agronomy

MDPI

Article

High-Throughput 3D Rice Chalkiness Detection Based on Micro-CT and VSE-UNet

Zhiqi Cai [1], Yangjun Deng [1], Xinghui Zhu [1], Bo Li [1], Chenglin Xu [2] and Donghui Li [1,*]

[1] College of Information and Intelligence, Hunan Agricultural University, Changsha 410128, China; onedecision@stu.hunau.edu.cn (Z.C.); dengyangjun@hunau.edu.cn (Y.D.); zhuxh@hunau.net (X.Z.); hunaulibo@hunau.edu.cn (B.L.)

[2] College of Computer Science and Technology, Hengyang Normal University, Hengyang 421002, China; xuchenglin@hynu.edu.cn

* Correspondence: doner@hunau.edu.cn

Abstract: Rice is a staple food for nearly half the global population and, with rising living standards, the demand for high-quality grain is increasing. Chalkiness, a key determinant of appearance quality, requires accurate detection for effective quality evaluation. While traditional 2D imaging has been used for chalkiness detection, its inherent inability to capture complete 3D morphology limits its suitability for precision agriculture and breeding. Although micro-CT has shown promise in 3D chalk phenotype analysis, high-throughput automated 3D detection for multiple grains remains a challenge, hindering practical applications. To address this, we propose a high-throughput 3D chalkiness detection method using micro-CT and VSE-UNet. Our method begins with non-destructive 3D imaging of grains using micro-CT. For the accurate segmentation of kernels and chalky regions, we propose VSE-UNet, an improved VGG-UNet with an SE attention mechanism for enhanced feature learning. Through comprehensive training optimization strategies, including the Dice focal loss function and dropout technique, the model achieves robust and accurate segmentation of both kernels and chalky regions in continuous CT slices. To enable high-throughput 3D analysis, we developed a unified 3D detection framework integrating isosurface extraction, point cloud conversion, DBSCAN clustering, and Poisson reconstruction. This framework overcomes the limitations of single-grain analysis, enabling simultaneous multi-grain detection. Finally, 3D morphological indicators of chalkiness are calculated using triangular mesh techniques. Experimental results demonstrate significant improvements in both 2D segmentation (7.31% improvement in chalkiness IoU, 2.54% in mIoU, 2.80% in mPA) and 3D phenotypic measurements, with VSE-UNet achieving more accurate volume and dimensional measurements compared with the baseline. These improvements provide a reliable foundation for studying chalkiness formation and enable high-throughput phenotyping.

Keywords: rice chalkiness; micro-CT; deep learning; 3D reconstruction

Academic Editor: Jayme Garcia Arnal Barbedo

Received: 16 January 2025
Revised: 9 February 2025
Accepted: 11 February 2025
Published: 12 February 2025

Citation: Cai, Z.; Deng, Y.; Zhu, X.; Li, B.; Xu, C.; Li, D. High-Throughput 3D Rice Chalkiness Detection Based on Micro-CT and VSE-UNet. *Agronomy* **2025**, *15*, 450. https://doi.org/10.3390/agronomy15020450

1. Introduction

Rice (*Oryza sativa* L.) is one of the three major staple crops worldwide, feeding nearly half of the global population. With the improvement in living standards, the demand for high-quality rice has been increasing [1]. The appearance quality of rice is primarily characterized by grain shape, chalkiness, and translucency; among which, chalkiness serves as a crucial indicator affecting the overall appearance quality [2]. It is characterized by

white opaque areas caused by the uneven distribution of starch granules within the grain and can be further classified (based on the location) into belly chalkiness, center chalkiness, and back chalkiness [3]. Chalkiness, a negative trait that seriously affects the milling, appearance, and eating quality of rice, plays an important role in the market value assessment of rice [4]. Therefore, improving rice quality and breeding new varieties with superior characteristics and lower chalkiness has become an urgent priority [5]. Although the research on the genetic mechanisms underlying rice chalkiness formation has made some progress [6], the lack of efficient phenotypic mutant screening methods for chalkiness poses significant technical challenges for a deeper understanding of its formation mechanism. Therefore, developing a fully automated, high-throughput, and high-precision non-destructive detection method for chalkiness is crucial to overcoming the current research bottleneck and advancing the study of chalkiness formation mechanisms.

In the field of chalkiness detection, scholars have conducted numerous studies. Wang et al. utilized convolutional neural network (CNN) models and gradient-weighted class activation mapping (Grad-CAM) visualization techniques to accurately distinguish between chalky and non-chalky grains [7]. Saha et al. employed advanced light transmission imaging technology to achieve quantitative evaluation of rice chalkiness under near-infrared (NIR) light sources [8]. Zhou et al [9]. innovatively proposed YO-LACTS, a novel algorithm combining You Only Look Once version 5 (YOLOv5) [10] and You Only Look At CoefficienTs (YOLACT) [11] networks for fine-grained segmentation of adhesive rice grains, laying the foundation for applications such as rice grain counting and physical/chemical index detection, including grain shape analysis. Zia et al. developed the "National Grain Tech" system, an efficient desktop application for rice quality assessment that leverages computer vision and machine learning to analyze key features, including grain length, width, and chalky kernels [12]. He et al. cleverly combined convolutional neural networks with image processing techniques by overlaying NIR and RGB images and integrating the watershed algorithm with the Otsu adaptive threshold function to achieve the intelligent detection of rice appearance quality [13]. Although machine learning techniques have brought significant breakthroughs to chalkiness detection, existing studies have critical limitations, e.g., scholars primarily rely on two-dimensional (2D) planar images for analysis, which inherently struggle to capture the complete three-dimensional (3D) morphological features of rice grains. The 2D image processing methods have inherent deficiencies in extracting depth and stereoscopic information, potentially leading to the incomplete recognition and understanding of chalkiness characteristics.

