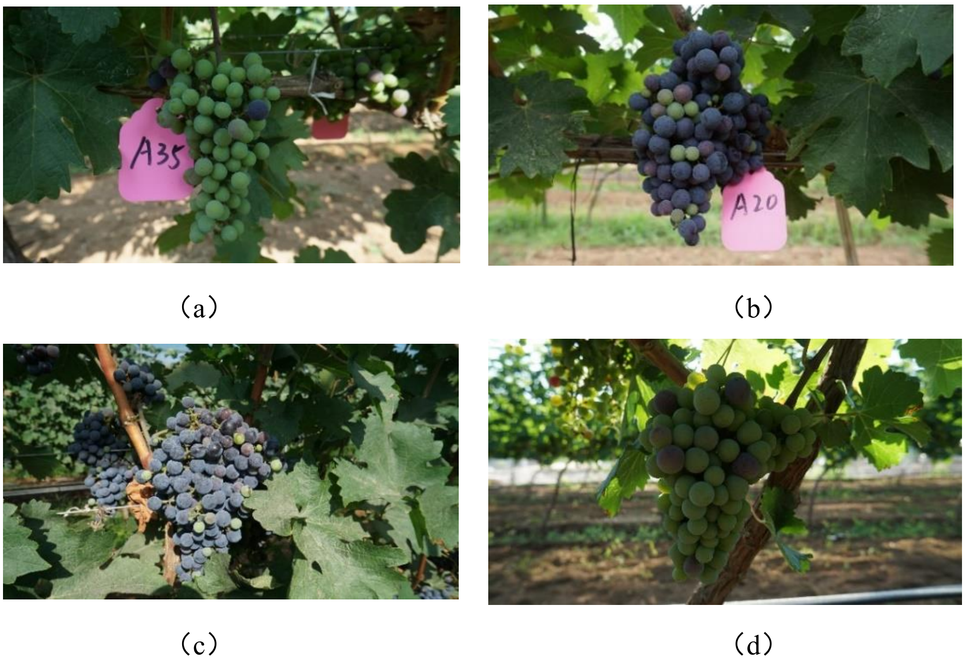

(a) (b) (c) (d)

Fig. 1. Example images of wine grapes in natural field environments: (a) grapes under normal light on a sunny day, (b) grapes under cloudy day, (c) grapes under direct sunlight, and (d) grapes under backlight conditions.

2.1.2 Semantic segmentation of grape cluster

Owing to the interference of the complex background and the small size and variation of individual berries, direct separation of berries could not meet the accuracy requirement, so the background was removed by first segmenting the grape clusters and then further extracting the berries. To this end, the improved PSPNet semantic segmentation model (Chen et al., 2021) was used to remove irrelevant backgrounds as shown in Fig. 2, thus constructing the berry instance segmentation dataset. The grape image data of different wine grape varieties under different weather conditions were selected, and the background was removed for these 85 images as the original dataset for grape berry instance segmentation.

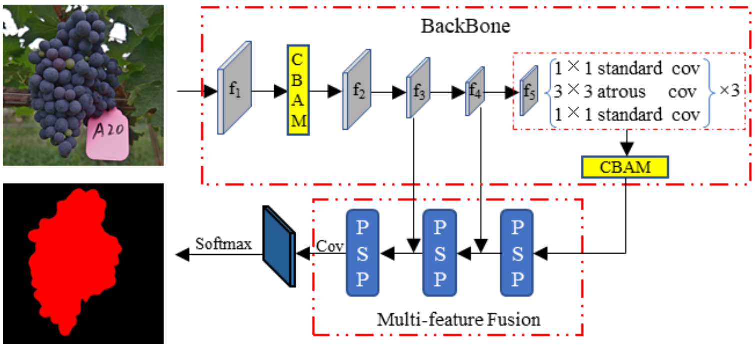
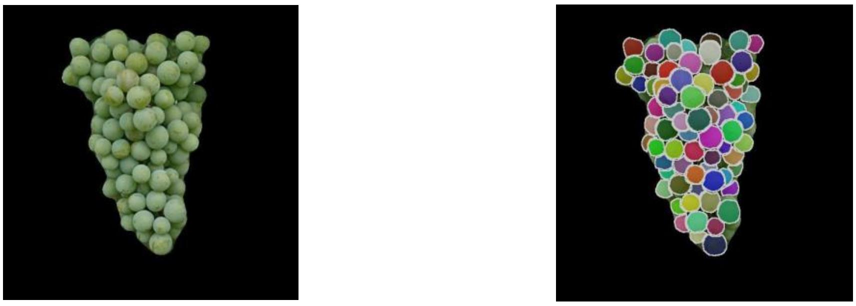


Fig. 2. Semantic segmentation model with irrelevant background removal.

2.1.3 Image annotation for instance segmentation

The berry segmentation dataset in Section. 2.1.2 was then annotated using the LabelMe software interactive polygon tool (Russell et al., 2008). The tool defines the berry outline by using a sequence of points. The label values were named uniformly as "berry", others were treated as background, and the annotation was saved as a JSON file. The criteria adopted in the annotation process included the creation of as accurate a mask as possible for each cluster shown in the image. When more than 80% of the berries were obscured, they were not annotated. An example of berry annotation visualization is shown in Fig. 3. The number of berries on each grape image ranged from 70 to 150, and a total of 5348 berry instances were annotated.



(a) (b)

Fig. 3. Berry annotation: (a) removal of irrelevant background image, (b) annotated individual berries.

2.2 Mask R-CNN based grape berry instance segmentation

Mask R-CNN (He et al., 2017a) is based on the Faster R-CNN (Ren et al., 2015) object detection network, and a branch of FCN is added after the basic feature extraction network to construct an advanced network that integrates object detection and semantic segmentation. It is a two-stage processing framework, where the first stage is to extract the proposals (i.e., regions that may contain an object) of the image using the RPN (Region proposal network). The second stage is to complete the three tasks of category classification, bounding box regression and binary mask generation for the proposal regions extracted in the first stage. The berry detection and segmentation pipeline based on Mask R-CNN is shown in Fig. 4.

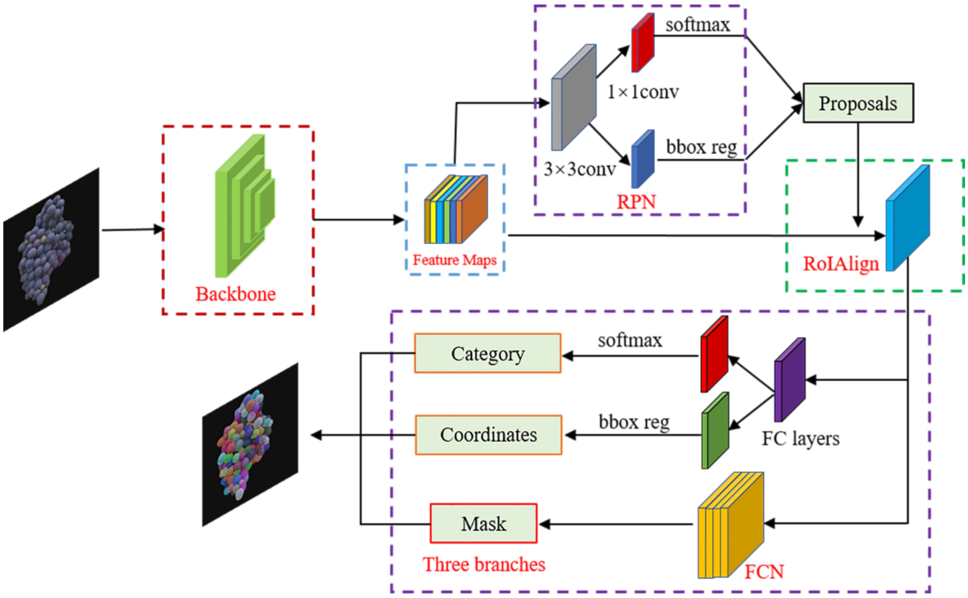


Fig. 4. Mask R-CNN based berry detection and segmentation pipeline.

