



GrapeCPNet: A self-supervised point cloud completion network for 3D phenotyping of grape bunches

Wenli Zhang ^{a,*}, Chao Zheng ^a, Chenhuizi Wang ^a, Pieter M. Blok ^b, Haozhou Wang ^b, Wei Guo ^b

^a Information Department, Beijing University of Technology, Beijing, China

^b Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

ARTICLE INFO

Dataset link: <https://github.com/I3-Laboratory/GrapeCPNet>

Keywords:

3D point cloud analysis
Instance segmentation
Shape completion

ABSTRACT

The measurement of phenotypic parameters of fresh grapes, especially at the individual berry level, is critical for yield estimation and quality control. Currently, these measurements are done by humans, making it costly, labor-intensive, and often inaccurate. Advances in 3D reconstruction and point cloud analysis allow extraction of detailed traits for grapes, yet current methods struggle incomplete point clouds due to occlusion. This study presents a novel deep-learning-based phenotyping pipeline designed specifically for 3D point cloud data. First, individual berries are segmented from the grape bunch using the SoftGroup deep learning network. Next, a self-supervised point cloud completion network, termed GrapeCPNet, addresses occlusions by completing missing areas. Finally, morphological analyses are applied to extract berry radius and volumes. Validation on a dataset of four fresh grape varieties yielded R^2 values of 85.5% for berry radius and 96.9% for berry volume, respectively. These results demonstrate the potential of the proposed method for rapid and practical extraction of 3D phenotypic traits in grape cultivation.

1. Introduction

Fresh grape is one of the most important agricultural products and economic crops (Alston and Sambucci, 2019). Its global production and consumption is increasing year by year. To effectively guide planting strategies and monitoring growing status, several geometric parameters, including the number of berries and their radius, are frequently measured. Traditional manual traits measuring is labor-intensive and time-consuming, and often destructive. To reduce the cost and increase the efficiency, the non-destructive and high-throughput acquisition of phenotypic traits based on computer vision has attracted extensive interest in recent years. Several simple geometry traits can be acquired from 2D images by using computer vision algorithms that use pattern recognition, machine learning, and deep learning (Chen et al., 2023; Zabawa et al., 2020). However, projecting the 2D image plane onto the 3D world inevitably results in information loss. The occlusion of berries often lead to incomplete representations of grape bunches. Additionally, clustered berries and their shadows create challenges for obtaining reliable measurements of phenotypic parameters through 2D image analysis, even under controlled lightning conditions.

Advances in 3D scanning and reconstruction techniques have made it possible to retain detailed spatial information about objects, enabling the acquisition of complex 3D morphological traits in grape

bunches (Ni et al., 2021). Researchers have developed 3D phenotyping processes for grapes by combining feature-engineered point cloud processing algorithms (Rose et al., 2016; Rist et al., 2018). These methods primarily rely on point cloud segmentation techniques, such as edge-based segmentation, region-growing segmentation, and sphere-fitting algorithms. Unfortunately, traditional point cloud processing algorithms heavily rely on prior knowledge of the target object and require extensive manual feature engineering and parameter tuning to achieve optimal results in a wide range of conditions, thereby limiting their generalization performance. Implementing state-of-the-art deep learning approaches, which automatically learn features from the point cloud data, can reduce the workload of designing robust algorithms for complex and diverse conditions. However, few studies have reported good performance deep learning networks and pipelines in berry level analysis.

Another challenge after the success of instance berry segmentation is the incomplete representation of berry surfaces. The aggregated cluster growth causes the berries to squeezed against each other, leading to self occlusion and resulting in an incomplete point cloud of berry surfaces (Du and Liu, 2023). Therefore, point cloud instance segmentation can only capture partial information rather than that of whole berries.

* Corresponding authors.

E-mail addresses: zhangwenli@bjut.edu.cn (W. Zhang), guowei@g.ecc.u-tokyo.ac.jp (W. Guo).

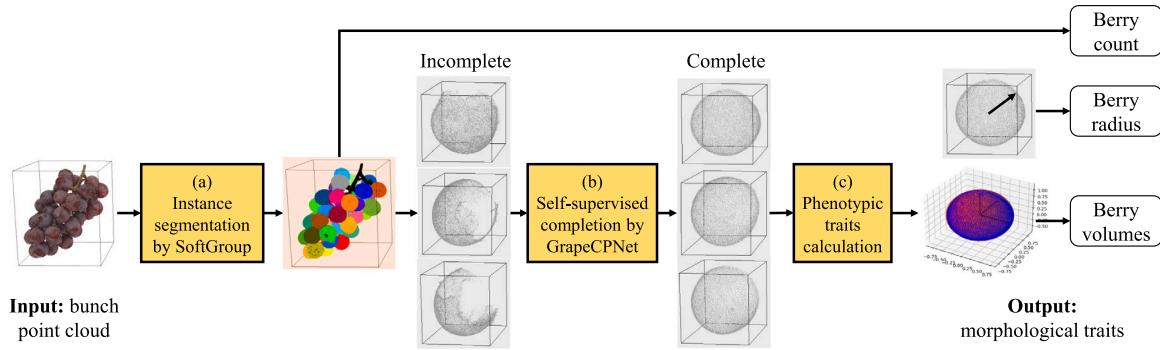


Fig. 1. Proposed pipeline for grape bunch 3D phenotyping. Taking the bunch point cloud data as input, it is realized through three steps: (1) berry instance segmentation by SoftGroup (Vu et al., 2022), (2) berry point cloud completion by proposed GrapeCPNet, and (3) phenotypic traits calculation.

Table 1
Comparison of the characteristics of the four species of fresh grape.

Grape species	Color	Sparsity	Shape
Red	Purple-red	Sparser	Ellipsoid
Green	Green	Sparse	Ellipsoid
Kyoho	Dark-purple	Dense	Sphere
Shine-Muscat	Green	Very dense	Sphere

To address this issue and consequently an underestimation of the berry traits, point cloud shape completion is one of the most commonly used solutions. Point cloud completion is the task of generating a complete 3D shape given partial point cloud input (Tesema et al., 2024). For instance, Ge et al. (2020) assume that strawberries are symmetrical, suggesting that their shape can be completed by mirroring the existing shape, provided an appropriate symmetrical plane can be identified. Marangoz et al. (2022) proposed an super-ellipsoids matching approach to generate cube-like point clouds for peppers and estimate their shape. However, these methods are limited to specific fruit characteristics and require manual design of algorithms according to the features of the fruit, lacking the potential for generalized application. Utilizing deep learning methods, researchers have designed point cloud completion networks (Wang et al., 2021). Park et al. (2019) introduced DeepSDF to achieve the completion of object point cloud models. It uses a learned continuous signed distance function to represent a class of shapes in latent space and is tolerant to noisy 3D input data. Magistri et al. (2022) utilized DeepSDF and its latent space in an encoder-decoder network that converted a single RGB-D frame to a 3D shape of a complete fruit. This encoder-decoder network was extended and optimized by Blok et al. (2025) for high-throughput 3D shape completion of potato tubers on a commercial harvester. For grape berry application, the shape of the berry can be roughly approached by the deformation of a sphere or ellipsoid depending on the species. Existing datasets from other fruits cannot be applied directly, and manually creating paired training datasets is time-consuming and labor-intensive. Therefore, there is a demand to develop an analysis pipeline, capable of completing the detailed surface of an incomplete berry point cloud while requiring less labor-intensive training data preparation.

In this study, we proposed a novel berry-level 3D grape phenotyping pipeline that will overcome the occlusion issues. The proposed pipeline includes: (1) an individual berry segmentation module that segment each berry from 3D point cloud of whole bunch; (2) a point cloud completion module that complete the missing part of segmented berry point cloud data through automatic training data generation; (3) a phenotyping module that extract berry phenotypic traits including radius and volume. The proposed pipeline was evaluated on four different fresh grape species (Red, Green, Kyoho, and Shine Muscat Grape).

Table 2
Image acquisition details for grape bunches and single berries.

Object	Distance	Camera view angles	Rotation angles	Total image #
Bunch	25–60 cm	0°, ±25°, ±45°	7.2° (50 images)	250
Single berry	20–40 cm	−15°, 0°, 25°, 45°	7.2° (50 images)	200

2. Materials and methods

In this section, we first introduce the plant material for our grape analysis, followed by the 3D reconstruction methods for obtaining 3D point cloud data. We then provide a brief overview of the Soft-Group (Vu et al., 2022) deep learning network, which was trained on our point cloud dataset for instance segmentation of individual berries (Fig. 1a). Next, we detail the proposed self-supervised completion deep learning module, GrapeCPNet, which predicts the complete berry point cloud from segmented partial berries (Fig. 1b). Subsequently, we describe the module and method for calculating phenotypic traits (Fig. 1c). Finally, we introduced the evaluation method for model training and results validation.

2.1. Plant materials and point cloud acquisition

2.1.1. Grape species

To ensure the generalization and broad application of our study, this paper considers the sparsity, shape, and color of the bunch, and chooses four common fresh grapes as the experimental materials, which are Red Grape, Green Grape, Kyoho Grape, and Shine-Muscat Grape. The characteristics of them are shown in Table 1 and the examples are shown in Fig. 2.

2.1.2. 3D reconstruction of grape

In this paper, the multi-view images acquisition and 3D reconstruction by Agisoft Metashape (Agisoft LLC, St. Petersburg, Russia) was used to obtain the 3D point cloud of bunch (Fig. 3a) and single complete berry (Fig. 3b). Their correspondence of the same berry was also labeled as shown in Fig. 3c.