In the field of 3D detection technology, Yu [14] attempted to use terahertz imaging technology to scan and generate cross-sectional images of polished rice, which were then converted into 3D images to ultimately improve the accuracy of rice chalkiness measurement. However, during the testing process, the limited penetration of the terahertz spectrometer's light source only allowed for the formation of rough contours of the rice grains, with the edges of the chalkiness regions being somewhat blurred. This posed challenges for edge detection. Su Yi et al. were the first to ingeniously combine X-ray micro-computed tomography (micro-CT) technology with the advanced medical image analysis software Mimics Innovation Suite (MIS) trial version (Materialise, Belgium). This groundbreaking integration enabled non-destructive three-dimensional quantitative analysis and detailed visualization of chalkiness traits in rice grains, opening up an innovative pathway for plant phenomics research [15]. Subsequently, we achieved the preliminary quantitative computation of chalkiness in individual rice grains but had yet to realize high-throughput processing, that is, the simultaneous processing of multiple rice grains [16]. While 3D imaging techniques show unique advantages in characterizing chalkiness morphology, several technical bottlenecks remain to be addressed. First, existing micro-CT-based approaches often require professional commercial software suites such as MIS for

data processing [15], which introduces substantial manual intervention and limits analytical throughput. Second, when extending 2D semantic segmentation to 3D reconstruction, conventional pixel-level segmentation struggles to maintain instance consistency across serial sections [16]. Although recent years have seen academia beginning to explore the synergistic innovation of deep learning and micro-CT technology, yielding certain advancements in other plant phenotypic detections [17–19], the existing research still faces significant challenges in high-throughput three-dimensional chalkiness detection. In particular, there is still no mature solution for the three-dimensional segmentation of chalkiness regions in multiple rice grains. Inspired by [20], which effectively applied density-based spatial clustering of applications with noise (DBSCAN) clustering for 3D point cloud segmentation in plant phenotyping, we propose a 3D point cloud clustering segmentation scheme tailored for high-throughput rice chalkiness detection, overcoming the bottleneck of high-throughput processing.

In summary, this paper proposes a rice chalkiness detection method based on micro-CT and VGG16-squeeze-and-excitation-UNet (VSE-UNet, where VGG16 represents the visual geometry group 16-layer network, SE denotes the squeeze-and-excitation attention mechanism, and UNet is the U-shaped network architecture). The specific contributions are as follows:

- Proposed a VSE-UNet semantic segmentation model for rice grain micro-CT images. By integrating the SE attention mechanism into the VGG16-UNet (VGG-UNet) base model, optimizing the loss function, and implementing dropout techniques, the model achieves enhanced accuracy in chalky region identification.
- Developed a unified 3D chalkiness detection framework by integrating the VSE-UNet model, marching cubes (MC) algorithm, DBSCAN, and Poisson reconstruction algorithm. This framework enables automated high-throughput 3D chalkiness detection for multiple rice grains simultaneously, overcoming the single-grain limitation of traditional methods.
- Established a comprehensive micro-CT image dataset of unhulled rice grains, comprising 2100 micro-CT images with their manually annotated labels, providing essential data support for chalkiness detection research and addressing the current dataset scarcity in rice chalkiness micro-CT imaging.

2. Materials and Methods

2.1. Dataset

This study focuses on the Xiangzao X220 (X220) rice variety, with samples provided by the Rice Research Institute of the Hunan Academy of Agricultural Sciences. CT images of the rice grains were obtained using the SkyScan 1172 micro-CT system manufactured by Bruker, Billerica, MA, USA. The device parameters were set as follows: image resolution of 2050 × 2050 pixels, voxel size of 5 × 5 × 5 μm^3, scanning voltage of 50 kV, and current of 140 μA. Under these settings, unhulled and active rice samples (i.e., rice grains) were scanned, resulting in the export of 2100 24-bit BMP format images.

The overall workflow of data collection and annotation is illustrated in Figure 1. Rice grain samples were first scanned using the SkyScan 1172 Micro-CT scanner to obtain micro-CT images, which were then manually annotated using LabelMe software (version 3.16.7) to mark the regions of interest.

LabelMe software was employed to manually annotate the kernels (specifically brown rice kernels) and chalky regions within the micro-CT images of the rice grain samples. Throughout this paper, all kernels are referred to simply as “kernels”. Kernels were marked in red, and chalky regions were marked in green. The annotations captured the boundary information of each component in the images, and the results were stored in a

JavaScript Object Notation (JSON) format. The JSON files recorded detailed information from the annotation process, including the coordinates, shapes, and attributes of the annotated regions, providing a foundation for subsequent data analysis. The JSON files were then converted into label images, where black areas represented the background, red areas indicated the kernels, and green areas denoted the chalky regions, as illustrated in Figure 2.