2.2.1 Backbone network

Mask R-CNN introduces the feature pyramid network (FPN) (Kim et al., 2018) in the backbone feature extraction network ResNet (He et al., 2016), which consists of three parts: bottom-up, top-down and lateral connection, so as to fuse the low-level features with high resolution and the high-level features with rich semantic information. This enables effective integration of low-level features and high-level features at multiple scales, thus making full use of the features extracted by the backbone feature network at each stage. The ResNet-FPN structure is shown in Fig. 5. The ResNet-FPN compresses the original image size to 1/4, 1/8, 1/16, 1/32 times of the original by the feature extraction network ResNet to obtain feature maps C2, C3, C4, C5 of image feature information at different scales. Then five effective feature layers P2, P3, P4, P5 and P6 are obtained by the feature pyramid structure. Finally, these five feature maps at different scales are used as the input of RPN to find the RoI.

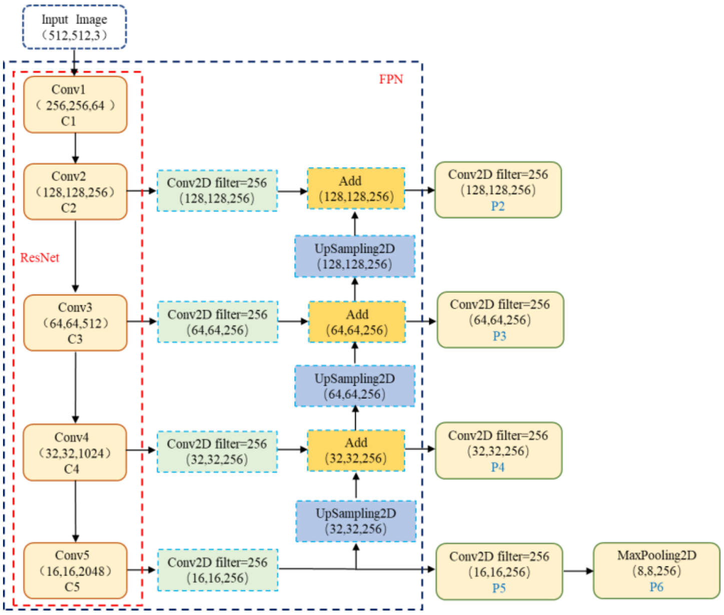

Input Image (512,512,3)
FPN
Conv1 (256,256,64) C1
Conv2 (128,128,256) C2
ResNet
Conv3 (64,64,512) C3
Conv4 (32,32,1024) C4
Conv5 (16,16,2048) C5
Conv2D filter=256 (128,128,256)
Conv2D filter=256 (64,64,256)
Conv2D filter=256 (32,32,256)
Conv2D filter=256 (16,16,256)
Add (128,128,256)
UpSampling2D (128,128,256)
Add (64,64,256)
UpSampling2D (64,64,256)
Add (32,32,256)
UpSampling2D (32,32,256)
Conv2D filter=256 (128,128,256) P2
Conv2D filter=256 (64,64,256) P3
Conv2D filter=256 (32,32,256) P4
Conv2D filter=256 (16,16,256) P5
MaxPooling2D (8,8,256) P6

Fig. 5. ResNet-FPN Structure for feature map generation.

2.2.2 RPN optimization

After the feature map is generated by the backbone feature extraction network, it is passed to the RPN module to generate the proposed regions. First, multiple anchor boxes are generated, and for each anchor box, a classification task and a regression task are performed on it. In the RPN, there are five object detection scales, respectively, 32, 64, 128, 256 and 512, which are anchored mainly to fit 80 different classes of object targets in the COCO2017 dataset (Lin et al., 2014). In this study, the above five detection scales are not fully suitable for the detection of grape berries. Therefore, in order to make the bounding box of grape berries more accurate, the anchor size of the original RPN is optimized. The anchor of the RPN is optimized by combining the size of the input image and the size of the berries, and five detection scales are designed, respectively, 8, 16, 32, 64 and 128, combined with three forms of aspect ratios of labeled rectangular frames, respectively, 0.5, 1 and 2. The final combination of 15 benchmark windows for predicting the region containing the target in the image makes the output more accurate for the region of interest.

2.2.3 Loss function

Mask R-CNN is a multi-task network with a loss function jointly composed of classification, bounding box regression and mask prediction branches. The overall loss calculation formula is as in Eq. 1.

$$L = L_{cls} + L_{box} + L_{mask} \tag{1}$$

where L_{cls} is the classification loss, L_{box} is the regression loss of the bounding box, and L_{mask} is the mask loss. In particular, for the loss in the mask, the Mask branch has an output of $k \times m \times m$ dimensions for each RoI (i.e., k $m \times m$ binary mask images), with k representing the total number of classes. For the predicted binary mask output, a sigmoid function is applied to each pixel point, and the obtained result is used as input to the cross-entropy loss function, and the overall loss is defined as the average binary cross-loss entropy. The calculation of L_{mask} is detailed in Eq. 2.

$$L_{mask} = -\frac{1}{m^2}\sum_{1 \le i,j \le m}\left[y_{ij}\log\hat{y}_{ij}^k + (1-y_{ij})\log(1-\log\hat{y}_{ij}^k)\right] \tag{2}$$

where y_{ij} is the coordinate point (i,j) in the true mask for the region of size $m \times m$, $\hat{y}_{ij}^k$ is the predicted value of the same coordinate in the mask learned for the ground truth class k.

2.3 Model training

The software and hardware configurations used for model training and testing in the experiments are shown in Table 2.

Table 2 Experimental software and hardware configuration details

Accessories	Parameters
Operating System	Linux（Ubuntu20.04）
CPU	Intel(R) CoreTM i9-11900K @ 3.50GHz×16
GPU	NVIDIA RTX 3090, 24GB
Development environments	Python 3.7, Detectron2(pytorch 1.7.1), CUDA 11.1

The dataset (images and annotated results) were divided into a training set and a test set with a ratio of 8:2. To accelerate the model convergence and improve the segmentation accuracy of the network, transfer learning was used to load pre-trained weights on the COCO dataset (Lin et al., 2014) to initialize the model parameters. The hyperparameters for model training were empirically set to 80 for epoch, 2 for bach size, 0.01 for the initial learning rate, and a decay rate of 0.1 times the initial value for every 1500 iterations. To prevent model overfitting, the weight decay was set to

10^{-4} and stochastic gradient descent (SGD) (Bottou, 2012) was used to update the parameters and optimize the training process.

Data enhancement techniques were used randomly during the training process, meaning that mirroring operations (horizontal and vertical), rotation, cropping, and color changes (brightness, contrast, and saturation with intensity between 0.9 and 1.1) were randomly applied online to the input images as each new batch of images was fed into the network for training, and the corresponding annotation files were transformed simultaneously. Meanwhile, a random scaling process was set for each batch of images with a minimum edge length from 448 to 512 pixels, in steps of 32, and a maximum edge size of no more than 512 pixels.

2.4 Identification of grape veraison process based on H component

Extraction of berries is achieved by establishing a berry segmentation pipeline (Fig. 6.), ensuring that image analysis can be performed on individual berries. The input raw image is first semantically segmented to remove the background, and then input into Mask R-CNN for instance segmentation of the berries to generate a mask. Each berry is separated and a connected component is generated by the mask. Finally, each berry is extracted and the number of each grade is counted by the connected component algorithm (He et al., 2017b).

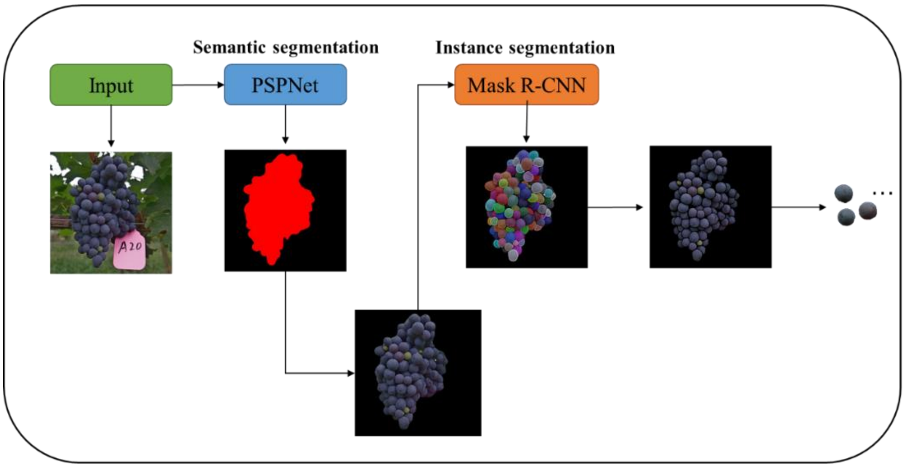


Fig. 6. Individual berry extraction pipeline.