For whole bunch multi-view images acquisition, the device was built as shown in Fig. 3a. One grape bunch was hung on the hook connected to propellers driven by a stepping motor (Fig. 3a1), which enabled rotation at a controlled speed. The entire device was placed in front of a solid-colored backdrop (Fig. 3a2), and multiple cameras were set up to photograph the bunch from different heights and view angles. In addition, a ruler with markers were used for scale correction. For single complete berry multi-view images acquisition in batch (Fig. 3b), nine berries were fixed floating by needle disk (Fig. 3b1) on a turntable (Fig. 3b2). The background, cameras, and marker settings are similar to the previous bunch reconstruction. The number of reconstructed images

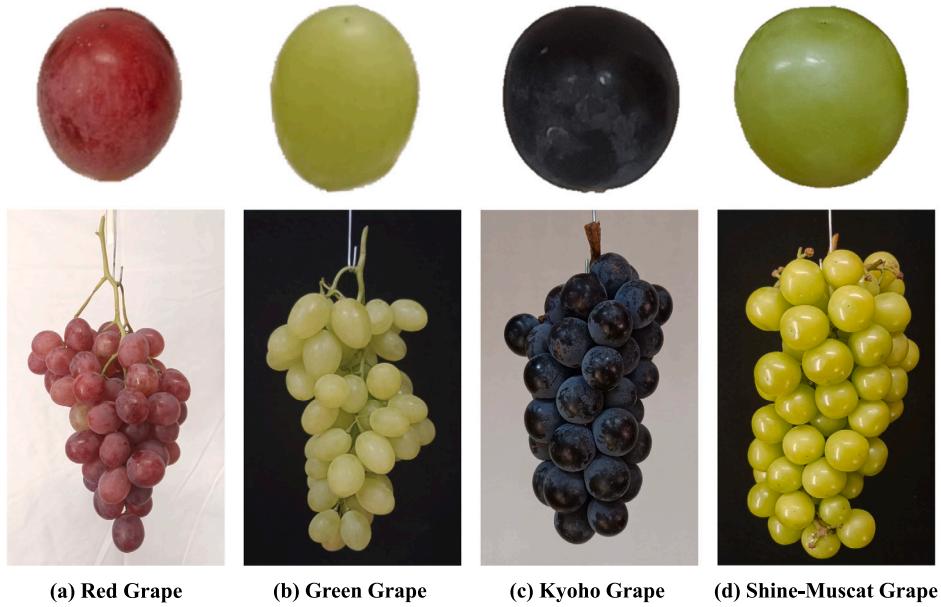


Fig. 2. Examples of four fresh grape bunches and their berries.

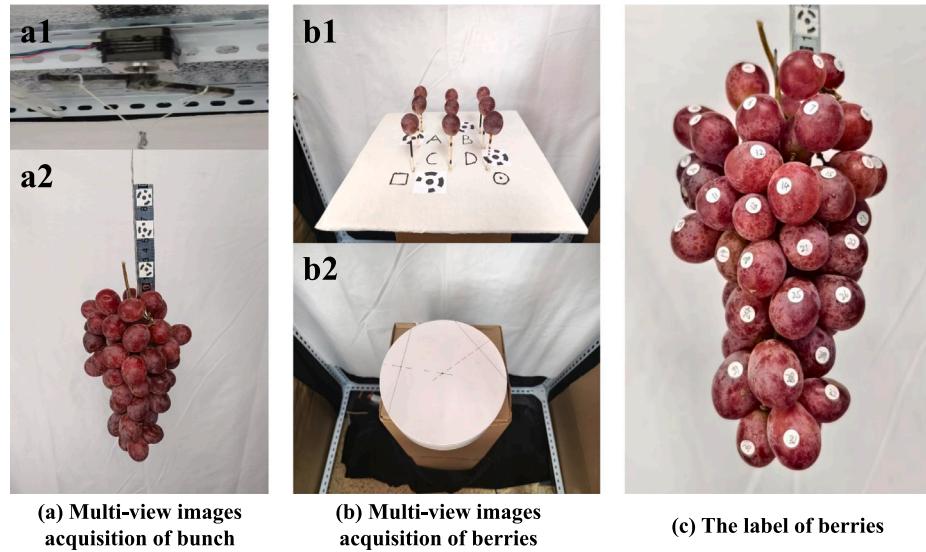


Fig. 3. Methods for obtaining multi-view images of bunch reconstruction and berry reconstruction, and the example of berry labeling.

and the image acquisition parameters are selected, as shown in Table 2. Examples of the point cloud of bunch, multiple berries, and single complete berries from multiple berries are shown in Fig. 4.

2.2. Berry instance segmentation

As Fig. 1a, the single-bunch shell point cloud was represented as $P_{bunch} = \{p_i \mid i = 1, \dots, m\}$, where $p_i = (x_i, y_i, z_i)$ denotes a point in 3D space and m denotes the total number of points in the point cloud of the bunch. The bunch point cloud segmentation task expects to segment P_{bunch} into a combination of several individual subsets of the berries point cloud and a subset of the stem point cloud, as shown in Eq. (1).

$$P_{bunch} = \sum_{i=1}^n P_{berry_i} + P_{stem} \quad (1)$$

where, P_{berry_i} denotes the subset point cloud of the i th berry in the bunch, n denotes the number of all the berries contained in the bunch,

and P_{stem} denotes the set of stem point cloud. Meanwhile, each subset $P_{berry_i} = \{p_i \mid i = 1, \dots, k\}$ consists of multiple points in 3D space, which can represent the location and morphology of the visible part of the berry within the bunch.

The task of instance berry segmentation from a bunch is complex. As shown in Fig. 5, it involves distinct localization features and complex boundary morphology. The distinct localization feature implies that individual berry has more geometrical and morphological details compared to the overall shape of the grape bunch, particularly when the object is a small-sized point cloud (Luo et al., 2022). The complex boundary morphology arises from berries pressing against each other, resulting in multiple curves with varying connecting depths. Therefore, accurate berry instance segmentation requires a model that can both capture detailed geometric features and identify instances at boundary point.

In this paper, SoftGroup (Vu et al., 2022) was selected for bunch point cloud segmentation after comparing different deep learning-based

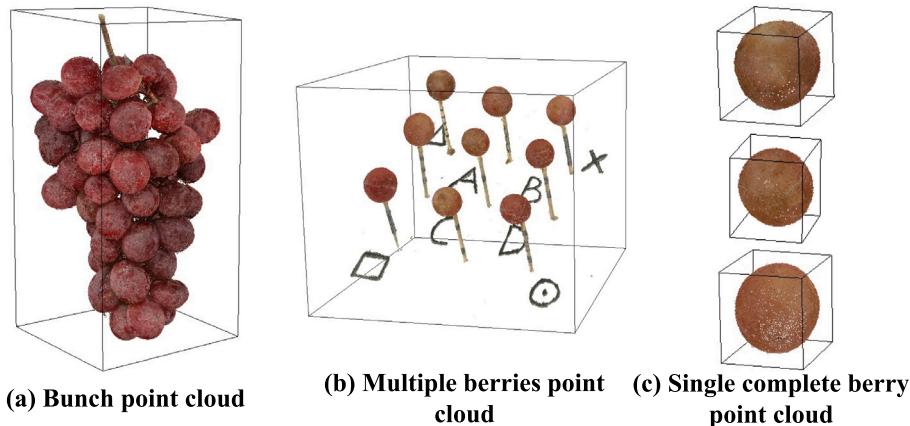


Fig. 4. Example of point cloud obtained based on multi-view images reconstruction.

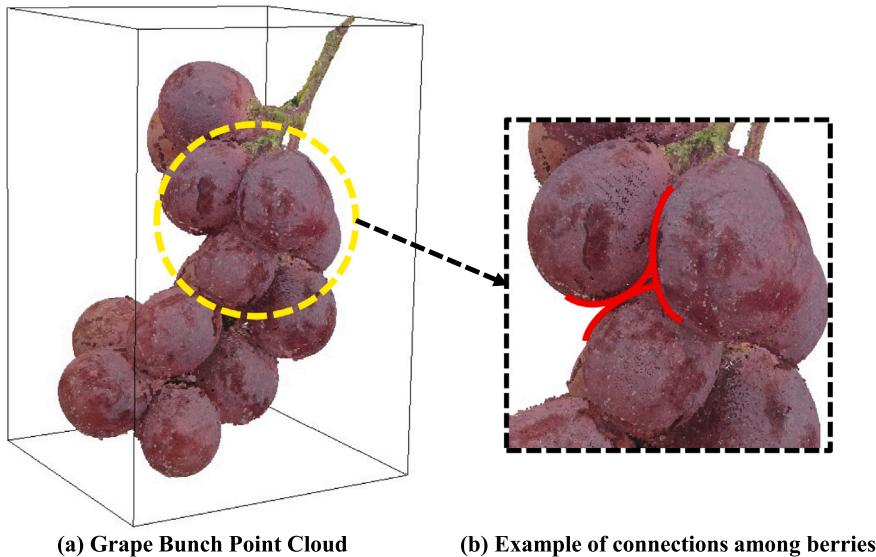


Fig. 5. Example of berry connection in a bunch (in the case of Red Grape). The red line in the figure shows the characteristics: multiple curves of different depths intersect each other.

point cloud instance segmentation models. SoftGroup used a two step approach: (1) the point-by-point prediction U-Net was used to capture the detailed geometric characteristics between points. (2) a threshold score was used to judge the boundary of the point to avoid point category misclassification. Therefore, it is suitable for use in a bunch where there is a tight connection between the berries.

In order to train and validate SoftGroup, each berry and stem in the bunch point cloud obtained in Section 2.1.2 was individually labeled (instance-level segmentation) by CloudCompare (<https://www.cloudcompare.org>) as shown in Fig. 6. This paper collected and labeled eight bunches for each of the four species (in total 32 bunches). Five grape bunches per grape species (in total 20) were randomly selected and used as training samples. The remaining three bunches were used for validation.

2.3. Self-supervised point cloud completion

The self-supervised training-based berry point cloud completion was used to predict their complete 3D model. Self-supervised learning is a way of training a model without annotations by clustering the data in unsupervised manners. The following section summarizes the self-

supervised strategy for the paired training data. It is followed by a description of the grape completion network, GrapeCPNet (Fig. 7).