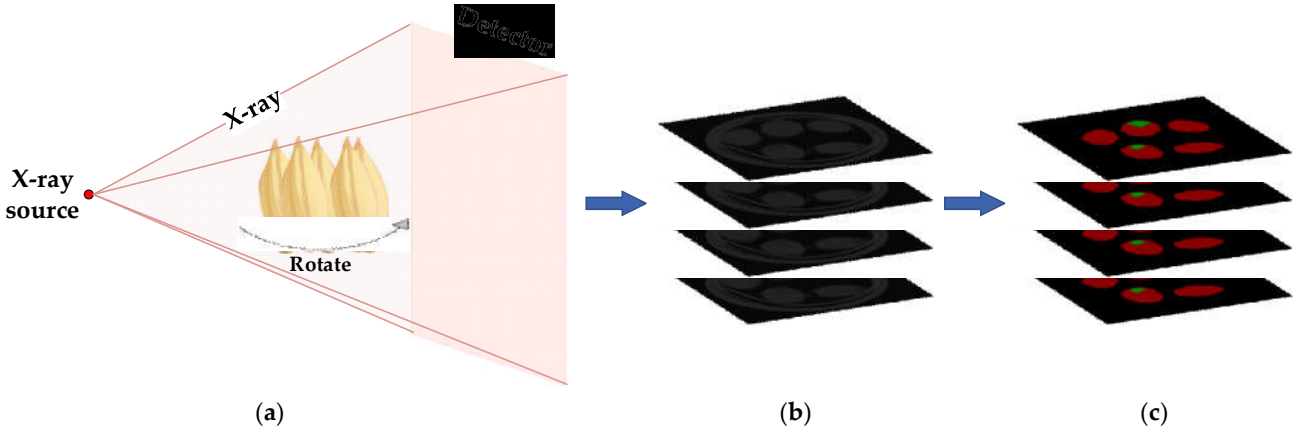


Figure 1. Schematic illustration of rice grain X-ray micro-CT imaging and annotation process: (**a**) Schematic illustration of X-ray micro-CT imaging principle; (**b**) raw micro-CT slice images; (**c**) manually annotated micro-CT slices.

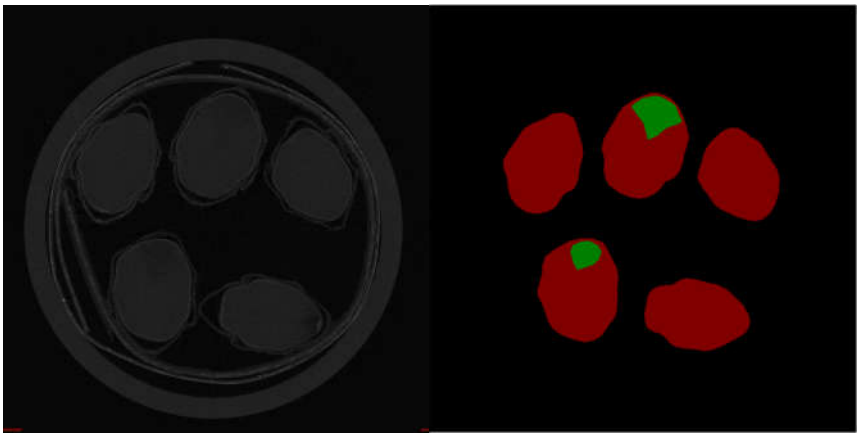


Figure 2. Image annotation status: (**a**) micro-CT image of original rice grain; (**b**) image annotation results. Black represents the background, red represents the rice kernel, and green represents the chalky areas.

For dataset partitioning, the 2100 micro-CT images of rice samples and their corresponding label images were randomly divided into training, validation, and test sets at a 7:2:1 ratio, resulting in 1470 training images, 420 validation images, and 210 test images. To enhance the model's learning capacity and improve its robustness and generalization ability, data augmentation techniques such as flipping and rotation were applied to the training set, increasing the number of training images to 5880.

To ensure fair comparison, all methods used identical dataset splits generated with a fixed random seed. The same validation set (420 images) and test set (210 images) were

used across all experiments for model selection and evaluation, respectively. This ensures that performance differences between methods reflect true algorithmic variations.

2.2. VSE-UNet Model

One of the primary challenges in this segmentation task is the significant imbalance in pixel distribution within the micro-CT images of rice grains. Specifically, the number of background pixels far exceeds that of the kernels and chalky regions, and the pixel count of the kernel areas is substantially higher than that of the chalky regions. This class imbalance restricts the model's effectiveness in feature extraction during the training process. These factors pose significant challenges for accurate semantic segmentation of rice grain micro-CT images.

To address these issues, this study enhances the VGG-UNet model [16] by incorporating the squeeze-and-excitation (SE) attention mechanism [21], and integrates the Dice focal loss (DFL) [22], a combined loss function merging Dice loss (DL) with focal loss (FL), along with a dropout mechanism [23], thereby constructing the VSE-UNet model (its architecture is illustrated in Figure 3).