2.4.1 Berry veraison grade and classification criterion

Some of the berries may have reached the mid to late stage of veraison, while others may have just started because of the slow color change. Considering the asynchronous veraison between different berries of the same berry cluster, it is necessary to accurately identify the veraison status of individual berries in order to accurately identify the veraison of the whole berry cluster. Therefore, it is necessary to classify the veraison grade of the wine grape berries.

The original image is an additive color mixing model consisting of R, G, and B light superimposed on each other, which is not suitable for grape berry grade of veraison determination because it is susceptible to light changes. The HSV color space has uniform color variation, where hue (H) only shows color information in the image, not intensity information in the image, with excellent light invariance (Hou et al., 2018; Seetharaman and Kamarasan, 2014; Zhang et al., 2017), which can better reflect the color information in the image. In this study, RGB is mapped to HSV based on the mathematical relationship between RGB and HSV space (Zhang et al., 2017), and the H component of HSV space is used to characterize the dynamics of the veraison of grape berries and thus determine the grade of berries. For the subsequent study, the value range of H is normalized from 0°-360° to between 0 and 1.

In this study, the veraison of berries was classified into four grades, G1, G2, G3, and G4, using the veraison of berries judged by wine viticulture experts as the standard. The G1 grade berries were basically green and had not changed color or the color change was slight; the G2 grade berries were in the transition stage from green to red; the G3 grade berries changed color completely but were still light in color; and the G4 grade berries changed color completely and were very dark, completely changing to dark blue. Using Cabernet Sauvignon as an example, 50 berries were selected for each grade. Since the berries are not a single pixel, the mean of the H value of the pixel area where the selected berries are located was calculated. The mean H values of these 200 berries were statistically analyzed and the results are shown in Fig. 7 (Supplementary Table S1). The range of H values taken for different grade of berries was derived from Fig. 8, and the berry grade of veraison were divided as shown in Table 3.

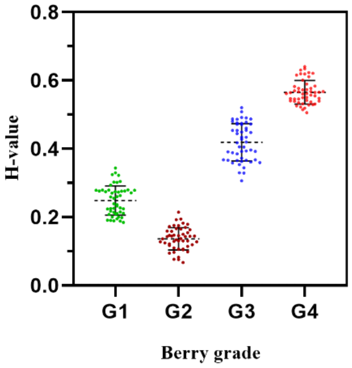


Fig. 7. Distribution of mean H component values of berries of different grades. (The black dashed line in the figure indicates the mean value and the black solid line indicates the error bar)

Table 3 Berry grade of veraison classification.

Grade	H average value range	Examples of berries
G1	0.167＜H≤0.333	
G2	0＜H≤0.167	
G3	0.333＜H≤0.5	
G4	0.5＜H≤0.667	

2.4.2 Classification of grape clusters veraison

In a study of red grape ripeness by Pothen and Nuske, they used the flame seedless red grape variety to classify grape ripeness into four classes based on the percentage of color changing berries in the grape bunches, mainly by determining the percentage of color changing berries in the whole grape bunches (Pothen and Nuske, 2016). Similarly, the veraison is divided into four stages, denoted by Stage1, Stage2, Stage3 and Stage4 respectively. Images of typical grape clusters at each stage of veraison are shown in Fig. 8.

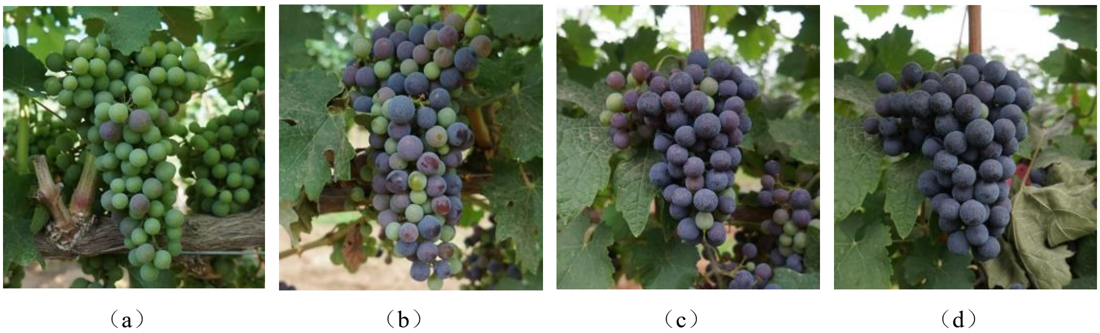


Fig. 8. Images of typical grape clusters at different stages of veraison: (a) Stage1, (b) Stage2, (c) Stage3, and (d) Stage4.

In the first stage of veraison, most of the berries on the grape bunches were G1 grade berries, and the total number of G2, G3, and G4 grade berries accounted for less than 20% of the total number of berries on the whole grape bunches. In the second stage of veraison, the grape bunches were still mainly G1 grade berries, and the total number of G2, G3, and G4 grade berries accounted for 20% to 50% of the total number of berries on the whole grape bunches. In the third stage of veraison, G2, G3, and G4 berries were the main ones on the bunches, with G2, G3, and G4 berries accounting for 50% to 80% of the total number of berries on the whole bunch. At the fourth stage of veraison, that is, at the end of veraison, most of the berries on the grape bunches were G3 and G4 grade berries, and the total number of G3 and G4 grade berries accounted for more than 80% of the total number of berries on the grape bunches. Alternatively, the total number of berries of G2, G3 and G4 grades is more than 80% of the total number of berries of the whole cluster and the total number of berries of G3 and G4 grades is more than 80% of the total number of berries of G2, G3 and G4 grades. The method for determining the veraison of grape clusters is shown in Algorithm 1.

$$scale1=\frac{\sum_{i=2}^{i=4} n_i}{\sum_{i=1}^{i=4} n_i},\quad scale2=\frac{\sum_{i=3}^{i=4} n_i}{\sum_{i=1}^{i=4} n_i},\quad scale3=\frac{\sum_{i=3}^{i=4} n_i}{\sum_{i=2}^{i=4} n_i} \tag{5}$$

Algorithm. 1

```
Algorithm. 1 Grapes cluster veraison determination
Input: $n_i$ , where $n_i$ is the total number of berries with grade $i$
        $(i = 1,2,3,4)$
Output: $V$ , $V \in \{s_1, s_2, s_3, s_4\}$
$k_1$ =scale1, $k_2$ =scale2, $k_3$ =scale3. Refer Eq.5.
if $k_1$ <0.2:
    $V = s_1$
elif 0.2<= $k_1$ <0.5:
    $V = s_2$
elif (0.5<= $k_1$ <0.8) or ( $k_1$ >=0.8 and $k_3$ <0.8):
    $V = s_3$
elif ( $k_2$ >=0.8) or ( $k_1$ >=0.8 and $k_3$ >=0.8):
    $V = s_4$
```

2.5 Evaluation Metrics

For berry instance segmentation, similar to the COCO competition metrics (Lin et al., 2014), average precision (AP) and average recall (AR) were used. The necessary metrics including precision (P) and recall (R) in the calculation of AP and AR are described by Eq. 8, and Eq. 9, respectively. It should be noted that precision and recall are dependent upon the IoU threshold. The IoU is calculated by the predicted segmentation mask (P_m) and ground truth (G) using Eq. 10.

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$IoU = \frac{mask(P_m) \cap mask(G)}{mask(P_m) \cup mask(G)} \tag{8}$$

where TP , FP , and FN means true positive, false positive, and false negative, respectively. The pixel size of the berry dataset is also considered. There are two available sizes (small and medium) according to the area conditions of each instance. The Table 4 details the definitions of the COCO metrics. In this case, the maximum number of detections per image in AR is set to 200, which is different from the original COCO metric, due to the wide distribution of the number of berries annotated in each image in the dataset, ensuring that every berry is detected.

Table 4 COCO Metrics Definition.