2.3.1. Self-supervised training data generation

Supervised point cloud completion training requires the incomplete-complete data pairs of the same object to build a mapping relationship. The common incomplete characteristics of the berry are shown in Fig. 8. In this paper, to decrease the labor cost of such data pair generation, we generated incomplete berries from complete berry scans automatically. The complete grapes were obtained by single berry reconstruction (Section 2.1.2) and were used as ground truth; The incomplete grapes were generated by removing parts overlapped with randomly generated 3D sphere regions from the complete grapes.

The pseudo-code of our proposed removal algorithm is shown in Algorithm 1. The inputs include a single complete berry point cloud and the maximum number of removal sphere regions. The output is a selected incomplete berry point cloud. First, the single complete berry point cloud was normalized and centralized (Fig. 9a). Second, randomly choose the number of removal sphere regions for current grape berry, with a maximum of five removal sphere regions specified in this paper. Then, for each removal sphere region, the collision portion with the

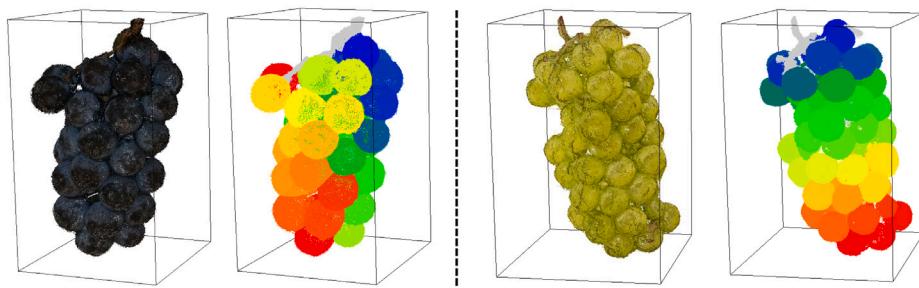


Fig. 6. Example of instance-level labeling of a bunch point cloud: each berry has a unique instance label, as shown by different colors of point sets.

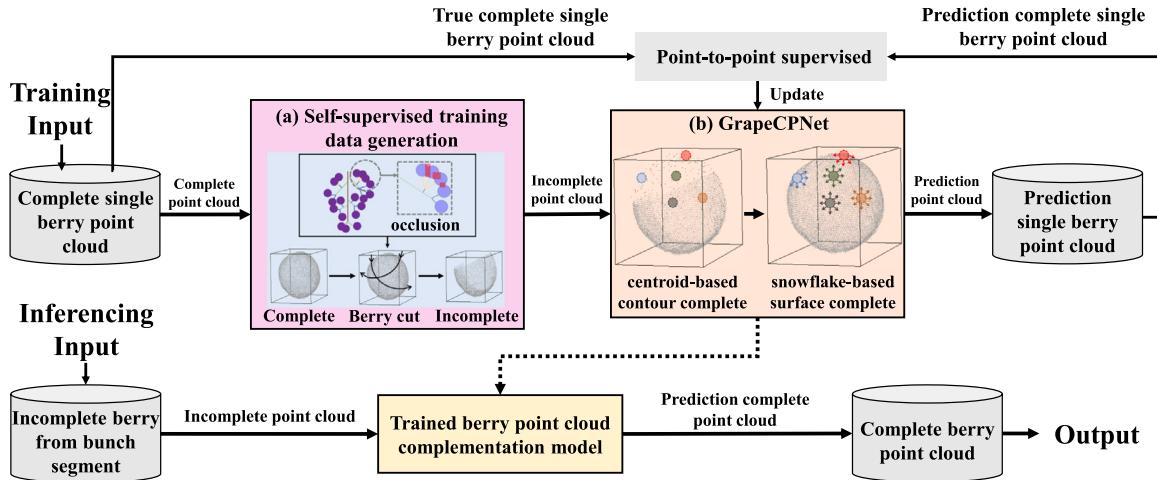


Fig. 7. Flowchart of the self-supervised berry point cloud completion, including model training and model inferencing phases. In the training phase, the paired incomplete training data was generated from the complete single berry point cloud, then GrapeCPNet was trained using self-supervised methods without manual annotation. In the inferencing phase, the incomplete berry point cloud is input to the trained model to obtain a complete prediction berry.

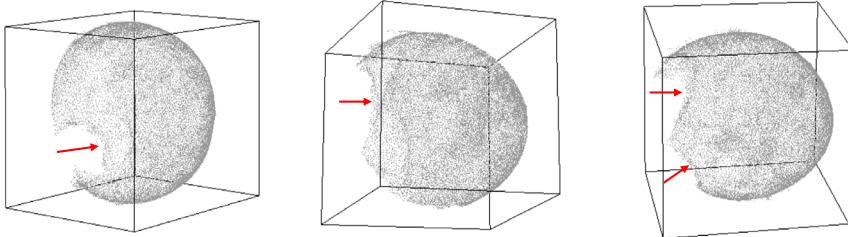


Fig. 8. Example of different perspectives of an incomplete berry point cloud segmented from the bunch. The edge of the absence shows multiple segments connected by circular curves, as shown by the red arrows.

berry point cloud was removed. The center and radius parameters of the 3D removal sphere region were randomly set according to the normalized berry size. In this paper, we set the sphere center positions within a random range of ± 0.75 relative to the grape center, and the sphere radii within a random range of 0 to 0.25. The incomplete berry point cloud was generated as training data by iteratively removing spherical regions (Fig. 9b).

For model training, 36 complete berries were collected for each grape species (in total 144). To further expand the data quantity, based on the single complete berry point cloud obtained by reconstruction, this paper uses the PointFlow (Yang et al., 2019), a point cloud generation model to generate the simulated samples (Fig. 9c). Then, 500 complete berry point clouds are generated from PointFlow method. Thus, a total of 644 complete samples are obtained. Each complete berry point cloud has 5 incomplete berry point clouds corresponding to it, and a total of 3220 training data pairs are obtained. For the training

and testing of the model, this paper divides the dataset according to a ratio close to 70%–30%, with a total of 450 complete berry point clouds (2250 training data pairs) to train the network and 194 complete berry point clouds (970 testing data pairs) to test.

For the model inference validation, from the 12 testing grape bunches of the berry instance, one bunch was randomly selected per grape species (in total 4 bunches), refer to (Section 2.2). For each bunch, 27 berries were randomly selected, yielding a total of 108 incomplete berries for testing the completion network inference validation. The labels of these berries (Section 2.1.2 & Fig. 3c), were used as ground truth.

2.3.2. GrapeCPNet

Since the entire contour of the incomplete berry is missing, the model is required to have the ability to predict the incomplete part based on the visible part and make the prediction part fit the real berry

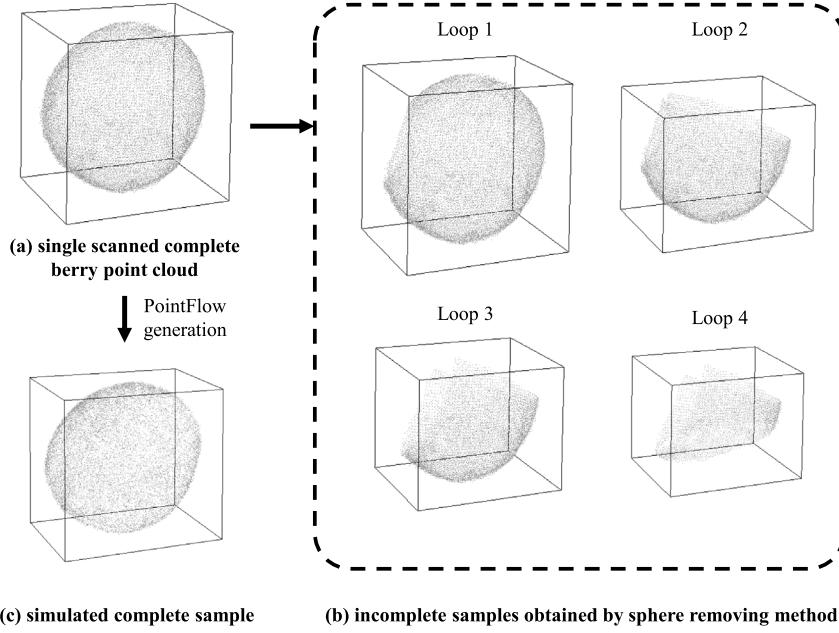


Fig. 9. Workflow of automatic paired training dataset generation for the berry point cloud completion network.

Algorithm 1: The selection method for generating training data of incomplete berries

```

Data:  $P_C$  // one complete berry point cloud
Data:  $n$  // number of incomplete berries to generate
Result:  $P_{\text{output}} = \{P_{o_i} \mid i = 1, \dots, n\}$  // incomplete berry point cloud sets
 $\hat{P}_C \leftarrow \text{Normalize}(\text{Translate}(P_C))$  // center to (0,0,0), range to [-0.5m,0.5m]
for  $i = 0 \rightarrow n$  do
    set  $m = \text{random}\{1, 2, 3, 4, 5\}$  // number of removal sphere regions
     $P_{o_i} \leftarrow \text{copy}(\hat{P}_C)$  // initialize one output
    for  $j = 0 \rightarrow m$  do
        set  $O = \{(x_o, y_o, z_o) \mid x, y, z \in \text{random}[-0.75, 0.75]\}$  // removal sphere center
        set  $R = \text{random}[0.25, 0.75]$  // removal sphere radius
        set  $M_S \leftarrow \text{sphere}(O, R)$  // generate removal sphere region
         $P_{o_i} = P_{o_i} - P_{o_i} \cap M_S$  // remove overlap within sphere region
    end
     $P_{\text{output}_i} \leftarrow P_{o_i}$  // add to output set
end

```

surface and avoid detail distortion. This paper combines the centroid-based sparse completion structure of PoinTr (Yu et al., 2021) with the diffusion-based dense completion structure of SnowFlakeNet (Xiang et al., 2021). Based on the characteristics of the berry, it has two parts: contour completion and surface completion. The network structure is shown in Fig. 10.