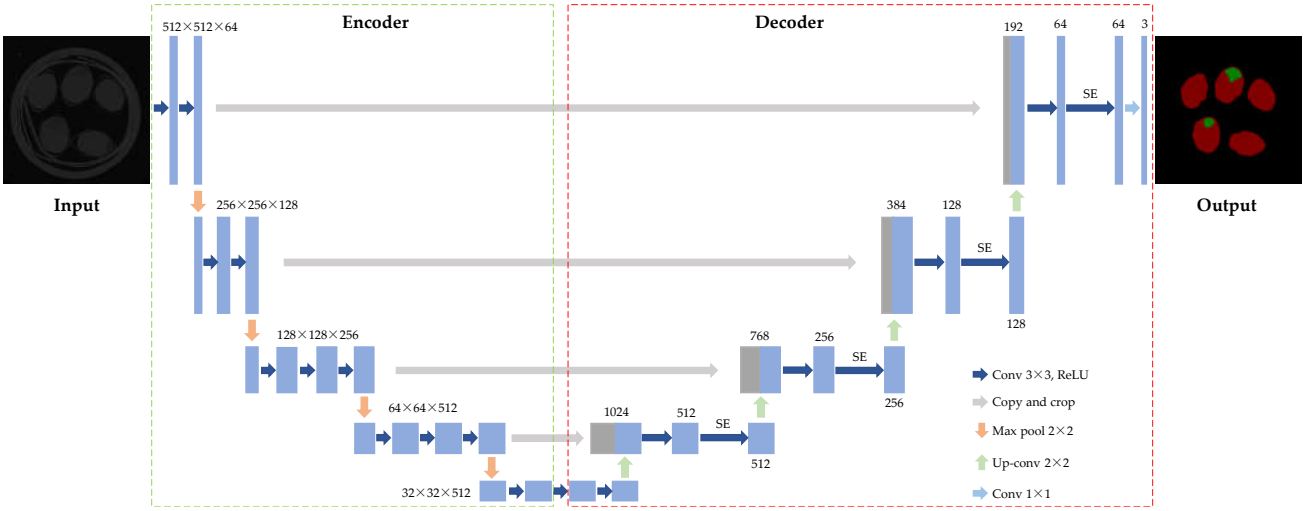


Figure 3. The architecture of VSE-UNet. SE refers to the squeeze-and-excitation attention mechanism.

The VSE-UNet architecture follows the classic U-Net [24] design, utilizing an encoder–decoder structure with skip connections for preserving spatial information. It employs the convolutional and pooling layers of VGG16 [25] as the backbone feature extraction network, leveraging their proven capability in hierarchical feature learning. Specifically, the encoder first resizes the input micro-CT images to 512 × 512 pixels, then progressively extracts features through multiple convolutional and pooling operations. In the decoder, feature maps are upsampled and merged with corresponding encoder features via skip connections, followed by SE attention mechanism integration to enhance feature representation.

To effectively mitigate the class imbalance issue, VSE-UNet employs DFL as its loss function. The DL component helps address class imbalance by calculating the overlap between predicted and ground truth segmentations, while the FL component automatically down-weights the contribution of easy examples during training, forcing the model to focus on harder cases. This combination enhances the model's ability to handle both class imbalance and challenging samples, thereby improving overall segmentation accuracy. Additionally, a dropout layer (rate = 0.1) is implemented after the SE attention mechanism in the decoder to prevent overfitting and enhance model generalization.

2.2.1. VGG-UNet

The VGG-UNet model integrates the strengths of U-Net and VGG16, aiming to enhance the segmentation accuracy of rice grain micro-CT images by effectively distinguishing between kernels and chalky regions. Rice grain micro-CT images exhibit complex internal structures, necessitating segmentation tasks that simultaneously capture both global semantic information and local details. Traditional segmentation methods, such as the standard U-Net, often struggle to balance global and local features when processing high-resolution, detail-rich, and noisy images. Although U-Net is renowned for its skip connection mechanism, which effectively fuses shallow and deep features, its feature extraction capability is limited when dealing with complex structures, particularly in high-resolution images.

VGG16 is a classic convolutional neural network that employs 3 × 3 convolutions and pooling layers to extract hierarchical features. Pre-trained on large-scale datasets, VGG16 provides robust support for transfer learning [26]. VGG-UNet utilizes the convolutional and pooling layers of VGG16 as the backbone feature extraction network and integrates them with the decoder component of U-Net. This combination effectively enhances the network's generalization capability and segmentation performance when handling complex images.

Compared with other commonly used segmentation networks, such as ResNet50-UNet [27] and the standard U-Net, VGG-UNet demonstrates significant advantages in segmentation accuracy and stability. The standard U-Net tends to exhibit blurred boundaries and loss of details when processing high-noise and detail-rich images. In contrast, VGG-UNet effectively mitigates these issues through deep feature extraction and multi-scale fusion. Additionally, with fewer parameters compared with ResNet50-UNet, VGG-UNet achieves a better balance between model capacity and computational complexity, making it more suitable for high-throughput data processing requirements.

2.2.2. SE Attention Mechanism

Attention mechanisms mimic the human visual system, guiding models to focus on crucial objects and their spatial locations. Various attention modules, such as CA (coordinate attention) [28], CBAM (convolutional block attention module) [29], and SE (squeeze-and-excitation), have been widely applied to tasks like image classification and semantic segmentation, achieving remarkable success [30].

The structure of the SE module is illustrated in Figure 4. Let **X** be the feature map after convolution, with spatial dimensions **H** (height), **W** (width), and **C** (channel count). The SE module first employs a squeeze operation to compress the feature map, followed by an excitation operation to reweight the channels, producing a recalibrated feature map $\tilde{\mathbf{X}}$ [31].

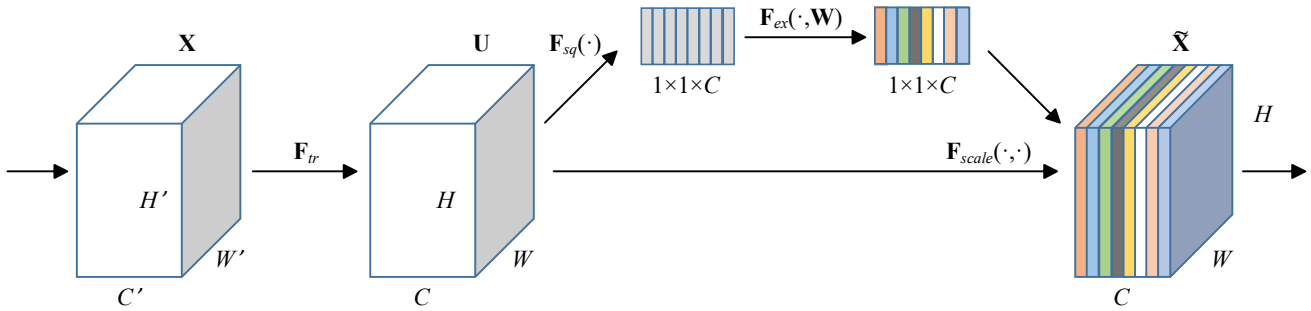


Figure 4. The structure of the SE attention mechanism.

