

RESEARCH ARTICLE

PanicleNeRF: Low-Cost, High-Precision In-Field Phenotyping of Rice Panicles with Smartphone

Xin Yang^{1,2}, Xuqi Lu^{1,2}, Pengyao Xie^{1,2}, Ziyue Guo^{1,2}, Hui Fang¹,
Haowei Fu³, Xiaochun Hu⁴, Zhenbiao Sun⁴, and Haiyan Cen^{1,2*}

¹College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China.

²Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China. ³Jiaxing Academy of Agricultural Science, Jiaxing, Zhejiang 314016, China. ⁴Yuan Longping High-Tech Agriculture Co. Ltd., Changsha 410128, China.

*Address correspondence to: hycen@zju.edu.cn

The rice panicle traits substantially influence grain yield, making them a primary target for rice phenotyping studies. However, most existing techniques are limited to controlled indoor environments and have difficulty in capturing the rice panicle traits under natural growth conditions. Here, we developed PanicleNeRF, a novel method that enables high-precision and low-cost reconstruction of rice panicle three-dimensional (3D) models in the field based on the video acquired by the smartphone. The proposed method combined the large model Segment Anything Model (SAM) and the small model You Only Look Once version 8 (YOLOv8) to achieve high-precision segmentation of rice panicle images. The neural radiance fields (NeRF) technique was then employed for 3D reconstruction using the images with 2D segmentation. Finally, the resulting point clouds are processed to successfully extract panicle traits. The results show that PanicleNeRF effectively addressed the 2D image segmentation task, achieving a mean F1 score of 86.9% and a mean Intersection over Union (IoU) of 79.8%, with nearly double the boundary overlap (BO) performance compared to YOLOv8. As for point cloud quality, PanicleNeRF significantly outperformed traditional SfM-MVS (structure-from-motion and multi-view stereo) methods, such as COLMAP and Metashape. The panicle length was then accurately extracted with the rRMSE of 2.94% for *indica* and 1.75% for *japonica* rice. The panicle volume estimated from 3D point clouds strongly correlated with the grain number ($R^2 = 0.85$ for *indica* and 0.82 for *japonica*) and grain mass (0.80 for *indica* and 0.76 for *japonica*). This method provides a low-cost solution for high-throughput in-field phenotyping of rice panicles, accelerating the efficiency of rice breeding.

Introduction

Rice (*Oryza sativa* L.) is a crucial crop globally, feeding more than half of the world's population [1,2]. Among the various factors influencing rice yield and quality, panicle traits such as panicle length, grain count, and grain mass are of great significance [3,4]. Panicle traits not only are closely associated with rice yield but also serve as essential indicators in rice breeding [5,6]. Therefore, approaches that enable high-precision measurements of rice panicle traits are crucial for accelerating rice breeding and improving overall crop productivity.

Numerous laboratory-based studies have been conducted on rice panicle phenotyping with machine vision technologies. Reported studies have employed different approaches, such as RGB scanning [7,8], x-ray computed tomography (CT) [9,10], structured light projection [11], hyperspectral imaging [12], and multi-view imaging [13], to extract several panicle traits, including grain number, grain dimensions (length, width, and perimeter), kernel dimensions (length, width, and perimeter), seed setting rate, and panicle health status based on spectral signatures. While these methods are capable of phenotyping of rice panicle

traits, they are limited to the controlled indoor environments and are still labor intensive as panicles have to be harvested from the field and manually processed, including spreading the branches and fixing them in place. Thereby, development of high-throughput in-field phenotyping of rice panicles would be highly desired for improving efficiency and monitoring panicle development over time [14,15].

In recent years, three-dimensional (3D) reconstruction methods have been increasingly employed for in-field plant phenotyping at different scales, offering more comprehensive and accurate information compared to 2D approaches. At the field scale, unmanned aerial vehicles (UAVs) equipped with RGB cameras [16] and airborne LiDAR [17] have been used to generate 3D point clouds of the entire field, which provide valuable insights into crop growth, canopy structure, and yield estimation. At the plot scale, terrestrial laser scanning (TLS) [18], depth cameras [19], and stereo vision systems [20] mounted on ground-based platforms have been utilized to reconstruct 3D models of plants within breeding plots, enabling the extraction of plant phenotypic traits and the assessment of genotypic differences. However, when it comes to the organ-level phenotyping, particularly for

Citation: Yang X, Lu X, Xie P, Guo Z, Fang H, Fu H, Hu X, Sun Z, Cen H. PanicleNeRF: Low-Cost, High-Precision In-Field Phenotyping of Rice Panicles with Smartphone. *Plant Phenomics* 2024;6:Article 0279. <https://doi.org/10.34133/plantphenomics.0279>

Submitted 2 July 2024

Revised 7 October 2024

Accepted 13 November 2024

Published 5 December 2024

Copyright © 2024 Xin Yang et al. Exclusive licensee Nanjing Agricultural University. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

rice panicles, the intricate structure and repetitive textures of rice panicles, combined with complex field conditions, present substantial challenges for accurate 3D reconstruction [21]. CT is impractical in field environments due to its bulky size and weight. Structured light cameras suffer from reduced measurement precision under strong outdoor lighting. Terrestrial laser scanners (TLS) and handheld laser scanners are limited by the need for coordination, with host computers hindering their field flexibility. Additionally, depth cameras can capture point cloud data quickly, but their low density is insufficient for detailed rice panicle analysis. With the fast development of 3D reconstruction algorithms, multi-view imaging using smartphones holds great promise in providing a low-cost and convenient solution for 3D reconstruction of plants at the organ level, making it accessible to almost everyone.

To process multi-view images captured by smartphones, various 3D reconstruction methods such as space carving, structure-from-motion [22] with multi-view stereo [23] (SfM-MVS), 3D gaussian splatting (3D-GS) [24], and neural radiance fields (NeRF) [25] can be employed. Space carving relies on controlled lighting and distinct textures, but uneven lighting and textures conditions in the field often lead to incomplete reconstructions. SfM-MVS, a traditional technique integrated into commercial software like COLMAP [26] and Metashape (Agisoft LLC, St. Petersburg, Russia), has been widely utilized to extract phenotypic parameters such as plant height, leaf length, and leaf width of soybean plants and sugar beets at two growth stages [27,28]. However, when applied to individual rice panicles, SfM-MVS often fails to generate complete and detailed point clouds due to limitations in feature matching and dense reconstruction algorithms. Similarly, 3D-GS, an emerging image rendering technology, enables models to generate high-quality images from specified viewpoints and produce 3D Gaussian point clouds, as demonstrated by developing a method based on 3D-GS to measure cotton phenotypic traits through instance segmentation [29]. Nonetheless, in the analysis of rice panicle phenotypes, the point clouds generated by 3D-GS remain too sparse because of the low density of Gaussian points. NeRF, a novel 3D reconstruction technique, has shown great potential in overcoming the limitations of traditional SfM-MVS methods by enabling the reconstruction of high-quality 3D models [30–32]. Recent advancements in NeRF have enhanced rendering quality and computational speed, making it effective for efficiently reconstructing detailed 3D models of rice panicles with reduced computational time. Hu et al. [33] explored the effectiveness of NeRF for 3D reconstruction in various crops, revealing the potential application of NeRF technology in the field of plant phenotyping. Zhang et al. [34] used NeRF technology to explore the robustness of the method at different scales in strawberry orchards, suggesting its high-performance rendering in real orchards. However, NeRF primarily focuses on scene reconstruction and lacks capabilities for object detection and segmentation. To accurately extract phenotypic traits of rice panicles, it is essential to remove background elements and segment the target panicles from the reconstructed 3D scene, necessitating the development of high-quality image segmentation methods that can be integrated with NeRF models for precise detection and segmentation in field conditions.

The overall goal of this study is to develop a novel in-field phenotyping method by combining NeRF with You Only Look Once version 8 (YOLOv8) [35] and Segment Anything Model (SAM) [36], called PanicleNeRF, to extract rice panicle traits in the field accurately using a smartphone. The specific objectives

are as follows: (a) to accurately segment target rice panicles and label from 2D images, (b) to reconstruct complete, high-precision, and low-noise 3D models of rice panicles, (c) to preprocess and calibrate the point clouds, and (d) to extract panicle length and volume traits from the calibrated point cloud models and predict panicle grain number and grain mass.

Materials and Methods

Field experimental design and rice plant materials

Two field experiments were conducted during the ripening stage of rice crops at the breeding sites of Longping High-Tech in Lingshui, Hainan Province on 2023 May 20 (Exp 1) and Jiaxing Academy of Agricultural Sciences in Jiaxing, Zhejiang Province on 2023 November 15 (Exp 2). *Indica* rice (S616-2261-4/Hua Hui 8612) and *japonica* rice crops (Pigm/Zhejiang *japonica* 99) were planted at Exp 1 and Exp 2, respectively. The *indica* rice was transplanted with a spacing of 8 inches in length and 5 inches in width, with two seedlings per hole, while the *japonica* rice was spaced 6 inches in length and 5 inches in width, with one seedling per hole, as shown in Fig. 1A and B.

Data acquisition

In-field data acquisition was conducted on 50 rice panicles each in Exp 1 and Exp 2. For Exp 1, data acquisition was performed between 1:30 PM and 3:30 PM under sunny conditions with wind speeds below 1.5 m/s. Similarly, for Exp 2, data acquisition occurred between 12:30 PM and 2:30 PM under sunny conditions with wind speeds below 1.5 m/s. The target rice panicles were randomly selected, and a label with known dimensions was affixed to each panicle for the size calibration. A 15-s video with the size of $1,920 \times 1,080$ pixels and a frame rate of 30 frames per second was then recorded by circling around the target rice panicle using a smartphone, as shown in Fig. 1C. Images were extracted from the collected video at a frequency of 15 frames per second to obtain multi-view images of each rice panicle. Two hundred twenty-five original images were finally obtained for each rice panicle sample. After video recording, rice panicles were cut and brought back to the laboratory for specific trait measurements as the ground truth. A steel ruler was used to measure the panicle length, and the number of grains and grain mass per panicle were measured using a Thousand-Grain Weight Instrument developed by the Digital Agriculture and Agricultural Internet-of-things Innovation (DAAI) team at Zhejiang University [37].