Metric	Definition
AP	Average of the ten AP calculated from IoU = 0.5 to IoU = 0.95 increasing in steps of 0.05
$AP_{IoU=0.5}$	AP at IoU =0.5
$AP_{IoU=0.75}$	AP at IoU =0.75
$AR_{\max=200}$	Recall considering the detection of up to 200 objects
APs	AP for small objects: area < 32^2
APm	for medium objects: 32^2 < area < 96^2

3. Results

3.1 Comparison of different Mask R-CNN backbone networks

Grape berries were extracted by constructing three different feature extraction structures of ResNet50-FPN, ResNet101-FPN, and ResNext101-FPN as the backbone feature extraction network of Mask R-CNN instance segmentation model. The performance of each backbone was tested on the grape berry instance segmentation test dataset, and the results given in Table 5 compare the three backbones. The results show that there is no significant difference in term of AP obtained by Mask R-CNN using deeper backbones for feature extraction. The AP of the ResNet50-FPN, ResNet101-FPN, and ResNext101-FPN backbones obtained 81.53%, 80.88%, and 81.94%, respectively. The segmentation effect for small targets is obviously not as high as the average precision of medium target segmentation, with the highest APm reaching 90.17%, which is related to the observed size presented by each berry in the image and some very small area targets were not annotated and detected by the model due to the error of dataset annotation. It should be noted that the deeper ResNet101-FPN backbone does not result in improved model performance, but rather increases the difficulty of training and convergence time as the parameters of the model and the complexity of the network increase. Although the Mask R-CNN with ResNext101-FPN has a slightly better average precision in instance segmentation, it has a longer inference time of 62.42ms/frame compared to the Mask R-CNN with ResNet50-FPN, which has an inference time of 45.70ms/frame.

Table 5 Comparison of Mask R-CNN test results with different backbones.

Backbone	ResNet50-FPN	ResNet101-FPN	ResNext101-FPN
AP	81.53	80.88	81.94
$AP_{IoU=0.5}$	97.63	96.70	97.62
$AP_{IoU=0.75}$	95.57	95.56	95.54
$AR_{\max=200}$	84.10	83.70	84.50
APs	77.91	77.36	78.11
APm	90.17	90.08	90.17
Inference time(ms/frame)	45.70	48.97	62.42

3.2 Influence of different light conditions on berry segmentation

In order to verify the model segmentation performance under different weather lighting conditions, grapes under four different lighting conditions were selected in the test set. As can be seen from the Fig. 9, the Mask R-CNN model has strong robustness under different weather lighting environments, which is partly attributed to the dataset augmentation operation during training, producing a rich variation set that potentially reflects the real field conditions. However, in the red boxed area, due to the gap in the camera field of view, some of the clusters vary more between each other because of occlusion and overlap making the light more variable, leading to blurring in some areas, which makes the accuracy of the predicted masks in these areas decrease and some missed detections occur. However, the overall segmentation effect is surprisingly good.

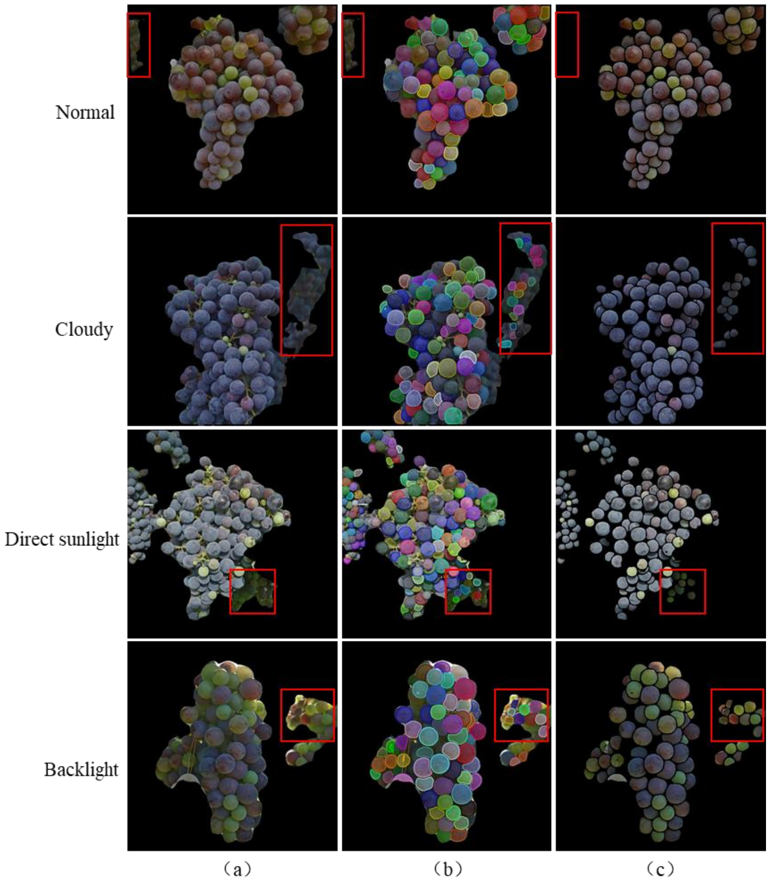


Fig. 9. Grape berries extraction results of Mask R-CNN model under different lighting conditions. (a) original input image. (b) grape berries segmentation result. (c) grape berries extraction result.

3.3 Comparison of different instance segmentation models

In addition, to further validate the effectiveness of the Mask R-CNN instance segmentation model, the same grape berry instance segmentation training dataset was used to train the advanced instance segmentation network SOLOv2 under the same training environment, and the model performance was tested using the test dataset. The comparative performance of the Mask R-CNN and SOLOv2 based berry extraction model for wine grapes is shown in Table 6.

The performance metrics of Mask R-CNN model are significantly better than the SOLOv2 model. The AP , $AP_{IoU=0.5}$, $AP_{IoU=0.75}$,and $AR_{\max=200}$ of Mask R-CNN model obtain 81.53%, 97.63%, 95.57%, and 84.10% higher than SOLOv2 model by 9.43%, 5.86%, 7.12%, and 10%, respectively. In addition, the Mask R-CNN model is also 15ms faster than the SOLOv2 model in terms of model computational efficiency.

Table 6 Comparison of Mask R-CNN and SOLOv2 on test dataset. (note: the one with better performance is highlighted in bold.)

Model	Mask R-CNN	SOLOv2
AP	**81.53**	72.10
$AP_{IoU=0.5}$	**97.63**	91.77
$AP_{IoU=0.75}$	**95.57**	88.45
$AR_{\max=200}$	**84.10**	74.10
APs	**77.91**	65.81
APm	**90.17**	86.55
Inference time(ms/frame)	**45.70**	60.70

Fig. 10 shows the comparison of the berry extraction by Mask R-CNN and SOLOv2. Compared with the Mask R-CNN model, the SOLOv2 model has more duplicate segmentation. As shown in the partially enlarged view of the red box content in Fig. 10(d), the SOLOv2 model misidentifies the same grape berry as multiple different individuals, which may be related to the structure and inference mechanism of the SOLOv2 network. SOLOv2 transforms the segmentation problem into a positional classification problem and directly deals with instance segmentation without relying on box detection, which does not facilitate the segmentation of mutually overlapping targets. Furthermore, the SOLOv2 model suffers from more missed segmentation problems, as shown in the partial enlarged view of the yellow box content in Fig.10 (d), where some of the grape berries are not segmented.