In the case of severe pixel distribution imbalance in rice grain micro-CT images, the SE attention mechanism can dynamically adjust the feature weights for the background,

kernels, and chalky regions, enhancing the feature response in key areas and effectively alleviating the feature extraction challenges caused by class imbalance. This adaptive feature recalibration mechanism provides more focused and discriminative feature representations for precise segmentation of rice grain micro-CT images.

2.2.3. Dice Focal Loss and Dropout Mechanism

In the segmentation task of rice grain micro-CT images, due to the class imbalance (the significant difference in pixel numbers between the background, kernels, and chalky regions), traditional cross-entropy loss (CEL) assigns equal weight to all samples. This leads to the model favoring the background class and easily neglecting the kernels and chalky regions. To address this issue, we adopt the DFL, a combined loss function that integrates DL [32] and FL [33].

The DL and FL components are defined in Equations (1) and (2), respectively, with their combination (DFL) shown in Equation (3):

$$DL = \frac{1}{C}\sum_{c=1}^{C}\left(1 - \frac{2|X_c \cap Y_c| + \epsilon}{|X_c| + |Y_c| + \epsilon}\right) \quad (1)$$

$$FL(p_t) = -(1 - p_t)^{\gamma}\log(p_t) \quad (2)$$

$$DFL = DL + \alpha FL \quad (3)$$

DL is an overlap-based loss function that measures the similarity between the predicted result and the true label. For multiclass problems, DL is calculated for each class individually and then averaged, as shown in Equation (1), where $C = 3$ corresponds to the three classes: background, kernel, and chalky region. The smoothing factor ϵ (set to 1e−5 in this study) ensures numerical stability by preventing division by zero errors.

FL, as defined in Equation (2), introduces a modulation factor $(1 - p_t)^{\gamma}$ to dynamically adjust sample weights, where pt represents the class probability of correct prediction. With the focusing parameter γ set to 2, a value determined based on prior work [33], FL directs the model's attention toward hard-to-classify samples, particularly beneficial for minority classes like kernels and chalky regions.

The combined DFL, as defined in Equation (3), integrates these advantages with a weighting factor $\alpha = 0.5$, a value determined based on prior work [22], to balance the contributions of DL and FL. This combination enables more effective learning across different classes, significantly improving overall segmentation performance.

Additionally, we implement a dropout mechanism (rate = 0.1) after the SE attention layers to prevent overfitting. By randomly deactivating neurons during training, dropout enhances the model's generalization ability, particularly improving its performance on fine textures and boundary features in imbalanced datasets. This is especially important for rice chalkiness detection, where accurate boundary detection between different regions is crucial, requiring the model to maintain robust feature extraction capabilities without overfitting to specific patterns.

2.3. Unified 3D Chalkiness Detection Framework

High-throughput 3D chalkiness detection methods for unhulled rice grains are still not fully mature, with most current approaches processing single grains individually. To address these limitations, this study establishes a unified 3D chalkiness detection framework that integrates the VSE-UNet model with clustering and reconstruction algorithms. The framework comprises the following steps: 2D segmentation, 3D reconstruction, multiple grain separation using DBSCAN, surface reconstruction, and metric calculation. The

entire process is illustrated in Figure 5. Below, we provide a detailed explanation using a set of micro-CT images containing five rice grains as an example.