Data preprocessing

Data preprocessing was performed to compute the camera position and orientation as well as filter out images that did not adequately direct toward the target rice panicle to reduce the image processing time. SfM was employed to calculate the position and orientation of the camera [22]. Pose-derived viewing direction (PDVD) and image alignment filtering (IAF) were proposed for image filtering. PDVD constructed a scene center point based on the camera's position and orientation. IAF calculated the angle between the line from the camera to the center point and the camera's orientation, filtering out images with an angle greater than 20° . For each sample, the number of images ranged from 168 to 225 after filtering.

2D image segmentation

In order to remove the background from 2D images and achieve an accurate generation of point clouds including only the target

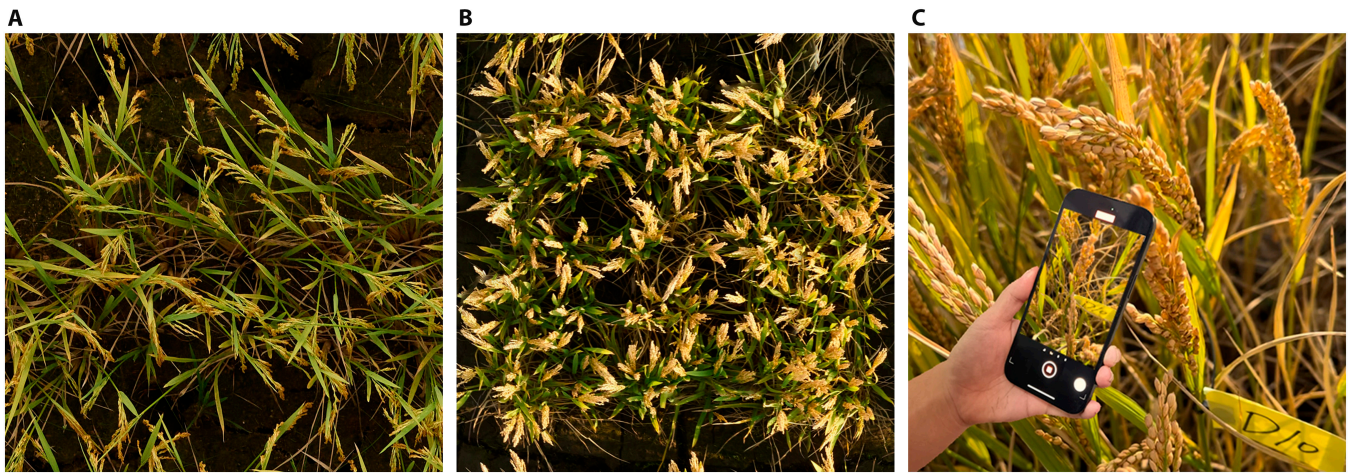


Fig. 1. Experimental fields and data acquisition. (A) Ripening stage of *indica* rice at the field of Longping High-Tech in Lingshui, Hainan Province, China. (B) Ripening stage of *japonica* rice at the field of Jiaxing Academy of Agricultural Sciences in Jiaxing, Zhejiang Province, China. (C) Data acquisition by circling around the target rice panicle using a smartphone.

rice panicle and the label, an ensemble model by integrating a large model with a small model was proposed in this study (Fig. 2B). The ensemble model utilized the SAM, a versatile image segmentation model for generic segmentation of the images [36]. However, SAM can only generate segmentation masks without providing pixel-level semantic information. The segmentation results were then filtered based on area and stability; regions with an area smaller than 10,000 pixels or with a stability score below 0.8 were removed to eliminate the noise. To compensate for the lack of semantic information in the segmentation results by SAM, the real-time object detection and instance segmentation model YOLOv8 was employed for instance segmentation, which generated rough masks of the target rice panicle and label [35]. These rough masks underwent edge erosion, resulting in red inner regions and white edge regions. For each red inner region, three to five positive points were randomly selected based on the pixel size of the region, and three negative points were randomly selected from the corresponding white outer region. These positive and negative points were then aligned to the mask regions in the SAM image after filter processing. Finally, the aligned mask regions were merged by taking their union to complete the matching process. The segmented images were processed to remove backgrounds, generating transparent PNG images and completing the 2D segmentation.

The specific parameters used for data training are shown in Table S2. The performance of 2D image segmentation was evaluated with F1 score, Intersection over Union (IoU), and boundary overlap (BO), which are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (4)$$

$$\text{BO} = \frac{|E_p \cap E_g|}{|E_p \cup E_g|} \quad (5)$$

where TP represents the number of true positive pixels, FP represents the number of false positive pixels, FN represents the number of false negative pixels, E_p represents the set of predicted edge pixels, E_g represents the set of ground-truth edge pixels, \cap denotes the intersection, and \cup denotes the union. In addition, the PanicleNeRF method was also compared with well-known instance segmentation methods, YOLOv8 and mask region-based convolutional neural network (Mask-RCNN) [38]. YOLOv8 is an object detection and instance segmentation method that utilizes a single-stage architecture for efficient and accurate predictions. It is an improved version of the YOLO series, offering enhanced performance and flexibility. While Mask-RCNN is a two-stage instance segmentation algorithm that extends the faster R-CNN framework by adding a branch for predicting segmentation masks. It has been widely used in both object detection and instance segmentation tasks by using a region proposal network (RPN) and a network head for mask prediction.

3D reconstruction of rice panicles

To develop high-quality 3D point cloud models of rice panicles, several NeRF methods were evaluated. Meanwhile, instant neural graphics primitives (Instant-NGP), which uses multi-resolution hash encoding and spherical harmonics for fast and efficient training and inference, was selected for 3D reconstruction [39]. The 3D reconstruction process is illustrated in Fig. 2C. The inputs comprised multi-view PNG images of rice panicles and their corresponding camera positions and orientations. The Instant-NGP algorithm was applied to learn a compact NeRF model representing the 3D structure of the target panicle and label. To convert the implicit density field generated by NeRF into an explicit 3D model and streamline the data processing pipeline, the marching cubes algorithm was combined with mesh point extraction, enabling direct export of point clouds from the trained NeRF model [40]. Specifically, the marching cubes algorithm was used to extract density isosurfaces, forming a closed triangular mesh model. By extracting the vertices from the

PanicleNeRF method

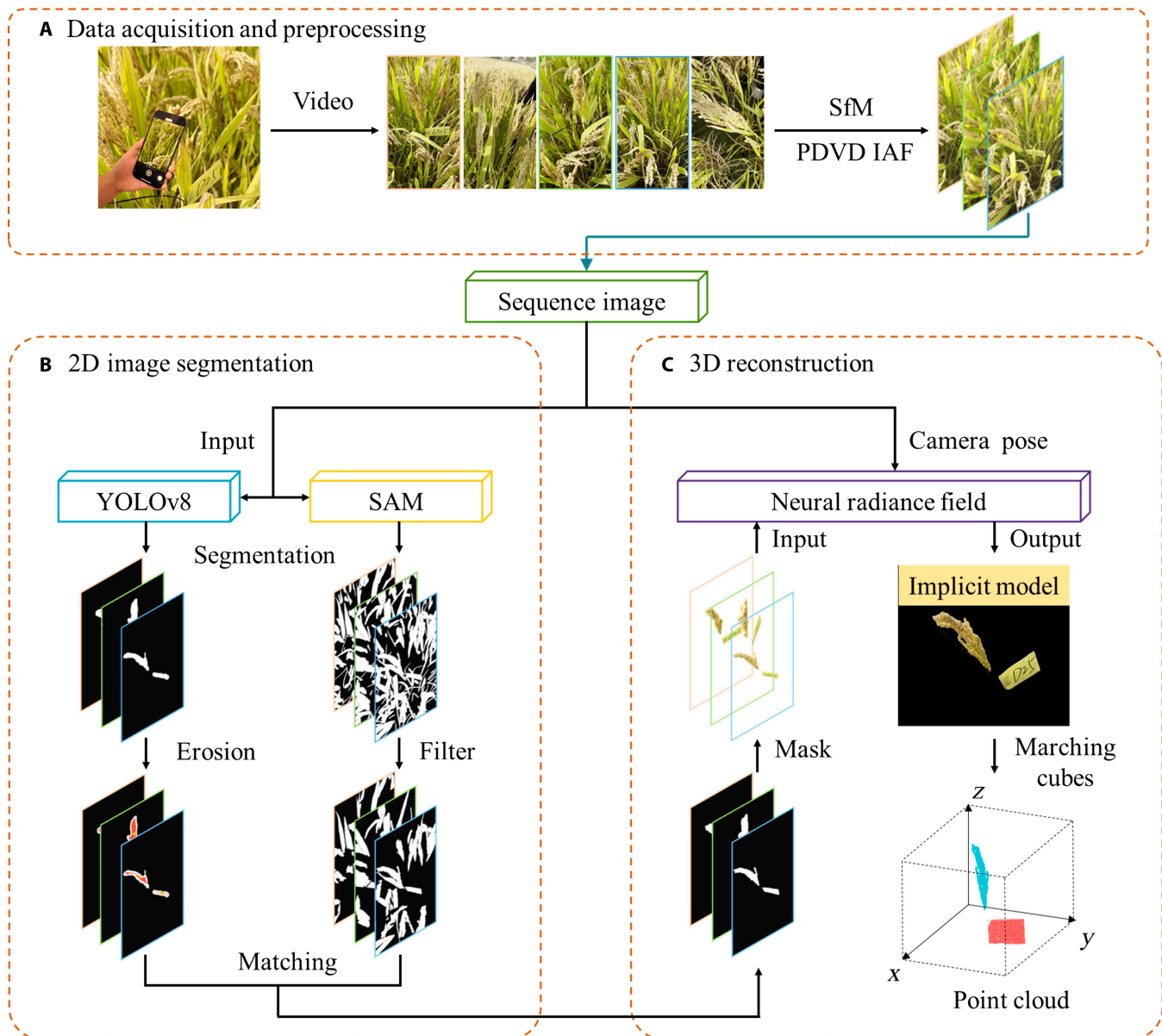


Fig. 2. Flowchart of PanicleNeRF method. (A) Data acquisition and preprocessing. (B) 2D image segmentation. (C) 3D reconstruction.

resulting mesh, the corresponding point clouds were obtained. The algorithm for optimized marching cubes was then applied to the trained NeRF model, producing dense point clouds of the rice panicle and label without background elements.