(a) Centroid-based berry contour completion

The visible portion of the incomplete berry belongs to a part of a sphere or ellipsoid, characterized by a distinctly spherical surface. In addition, the point cloud is unstructured and dense, which makes feature extraction less effective. Based on the above analysis, this paper utilizes the centroid set of the visible part to infer the centroid set of

the invisible part. Specifically, GrapeCPNet uses centroid-based feature extraction in PoinTr (Yu et al., 2021) to accomplish berry point cloud contour completion. The process is described below:

Assume that the input to the network is represented by a set of incomplete berry point clouds, denoted as $P_{\text{partial}} = \{p_i \mid i = 1, \dots, m\}$, where $p_i = (x_i, y_i, z_i)$ denotes a point in 3D space and m denotes the number of points. The berry point cloud completion task aims to output the complete berry point set $P_{\text{complete}} = \{p_i \mid i = 1, \dots, n\}$, where n denotes the number of points of the output, and $n > m$, thereby it can represent the complete 3D morphology of the berry.

After the incomplete berry point cloud set $P_{\text{partial}} = \{p_i \mid i = 1, \dots, m\}$ was fed into the network, it was firstly down-sampled to a fixed number of points M (M is often less than m , in this paper, M is set to 2048) through the farthest point sampling (FPS) to obtain the center point set $P_{\text{partial-center}} = \{p_i \mid i = 1, \dots, M\}$ of the visible part of the berry. This set was used as a “representative” to avoid huge computational effort while characterizing the complete morphology of the visible part.

Then, the 3D convolutional neural networks DGCNN (Wang et al., 2019) and MLP (Tolstikhin et al., 2021) are used to obtain the local structural features of each point p_i in the center point set, as shown in Eq. (2).

$$F_i = F'_i + \varphi(p_i) \quad (2)$$

where, F'_i denotes the feature around p_i obtained by DGCNN, which is the local feature of the point in the point cloud; and $\varphi(p_i)$ denotes the position information of p_i obtained by MLP, which is the global feature of the point.

(b) Transformer network

Since the $\varphi = \{\varphi(p_i) \mid i = 1, \dots, M\}$ contains the relative position information of each point, it can be used as the position embedding and the serialized feature $F = \{F_i \mid i = 1, \dots, M\}$ was fed into the geometrically-aware Transformer (Vaswani et al., 2017) based encoder. The process is Eq. (3).

$$V = T_e(F) \quad (3)$$

where, T_e denotes the encoder of the Transformer and $V = \{V_i \mid i = 1, \dots, M\}$ denotes the encoded feature. Next, the predicted centroid point set of the missing part was output by the decoder, and the process as Eq. (4).

$$P_{\text{pred-center}} = T_d(V) \quad (4)$$

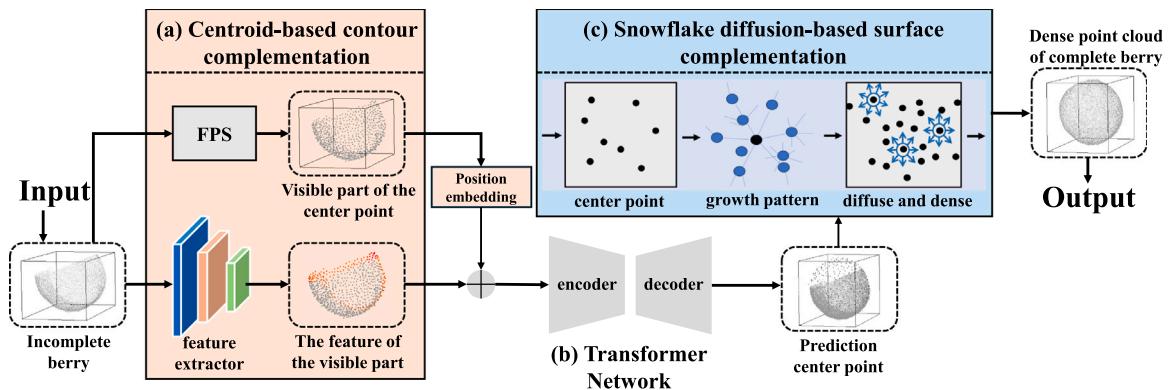


Fig. 10. GrapeCPNet structure. (a) Centroid-based completion and (b) Transformer network for contour completion of incomplete berry, using sparse point cloud to determine the entire shape; (c) Snowflake diffusion-based completion for dense surface completion to obtain final prediction.

where, T_d denotes the decoder of the Transformer, $P_{pred-center} = \{p_i \mid i = 1, \dots, N\}$ denotes the predicted centroid point set, and N denotes the number of points in the set. Through the above steps, the set of center points of the complete berry can be initially obtained, which was the result of sparse contour completion.

(c) Snowflake diffusion-based berry surface completion

The results of the contour completion can determine the overall 3D morphology. However, the low density of this sparse point cloud requires the model to accurately capture detailed surface geometry variations between points to avoid distortion. Therefore, this paper utilizes the snowflake diffusion-based generation method to regard each point in the contour completion as the center point of the local region. Its diffusion is based on the centroid points and generates further at the generated location, densification by point-to-point splitting. Specifically, as follows:

First, GrapeCPNet sets each point p of the sparse complement $P_{pred-center} = \{p_i \mid i = 1, \dots, N\}$ as a “seed point”. Then, $P_{pred-center}$ was fed into multiple SPD (Xiang et al., 2021) connected serially of SnowFlakeNet, which are used to realize point splitting and generation. The SPD expands from the center point in a tree-like structure, capturing the detailed geometric of the local region. This method effectively represents the surface geometric features composed between the points in the localized region of the berry. Additionally, the iteration of multiple point splitting steps delivers information between consecutive steps. This approach effectively characterizes the curvature changes on the surface of the berry. The process is Eq. (5).

$$P_{i+1} = SPD(P_i) \quad (5)$$

where, P_i denotes the input point cloud of the i^{th} SPD, P_{i+1} denotes the output point cloud of the $(i+1)^{\text{th}}$ SPD, and $P_0 = P_{pred-center}$. In this paper, when $i = 2$ (an empirical value), the final predicted dense point cloud on the surface of the complete berry $P_3 = P_{complete}$ was output and $P_{complete} = \{p_i \mid i = 1, \dots, n\}$.

Through the above steps, the complete 3D point cloud model of the berry can be obtained after contour completion and surface completion of GrapeCPNet.

2.4. Phenotypic traits calculation and validation

To describe the 3D morphology of berries, this paper used ellipsoidal surface fitting with three axes, for the completed berry point cloud (Fig. 11).

Suppose the point cloud of a complete berry is denoted as $P_{complete} = \{p_i \mid i = 1, \dots, n\}$, where $p_i = (x_i, y_i, z_i)$ denotes a point in the 3D space, and n denotes the number of points. In this paper, it is fitted to an ellipsoid surface with three unequal long axes by the least squares method, and the function is shown in Eq. (6).

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} + \frac{(z - z_0)^2}{c^2} = 1 \quad (6)$$

where, (x, y, z) denotes a point on the ellipsoid surface, (x_0, y_0, z_0) denotes the coordinates of the ellipsoid center point, and a, b, c are the half-long, half-middle, and half-short axes, respectively (assuming that $a > b > c > 0$).

Thus, the values of a, b, c can represent the berry radius parameters. At the same time, the volume of the berry can be calculated as shown in Eq. (7).

$$V = \frac{4}{3}\pi abc \quad (7)$$

To validate the accuracy of extracted morphological traits, the 12 validation bunches with 108 berry samples were selected. The berries were first scanned while still attached to bunch (Fig. 3a2). After instance segmentation, these scans of each berry are incomplete caused by occlusion. We then applied our proposed self-supervised completion method to get completion predictions. Finally get ellipsoidal surface fitting results as prediction values. For ground truth measurement, we separated berries from the bunch to get complete scans of corresponding berry (Fig. 3b1). Then applied the ellipsoidal surface fitting to these completely scans.

2.5. Evaluation metrics

In this paper, the metrics for the bunch segmentation, berry completion, berry counting, and the radius and volume calculation were used for evaluating the performance of our analysis pipeline.

2.5.1. Berry instance segmentation

We used average precision (AP) as the evaluation metric for berry instance segmentation, similar to the original paper of SoftGroup (Vu et al., 2022). The average APs with IoU (Intersection over Union) ranging from 50% to 95% in steps of 5% were used and noted as mAP@[.5:.95].

2.5.2. Berry self-supervised completion

We used the chamfer distance (CD-distance, d_{CD}) and F-Score to measure the effect of berry completion, similar to the original paper of Yu et al. (2021). For each sample, d_{CD} between the predicted point set P and the true point set G was calculated as shown in Eq. (8)

$$d_{CD}(P, G) = \frac{1}{|P|} \sum_{p \in P} \min_{g \in G} \|p - g\|_2^2 + \frac{1}{|G|} \sum_{g \in G} \min_{p \in P} \|g - p\|_2^2 \quad (8)$$

where, $|P|$ denotes the number of points of P , $|G|$ denotes the number of points of G , and $\|p - g\|$ denotes the distance between p and g , the L1 norm was used to calculate the closest distance between two points.