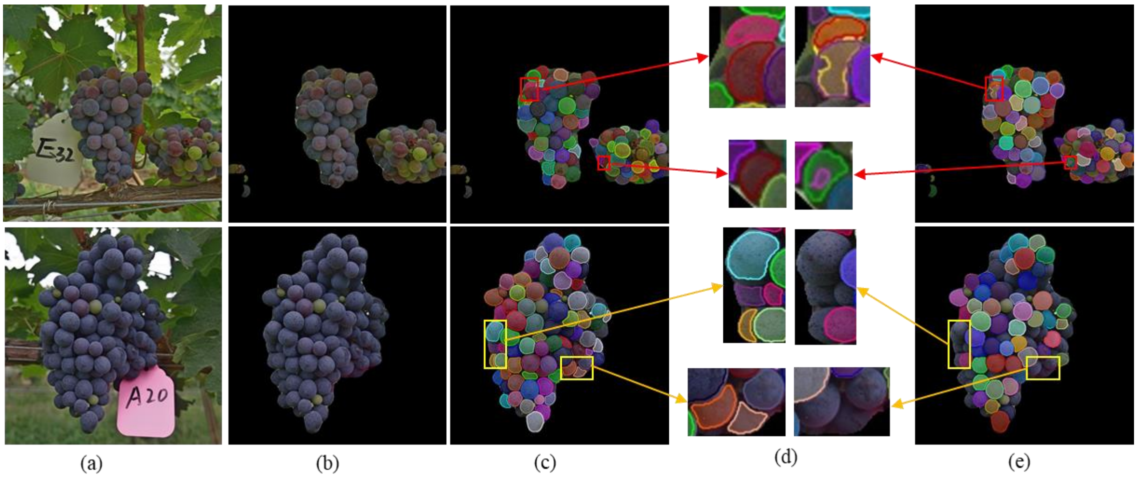
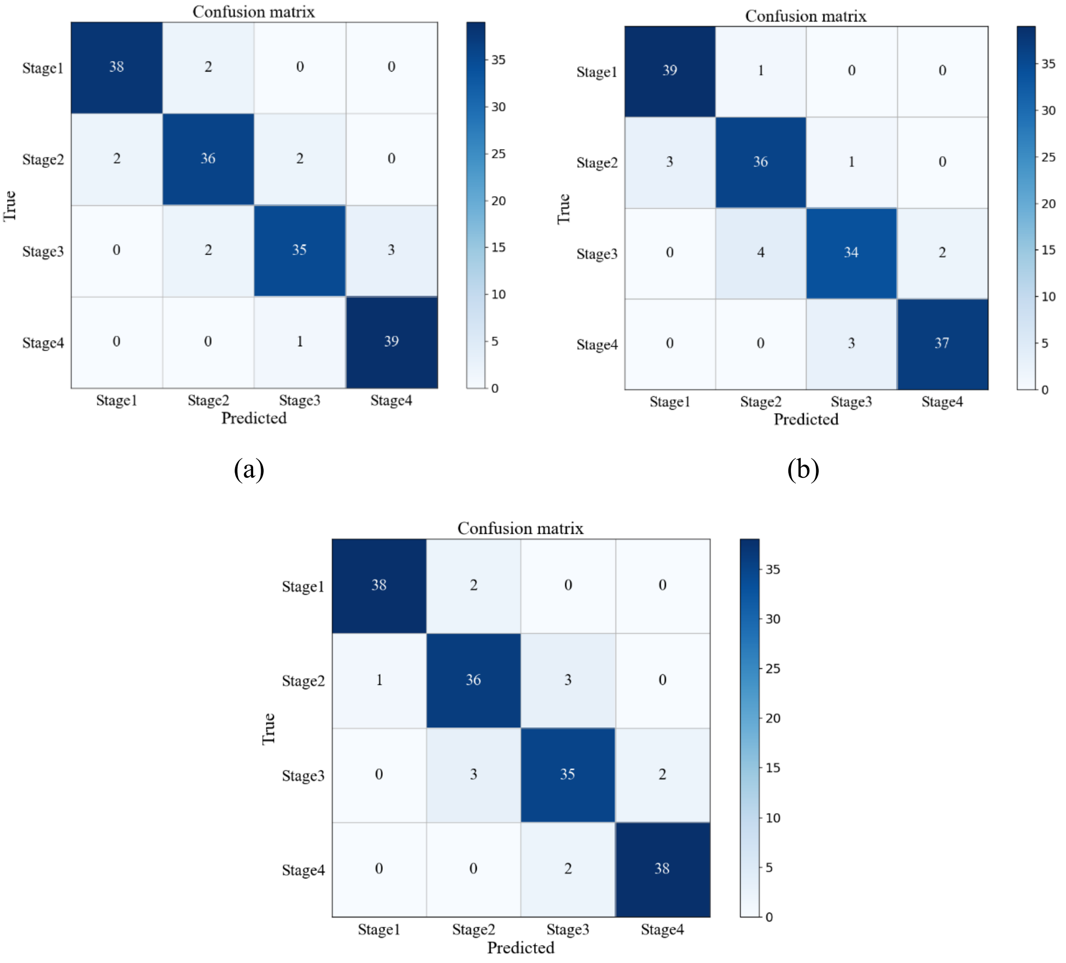


Fig. 10. Comparisons of Mask R-CN and SOLOv2 for some grape berry segmentation examples: (a) original image, (b) results of background removal by semantic segmentation, (c) segmentation results of Mask R-CN grape berries, (d) indicates local zoomed image, (e) segmentation results of SOLOv2 grape berries.

3.4 Results of identification of the colored wine grape veraison process

The images of three colored wine grape varieties including Cabernet Sauvignon, Matheran and Syrah, were randomly selected at different stages of veraison under the guidance of wine viticulture experts. 40 images were selected at each stage for each variety. The proposed algorithm was used to identify the 160 images of each wine grape variety, and the results are shown in Fig. 11.



(a)

(b)

(c)

Fig. 11. Confusion matrix for identifying the veraison of three wine grapes. (a) Cabernet Sauvignon. (b) Matheran. (c) Syrah.

The overall accuracy of the identification results for Cabernet Sauvignon, Matheran, and Syrah wine grapes at each stage of veraison was 92.50%, 91.25%, and 91.88%, respectively. The proposed algorithm has a high accuracy in determining the veraison process for all three colored wine grape varieties. It is also evident from the three confusion matrices that the highest precision was found in the Stage1 and Stage4 for all three wine grapes. It seems possible that these results are due to the berry color of each wine grape variety, which was more obvious in these two stages. The accuracy of the proposed algorithm was slightly lower in the Stage2 and Stage3 of the veraison process, especially in the Stage3, where it resulted in the least precision. These results are likely to be related to the grape berry color, which is closer in these two stages, and it is relatively difficult for the human eye to distinguish the grape berry classes, leading to errors. However, the precision and recall of all three varieties at all four stages was all above 85%, which is encouraging for practical applications. Detailed precision and recall results for all three varieties under four stages are shown in Table 7.

Table 7 Precision and recall of the veraison identification for three different grape varieties.

	Cabernet Sauvignon		Matheran		Syrah	
	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)
Stage1	95.00	95.00	92.86	97.50	97.44	95.00
Stage2	90.00	90.00	87.80	90.00	87.80	90.00
Stage3	92.11	87.50	89.47	85.00	87.50	87.50
Stage4	92.86	97.50	94.87	92.50	95.00	95.00

4. Discussion

4.1 Segmentation error analysis

There are four main types of berry segmentation errors: missed detection, repeated detection, two berries detected as one and one berry detected as two, as shown in Fig. 12. For missed detection, some berries are obscured by other berries, and these berries are difficult to detect even by the human eye. For repeated detection, the segmentation is repeated on top of the correct segmentation of two berries, which may be related to the fact that the model learns some features with adversarial nature during feature learning. Because the dataset was annotated manually, it is difficult to avoid individual berry annotation errors, and then the small pixel size of the berries and the resolution of the image make the contour information between the berries unclear. Due to this error, the mask accuracy decreases. When two berries are not clearly separated and one berry is partially covered by the other berry, they are more likely to be detected as one berry. For a berry detected as two, the example shows that a small portion of the berry is incorrectly detected as a single berry, while another portion of that berry is detected as a separate berry. The most likely reason for these inaccurate detections is that there are many berries covering each other, resulting in only a small portion of some berries being visible. In fruit segmentation studies, Perez-Borrero et al. (Perez-Borrero et al., 2020) used deep learning techniques to segment strawberry instances with an AP of only 45.35%, and Ni et al. (2020) segmented blueberries by developing a image segmentation technique to extract fruit traits with an average precision of 71.6% in the test set under an IoU=0.5 threshold. The AP of our proposed method for segmenting berries was able to reach 81.53%. Although there were some detection errors due to the inherent limitations of 2D images, the overall results were promising.