The 3D point clouds of rice panicle generated by PanicleNeRF were evaluated by the correlation with ground-truth phenotypic traits. Specifically, the coefficient of determination (R^2), root mean square error (RMSE), and relative root mean square error (rRMSE) between the predicted panicle length and the ground truth were calculated to validate the accuracy of the point cloud. R^2 measures the proportion of variance in the ground truth explained by the predicted values, while RMSE and rRMSE quantify the average magnitude of prediction errors in absolute and relative terms, respectively. These metrics provide a comprehensive evaluation of the accuracy and reliability of the PanicleNeRF method for 3D reconstruction of rice panicles.

Extraction of rice panicle traits

Point cloud processing

Point cloud clustering was conducted using the density-based spatial clustering of applications with noise (DBSCAN) method [41]. It started with identifying core points by comparing the number of points within a predefined neighborhood to a set threshold. Then, points were assessed to determine if they were in the same cluster as a core point based on density-reachable relationships, effectively clustering the point cloud and removing insufficiently sized clusters deemed as outliers. After clustering, principal components analysis (PCA) was utilized to differentiate between the label and rice panicle clusters based on normal vector features, with the former marked in red and the latter in blue [42].

The size calibration of point clouds was performed based on a reference label with known dimensions. A bounding box algorithm was employed to locate the smallest enclosing rectangle

of the point cloud representing the label, which had three faces with different areas. The point cloud was then projected onto the face with the median area and was fitted to calculate the length of the label in the point cloud scene. The size calibration was ultimately achieved by calculating the ratio between the length of the label and its corresponding length in the point cloud scene.

Extraction of panicle length trait

The point cloud of the target rice panicle was downsampled for rapid skeleton extraction and length calculation. As the standard Laplacian-based contraction (LBC) skeleton method had difficulty in identifying the start and end points of the rice panicle skeleton, a multi-tangent angle constraint was incorporated to avoid treating the tips of the rice panicle branches as endpoints [43]. After that, the skeleton was fitted with a curve to calculate the length of the panicle within the point cloud scene, and the panicle length trait was determined using the following formula:

$$L = L_1 \times \frac{X}{X_1} \quad (6)$$

where L is the predicted length of the rice panicle, L_1 is the length of the rice panicle in the point cloud scene, X is the length of the label, which equals 7.5 cm, and X_1 is the length of the label in the point cloud scene.

Extraction of panicle volume trait

After the size calibration of point clouds, the rice panicle point cloud was voxelized with a voxel size of 0.01 units. The number of voxel units encompassed by the rice panicle point cloud was then calculated to determine the panicle volume using the following formula:

$$V = Num \times (0.01)^3 \times X^3 \quad (7)$$

where V is the predicted volume of the rice panicle, Num is the number of voxel units encompassed by the rice panicle point, and X is the length of the label, which equals to 7.5 cm.

Development of a web-based platform

To enhance the accessibility and user-friendliness of our proposed PanicleNeRF method, we developed a web-based platform (<http://www.paniclenerf.com>). The platform offers researchers an intuitive interface to utilize our method without requiring extensive technical expertise or specialized computer hardware. Users can upload videos of rice panicles recorded under windless or gentle wind conditions using smartphones or other RGB cameras, with a 7.5-cm label attached to the panicle for accurate size calibration. Detailed images and dimensions of the label are provided in Fig. S1 to facilitate accurate printing and usage. The video should be captured by circling around the target rice panicle, providing a comprehensive view from multiple angles. An example of the recording process is available in Movie S1. Upon uploading the video, the platform processes the data automatically using our PanicleNeRF method and generates a rotating demonstration video of the reconstructed point cloud, along with the predicted phenotypic traits for the given rice panicle.

Results

2D image segmentation

Table 1 presents the results of three different 2D image segmentation methods for *indica* and *japonica* rice varieties. The

Table 1. The comparison of 2D image segmentation methods for *indica* and *japonica* rice varieties. The best results are in boldface.

Variety	Method	F1 score (%)		IoU (%)	
		Val	Test	Val	Test
<i>Indica</i> rice	Mask-RCNN	85.0	82.6	76.3	73.5
	YOLOv8	86.1	82.9	79.2	76.4
	PanicleNeRF	87.3	84.0	81.1	78.6
<i>Japonica</i> rice	Mask-RCNN	85.5	85.8	75.7	75.9
	YOLOv8	88.7	85.3	80.9	77.4
	PanicleNeRF	89.2	87.0	81.3	78.2

proposed method, PanicleNeRF, outperformed the baseline method Mask-RCNN. For *indica* rice, PanicleNeRF achieved an F1 score of 87.3% and 84.0% on the validation and test datasets, representing an improvement of 2.3% and 1.4%, respectively. Similarly, PanicleNeRF attained an IoU of 81.1% and 78.6%, surpassing Mask-RCNN by 4.8% and 5.1%. PanicleNeRF also demonstrated superior performance on *japonica* rice, with an F1 score of 89.2% and 87.0%, exceeding Mask-RCNN by 3.7% and 1.2%. The IoU reached 81.3% and 78.2%, outperforming Mask-RCNN by 5.6% and 2.3%. Moreover, PanicleNeRF performed slightly better than YOLOv8 across all metrics for both rice varieties.

The boundary segmentation performance is presented in Fig. 3. PanicleNeRF achieved the best BO results in *indica* with 8.5% for both validation and test, and in *japonica* with 9.5% for validation and 10.0% for test. These results showed improvements of 4.0%, 3.9%, 5.1%, and 5.6% over the second-best results obtained by YOLOv8. Notably, the interquartile range (IQR) of PanicleNeRF's performance was entirely above that of Mask-RCNN and YOLOv8 across all four categories. Figure 4 illustrates the representative results of the three methods on the testing set. Despite the complex background and branching structure of rice panicles, PanicleNeRF achieved excellent segmentation performance, while Mask-RCNN displayed significant fluctuations in edge segmentation and YOLOv8 produced noticeable jagged edges when segmenting the branches. The substantial improvement in the BO value distribution, along with the enhanced segmentation results, demonstrated the superior performance and robustness of the PanicleNeRF method.

3D reconstruction of rice panicle

Figure 5 presents the background removal results in the NeRF model after using the PanicleNeRF method for representative samples of *indica* and *japonica* rice. Figure 5A and C shows the NeRF model obtained by inputting the original image set, while Fig. 5B and D displays the NeRF model obtained by inputting the same set of images after 2D segmentation. The results demonstrate that PanicleNeRF effectively removed nonregions of interest and extracts the target rice panicles and labels in the NeRF model. Furthermore, the clustering results of the panicle and label point clouds are illustrated in Fig. 6, where panels (A) and (B) represent *indica* and *japonica* rice, respectively. In both cases, the blue points indicate the panicle semantics, while the

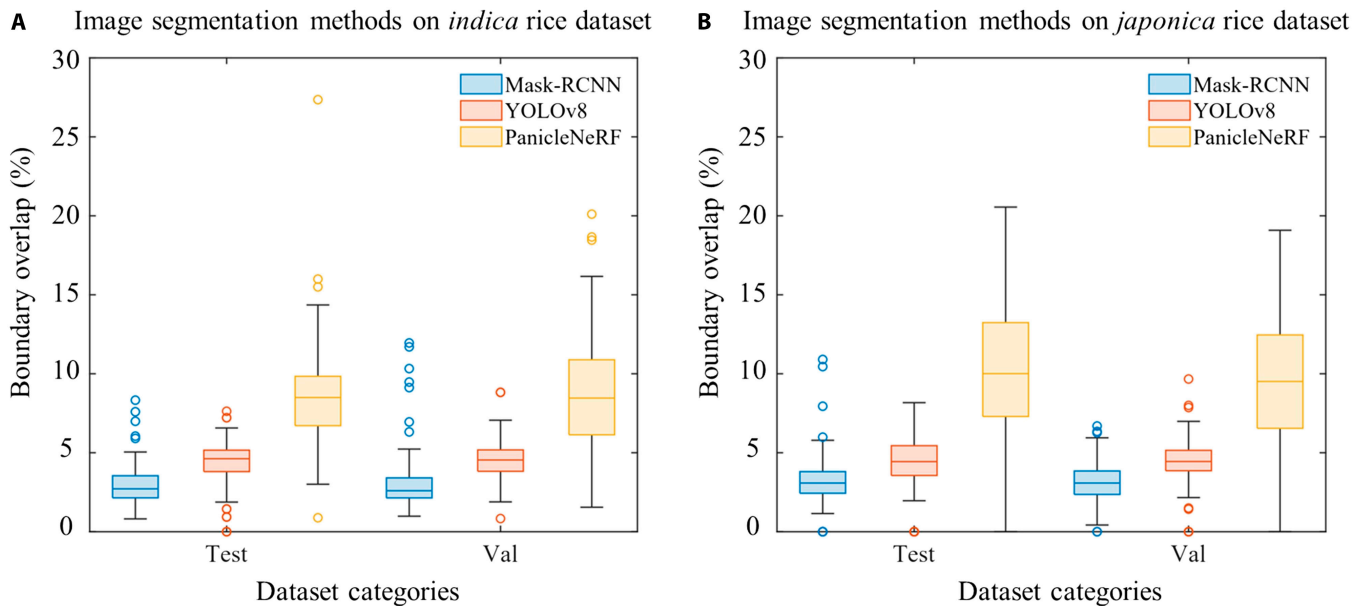


Fig. 3. The boundary overlap performance of different methods on rice varieties. (A) Performance on *indica* rice dataset. (B) Performance on *japonica* rice dataset.