The F-Score measured the similarity between P and G by the harmonic mean between Precision and Recall (Eq. (9)). Precision (Eq. (10)) is the percentage of predicted points within a certain distance from the truth point, i.e. the accuracy of the prediction. Recall (Eq. (11)) is the

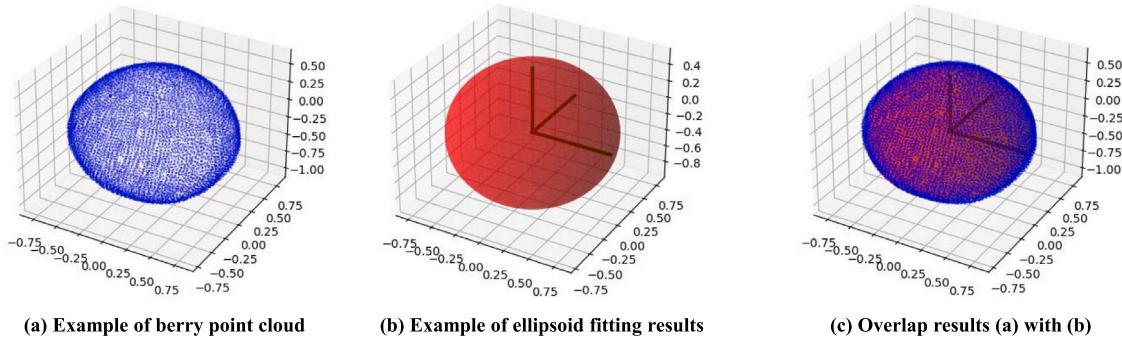


Fig. 11. Example of berry point cloud fitted to its ellipsoid surface.

percentage of ground truth points within a certain distance from the predicted point, i.e. the completeness of the prediction.

$$F\text{-Score}(d) = \frac{2 \times \text{Precision}(d) \times \text{Recall}(d)}{\text{Precision}(d) + \text{Recall}(d)} \quad (9)$$

$$\text{Precision}(d) = \frac{1}{|P|} \sum_{p \in P} \left[\min_{g \in G} \|p - g\| < d \right] \quad (10)$$

$$\text{Recall}(d) = \frac{1}{|G|} \sum_{g \in G} \left[\min_{p \in P} \|g - p\| < d \right] \quad (11)$$

The distance threshold d was set to 0.01 m. Given that the point cloud was initially normalized to the range [-0.5 m, 0.5 m]. From the above Equation, the higher the F-Score the better the match between the two point sets.

To verify the efficiency of GrapeCPNet, we compared it to other point cloud completion networks: PoinTr (Yu et al., 2021) SnowFlakeNet (Xiang et al., 2021), TopNet (Tchapmi et al., 2019), and GR-Net (Xie et al., 2020).

2.5.3. Phenotypic traits calculation

To evaluate the efficiency of the number of berry statistics and the calculation of berry radius and volume, the MAE (Mean Absolute Error) and the coefficient of determination R^2 (R-Square) are used to calculate the error between the prediction value and the true value of the sample.

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (\bar{y}_i - y_i)^2} \quad (13)$$

where, m is the number of samples, y_i is the true value of the i th sample, \hat{y}_i is the prediction value, and \bar{y}_i represents the mean of all true values. The lower the MAE, the closer the prediction value is to the true value. The values of R^2 range between 0 and 1, and the closer it is to 1 means the better the fit with the predicted values.

3. Results

To validate the efficiency of our method, the experiments were conducted on each component separately, including the result of bunch point cloud segmentation, the result of berry point cloud completion based on self-supervised training, and the result of obtaining the phenotypic traits.

3.1. Berry instance segmentation results

Our controlled indoor environment (Fig. 3) with high-overlap scanning enabled high-quality grape point cloud reconstruction, featuring clear boundaries and minimal noise. Some reconstructed results are shown in Figs. 4, 5, 6, and 12. These high-quality input data reduced the challenges in annotating instance berry segmentation training data.

They also simplified the segmentation model training by minimizing concerns about noise and unclear boundaries.

The segmentation task divides the grape bunch into two parts: berries and stems. The mAP@[.5:.95] values of berry and stem class are 0.995 and 0.941, respectively, which showed a high accuracy of the berry segmentation from the bunch. Since mAP@[.5:.95] is a strict evaluation metric, it represents the mean values of Average Precision (AP) scores calculated from AP50 to AP95 at 5% IoU intervals. Compared to other studies, Ni et al. (2021) reported that their proposed segmentation network achieved a mAP@[.5:.95] of 0.877 on their testing dataset, while Du and Liu (2023) achieved a mAP[.5:.95] of 0.657 and an AP50 of 0.952. In our study, the AP50 value exceeded 0.999, with both recall and precision values at this level also over 0.999. When compared to traditional berry segmentation algorithms, Rose et al. (2016) reported a mean recall of 0.776 and precision of 0.978; The results of four fresh grapes for bunch point cloud segmentation are visualized in Fig. 12. According that figure, we can also see an almost perfect instance berry segmentation results.

3.2. Berry self-supervised completion results

The berry point cloud completion experiments are divided into two parts according to the different phases of the network: the results of the self-supervised training (Test) and the results of the practical application (Inference). As shown in Table 3, GrapeCPNet achieves the best results under the self-supervised training method in both the training phase and the application phase, better than the comparative algorithms under the same condition of our grape dataset. Fig. 13 visualizes the shape completion result of GrapeCPNet on the incomplete berry segmented from the bunch in the application phase.

The results of the “Inference” experiments are worse than those of the “Test” experiments for two main reasons. First, there are differences between the dataset of the training phase and the application phase of self-supervised training, resulting in a drop in the accuracy. In addition, when evaluating the actual application efficiency of the model, the true value of the reconstructed complete point cloud of a single berry and the prediction value of the complemented point cloud suffer from the drop in the accuracy after aligning them. Through the above results, the experiments of berry point cloud completion based on self-supervised training reflected the efficacy of GrapeCPNet, meanwhile, this paper further validated the efficacy of the 3D phenotypic traits of grape by the proposed method.

Since GrapeCPNet combines features from two existing models (PoinTr and SnowFlakeNet) rather than introducing a completely new architecture, ablation studies may not align with our research focus. However, we conducted comparative experiments by treating each network as an independent module. As shown in Table 3, we evaluated three approaches: (1) PoinTr alone, (2) SnowFlakeNet alone, and (3) our combined GrapeCPNet. The results show that GrapeCPNet achieves better performance than either individual network.

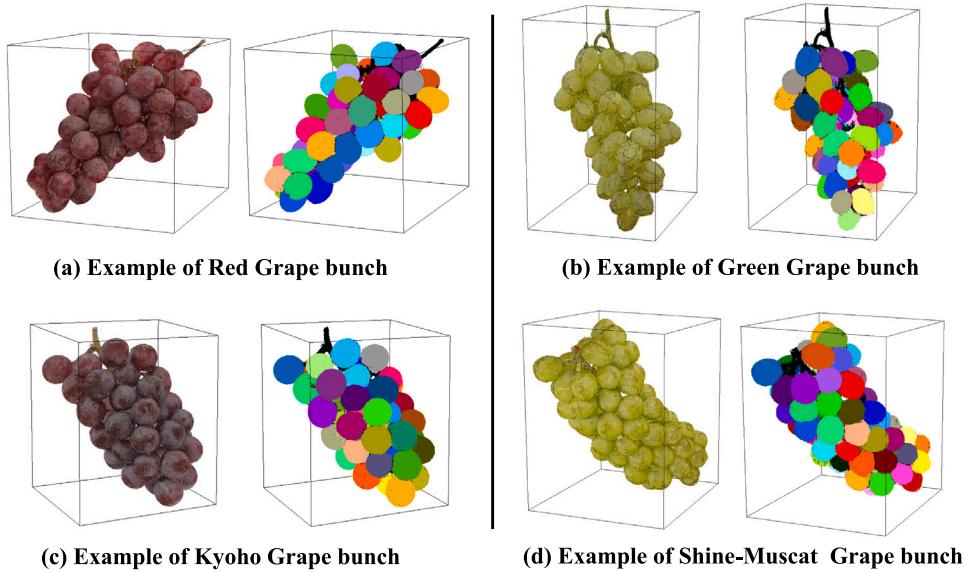


Fig. 12. Segmentation results of the bunch of four species of fresh grape, the stems point cloud and the noise point cloud are set to black, and the different color point sets indicate the point clouds of different berry instances obtained.

Table 3
Comparing proposed method with different completion networks for berry point cloud. Best performing results are visualized in bold.

Model	DataSet	d_{CD} (mm) ↓	Precision ↑	Recall ↑	F-score ↑
PoinTr (Yu et al., 2021)	Test	9.036	0.721	0.705	0.713
	Inference	10.542	0.623	0.601	0.612
SnowFlakeNet (Xiang et al., 2021)	Test	11.235	0.512	0.534	0.521
	Inference	12.384	0.415	0.423	0.419
TopNet (Tchapmi et al., 2019)	Test	23.019	0.201	0.200	0.201
	Inference	31.880	0.155	0.154	0.154
GRNet (Xie et al., 2020)	Test	19.240	0.402	0.392	0.397
	Inference	21.250	0.365	0.363	0.364
GrapeCPNet (Ours)	Test	8.315	0.729	0.772	0.751
	Inference	9.158	0.685	0.696	0.689

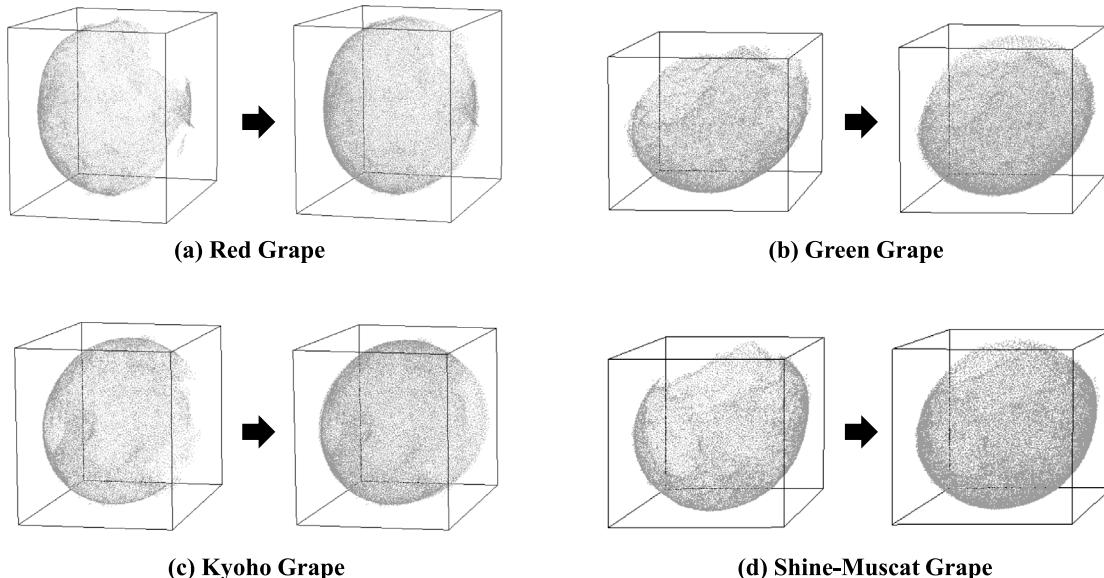


Fig. 13. Example of four fresh grape berry point cloud completion results: in subfigures a-d, the left side shows incomplete berry while the right side shows the completion results using proposed GrapeCPNet.