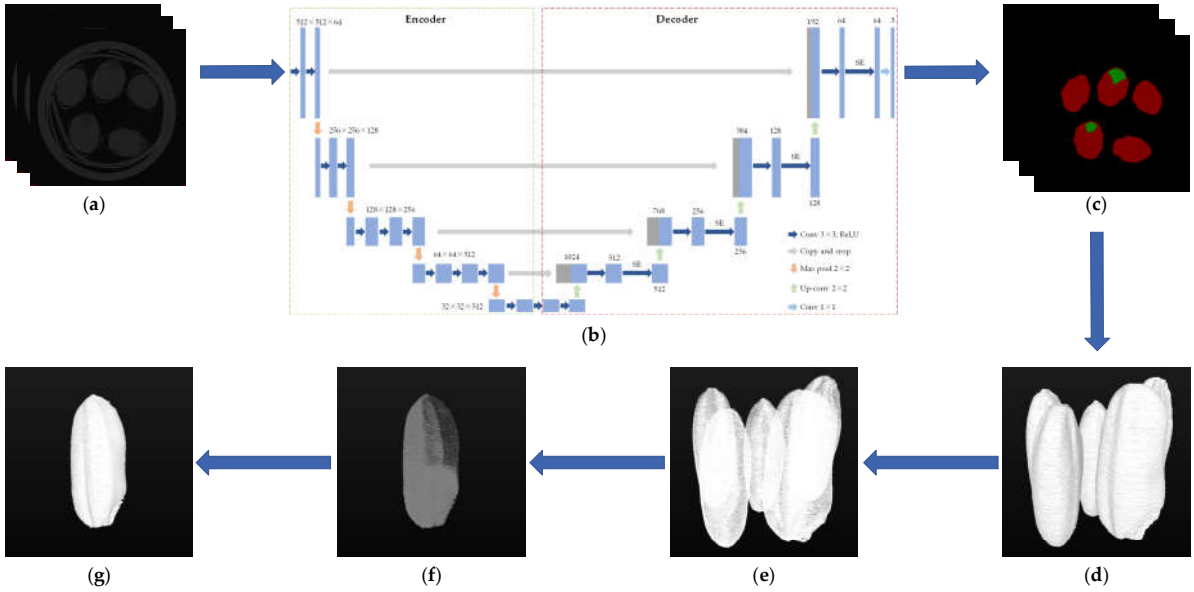


Figure 5. Unified 3D chalkiness detection framework demonstrates the complete process of rice grain analysis. (**a**) Original micro-CT scan image of rice grains showing the raw input data; (**b**) implementation of VSE-UNet for image segmentation; (**c**) generated mask image highlighting the kernel and chalkiness regions; (**d**) 3D surface reconstruction model showing 5 rice kernels; (**e**) point cloud representation of 5 rice kernels demonstrating the spatial distribution; (**f**) detailed point cloud model focusing on a single rice kernel; (**g**) surface reconstruction model of an individual rice kernel showing detailed morphological features.

First, the trained VSE-UNet model is employed to perform 2D segmentation on the micro-CT images, generating mask images that clearly delineate the contours of the rice kernels (red) and the chalkiness regions (green), providing a foundation for subsequent 3D reconstruction. Next, based on these mask images, the MC algorithm [34,35] is utilized for 3D reconstruction, resulting in 3D surface models of the five kernels. The Open3D library [36] is utilized for 3D visualization and surface processing throughout the reconstruction pipeline, providing efficient point cloud manipulation and mesh rendering capabilities.

To separate the clustered kernels, the vertices of the surface models are extracted to form point clouds. The DBSCAN clustering algorithm [37] is then applied to isolate each kernel into individual point cloud models. Unlike previous studies that could only handle single rice kernels, our innovative approach successfully achieves high-throughput chalkiness detection for multiple rice kernels, a feat not accomplished in the prior research. Subsequently, the Poisson reconstruction algorithm [38] is used to perform surface reconstruction on the separated point clouds, with a hyperparameter depth of 8 following the optimal settings established in [39]. This algorithm demonstrates excellent robustness, capable of automatically filling gaps and smoothing surface details to a certain extent. Notably, the mesh models produced by the Poisson reconstruction algorithm are watertight, meaning there are no leaks or unclosed gaps within the models. This characteristic makes the reconstructed models suitable for volume calculations and subsequent analyses.

Chalkiness processing: The processing of chalkiness regions follows a parallel workflow but requires special consideration for accurate association with individual kernels. The minimum oriented bounding box (MOBB) method is essential here, as chalkiness may not be present in all kernels, and, without MOBB, it would be challenging or even require

manual intervention to correctly associate chalkiness regions with their corresponding kernels. After kernel segmentation, the MOBB of each kernel is used to extract the corresponding chalkiness point clouds. These extracted point clouds then undergo the same surface reconstruction process as the kernels.

Finally, based on the 3D reconstructions, multiple phenotypic indicators of the rice kernels and chalkiness areas are calculated. First, the volumes of the rice kernels (VK) and chalkiness areas (VC) are computed, and the 3D chalkiness rate (CD3D) is determined by the volume ratio:

$$\mathrm{CD3D} = \frac{\mathrm{VC}}{\mathrm{VK}} \tag{4}$$

where CD3D represents the 3D chalkiness rate, VC denotes the volume of the chalkiness area, and VK signifies the volume of the rice kernel.

Using the MOBB method, the lengths (L: the longest edge of the minimum oriented bounding box of the rice kernel), widths (W: the second longest edge of the minimum oriented bounding box), and thicknesses (T: the shortest edge of the minimum oriented bounding box) of the rice kernels are accurately calculated. Additionally, the length-to-width ratio (L/W: the ratio of the longest edge to the second longest edge of the minimum oriented bounding box) is used to assess the morphological characteristics of the kernels.

2.4. Training Environment and Evaluation Metrics

The VSE-UNet model is trained on a desktop computer running the Windows platform, utilizing the PyTorch deep learning framework and Python 3.9.0. The hardware configuration includes an Intel Core i5-13600K CPU (Intel Corporation, Santa Clara, CA, USA), NVIDIA GeForce RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA), and 64 GB of RAM. The Compute Unified Device Architecture (CUDA) version is 12.2, and the CUDA Deep Neural Network library (cuDNN) version is 8.9.7. The training is set for 150 iterations, initialized using pre-trained weights. During training, the network parameters are frozen for the first 50 iterations with a batch size of 8, and then unfrozen for the remaining 100 iterations with a batch size of 4. The Adam optimizer is employed with an initial learning rate of 0.0001, utilizing a cosine annealing strategy to adjust the learning rate. The momentum is set to 0.9, and the minimum learning rate is 0.01 times the initial learning rate.

This study evaluates the model's accuracy using the following metrics: mean intersection over union (mIoU), mean pixel accuracy (mPA), and accuracy. The calculation formulas are as follows:

$$\mathrm{IoU_i} = \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FP_i} + \mathrm{FN_i}} \tag{5}$$

$$\mathrm{mIoU} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{IoU_i} \tag{6}$$

$$\mathrm{mPA} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FN_i}} \tag{7}$$

$$\mathrm{Accuracy} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{TP_i} + \mathrm{TN_i})}{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{TP_i} + \mathrm{TN_i} + \mathrm{FP_i} + \mathrm{FN_i})} \tag{8}$$

where N represents the number of classes to be identified; i denotes the class; $\mathrm{TP_i}$ is the true positive for class i, indicating the number of correctly predicted pixels for class i; $\mathrm{TN_i}$ is the true negative for class i, indicating the number of correctly predicted non-class i pixels; $\mathrm{FP_i}$ is the false positive for class i, indicating the number of pixels incorrectly

predicted as class i; and FN_i is the false negative for class i, indicating the number of pixels that are actually class i but are not predicted as such.