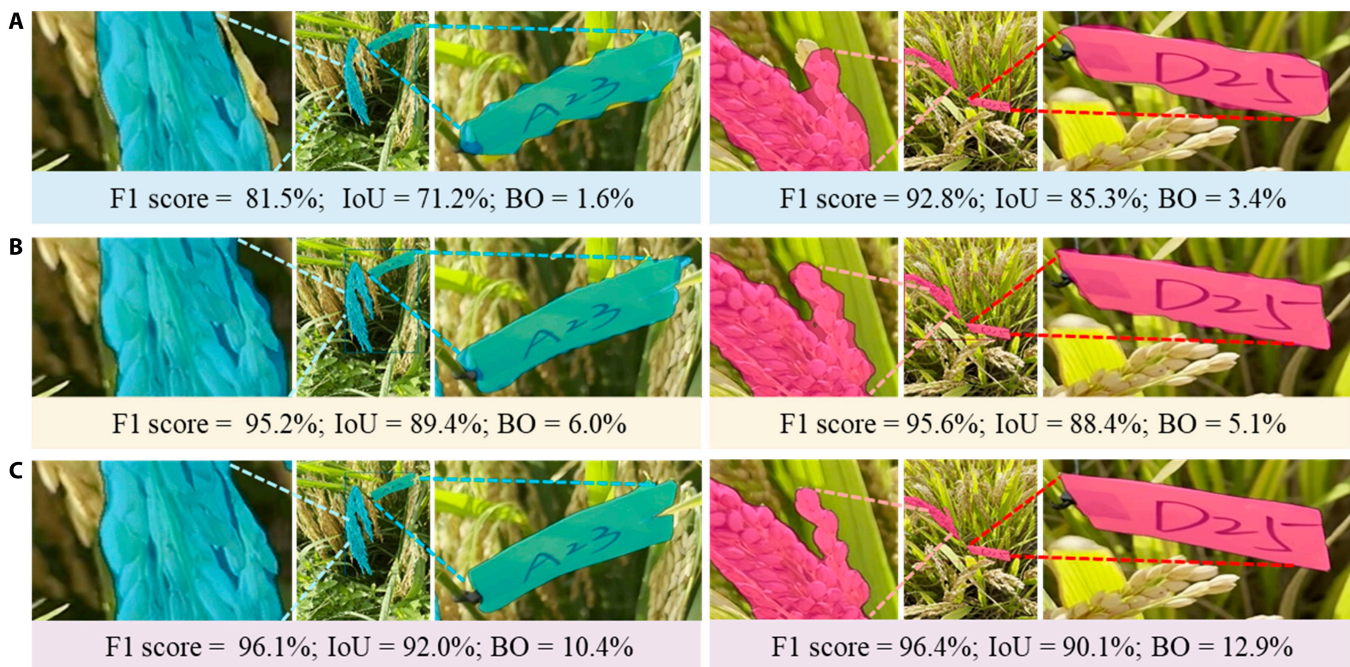


Fig. 4. Illustration of the representative image segmentation results on *indica* rice (left column) and *japonica* rice (right column) by different methods. (A) Mask-RCNN segmentation results. (B) YOLOv8 segmentation results. (C) PanicleNeRF segmentation results.

red points indicate the label semantics. The clear separation of the panicle and label point clouds demonstrates the effectiveness of PanicleNeRF in distinguishing between the two semantic categories. This effective clustering lays the foundation for subsequent extraction of rice panicle traits.

Figure 7 presents a comparison of the point clouds generated by PanicleNeRF with those generated by COLMAP and Metashape. Figure 7A and B shows the front and side views of the *indica* rice point cloud, while Fig. 7C and D displays the front and side views of the *japonica* rice point cloud. Overall, the point cloud generated by PanicleNeRF was the

most complete, with the fewest point cloud holes and the highest resolution. The side view of the *indica* rice point cloud (Fig. 7B) clearly illustrates that the point cloud generated by COLMAP had more noise and could not distinguish the branches of the rice panicle, while the point cloud generated by Metashape was more fragmented and had poor completeness. The other viewpoints also demonstrate the fragmentation and incompleteness of COLMAP and Metashape point clouds in the reconstruction of rice panicles and labels, which leads to difficulties in their practical application for field phenotype extraction.

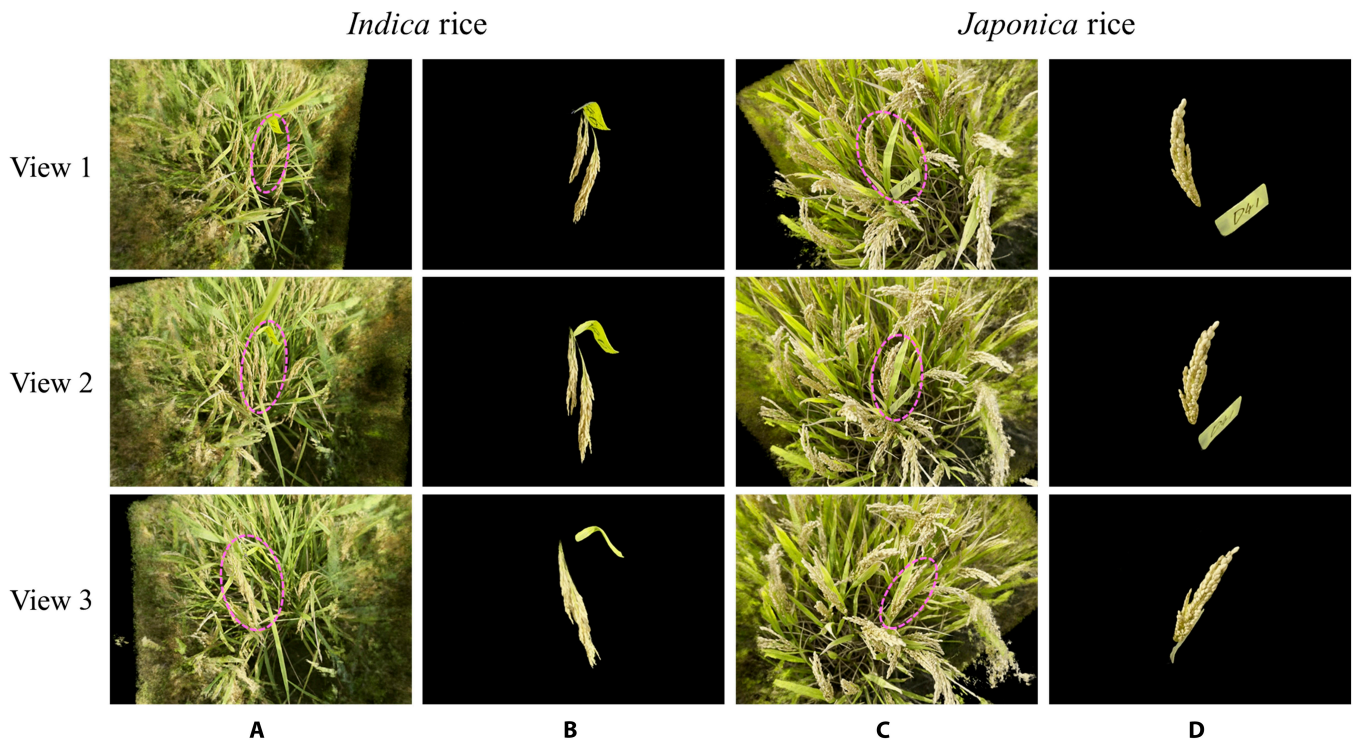


Fig. 5. NeRF models of *indica* and *japonica* rice reconstructed using original image set and 2D segmented images, viewed from three different perspectives of the NeRF model. (A) *Indica* NeRF model obtained from the original image set. (B) *Indica* NeRF model obtained from images with 2D segmentation. (C) *Japonica* NeRF model obtained from the original image set. (D) *Japonica* NeRF model obtained from images with 2D segmentation.