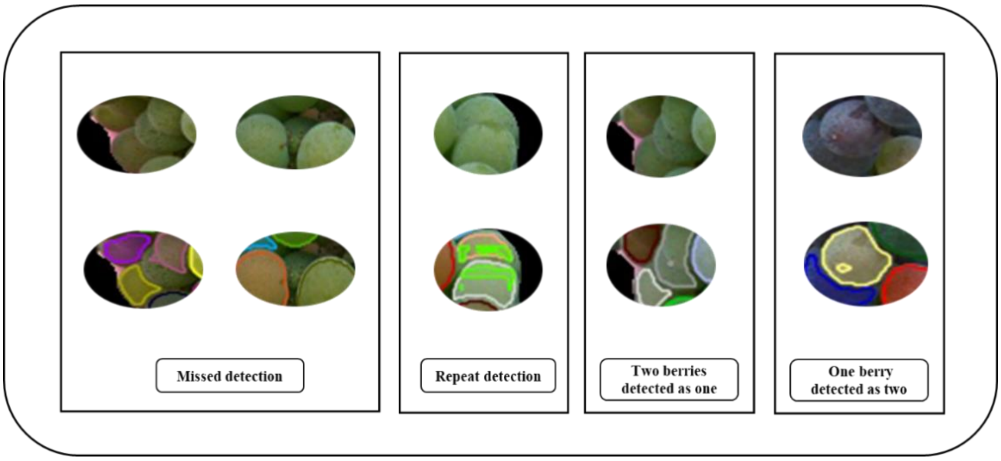


Fig. 12. Four examples of berry segmentation errors.

4.2 H component light invariance

To further validate the light invariance of the H component in the HSV color space of wine grapes. Statistical analyses of H values were performed for each colored wine grape variety at different stages of veraison under different environmental conditions (normal, cloudy, direct sunlight, and backlight). Berries of each variety were randomly selected from 480 images under four environmental conditions at each color change stage, 60 berries were selected for each environment. The total of 2880 individual berry images were selected for analysis (Supplementary Table S2, Table S3 and Table S4). Some selected images of the berries in the different environments are shown in Fig. 13 (Cabernet Sauvignon for example).

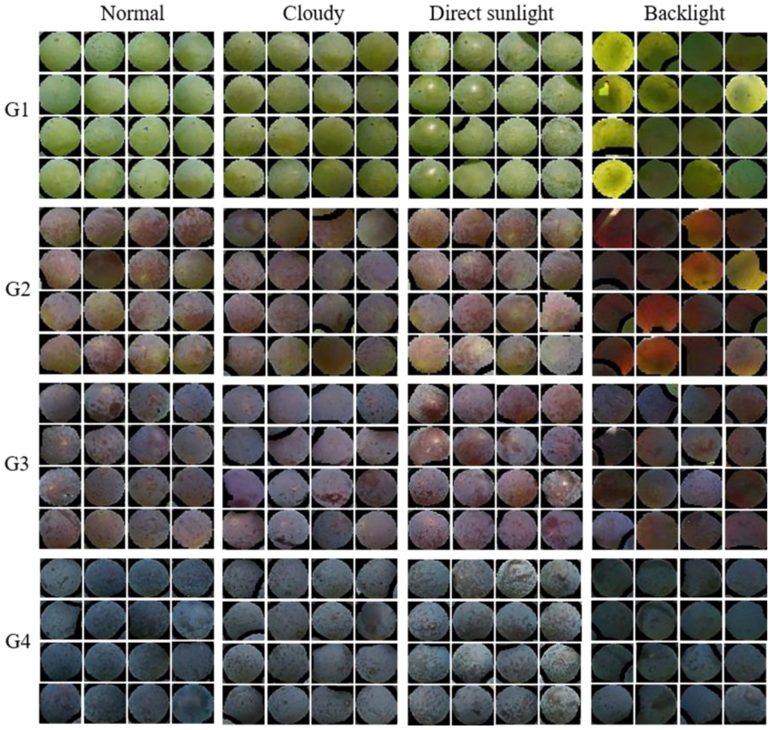
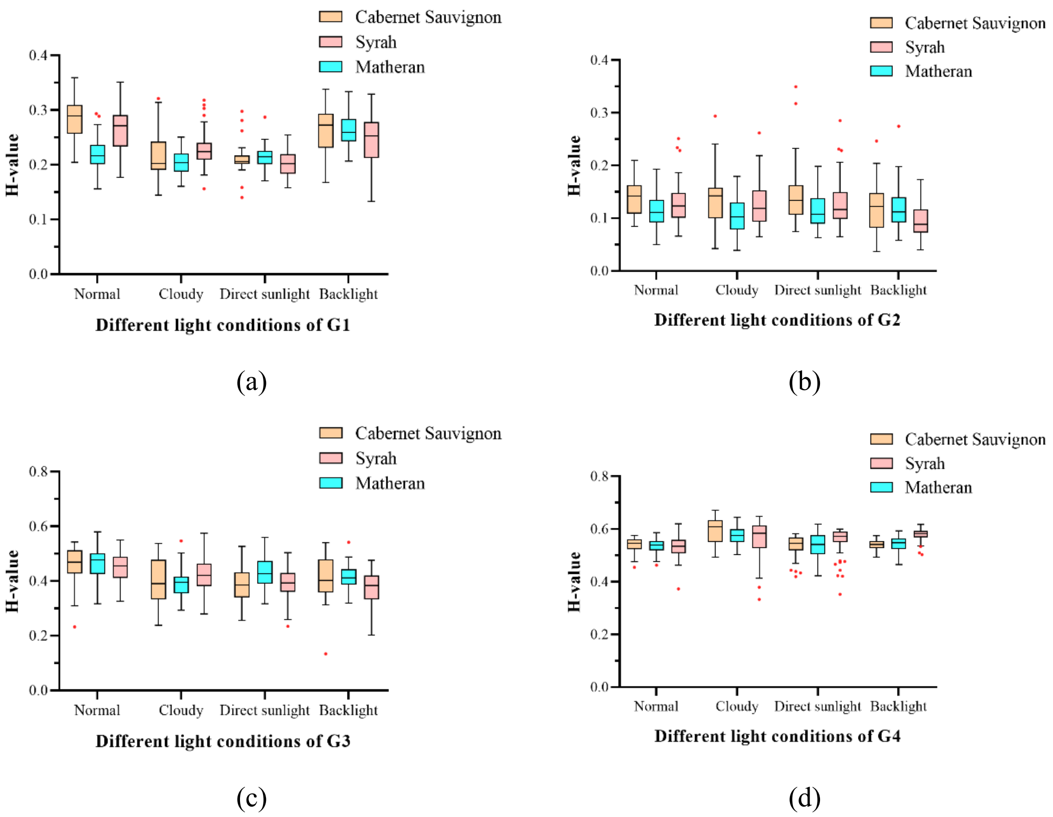


Fig. 13. Images of grapevine berries at four grades of veraison under different light conditions.

Fig. 14 shows that the mean H component values of the G1 grade berries ranging from 0.2 to 0.3 under different weather conditions, 0.05 to 0.15 for the G2 grade, 0.4 to 0.45 for the G3 grade, and 0.5 to 0.65 for the G4 grade. Surprisingly, this is consistent with the method mentioned in section 2.3.1 for the range of grades. An implication of this is the possibility that the H component in HSV was able to determine the veraison of berries of different colored wine grape varieties under both sunny and cloudy weather conditions, and under both direct sunlight and backlight conditions. It was demonstrated that the mean values of the H component of the berries of different veraison of wine grape varieties had good light invariance under various weather conditions, which further demonstrated the feasibility of the proposed method to characterize the berry veraison of colored wine grapes using the H component of HSV.

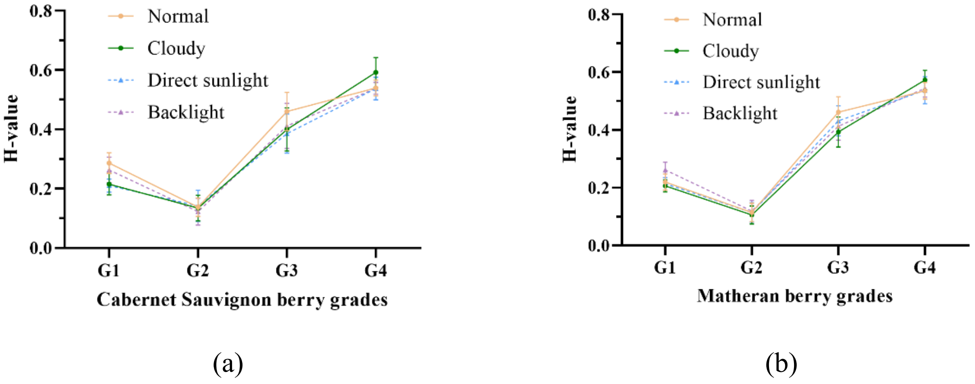
The variation of H component in berries of the three colored wine grape varieties with different grades showed a consistent pattern under different light conditions (Fig. 15). The lowest H component values were taken in G2, which was related to the spatial color distribution of HSV. The H component values of the Matheran variety fluctuated less in a certain range of the four grades G1-G4 under different weather conditions. The berries of the G4 of the three varieties showed more stable H component values under different weather conditions compared to the other three grades, while at the same time, there were more outliers, with Syrah showing the most significant (Fig. 14). This probably relates to the mask accuracy of the segmentation, where the model failed to account for some of the berry edge pixels. Moreover, since the berries are relatively compact, there potentially exists multiple berry parts within the same pixel point.