3. Results

3.1. Semantic Segmentation Model Accuracy Evaluation

3.1.1. Comparison of Different Segmentation Models

To validate the superiority of the VSE-UNet model, we compare it with other common segmentation models. The experimental results are presented in Table 1.

Table 1. Comparison results of different segmentation models.

Model	mIoU (%)	mPA (%)	Accuracy (%)
DeepLabV3+ [40]	90.32	93.72	98.69
PSPNet [41]	90.29	93.45	99.67
HRNet [42]	90.36	93.56	99.70
ResNet50-UNet [27]	90.35	93.43	98.25
VGG-UNet [16]	90.61	93.63	99.77
VSE-UNet	**93.15**	**96.43**	**99.80**

Note: Bold values indicate the best performance for each item. mIoU represents mean intersection over union. mPA represents mean pixel accuracy.

From Table 1, it is evident that VSE-UNet outperforms DeepLabV3+ [40], PSPNet [41], HRNet [42], ResNet50-UNet [27], and VGG-UNet [16] in key metrics such as mean intersection over union (mIoU, 93.15%), mean pixel accuracy (mPA, 96.43%), and overall accuracy (99.80%). This demonstrates the significant superiority of VSE-UNet in the rice chalkiness segmentation task. VSE-UNet achieves the highest performance across all evaluated metrics, highlighting its effectiveness and robustness compared with other state-of-the-art models.

Overall, VSE-UNet not only leads other common models in segmentation accuracy but also ensures computational efficiency. This makes it suitable for high-throughput automated rice kernel chalkiness detection methods. Its excellent performance provides reliable technical support for practical deployment and holds broad application prospects.

3.1.2. Comparison of Different Attention Mechanisms in VGG-UNet

In rice grain micro-CT images, the chalky regions occupy a small area and have subtle differences compared with the kernels. To enhance the model's focus on kernels and chalky regions, thereby improving the accuracy of chalkiness segmentation, we incorporate different attention modules into the decoder part of VGG-UNet (positioned similarly to the SE attention mechanism in Figure 3) for comparison. The results are shown in Table 2.

Table 2. Performance comparison of VGG-UNet with different attention mechanisms.

Method	Background IoU (%)	Kernel IoU (%)	Chalkiness IoU (%)	mIoU (%)	mPA (%)	Accuracy (%)
VGG-UNet	99.86	97.21	74.75	90.61	93.63	99.77
VGG-UNet + CBAM	**99.87**	97.42	77.57	91.62	94.35	**99.79**
VGG-UNet + CA	99.85	97.14	75.44	90.81	93.84	99.76
VGG-UNet + SE	**99.87**	**97.50**	**78.63**	**92.00**	**94.73**	**99.79**

Note: + indicates the introduction of the corresponding attention mechanism to the base VGG-UNet model. Bold values indicate the best performance for each item. IoU represents intersection over union. mIoU represents mean intersection over union. mPA represents mean pixel accuracy. CBAM

represents convolutional block attention module. CA represents coordinate attention. SE represents squeeze-and-excitation.

From Table 2, it can be seen that the introduction of attention modules significantly improves the model's performance. Specifically, the VGG-UNet model with the SE (squeeze-and-excitation) module performs the best in chalkiness IoU (78.63%), mIoU (92.00%), mPA (94.73%), and accuracy (99.79%). This demonstrates that the SE module effectively enhances the model's focus on rice kernels and chalky regions, thereby significantly improving segmentation accuracy. In comparison, the CBAM and CA modules also improve chalkiness IoU and other metrics, but the SE module achieves the best results across all metrics, indicating its superior effectiveness in feature recalibration.

3.1.3. Comparison of Different Loss Functions

In this experiment, we compare the impact of different loss functions on the performance of the baseline model VGG-UNet. The results are shown in Table 3.

Table 3. Comparison results of different loss functions.

Loss Function	Background IoU (%)	Kernel IoU (%)	Chalkiness IoU (%)	mIoU (%)	mPA (%)	Accuracy (%)
CEL	**99.86**	**97.21**	74.75	90.61	93.63	**99.77**
FL	99.85	97.12	75.83	90.93	94.44	99.76
DL	99.83	96.86	76.67	91.12	95.70	99.74
DFL	99.83	96.98	**77.39**	**91.40**	**95.81**	99.75

Note: Bold values indicate the best performance for each item. IoU represents intersection over union. mIoU represents mean intersection over union. mPA represents mean pixel accuracy. CEL represents cross-entropy loss. FL represents focal loss. DL represents Dice loss. DFL represents Dice focal loss.

From Table 3, it is evident that, while CEL achieves marginally better performance in background IoU, kernel IoU, and accuracy, the DFL shows notable improvements in chalkiness IoU (77.39%), mIoU (91.40%), and mPA (95.81%), effectively addressing the class imbalance challenge. This indicates that DFL effectively alleviates the class imbalance problem and enhances the model's segmentation accuracy. Specifically, FL and DL show slight improvements in chalkiness IoU and other metrics, but their effects are not as significant as DFL. These results are consistent with previous studies [26], demonstrating that combined loss functions can integrate the advantages of individual losses to optimize the model's overall performance.

3.1.4. Comparison of Different Dropout Rates

To further optimize the model's performance, we experiment with different dropout rates and evaluate their impact on the baseline model VGG-UNet. The experimental results are presented in Table 4.

Table 4. Comparison results of different dropout rates.