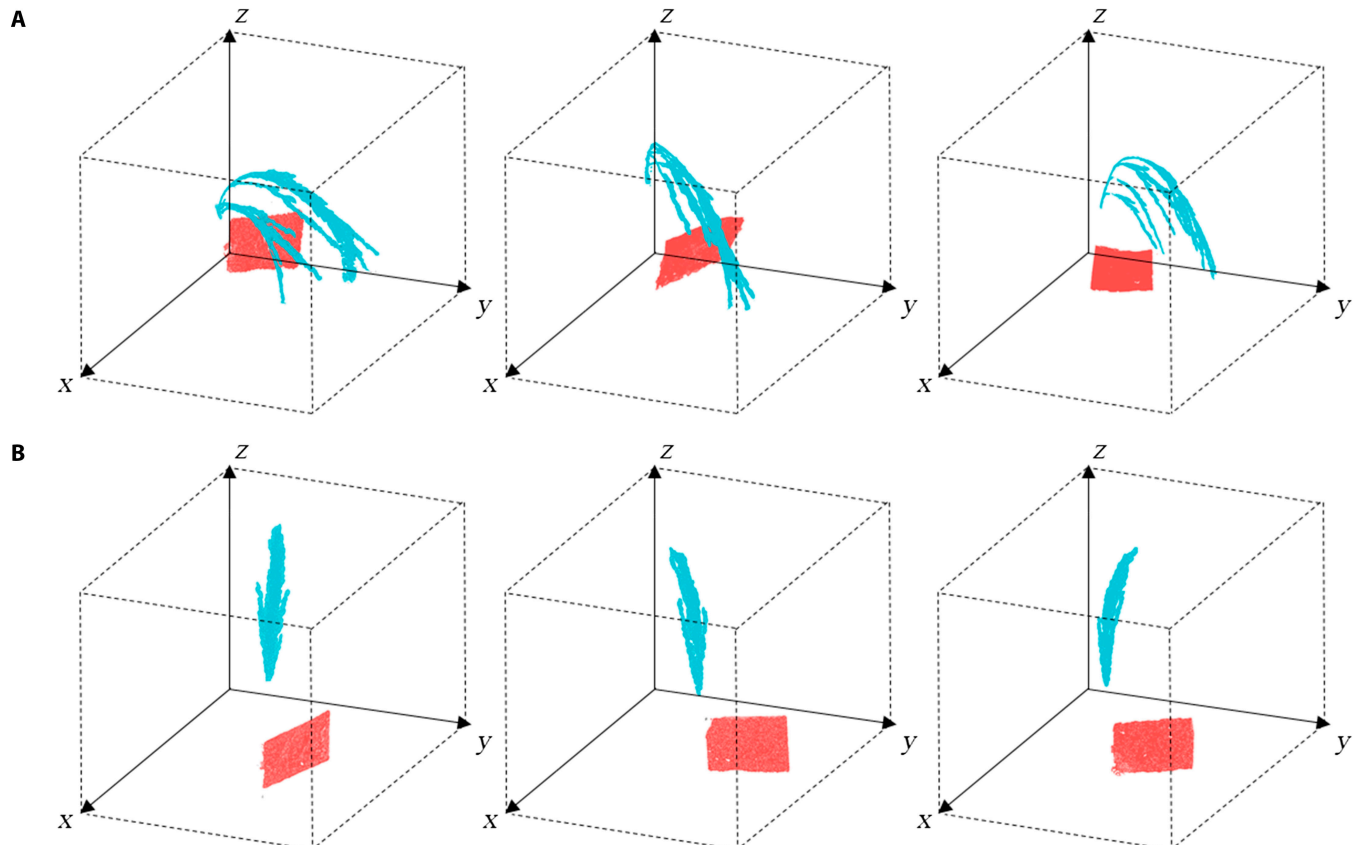


Fig. 6. Illustration of the representative clustering results of rice panicle and label point clouds for (A) *indica* rice and (B) *japonica* rice, where blue indicates the panicle semantics and red indicates the label semantics.

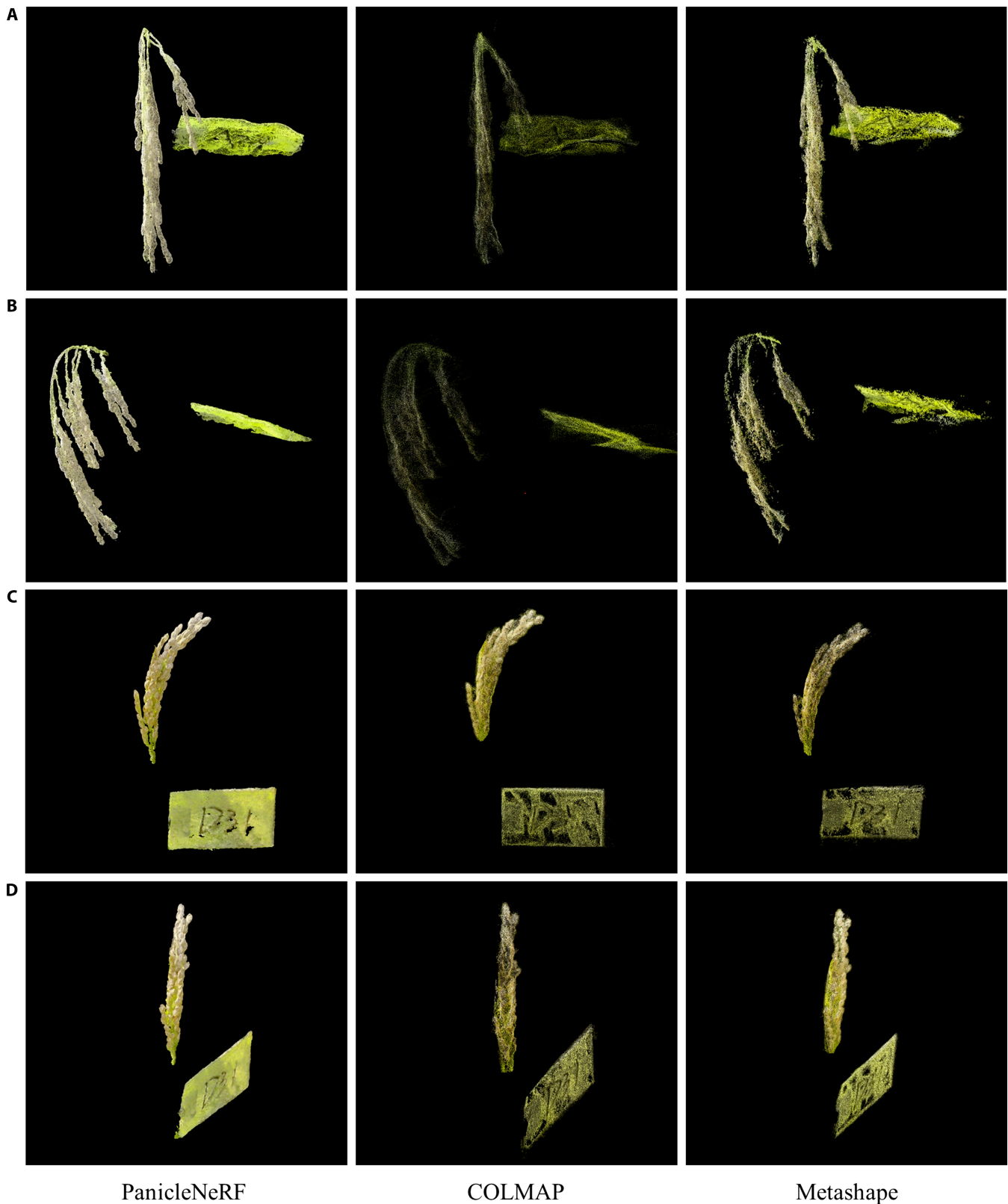


Fig. 7. The comparison of representative rice panicle point clouds reconstructed by PanicleNeRF and traditional 3D reconstruction methods (COLMAP and Metashape). (A) Front view of *indica* rice. (B) Side view of *indica* rice. (C) Front view of *japonica* rice. (D) Side view of *japonica* rice.

Panicle trait extraction

Figure 8 shows the correlation between the predicted and measured panicle lengths for *indica* and *japonica* rice. For *indica*

rice, the R^2 value was 0.90, the RMSE was 0.71, and the rRMSE was 2.94%, while for *japonica* rice, the R^2 value was 0.95, the RMSE was 0.26, and the rRMSE was 1.75%. These results validated the

effectiveness of the size calibration and the accuracy of the 3D point clouds. The results also indicated that the prediction accuracy for *japonica* rice was higher than that for *indica* rice, as the R^2 value was higher and the rRMSE was lower for *japonica* rice.

Figure 9 shows the correlation between the predicted panicle volume and the measured grain count and grain mass for *indica* and *japonica* rice. For *indica* rice, the R^2 , RMSE, and rRMSE values were 0.85, 15.61, and 7.45%, respectively, for grain count and 0.80, 0.45, and 8.51%, respectively, for grain mass. For *japonica* rice, the corresponding values were 0.82, 15.42, and 9.74% for grain count and 0.76, 0.47, and 10.33% for grain mass. These results demonstrated a strong correlation between the predicted panicle volume and grain count, as well as grain mass. Besides, the prediction accuracy for both grain count and grain mass was higher in *indica* rice than in *japonica* rice, with *indica* rice having higher R^2 values and lower rRMSE values for both traits.

The correlations between panicle length, panicle volume, grain count, and grain mass in the *indica* and *japonica* rice datasets were calculated and visualized in Fig. 10. For *indica* rice, the correlations between panicle length and the other three phenotypes were only 0.37, 0.32, and 0.43, while the correlations between panicle volume and grain count and grain mass were 0.85 and 0.80, respectively. A similar pattern was observed for *japonica* rice.

The average time spent on the entire PanicleNeRF workflow is summarized in Table 2. Data acquisition takes 15 s, while data preprocessing requires 15 min. The 2D image segmentation and 3D reconstruction steps each take 2 min. Finally, the extraction of rice panicle traits takes 4 min.

Discussion

Multi-model fusion outperforms single models in rice panicle 2D segmentation

The results demonstrated that the proposed 2D rice panicle segmentation method, which combined the advantages of the large model SAM and the small model YOLOv8, achieved superior performance compared to using a single model. As illustrated in Fig. 4, utilizing a single model for segmentation often led to incomplete target contours, while the accuracy of the 3D reconstruction method employed in this research relied on the

segmentation quality of 2D images. Existing models struggled to achieve satisfactory results because they were primarily designed to perform well on public datasets such as ImageNet, where crop categories accounted for only around 2.4% of the total image categories, with the majority of targets being common rigid objects encountered in daily life and exhibiting distinct features compared to crops [44]. It is unsurprising that using transfer learning to fine-tune existing models on crop datasets caused suboptimal segmentation contours [45]. However, training a new network specifically tailored for crop segmentation requires a substantial amount of samples, and currently, there is a lack of sufficient data to train a model capable of achieving high-precision contour segmentation.

With the advancement of ultra-large models such as SAM, their generalization ability has significantly improved, enabling accurate contour segmentation even for targets not included in the training set. Nevertheless, these models could not perform individual segmentation for specific targets without appropriate prompts. It is precisely by leveraging the strengths of both large and small models that our proposed method achieved superior segmentation results. Targets with inaccurate segmentation can be used as prompts and fed into large models, thereby obtaining targets with precise contour segmentation.