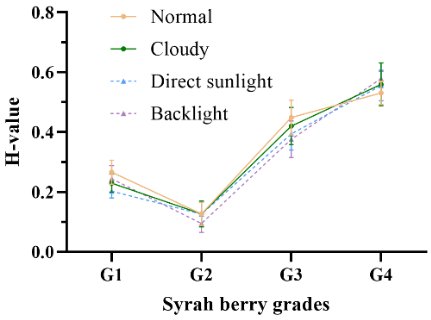
(a) (b)

(c) (d)

Fig. 14. H values of three wine grapes under different light conditions for four grades of berries. (a), (b), (c) and (d) represent G1, G2, G3 and G4 grape berries, respectively (In each boxplot, the top edge, black line inside, and the bottom edge of the box represent the upper (Q3), median (Q2), and lower (Q1) quartiles, respectively. The whiskers represent the maximum (Q3 + 1.5*IQR) and minimum (Q1–1.5*IQR) valid values defined by interquartile ranges (IQR = Q3-Q1), respectively. The red dots outside the box plot represent outliers.)



(a) (b)



(c)

Fig. 15. Variation in H values under different light conditions for different grades of berries. (a), (b) and (c) for Cabernet Sauvignon, Matheran and Syrah berries, respectively.

4.3 Difference between the proposed algorithm and manual identification

Fig. 16 shows images of the same grape bunches and of the same wine grape variety taken on August 10, 2021 and August 11, 2021, and their veraison by wine viticulture experts are that both images are at Stage3. Using the proposed algorithm, the results were that the grape bunches on August 10, 2021 had Scale1, Scale2 and Scale3 of 73.49%, 69.88% and 95.08%, respectively, at Stage3. The results of the wine grape cultivation expert's determination were the same as those of the Stage4 image, which were 82.89%, 80.26%, and 96.83% for the August 11, 2021 grape clusters. It is interesting to note that the results of the manual determination and the proposed algorithm are not consistent. This discrepancy is attributed to the fact that the colors of the two images are close, but there are actual potential differences. As shown in the red box in the partial magnification, it is hard for the human eye to perceive the variance. This finding was unexpected and suggests that the proposed method is in some way more objective than the manual determination. It is a promising method that uses the H component feature to describe the veraison of the whole cluster in terms of dimension of individual berries, which fundamentally explains the dynamics of the whole veraison. It provides sufficient data support for an accurate determination of the veraison.

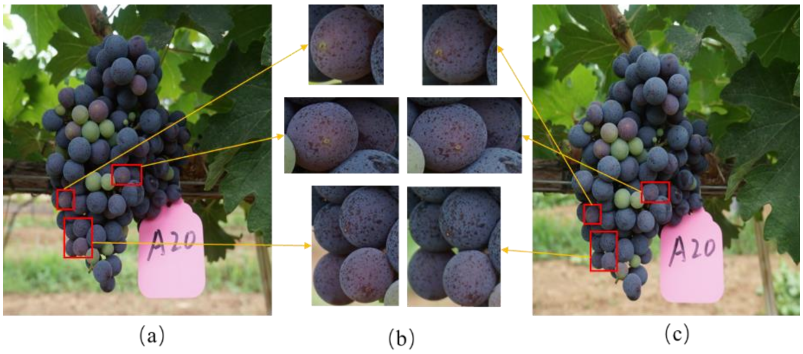


Fig. 16. Images of the same cluster on different dates. (a) indicates images taken on August 10, 2021, (b) indicates partial zoom and (c) indicates images taken on August 11, 2021.

5. Conclusions and future work

The veraison process varies among different clusters and among different berries of the same cluster. The traditional manually identifying method is too subjective, inaccurate and inefficient. In this study, berry segmentation dataset was first constructed using semantic segmentation model to remove irrelevant background. Three different feature extraction structures, ResNet50-FPN, ResNet101-FPN and ResNext101-FPN, were constructed as the backbone feature extraction network of Mask R-CNN instance segmentation model to extract berries from wine grape clusters by optimizing the relevant parameters of the model RPN network. The results show that the Mask R-CNN with ResNet50-FPN structure as the backbone feature extraction network performs relatively well, obtaining AP, $AP_{IoU=0.5}$, $AP_{IoU=0.75}$ and $AR_{\max=200}$ on the test set with 81.53%, 97.63%, 95.57% and 84.10%, respectively, which is higher than the advanced SOLOv2 by 9.43%, 5.86% ,7.12% and 10%, respectively. In addition, the model has good robustness under different weather and lighting conditions.

The H component was proposed to characterize berry veraison grade, and the invariance of the H component of different colored wine grape berries under different light conditions was verified and discussed. The algorithm was developed by calculating the proportion of the total number of berries of different veraison levels in the total number of berries of the whole grape bunches and compared with the results of cultivation experts. This is a promising method to more objectively describe the veraison of the whole bunches in terms of the veraison grade dimension of individual berries, and to provide certain research references to promote the wine grape industry in the direction of refinement, intelligence and automation.

There is also further room for improvement. Firstly, a two-step approach was adopted for individual berry segmentation including semantic segmentation for background removal (e.g. grape clusters segmentation) and instance segmentation for berry segmentation. It is worthy developing a direct berry instance segmentation model with good performance in field conditions. Moreover, although the use of 2D images for berry extraction and the calculation of the percentage of berries with different veraison levels is sufficient to characterize the veraison process of the whole cluster, the inherent limitations of 2D images exist such as making some berries invisible and resulting in segmentation errors. Therefore, it is necessary to extract features from 3D images in the future. Meanwhile, the annotation of the berry dataset was labor-intensive and therefore only three varieties were explored. Therefore, future work also focuses on how to make the developed method generalizable to other wine grape varieties.

CRediT authorship contribution statement

Shen Lei: Conceptualization, Methodology, Software, Formal analysis, Resources, Visualization, Writing – original draft. **Shan Chen:** Conceptualization, Methodology, Software, Resources, Formal analysis. **Zhiwen Mi:** Methodology, Investigation. **Jinya Su:** Investigation, Writing – review & editing. **Rong Huang:** Investigation. **Yuyang Song:** Investigation, Validation, Supervision. **Yulin Fang:** Investigation, Validation. **Baofeng Su:** Writing – review & editing, Project administration, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key R&D Program Project of China (Grant No. 2019YFD1002500) and Guangxi Key R&D Program Project (Grant No. Gui Ke AB21076001) The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

Appendix A. Supplementary data

The following are Supplementary data to this article:

Supplementary data.xlsx (Supplementary tables)

References

Bottou, L., 2012. Stochastic gradient descent tricks, Neural networks: Tricks of the trade. Springer, pp. 421-436.