To facilitate future research, we will release our source code and training data at <https://github.com/I3-Laboratory/GrapeCPNet>. This

will enable interested researchers to examine how preprocessing parameters influence the network's performance.

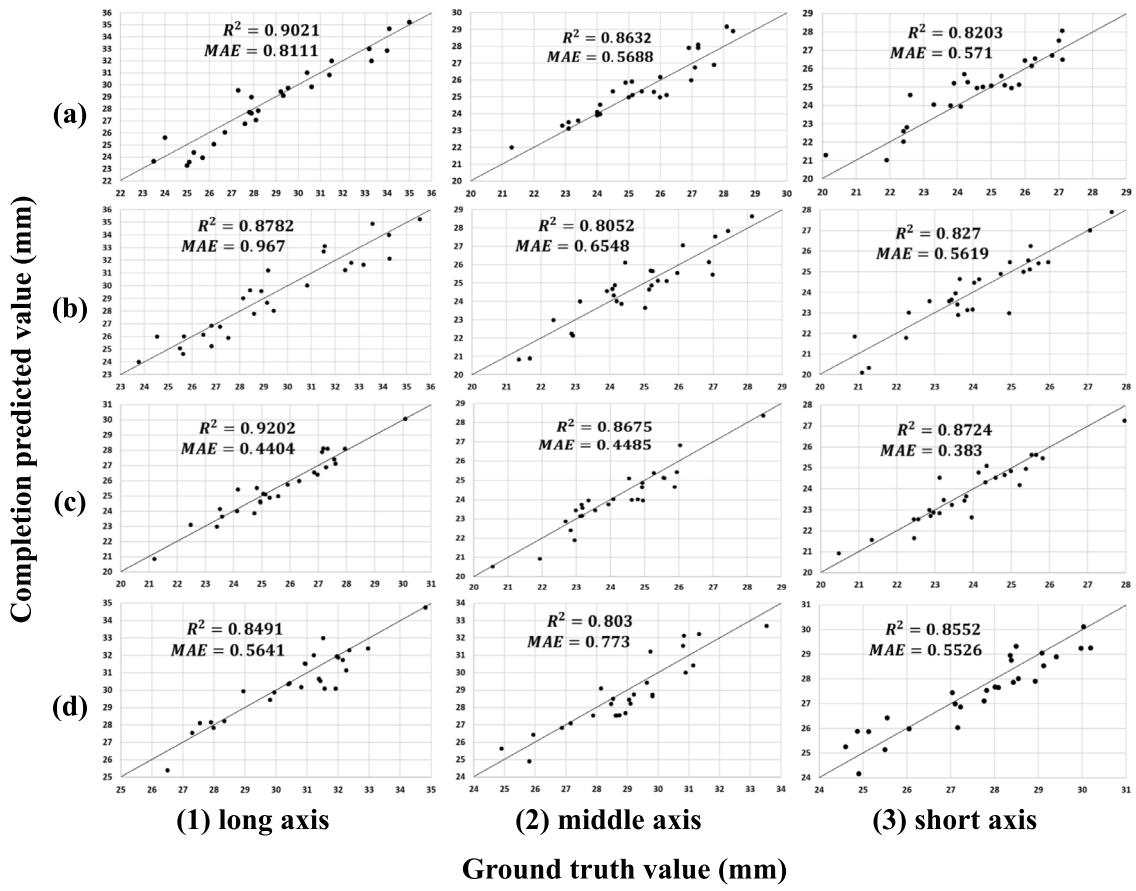


Fig. 14. Experimental results of axis length calculation for four species of fresh grape. From (a)–(d): Red Grape, Green Grape, Kyoho Grape, and Shine-Muscat Grape, respectively. The ground truth values are obtained by completely scanning berries that were separated from the bunch (Fig. 3b1); The predicted values are obtained from our proposed self-supervised completion method, which addresses incomplete scans of berries that occur due to occlusion while they were still attached to the bunch (Fig. 3a2).

3.3. Phenotypic traits calculation results

The radius parameters were calculated by fitting the berry to the surface of an ellipsoid. From Fig. 14, it can be seen that the proposed method achieves an average R^2 of 0.86 and average MAE of 0.608 mm for the three axis lengths of the four species of grape. This result can be dissected as follows:

(1) The different species: Red Grape, Green Grape, Kyoho Grape, and Shine-Muscat Grape axis length prediction reached an average R^2 of 0.86, 0.84, 0.89, and 0.84, respectively. Among them, Kyoho Grape achieved the best results due to its spherical berries that are not too densely clustered and the lower difficulty in berry completion and morphology fitting. Secondly, Red Grape and Green Grape both present a typical ellipsoid shape, but the Red is less precise than the Green due to its larger ratio of long axis to short axis, which provides a more complex presentation. As also reported by Schneider et al. (2020), the mean length-to-width ratio of Green Grape is 21.18 mm to 18.75 mm, while for Red Grape, it is 24.11 mm to 18.41 mm. Shine-Muscat Grape had the lowest precision, which was due to the fact that its berry grows tightly and occlude each other heavily, making it harder to accurately complete its shape.

(2) The different axis: the average R^2 of the four species of grape were 0.89, 0.83, and 0.84 for the long, middle, and short axis, respectively. It can be noted that the prediction accuracy of the long axis was significantly higher than that of the others, in which the R^2 of Red Grape and Kyoho Grape reached 0.90 and 0.92, respectively, achieving a high accuracy. This was due to the fact that in berry completion and fitting, the long axis is the easiest to determine, with less fitting errors compared to the other two axes.

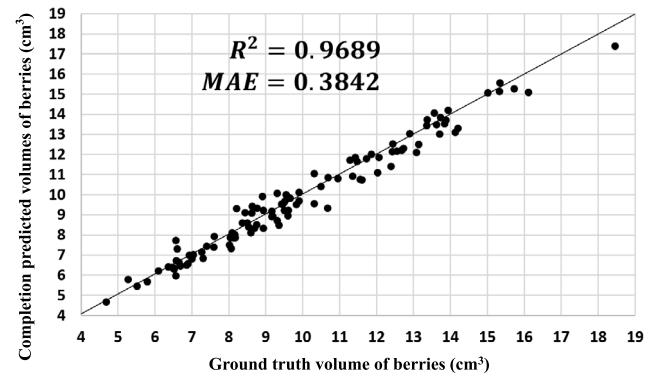


Fig. 15. Experimental results of berry volume calculation. The ground truth values are obtained by completely scanning berries that were separated from the bunch (Fig. 3b1); The predicted values are obtained from our proposed self-supervised completion method, which addresses incomplete scans of berries that occur due to occlusion while they were still attached to the bunch (Fig. 3a2).

In addition, based on the radius prediction results, the volume of all 108 berry samples was calculated in this paper, and the experimental results are shown in Fig. 15. For the berry volume calculation of the four species, the R^2 was 0.97, and the MAE was 0.3842 cm³.

The experimental results showed that the proposed method achieved both high accuracy in the calculation of the radius and volumes of different grape species. In our study, we achieved an average R^2 of 0.86 and an average MAE of 0.608 mm across the three axis of berry, as well as an R^2 of 0.9689 and an MAE of 0.3842 cm³ for berry

volumes. By comparison, Rist et al. (2018, Fig. 4) reported R^2 values of 0.54, 0.85, and 0.83 for berry width, length, and volume, respectively. Botturi et al. (2023) reported a best-case MAE of 21.7 mm for berry ratio estimation using deep learning network on 2D images. Although the method by Ni et al. (2021) achieved a lower errors for volume estimation (RMSE of 0.292 cm³), it relies on sphere fitting that are primarily applicable to clustered, round fruits. In contrast, our proposed method not only accounts for occluded areas but is also adaptable to fruits of varying shapes, if appropriate paired training data is available.

4. Discussion

In this paper, an automated deep learning pipeline for acquisition of 3D phenotypic traits of grape bunch was presented to address the problems of high labor cost, poor accuracy, and inability to realize end-to-end analysis. As a major contribution, to address the problem of incomplete point clouds of berries due to occlusion, we introduced GrapeCPNet, a berry point cloud completion model based on the self-supervised training method.

Grape berry counting and size estimation plays a critical role in vineyard management and yield prediction. By using common digital cameras alongside deep learning networks, several studies have achieved automatic berry detection and instance segmentation (Jadhav and Barbole, 2021; Chen et al., 2023), and then forward the analyses of geometry traits and yield estimation (Quiñones et al., 2025). However, because of the dense and compact arrangement of berries within a grape bunch, some portions are often occluded and not visible when relying solely on single 2D images. To reduce the effects of occlusion, Du and Liu (2023) reported accurate berry counting estimations by correlating image-based detections with ground measurements. However, applying this method to other vineyards requires extensive, labor-intensive ground measurements. Similarly, Botturi et al. (2023) attempted to extract feature maps, such as berry radii, directly from deep networks applied to 2D images. Despite this approach, it still suffers from information loss and resulted in a high MAE of 21.7 mm. Applying multi-view images for 3D reconstruction can significantly reduce the extent of occlusion caused by viewpoint limitations. But unlike 2D images, which have well-structured data in the form of organized pixel matrices, the unordered nature of points in a 3D point cloud introduces significant challenges. Meanwhile, the inner portion of berry is still invisible, only the surface of exposed berry surfaces can be scanned or reconstructed. This complicates the algorithm design for individual berry segmentation and accurate berry size estimation.