Superiority of PanicleNeRF in fine-scale

Traditional SfM-MVS-based methods, such as COLMAP and Metashape, produce noisy and incomplete point clouds when dealing with in-field rice panicle scenes. This suboptimal performance could be attributed to the following reasons. First, objects with repetitive textures, such as rice panicles, are prone to feature matching errors due to the high similarity of feature points at different positions, leading to incorrect matching and large deviations in the recovered 3D point positions [46]. Second, small objects present challenges for accurate 3D structure recovery due to inadequate parallax information. The small disparities of small objects across different views lead to large depth estimation errors in triangulation and 3D point recovery. Lastly, SfM-MVS relies on geometric constraints between images, which are often not satisfied when objects have occlusions and complex shapes [47]. Complex shapes reduce the correspondence between images, while occlusions obstruct feature matching, resulting in

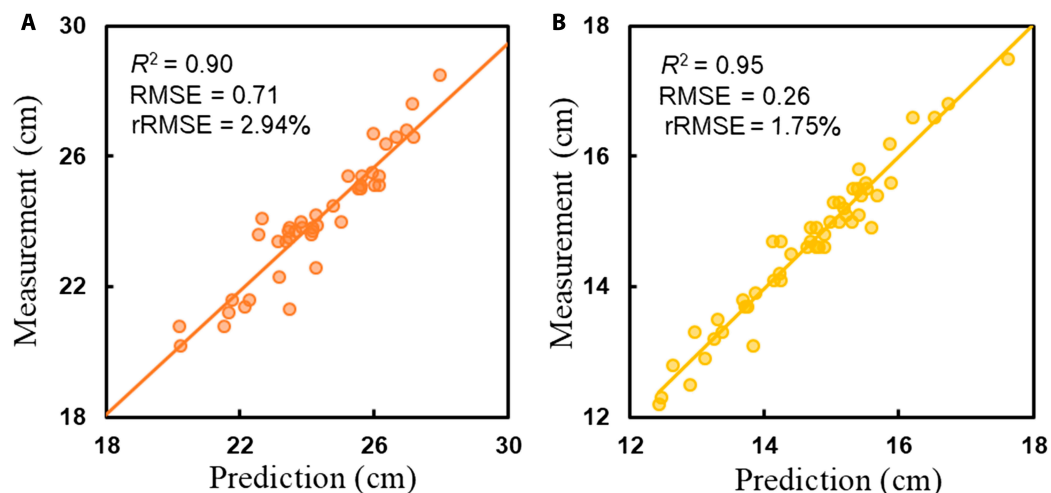


Fig. 8. Correlation analysis between predicted and measured panicle lengths for (A) *indica* rice and (B) *japonica* rice.

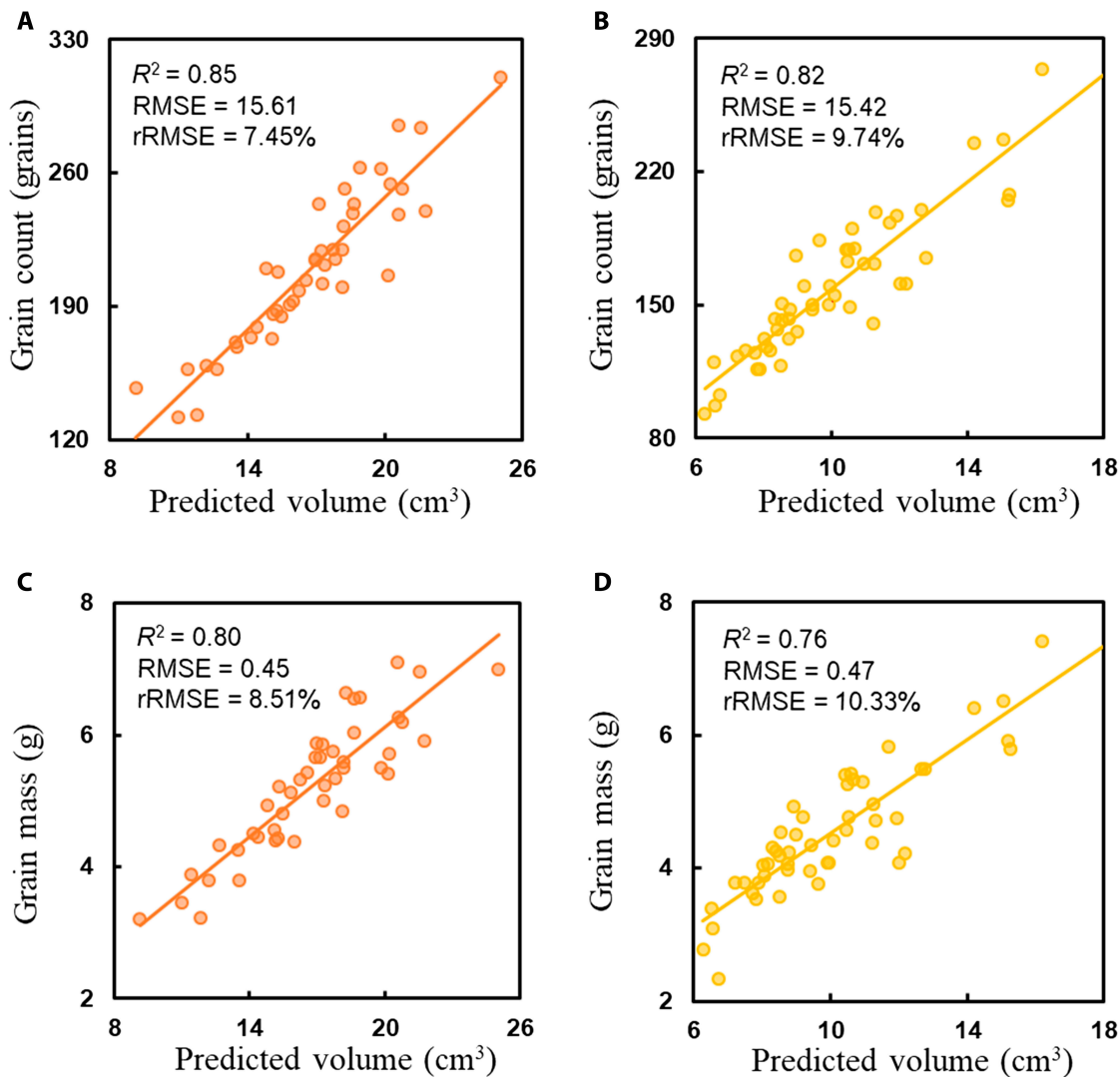


Fig. 9. Correlation analysis between predicted panicle volume and measured grain count and grain mass for *indica* and *japonica* rice. (A) Predicted volume versus measured grain count for *indica* rice. (B) Predicted volume versus measured grain count for *japonica* rice. (C) Predicted volume versus measured grain mass for *indica* rice. (D) Predicted volume versus measured grain mass for *japonica* rice.

a lack of effective geometric constraints and compromising the completeness and precision of the reconstructed objects.

In contrast, PanicleNeRF achieved the least noisy and most complete point cloud compared to the SfM-MVS-based method. The superiority of PanicleNeRF can be attributed to its ability to learn the geometric structure and appearance information of the scene through neural networks. Unlike traditional methods that rely on feature point matching and triangulation, PanicleNeRF directly learns the 3D structure from images, enabling it to handle situations with similar features, repetitive textures, and insufficient parallax. Furthermore, by representing the scene through continuous density and color fields, PanicleNeRF can generate more complete and detail-rich reconstruction results. The significant advantages demonstrated by PanicleNeRF in fine-grained 3D reconstruction tasks under complex scenes highlight its potential to advance plant phenotypic analysis.

Panicle trait extraction performance and comparisons

The results demonstrate that high R^2 and low rRMSE were achieved in panicle length, grain count, and grain mass prediction.

During this process, we discovered that the panicle length prediction for *indica* rice was not as accurate as that for *japonica* rice, possibly due to the more compact and convergent branches of *japonica* rice compared to the looser and more dispersed branches of *indica* rice. When extracting the skeleton, the central skeleton extraction of multi-branched panicles is more challenging than that of panicles with branches clustered together. Additionally, we found that when using panicle volume to predict grain count and grain mass, the *indica* rice dataset performed better than the *japonica* rice dataset. This is primarily attributed to the compact branches of *japonica* rice, which might have created gaps within the panicle that were treated as solid matter, potentially limiting the accuracy of the model. Such limitations are common in visible image-based reconstructions and may only be addressed by adopting other techniques like CT or magnetic resonance imaging [48,49]. Figure 10 demonstrates that panicle volume, extractable through 3D reconstruction methods, exhibits significantly stronger correlations with grain count and grain mass than panicle length. This underscores the importance of accurate 3D models and