Buayai, P., Saikaew, K.R., Mao, X., 2020. End-to-end automatic berry counting for table grape thinning. IEEE Access 9, 4829-4842. http://doi.org/10.1109/ACCESS.2020.3048374

Chen, S., Song, Y., Su, J., Fang, Y., Shen, L., Mi, Z., Su, B., 2021. Segmentation of field grape bunches via an improved pyramid scene parsing network. International Journal of Agricultural and Biological Engineering 14(6), 185-194. http://doi.org/10.25165/j.ijabe.20211406.6903

Costa, R., Fraga, H., Fonseca, A., García de Cortázar-Atauri, I., Val, M.C., Carlos, C., Reis, S., Santos, J.A., 2019. Grapevine phenology of cv. Touriga Franca and Touriga Nacional in the Douro wine region: Modelling and climate change projections. Agronomy 9(4), 210. http://doi.org/10.3390/AGRONOMY9040210

Grimm, J., Herzog, K., Rist, F., Kicherer, A., Toepfer, R., Steinhage, V., 2019. An adaptable approach to automated visual detection of plant organs with applications in grapevine breeding. Biosystems Engineering 183, 170-183. http://doi.org/10.1016/J.BIOSYSTEMSENG.2019.04.018

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017a. Mask r-cnn, Proceedings of the IEEE international conference on computer vision, pp. 2961-2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

He, L., Ren, X., Gao, Q., Zhao, X., Yao, B., Chao, Y., 2017b. The connected-component labeling problem: A review of state-of-the-art algorithms. Pattern Recognition 70, 25-43. http://doi.org/10.1016/j.patcog.2017.04.018

Hou, G., Pan, Z., Huang, B., Wang, G., Luan, X., 2018. Hue preserving-based approach for underwater colour image enhancement. IET Image Processing 12(2), 292-298. http://doi.org/10.1049/iet-ipr.2017.0359

Ji, W., Zhao, D., Cheng, F., Xu, B., Zhang, Y., Wang, J., 2012. Automatic recognition vision system guided for apple harvesting robot. Computers & Electrical Engineering 38(5), 1186-1195. http://doi.org/10.1016/j.compeleceng.2011.11.005

Kalt, W., McRae, K., Hamilton, L., 1995. Relationship between surface color and other maturity indices in wild lowbush blueberries. Canadian journal of plant science 75(2), 485-490. http://doi.org/10.4141/CJPS95-085

Kang, H., Chen, C., 2019. Fruit detection and segmentation for apple harvesting using visual sensor in orchards. Sensors 19(20), 4599. http://doi.org/10.3390/s19204599

Keller, M., Hrazdina, G., 1998. Interaction of nitrogen availability during bloom and light intensity during veraison. II. Effects on anthocyanin and phenolic development during grape ripening. American Journal of Enology and Viticulture 49(3), 341-349.

Kestur, R., Meduri, A., Narasipura, O., 2019. MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. Engineering Applications of Artificial Intelligence 77, 59-69. http://doi.org/10.1016/j.engappai.2018.09.011

Khoshroo, A., Arefi, A., Khodaei, J., 2014. Detection of red tomato on plants using image processing techniques. Agricultural Communications 2(4), 9-15.

Kim, S.-W., Kook, H.-K., Sun, J.-Y., Kang, M.-C., Ko, S.-J., 2018. Parallel feature pyramid network for object detection, Proceedings of the European Conference on Computer Vision (ECCV), pp. 234-250.

Liang, C., Xiong, J., Zheng, Z., Zhong, Z., Li, Z., Chen, S., Yang, Z., 2020. A visual detection method for nighttime litchi fruits and fruiting stems. Computers and Electronics in Agriculture 169, 105192. http://doi.org/10.1016/j.compag.2019.105192

Lin, G., Tang, Y., Zou, X., Xiong, J., Li, J., 2019. Guava detection and pose estimation using a low-cost RGB-D sensor in the field. Sensors 19(2), 428. http://doi.org/10.3390/s19020428

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context, European conference on computer vision. Springer, pp. 740-755.

Llerena, W., Samaniego, I., Angós, I., Brito, B., Ortiz, B., Carrillo, W., 2019. Biocompounds content prediction in ecuadorian fruits using a mathematical model. Foods 8(8), 284. http://doi.org/10.3390/foods8080284

Marani, R., Milella, A., Petitti, A., Reina, G., 2021. Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera. Precision Agriculture 22(2), 387-413. http://doi.org/10.1007/s11119-020-09736-0

Martins, V., Cunha, A., Gerós, H., Hanana, M., Blumwald, E., 2012. Mineral compounds in grape berry. The Biochemistry of the grape berry, 23-43. http://doi.org/10.2174/978160805360511201010023

Meng, J.-F., Xu, T.-F., Song, C.-Z., Yu, Y., Hu, F., Zhang, L., Zhang, Z.-W., Xi, Z.-M., 2015. Melatonin treatment of pre-veraison grape berries to increase size and synchronicity of berries and modify wine aroma components. Food chemistry 185, 127-134. http://doi.org/10.1016/j.foodchem.2015.03.140

Ni, X., Li, C., Jiang, H., Takeda, F., 2020. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. Horticulture research 7. http://doi.org/10.1038/s41438-020-0323-3

Parker, A., de Cortázar-Atauri, I.G., Chuine, I., Barbeau, G., Bois, B., Boursiquot, J.-M., Cahurel, J.-Y., Claverie, M., Dufourcq, T., Gény, L., 2013. Classification of varieties for their timing of flowering and veraison using a modelling approach: A case study for the grapevine species Vitis vinifera L. Agricultural and Forest Meteorology 180, 249-264. http://doi.org/10.1016/J.AGRFORMET.2013.06.005

Parker, A.K., DE CORTÁZAR‐ATAURI, I.G., van Leeuwen, C., Chuine, I., 2011. General phenological model to characterise the timing of flowering and veraison of Vitis vinifera L. Australian Journal of Grape and Wine Research 17(2), 206-216. http://doi.org/10.1111/J.1755-0238.2011.00140.X

Perez-Borrero, I., Marin-Santos, D., Gegundez-Arias, M.E., Cortes-Ancos, E., 2020. A fast and accurate deep learning method for strawberry instance segmentation. Computers and Electronics in Agriculture 178, 105736. http://doi.org/10.1016/j.compag.2020.105736

Pothen, Z., Nuske, S., 2016. Automated assessment and mapping of grape quality through image-based color analysis. IFAC-PapersOnLine 49(16), 72-78. http://doi.org/10.1016/J.IFACOL.2016.10.014

Rahman, A., Hellicar, A., 2014. Identification of mature grape bunches using image processing and computational intelligence methods, 2014 IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP). IEEE, pp. 1-6. http://doi.org/10.1109/CIMSIVP.2014.7013272

Ren, S., He, K., Girshick, R.B., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1137-1149. http://doi.org/10.1109/TPAMI.2016.2577031

Rienth, M., Vigneron, N., Darriet, P., Sweetman, C., Burbidge, C., Bonghi, C., Walker, R.P., Famiani, F., Castellarin, S.D., 2021. Grape berry secondary metabolites and their modulation by abiotic factors in a climate change scenario–a review. Frontiers in Plant Science 12, 643258. http://doi.org/10.3389/fpls.2021.643258

Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T., 2008. LabelMe: a database and web-based tool for image annotation. International journal of computer vision 77(1), 157-173. http://doi.org/10.1007/s11263-007-0090-8

Sadras, V.O., Petrie, P.R., 2012. Predicting the time course of grape ripening. Australian journal of grape and wine research 18(1), 48-56. http://doi.org/10.1111/J.1755-0238.2011.00169.X

Santesteban, L.G., 2019. Precision viticulture and advanced analytics. A short review. Food chemistry 279, 58-62. http://doi.org/10.1016/j.foodchem.2018.11.140

Santos, T.T., de Souza, L.L., dos Santos, A.A., Avila, S., 2020. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. Computers and Electronics in Agriculture 170, 105247. http://doi.org/10.1016/J.COMPAG.2020.105247

Seetharaman, K., Kamarasan, M., 2014. Statistical framework for image retrieval based on multiresolution features and similarity method. Multimedia tools and applications 73(3), 1943-1962. http://doi.org/10.1007/s11042-013-1637-z

Wang, D., He, D., 2022. Fusion of Mask RCNN and attention mechanism for instance segmentation of apples under complex background. Computers and Electronics in Agriculture 196, 106864. http://doi.org/10.1016/j.compag.2022.106864

Wang, Y., Zhang, X., 2014. Segmentation algorithm of muskmelon fruit with complex background. Transactions of the Chinese Society of Agricultural Engineering 30(2), 176-181.

Xu, Y., Imou, K., Kaizu, Y., Saga, K., 2013. Two-stage approach for detecting slightly overlapping strawberries using HOG descriptor. Biosystems engineering 115(2), 144-153. http://doi.org/10.1016/J.BIOSYSTEMSENG.2013.03.011

Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Kuhlmann, H., Roscher, R., 2020. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing 164, 73-83. http://doi.org/10.1016/j.isprsjprs.2020.04.002

Zhang, M., Zou, F., Zheng, J., 2017. The linear transformation image enhancement algorithm based on HSV color space, Advances in Intelligent Information Hiding and Multimedia Signal Processing. Springer, pp. 19-27. http://doi.org/10.1007/978-3-319-50212-0_3