For individual berry segmentation, early studies relied on the circular characteristics of the visible portions of berries for segmentation. Building on the work of Schnabel et al. (2007), who introduced point cloud shape detection using RANSAC, researchers were able to cluster partial berries based on their rounded shapes (Schöler and Steinhage, 2015; Rist et al., 2018; Schneider et al., 2020). However, this feature-based approach performs well only in sparse grape arrangements and proves ineffective in more complex, dense conditions. The reliance on manually designed features for algorithm development has now reached its limitations. With advancements in machine learning, the focus has shifted to a data-driven era, where algorithms can learn complex feature patterns automatically. For instance, Rose et al. (2016) applied machine learning to distinguish features between grape berry clusters and background leaves automatically. Similarly, Ni et al. (2021) implemented a two-step approach by first using mature 2D instance segmentation techniques on 2D images and then projecting the resulting segmentation masks onto reconstructed 3D models to complete the challenging task of 3D instance segmentation. In another approach, Luo et al. (2023) Mask R-CNN for 2D image segmentation and projected the results onto RGB-D point clouds. The normal directions derived from the point clouds were then packed as an additional feature in the PointNet++ network for point cloud segmentation. In this study, we applied the state-of-the-art point cloud segmentation

network, SoftGroup (Vu et al., 2022) directly to the problem. The primary time-consuming work of our approach was the annotation of 3D data for instance segmentation. We acknowledge that manual annotation of 3D data on 2D displays may introduce some human errors, meaning even the ground truth segmentation annotations are not perfectly accurate. However, due to the high-quality 3D reconstruction with clear berry boundaries and minimal noise effects, we could train a highly accurate SoftGroup segmentation network, achieving an mAP@[.5:.95] of 0.995 and AP50 of 0.999. Compared to similar studies on grape berries, our proposed method demonstrated superior performance, achieving the best results among the reviewed methods (Section 3.1).

For berry completion using deep learning, the most challenging and time-consuming task is preparing the paired training datasets of complete and incomplete point clouds. In our recent work on potato tuber completion networks (Blok et al., 2025), a distinguishable pin was used as a reference point to align the complete tuber scans obtained via photogrammetry with the incomplete tuber scans captured using an RGB-D camera. A pin-guided point cloud matching algorithm was then applied to calculate the transformation matrix for pairing the datasets. Potato tubers, being single objects with clear textures and shape variation, presented fewer challenges when aligning data collected by two different types of sensors. However, grape berries pose a significantly more difficult problem due to their highly rounded shapes, weak texture features, and strong reflectance, which complicates the creation of paired data. As a result, the previous pin-guided point cloud pairing method used for potato tuber datasets is not effective for grape berries. Thus, we simulated random sphere cutting on the complete berry point clouds to generate thousands of paired incomplete point clouds, as described in Section 2.3.1. As demonstrated in Section 3.2, these automatically generated paired training datasets were successfully used to train several completion models (Table 3).

Although our recent work proposed a completion network for partial potato tubers (Blok et al., 2025), the purposes and input data types differ. This grape completion study uses the same sensor type, whereas the potato network focuses on fusing different sensors with an emphasis on processing speed and real-time performance on fast-moving conveyors. Specifically, the tuber completion study processes incomplete point clouds from RGB-D sensors and complete point clouds from SfM, which show significant differences in shape and texture. Additionally, the grape completion targets smooth ellipsoid berry surfaces, while potato tubers have more irregular and elongated shapes. For future work, it may be interesting to compare the two networks on inverted crops. We also found that some researchers (Pan et al., 2023; Magistri et al., 2024) have improved the completion effect by encoding a prior knowledge of the geometry of the object into the weights of the neural network. Considering the agricultural plant completion task, our proposed GrapeCPNet, which combines centroid-based sparse completion with diffusion-based dense completion, outperformed other networks in terms of performance. We found that GrapeCPNet performed a little bit worse in the case of strange berry shapes, such as drop-shaped berries, particularly when berries were severely occluded. We subsequently consider utilizing the prior knowledge of berry geometry to achieve more detailed morphological feature extraction and representation to improve the quality of completion.

This study has another limitation: it focuses only on four table grapes, which are currently common and available in Chinese markets. Although we examined a wide range of cluster densities (from sparse to very dense) and berry sizes (20–35 mm axis length, as shown in Table 1 and Fig. 14), the study did not include wine grapes, which account for over 60% of U.S. grape production and typically have smaller berries and denser clusters. As Manso-Martínez et al. (2021) and Melo et al. (2015) report, wine grape berries are often smaller than 13–15 mm, a size range not covered in our dataset. From a deep learning perspective, our model architecture should be able to detect smaller wine grapes with appropriate training data, as the basic detection principles are

similar across grape varieties. Future work should involve collaboration with U.S. researchers to collect wine grape datasets and assess whether models trained on table grapes can generalize well to smaller wine grape berries without additional training.

The design philosophy focuses on ensuring input quality and assigning specialized networks to their respective tasks, which provides the most efficient approach. Rather than developing complex algorithms to process low-quality inputs with unclear berry boundaries and noisy 3D models, improving input quality simplifies the problem and enhances the results. Similarly, assigning each specialized network to its specific task is essential. In this study, the segmentation network reliably performs instance separation, while the completion network specializes in partial point cloud completion. Although the automatic training data generation strategy can be adjusted to produce poor segmentation outputs — for example, by adding noise or partially cutting neighboring berries — it remains difficult to cover all potential error patterns to improve robustness. Additionally, training with too many low-quality samples may negatively affect network performance. However, this design philosophy also faces challenges in practical applications. For example, this study focused on a single berry bunch in a controlled environment. The noise in the data is also related to the effectiveness of the method proposed in this paper. For practical applications in vineyards, collecting the same point cloud quality as that achieved in controlled indoor environments can be challenging, and multi-view data collection may not be as efficient. How to effectively implement the method in practical application scenarios is another issue to be considered in future research. To increasing data collection efficiency, devices such as RGB-D cameras (Blok et al., 2025), fisheye cameras (Tamura et al., 2025), or even 360 cameras can be used. Moreover, as in previous studies (Rose et al., 2016; Luo et al., 2023), investigating methods to segment individual grape bunches directly from vineyard scans using deep learning techniques should be an area of future exploration. Additionally, Schöler and Steinlage (2015, Fig. 9) successfully demonstrated the construction of virtual 3D models of grape bunches, including the hidden branching architecture. By collecting time-series growth data, the branching structure could potentially be captured during early stages of thinning, as suggested by Du and Liu (2023). Lastly, with the development of lidar scanning techniques, some companies also provide low-cost products to improve 3D scanning results in the field, such as XGrids (XGRIDS LIMITED, <https://www.xgrids.com>). Such data could enable studies on growth prediction for optimizing harvesting times and assessments of branching structure deviations within crowded berry bunches. Finally, applying the latest procedural computer graphics (CG) modeling techniques or functional-structural plant models represents another promising approach to advancing practical applications in this field.

Through the method proposed in this paper, the automated acquisition of 3D phenotypic traits of grape can be realized, which is of great significance in the field of agricultural production and scientific research. First, there are large differences in berry morphology among different species of grape, and accurate reconstruction and analysis at the berry-level can help researchers analyze character differences more accurately. This plays an important role in plant breeding, cultivation management, and yield prediction. Secondly, the point cloud data analysis technology based on 3D reconstruction has a wide application of field crop analysis. Through the designed data acquisition method and related technologies, combined with the actual field crop growing environment, the growth process can be monitored and analyzed in real-time. Further, it can help understanding the phenotypic change of crops in natural growing environments, providing better and objective measures for crop production and quality control.

5. Conclusion

In this study, we proposed a deep-learning-based pipeline for processing 3D phenotyping grape bunches reconstructed by photogrammetry. The major contribution is the self-supervised completion network,

named GrapeCPNet, designed to address the problem of missing parts in berries occluded within grape bunches with few manual annotations. We utilized state-of-the-art, data-driven deep learning approaches to address these challenges. First, the SoftGroup network, utilizing segmentation annotations from 8 grape bunches, was used to train a data-driven berry instance segmentation model. The proposed GrapeCPNet also achieved better performance than existing completion networks, with Precision of 0.729, Recall of 0.772 and F-score of 0.751. The ellipsoid surface fitting, used to calculate three axis lengths and volumes, achieved an average R^2 of 0.86 and 0.97 respectively, when comparing completed berries with their ground truth. This study clearly demonstrates the potential of using data-drive strategies with minimal annotation effort to address the 3D occlusion challenge in agricultural applications. Although this network has only been tested on grape bunches collected in a controlled indoor environment, the proposed deep learning pipelines and training data generation approach have broad potential applications for in-field vineyards and even other crops in the future.

CRediT authorship contribution statement

Wenli Zhang: Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Chao Zheng:** Writing – original draft, Validation, Software, Resources, Methodology. **Chenhuizi Wang:** Software, Data curation. **Pieter M. Blok:** Writing – review & editing, Supervision, Methodology. **Haizhou Wang:** Writing – review & editing, Software, Methodology. **Wei Guo:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to check grammar, spelling, and improve fluency. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This study was partially supported by the National Natural Science Foundation of China (NSFC) Program 62276009, the Japan Science and Technology Agency (JST) AIP Acceleration Research JPMJCR21U3, and the Hokkaido Sarabetsu Village “Endowed Chair for Field Phenomics” project.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Wenli Zhang reports financial support was provided by National Natural Science Foundation of China. Wei Guo reports financial support was provided by Japan Science and Technology Agency, and Hokkaido Sarabetsu Village. The other 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.

Data availability

The data and code in manuscript is shared at <https://github.com/I3-Laboratory/GrapeCPNet>.