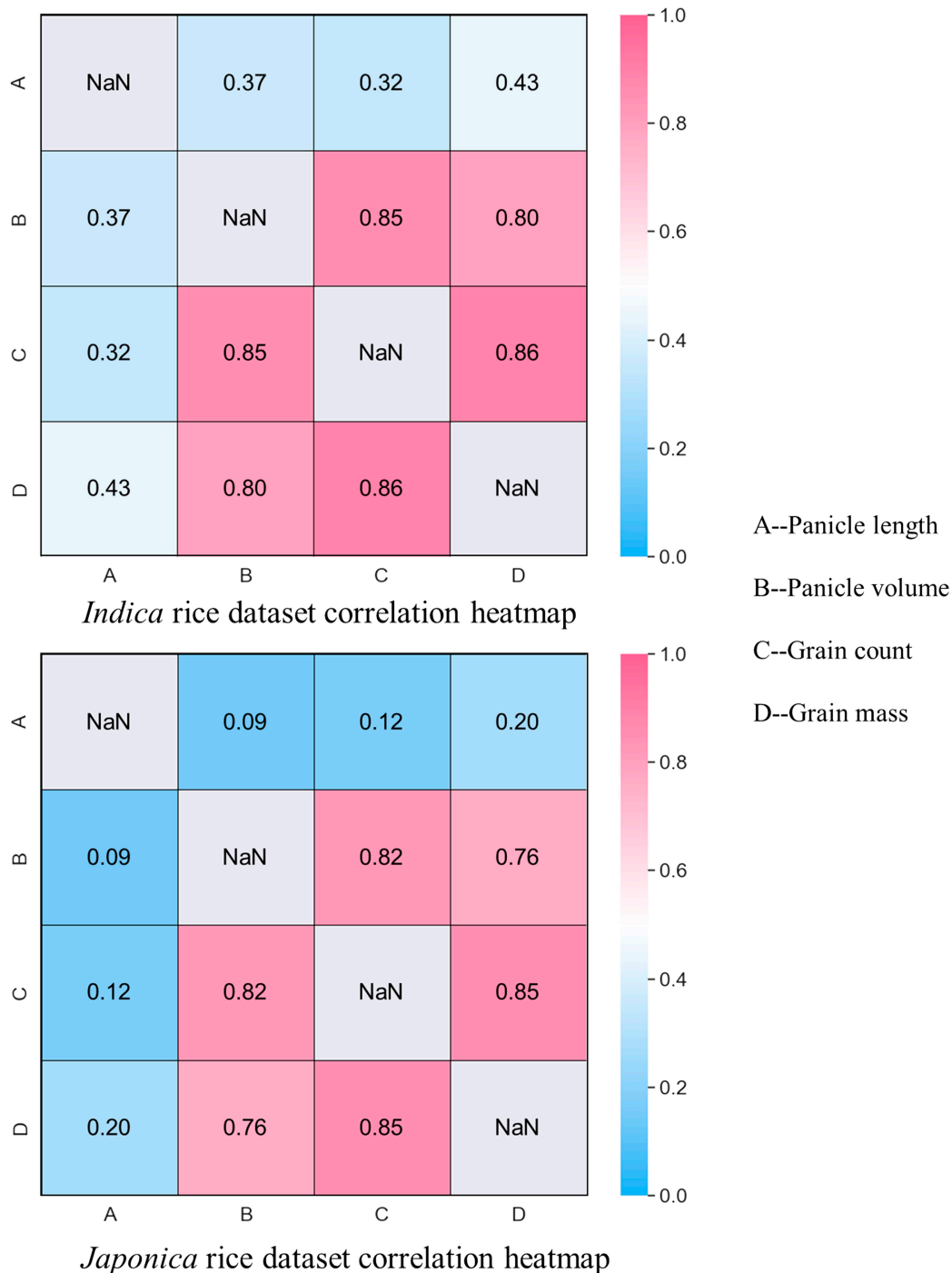


Fig. 10. Heatmap of correlations among panicle length, panicle volume, grain count, and grain mass for different rice panicle types in the *indica* and *japonica* datasets.

their unique capability to extract panicle volume, providing valuable information that cannot be achievable with 2D models.

In recent years, researchers have utilized 2D images captured by smartphones to perform phenotyping of panicles both at the indoor and infield conditions. Qiu et al. [50] fixed a smartphone and detached panicles in relative positions, enabling grain counting indoors. Sun et al. [51] used a smartphone with a black background board to achieve panicle segmentation and phenotyping of rice grains in the field. However, because rice panicles are inherently 3D structures, 2D images inevitably

result in information loss, such as difficulties in accurately analyzing panicle volume, while Sandhu et al. [52] fixed panicles in the isolated environments and reconstructed 3D point clouds indoors, achieving the R^2 values ranging from 0.70 to 0.76 for grain count and mass prediction. In comparison, the PanicleNeRF method proposed in this study enables fast evaluation of rice panicle traits in the field with better performance than that of other 3D methods, with the R^2 values of 0.83 for the grain count and 0.78 for the grain mass, demonstrating its effectiveness in field rice panicle phenotyping.

Table 2. Time consumption analysis for each step in the PanicleNeRF workflow. The hardware used in this study can be found in Table S1.

Workflow step	Time consumed
Data acquisition	15 s
Data preprocessing	15 min
2D image segmentation	2 min
3D reconstruction	2 min
Extraction of rice panicle traits	4 min

Limitation and future prospects

PanicleNeRF had achieved promising performance in rice panicle phenotyping. However, there were still some limitations of our method. One limitation was that the throughput was limited as it focused on extracting detailed 3D models of individual rice panicle. In the future, by extending PanicleNeRF, low-cost remote sensing equipment, such as UAVs, could potentially be employed to analyze high-throughput and large-scale fine-grained 3D models of rice panicles in the field [53–55]. In addition, the method required data acquisition under windless or light breeze conditions, which restricted its convenience of application in actual paddy fields. To address this issue, multiple cameras could be utilized for synchronous exposure, enabling simultaneous acquisition of multi-view high-quality imaging of rice plants, effectively reducing the interference of natural wind. Moreover, the camera extrinsic parameter inference in data preprocessing was time consuming, taking up to 15 min (Table 2). This could be improved in the future by designing a specialized data acquisition device that enables consistent relative positions and orientations of cameras in each acquisition, thereby utilizing fixed camera extrinsic parameters to avoid this time-consuming step.

Conclusion

The conventional 3D reconstruction and segmentation methods often generate noisy and fragmented point clouds when dealing with the complex structure and repetitive texture of rice panicles, which is not suitable for phenotyping panicles in the field. To address this challenging problem, we propose PanicleNeRF, a novel method that enables high-precision and low-cost reconstruction of rice panicle 3D models in the field based on the video acquired by the smartphone. The proposed method combined the large model SAM and the small model YOLOv8 to achieve high-precision segmentation of rice panicle images. The NeRF technique was then employed for 3D reconstruction using the images with 2D segmentation. Finally, the resulting point clouds are processed to successfully extract and analyze panicle phenotypes. The results show that PanicleNeRF effectively addressed the task of 2D image segmentation, achieving a mean F1 score of 86.9% and a mean IoU of 79.8%, with nearly double the BO performance compared to YOLOv8. In terms of point cloud quality, PanicleNeRF significantly outperformed traditional SfM-MVS methods, such as COLMAP and Metashape. The panicle length was then accurately extracted with the rRMSE of 2.94% for *indica* and 1.75% for *japonica* rice. The panicle

volume estimated from high-quality 3D point clouds strongly correlated with the actual grain number ($R^2 = 0.85$ for *indica* and $R^2 = 0.82$ for *japonica*) and grain mass ($R^2 = 0.80$ for *indica* and $R^2 = 0.76$ for *japonica*). This work is expected to contribute to the advancement of high-quality in-field rice panicle phenotyping, facilitating the progress of rice phenotyping and breeding efforts.

Acknowledgments

Funding: This work was supported by the Fundamental Research Funds for the Central Universities (226-2022-00217), Key R&D Program of Zhejiang Province (2021C02057), and Zhejiang University Global Partnership Fund (188170+194452208/005).

Author contributions: X.Y., X.L., P.X., and Z.G. designed the study, performed the experiment, and wrote the manuscript. X.Y. developed the algorithm and analyzed the data. H.F. developed a web platform for the algorithms. H.F., X.H., and Z.S. cultivated two different rice varieties. H.C. supervised experiments at all stages and revised the manuscript.

Competing interests: The authors declare that they have no competing interests.

Data Availability

The data are freely available upon reasonable request.

Supplementary Materials

Figs. S1 and S2
Tables S1 to S3
Movie S1

References

1. He X, Batary P, Zou Y, Zhou W, Wang G, Liu Z, Bai Y, Gong S, Zhu Z, Settele J, et al. Agricultural diversification promotes sustainable and resilient global rice production. *Nat Food*. 2023;4(9):788–796.
2. Shin D, Lee S, Kim T-H, Lee J-H, Park J, Lee J, Lee JY, Cho L-H, Choi JY, Lee W, et al. Natural variations at the Stay-Green gene promoter control lifespan and yield in rice cultivars. *Nat Commun*. 2020;11(1):2819.
3. Guo T, Lu Z-Q, Xiong Y, Shan J-X, Ye W-W, Dong N-Q, Kan Y, Yang Y-B, Zhao H-Y, Yu H-X, et al. Optimization of rice panicle architecture by specifically suppressing ligand–receptor pairs. *Nat Commun*. 2023;14(1):1640.
4. Agata A, Ando K, Ota S, Kojima M, Takebayashi Y, Takehara S, Doi K, Ueguchi-Tanaka M, Suzuki T, Sakakibara H, et al. Diverse panicle architecture results from various combinations of Prl5/GA20ox4 and Pbl6/APO1 alleles. *Commun Biol*. 2020;3(1):302.
5. Sun X, Xiong H, Jiang C, Zhang D, Yang Z, Huang Y, Zhu W, Ma S, Duan J, Wang X, et al. Natural variation of DROT1 confers drought adaptation in upland rice. *Nat Commun*. 2022;13(1):4265.
6. Ye J, Zhang M, Yuan X, Hu D, Zhang Y, Xu S, Li Z, Li R, Liu J, Sun Y, et al. Genomic insight into genetic changes and shaping of major inbred rice cultivars in China. *New Phytol*. 236(6):2311–2326.
7. Wu W, Liu T, Zhou P, Yang T, Li C, Zhong X, Sun C, Liu S, Guo W. Image analysis-based recognition and quantification

- of grain number per panicle in rice. *Plant Methods*. 2019;15(1):122.
8. Lu Y, Wang J, Fu L, Yu L, Liu Q. High-throughput and separating-free phenotyping method for on-panicle rice grains based on deep learning. *Front Plant Sci*. 2023;14:1219584.
 9. Su L, Chen P. A method for characterizing the panicle traits in rice based on 3D micro-focus X-ray computed tomography. *Comput Electron Agric*. 2019;166:104984.
 10. Wu D, Guo Z, Ye J, Feng H, Liu J, Chen G, Zheng J, Yan D, Yang X, Xiong X, et al. Combining high-throughput micro-CT-RGB phenotyping and genome-wide association study to dissect the genetic architecture of tiller growth in rice. *J Exp Bot*. 2019;70(2):545–561.
 11. Gong L, Du X, Zhu K, Lin K, Lou Q, Yuan Z, Huang G, Liu C. Panicle-3D: Efficient phenotyping tool for precise semantic segmentation of rice panicle point cloud. *Plant Phenomics*. 2021;2021:9838929.
 12. Liu Z-Y, Wu H-F, Huang J-F. Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. *Comput Electron Agric*. 2010;72(2):99–106.
 13. Wu D, Yu L, Ye J, Zhai R, Duan L, Liu L, Wu N, Geng Z, Fu F, Huang C, et al. Panicle-3D: A low-cost 3D-modeling method for rice panicles based on deep learning, shape from silhouette, and supervoxel clustering. *Crop J*. 2022;10(5):1386–1398.
 14. Fiorani F, Schurr U. Future scenarios for plant phenotyping. *Annu Rev Plant Biol*. 2013;64:267–291.
 15. Janni M, Pieruschka R. Plant phenotyping for a sustainable future. *J Exp Bot*. 2022;73(15):5085–5088.
 16. Xiao S. 3D reconstruction and characterization of cotton bolls in situ based on UAV technology. *ISPRS J Photogramm Remote Sens*. 2024;209:101–116.
 17. Jin S, Sun X, Wu F, Su Y, Li Y, Song S, Xu K, Ma Q, Baret F, Jiang D, et al. Lidar sheds new light on plant phenomics for plant breeding and management: Recent advances and future prospects. *ISPRS J Photogramm Remote Sens*. 2021;171:202–223.
 18. Esser F, Rosu RA, Cornelißen A, Klingbeil L, Kuhlmann H, Behnke S. Field robot for high-throughput and high-resolution 3D plant phenotyping: Towards efficient and sustainable crop production. *IEEE Robot Autom Mag*. 2023;30(4):20–29.
 19. Lin G, Tang Y, Zou X, Li J, Xiong J. In-field citrus detection and localisation based on RGB-D image analysis. *Biosyst Eng*. 2019;186:34–44.
 20. Xiang L, Gai J, Bao Y, Yu J, Schnable PS, Tang L. Field-based robotic leaf angle detection and characterization of maize plants using stereo vision and deep convolutional neural networks. *J Field Robot*. 2023;40(5):1034–1053.
 21. Ming L, Fu D, Wu Z, Zhao H, Xu X, Xu T, Xiong X, Li M, Zheng Y, Li G, et al. Transcriptome-wide association analyses reveal the impact of regulatory variants on rice panicle architecture and causal gene regulatory networks. *Nat Commun*. 2023;14(1):7501.
 22. Schonberger JL, Frahm J-M. Structure-from-motion revisited. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, NV, USA.
 23. Zhu R, Sun K, Yan Z, Yan X, Yu J, Shi J, Hu Z, Jiang H, Xin D, Zhang Z, et al. Analysing the phenotype development of soybean plants using low-cost 3D reconstruction. *Sci Rep*. 2020;10(1):7055.
 24. Kerbl B, Kopanas G, Leimkuehler T, Drettakis G. 3D gaussian splatting for real-time radiance field rendering. *ACM Trans Graph*. 2023;42(4):1–14.
 25. Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R. NeRF: Representing scenes as neural radiance fields for view synthesis. *Commun ACM*. 2021;65(1):99–106.
 26. Schönberger JL, Zheng E, Frahm J-M, Pollefeys M. Pixelwise view selection for unstructured multi-view stereo. Paper presented at: Computer Vision – ECCV 2016; 2016; Cham, Switzerland.
 27. He W, Ye Z, Li M, Yan Y, Lu W, Xing G. Extraction of soybean plant trait parameters based on SfM-MVS algorithm combined with GRNN. *Front Plant Sci*. 2023;14:1181322.
 28. Chen H, Zhang M, Xiao S, Wang Q, Cai Z, Dong Q, Feng P, Shao K, Ma Y. Quantitative analysis and planting optimization of multi-genotype sugar beet plant types based on 3D plant architecture. *Comput Electron Agric*. 2024;225:109231.
 29. Jiang L, Li C, Sun J, Chee P, Fu L. Estimation of cotton boll number and main stem length based on 3D gaussian splatting. Paper presented at: 2024 ASABE Annual International Meeting; 2024; St. Joseph, MI.
 30. Smitt C, Halstead M, Zimmer P, Läbe T, Guclu E, Stachniss C, McCool C. PAG-NeRF: Towards fast and efficient end-to-end panoptic 3d representations for agricultural robotics. *IEEE Robot Autom Lett*. 2024;9(1):907–914.
 31. Kelly S, Riccardi A, Marks E, Magistri F, Guadagnino T, Chli M, Stachniss C. Target-aware implicit mapping for agricultural crop inspection. Paper presented at: 2023 IEEE International Conference on Robotics and Automation (ICRA); May 2023; London, UK.
 32. Saeed F, Sun J, Ozias-Akins P, Chu YJ, Li CC. PeanutNeRF: 3D radiance field for peanuts. Paper presented at: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); Jun. 2023; Vancouver, BC, Canada.
 33. Hu K, Ying W, Pan Y, Kang H, Chen C. High-fidelity 3D reconstruction of plants using Neural Radiance Fields. *Comput Electron Agric*. 2024;220:108848.
 34. Zhang J, Wang X, Ni X, Dong F, Tang L, Sun J, Wang Y. Neural radiance fields for multi-scale constraint-free 3D reconstruction and rendering in orchard scenes. *Comput Electron Agric*. 2024;217:108629.
 35. Jocher G, Chaurasia A, Qiu J. Ultralytics YOLO. Jan. 2023. <https://github.com/ultralytics/ultralytics>.
 36. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, et al. Segment Anything. arXiv. 2024. <http://arxiv.org/abs/2304.02643>.
 37. Liang N, Sun S, Yu J, Farag Taha M, He Y, Qiu Z. Novel segmentation method and measurement system for various grains with complex touching. *Comput Electron Agric*. 2022;202:107351.
 38. Abdulla W. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow, GitHub repository. Github. 2017. https://github.com/matterport/Mask_RCNN.
 39. Müller T, Evans A, Schied C, Keller A. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans Graph*. 2022;41(4):1–15.
 40. Cole F, K Genova, A Sud, D Vlasic, Z Zhang. Differentiable surface rendering via non-differentiable sampling. Paper presented at: 2021 IEEE/CVF International Conference on Computer Vision (ICCV); Oct. 2021; Montreal, QC, Canada.
 41. Ester M, Kriegel H-P, Xu X. *A density-based algorithm for discovering clusters in large spatial databases with noise*. Portland (OR): AAAI Press; 1996. p. 226–231.
 42. Greenacre M, PJF G, Hastie T, D'Enza AI, Markos A, Tuzhilina E. Principal component analysis. *Nat Rev Methods Primer*. 2022;2(1):100.
 43. Cao J, Tagliasacchi A, Olson M, Zhang H, Su Z. Point cloud skeletons via Laplacian based contraction. Paper presented at:

- in 2010 Shape Modeling International Conference; Jun. 2010; Aix-en-Provence, France.
44. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. Paper presented at: 2009 IEEE Conference on Computer Vision and Pattern Recognition; Jun. 2009; Miami, FL, USA.
 45. Kornblith S, Shlens J, Le QV. Do Better ImageNet Models Transfer Better? Paper presented at: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Jun. 2019; Long Beach, CA, USA.
 46. Jiang S, Jiang C, Jiang W. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. *ISPRS J Photogramm Remote Sens.* 2020;167:230–251.
 47. Furukawa Y, Hernández C. Multi-view stereo: A tutorial. *Found Trends Comput Graph Vis.* 2015;9(1–2):1–148.
 48. Harandi N, Vandenberghe B, Vankerschaver J, Depuydt S, Van Messem A. How to make sense of 3D representations for plant phenotyping: A compendium of processing and analysis techniques. *Plant Methods.* 2023;19(1):60.
 49. Yu L, Shi J, Huang C, Duan L, Wu D, Fu D, Wu C, Xiong L, Yang W, Liu Q. An integrated rice panicle phenotyping method based on X-ray and RGB scanning and deep learning. *Crop J.* 2021;9(1):42–56.
 50. Qiu R, He Y, Zhang M. Automatic detection and counting of wheat spikelet using semi-automatic labeling and deep learning. *Front Plant Sci.* 2022;13:872555.
 51. Sun J, Ren Z, Cui J, Tang C, Luo T, Yang W, Song P. A high-throughput method for accurate extraction of intact rice panicle traits. *Plant Phenomics.* 2024;6:0213.
 52. Sandhu J, Zhu F, Paul P, Gao T, Dhatt BK, Ge Y, Staswick P, Yu H, Walia H. PI-Plat: A high-resolution image-based 3D reconstruction method to estimate growth dynamics of rice inflorescence traits. *Plant Methods.* 2019;15(1):162.
 53. Xiao S, Ye Y, Fei S, Chen H, Zhang B, Li Q, Cai Z, Che Y, Wang Q, Ghafoor A, et al. High-throughput calculation of organ-scale traits with reconstructed accurate 3D canopy structures using a UAV RGB camera with an advanced cross-circling oblique route. *ISPRS J Photogramm Remote Sens.* 2023;201:104–122.
 54. Zhu B, Zhang Y, Sun Y, Shi Y, Ma Y, Guo Y. Quantitative estimation of organ-scale phenotypic parameters of field crops through 3D modeling using extremely low altitude UAV images. *Comput Electron Agric.* 2023;210:107910.
 55. Dong X, Kim W-Y, Zheng Y, Oh J-Y, Ehsani R, Lee K-H. Three-dimensional quantification of apple phenotypic traits based on deep learning instance segmentation. *Comput Electron Agric.* 2023;212:108156.