References

- Alston, J.M., Sambucci, O., 2019. Grapes in the world economy. In: Cantu, D., Walker, M.A. (Eds.), *The Grape Genome*. Springer International Publishing, Cham, pp. 1–24. http://dx.doi.org/10.1007/978-3-030-18601-2_1.
- Blok, P.M., Magistri, F., Stachniss, C., Wang, H., Burridge, J., Guo, W., 2025. High-throughput 3D shape completion of potato tubers on a harvester. *Comput. Electron. Agric.* 228, 109673. <http://dx.doi.org/10.1016/j.compag.2024.109673>.
- Botturi, D., Gnutti, A., Nuzzi, C., Lanza, B., Pasinetti, S., 2023. STEWIE: estimating grape berries number and radius from images using a weakly supervised neural network. In: 2023 IEEE International Workshop on Metrology for Agriculture and Forestry. MetroAgriFor, IEEE, Pisa, Italy, pp. 277–282. <http://dx.doi.org/10.1109/MetroAgriFor58484.2023.10424094>.
- Chen, Y., Li, X., Jia, M., Li, J., Hu, T., Luo, J., 2023. Instance segmentation and number counting of grape berry images based on deep learning. *Appl. Sci.* 13 (11), 6751. <http://dx.doi.org/10.3390/app13116751>.
- Du, W., Liu, P., 2023. Instance segmentation and berry counting of table grape before thinning based on AS-Swint. *Plant Phenomics* 5, 0085. <http://dx.doi.org/10.34133/plantphenomics.0085>.
- Ge, Y., Xiong, Y., From, P.J., 2020. Symmetry-based 3D shape completion for fruit localisation for harvesting robots. *Biosyst. Eng.* 197, 188–202. <http://dx.doi.org/10.1016/j.biosystemseng.2020.07.003>.
- Jadhav, P., Barbole, D., 2021. Comparative analysis of deep learning architectures for grape cluster instance segmentation. *Inf. Technol. Ind.* 9 (1), 344–352. <http://dx.doi.org/10.17762/iti.v9i1.138>.
- Luo, L., Yin, W., Ning, Z., Wang, J., Wei, H., Chen, W., Lu, Q., 2022. In-field pose estimation of grape clusters with combined point cloud segmentation and geometric analysis. *Comput. Electron. Agric.* 200, 107197. <http://dx.doi.org/10.1016/j.compag.2022.107197>.
- Luo, J., Zhang, D., Yi, T., 2023. 3D semantic segmentation for grape bunch point cloud based on feature enhancement. In: 2023 IEEE International Conference on Robotics and Biomimetics. ROBIO, IEEE, Koh Samui, Thailand, pp. 1–6. <http://dx.doi.org/10.1109/ROBIO58561.2023.10354793>.
- Magistri, F., Marcuzzi, R., Marks, E., Sodano, M., Behley, J., Stachniss, C., 2024. Efficient and accurate transformer-based 3D shape completion and reconstruction of fruits for agricultural robots. In: 2024 IEEE International Conference on Robotics and Automation. ICRA, IEEE, Yokohama, Japan, pp. 8657–8663. <http://dx.doi.org/10.1109/ICRA57147.2024.10611717>.
- Magistri, F., Marks, E., Nagulavancha, S., Vizzo, I., Labe, T., Behley, J., Halstead, M., McCool, C., Stachniss, C., 2022. Contrastive 3D shape completion and reconstruction for agricultural robots using RGB-D frames. *IEEE Robot. Autom. Lett.* 7 (4), 10120–10127. <http://dx.doi.org/10.1109/LRA.2022.3193239>.
- Manso-Martínez, C., Sáenz-Navajas, M.P., Menéndez, C.M., Hernández, M.M., 2021. Wine quality and berry size: A case study with Tempranillo Tinto progenies. *J. Sci. Food Agric.* 101 (9), 3952–3960. <http://dx.doi.org/10.1002/jsfa.11035>.
- Marangoz, S., Zaenker, T., Menon, R., Bennewitz, M., 2022. Fruit mapping with shape completion for autonomous crop monitoring. In: 2022 IEEE 18th International Conference on Automation Science and Engineering. CASE, IEEE, Mexico City, Mexico, pp. 471–476. <http://dx.doi.org/10.1109/CASE49997.2022.9926466>.
- Melo, M., Schultz, H., Volschenk, C., Hunter, J., 2015. Berry size variation of vitis vinifera L. Cv. Syrah: Morphological dimensions, berry composition and wine quality. *South Afr. J. Enol. Vitic.* 36 (1), <http://dx.doi.org/10.21548/36-1-931>.
- Ni, X., Li, C., Jiang, H., Takeda, F., 2021. Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits. *ISPRS J. Photogramm. Remote Sens.* 171, 297–309. <http://dx.doi.org/10.1016/j.isprsjprs.2020.11.010>.
- Pan, Y., Magistri, F., Läbe, T., Marks, E., Smitt, C., McCool, C., Behley, J., Stachniss, C., 2023. Panoptic mapping with fruit completion and pose estimation for horticultural robots. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, Detroit, MI, USA, pp. 4226–4233. <http://dx.doi.org/10.1109/IROS55552.2023.10342067>.
- Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S., 2019. DeepSDF: Learning continuous signed distance functions for shape representation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, Long Beach, CA, USA, pp. 165–174. <http://dx.doi.org/10.1109/CVPR.2019.00025>.
- Quiñones, R., Banu, S.M., Gültepe, E., 2025. GCNet: A deep learning framework for enhanced grape cluster segmentation and yield estimation incorporating occluded grape detection with a correction factor for indoor experimentation. *J. Imaging* 11 (2), 34. <http://dx.doi.org/10.3390/jimaging11020034>.
- Rist, F., Herzog, K., Mack, J., Richter, R., Steinhage, V., Töpfer, R., 2018. High-precision phenotyping of grape bunch architecture using fast 3D sensor and automation. *Sensors* 18 (3), 763. <http://dx.doi.org/10.3390/s18030763>.
- Rose, J., Kicherer, A., Wieland, M., Klingbeil, L., Töpfer, R., Kuhlmann, H., 2016. Towards automated large-scale 3D phenotyping of vineyards under field conditions. *Sensors* 16 (12), 2136. <http://dx.doi.org/10.3390/s16122136>.
- Schnabel, R., Wahl, R., Klein, R., 2007. Efficient RANSAC for point-cloud shape detection. *Comput. Graph. Forum* 26 (2), 214–226. <http://dx.doi.org/10.1111/j.1467-8659.2007.01016.x>.
- Schneider, T., Paulus, G., Anders, K.-H., 2020. Towards predicting vine yield: Conceptualization of 3D grape models and derivation of reliable physical and morphological parameters. *GI Forum* 1, 73–88. http://dx.doi.org/10.1553/giscience2020_01_s73.
- Schöler, F., Steinhage, V., 2015. Automated 3D reconstruction of grape cluster architecture from sensor data for efficient phenotyping. *Comput. Electron. Agric.* 114, 163–177. <http://dx.doi.org/10.1016/j.compag.2015.04.001>.
- Tamura, Y., Utsumi, Y., Miwa, Y., Iwamura, M., Kise, K., 2025. Unsupervised monocular depth estimation with omnidirectional camera for 3D reconstruction of grape berries in the wild. *PLoS One* 20 (2), e0317359. <http://dx.doi.org/10.1371/journal.pone.0317359>.
- Tchapmi, L.P., Kosaraju, V., Rezatofighi, H., Reid, I., Savarese, S., 2019. TopNet: Structural point cloud decoder. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 383–392.
- Tesema, K.W., Hill, L., Jones, M.W., Ahmad, M.I., Tam, G.K., 2024. Point cloud completion: A survey. *IEEE Trans. Vis. Comput. Graphics* 30 (10), 6880–6899. <http://dx.doi.org/10.1109/TVCG.2023.3344935>.
- Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., Dosovitskiy, A., 2021. MLP-mixer: An all-MLP architecture for vision. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (Eds.), *Advances in Neural Information Processing Systems*. vol. 34, Curran Associates, Inc., pp. 24261–24272.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*. vol. 30, Curran Associates, Inc., pp. 1–11.
- Vu, T., Kim, K., Luu, T.M., Nguyen, T., Yoo, C.D., 2022. SoftGroup for 3D instance segmentation on point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2708–2717. <http://dx.doi.org/10.48550/arXiv.2203.01509>.
- Wang, H., Liu, Q., Yue, X., Lasenby, J., Kusner, M.J., 2021. Unsupervised point cloud pre-training via occlusion completion. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, pp. 9782–9792. <http://dx.doi.org/10.48550/arXiv.2010.01089>.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph.* 38 (5), 1–12. <http://dx.doi.org/10.1145/3326362>.
- Xiang, P., Wen, X., Liu, Y.-S., Cao, Y.-P., Wan, P., Zheng, W., Han, Z., 2021. SnowflakeNet: Point cloud completion by snowflake point deconvolution with skip-transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, pp. 5499–5509. <http://dx.doi.org/10.48550/arXiv.2108.04444>.
- Xie, H., Yao, H., Zhou, S., Mao, J., Zhang, S., Sun, W., 2020. GRNet: Gridding residual network for dense point cloud completion. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), *Computer Vision – ECCV 2020*. vol. 12354, Springer International Publishing, Cham, pp. 365–381. http://dx.doi.org/10.1007/978-3-03-58545-7_21.
- Yang, G., Huang, X., Hao, Z., Liu, M.-Y., Belongie, S., Hariharan, B., 2019. PointFlow: 3D point cloud generation with continuous normalizing flows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, pp. 4541–4550.
- Yu, X., Rao, Y., Wang, Z., Liu, Z., Lu, J., Zhou, J., 2021. PoinTr: Diverse point cloud completion with geometry-aware transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, pp. 12498–12507. <http://dx.doi.org/10.48550/arXiv.2108.08839>.
- Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Kuhlmann, H., Roscher, R., 2020. Counting of grapevines berries in images via semantic segmentation using convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* 164, 73–83. <http://dx.doi.org/10.1016/j.isprsjprs.2020.04.002>.