Dropout Rates	Background IoU (%)	Kernel IoU (%)	Chalkiness IoU (%)	mIoU (%)	mPA (%)	Accuracy (%)
0	**99.86**	**97.21**	74.75	90.61	93.63	**99.77**
0.1	**99.86**	97.16	**76.23**	**91.08**	95.24	**99.77**
0.2	**99.86**	**97.21**	76.01	91.03	95.17	**99.77**
0.3	**99.86**	97.12	75.63	90.87	**95.65**	99.76

Note: Bold values indicate the best performance for each item. IoU represents intersection over union. mIoU represents mean intersection over union. mPA represents mean pixel accuracy.

From Table 4, it can be observed that when the dropout rate is set to 0.1, the model achieves peak performance in the key metrics—chalkiness IoU (76.23%) and mIoU (91.08%)—while the mPA (95.24%) is slightly lower than the 95.65% achieved, with a dropout rate of 0.3, but remains at a high level. This indicates that, with other performance metrics remaining similar, a dropout rate of 0.1 optimally balances the model's ability to recognize key features (chalkiness) and its overall segmentation accuracy. In contrast, higher dropout rates (0.2 and 0.3) can improve some metrics but sacrifice the critical chalkiness detection performance, and may negatively affect the model's learning capability due to excessive feature dropout. Therefore, a dropout rate of 0.1 is ultimately selected as the optimal configuration.

3.1.5. Ablation Experiments

To evaluate the contributions of the improvement modules and optimization strategies to the performance of the VSE-UNet model, we conduct ablation experiments analyzing the effects before and after the introduction of different improvement methods. The experimental results are shown in Table 5.

Table 5. Ablation study of attention mechanism and optimization strategies.

VGG-UNet	SE	DFL	Dropout	Background IoU (%)	Kernel IoU (%)	Chalkiness IoU (%)	MIoU (%)	mPA (%)	Accuracy (%)
√				99.86	97.21	74.75	90.61	93.63	99.77
√	√			**99.87**	97.50	78.63	92.00	94.73	99.79
√	√	√		**99.87**	**97.57**	81.13	92.86	95.42	**99.80**
√	√	√	√	99.86	97.54	**82.06**	**93.15**	**96.43**	**99.80**

Note: √ indicates the introduction of the corresponding method, while a blank space denotes the absence of that method. Bold values indicate the best performance for each item. SE represents squeeze-and-excitation. DFL represents Dice focal loss. IoU represents intersection over union. mIoU represents mean intersection over union. mPA represents mean pixel accuracy.

From the ablation study results presented in the table, it is evident that the gradual introduction of the SE module, dropout regularization, and DFL leads to significant improvements in chalkiness IoU, mIoU, mPA, and accuracy; specifically, as follows:

- Introduction of SE attention mechanism: Compared with the baseline VGG-UNet model, adding the SE module enhances chalkiness IoU by 3.88 percentage points, mIoU by 1.39 percentage points, and mPA by 1.10 percentage points. These results indicate that the SE module significantly improves the model's ability to extract features from key regions.
- Introduction of DFL: Building upon the VGG-UNet + SE model, incorporating the DFL further increases chalkiness IoU by 2.50 percentage points, mIoU by 0.86 percentage points, and mPA by 0.69 percentage points. This demonstrates that DFL effectively optimizes the training process and enhances the model's performance on hard-to-classify samples.
- Introduction of dropout regularization: In the VGG-UNet + SE + DFL model, adding dropout (with a rate of 0.1) leads to a further increase in chalkiness IoU by 0.93 percentage points, mIoU by 0.29 percentage points, and mPA by 0.99 percentage points. The accuracy remains stable at 99.80%. This indicates that dropout regularization effectively mitigates model overfitting during training, enhancing the model's generalization capability.

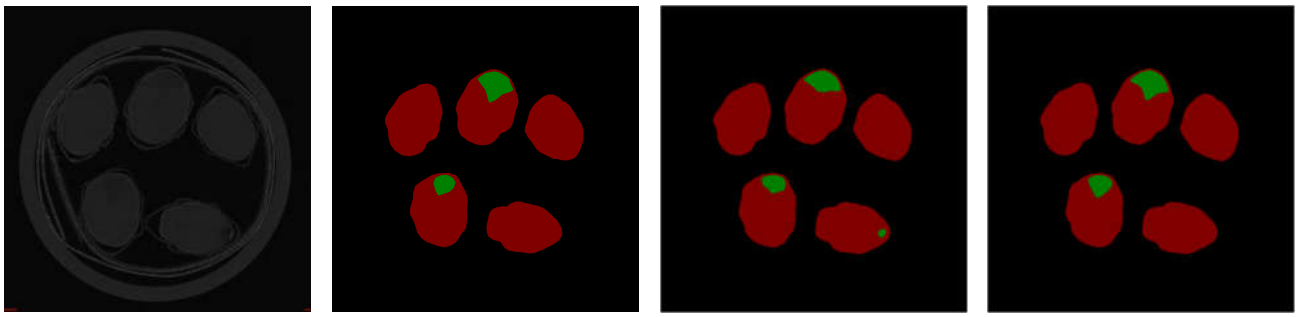
Overall, the integration of the SE module, dropout regularization, and DFL significantly improves the model's performance in the rice chalkiness segmentation task. By

leveraging attention mechanisms, regularization strategies, and a combined loss design, the VSE-UNet model attains new levels of segmentation accuracy and robustness.

3.2. Results and Analysis of Unified 3D Chalkiness Detection Framework

3.2.1. Comparison of 2D Segmentation Results

To validate the effectiveness of our proposed improvements, we compare the 2D segmentation results of VSE-UNet with those of the baseline VGG-UNet model. The comparison results are shown in Figure 6. As illustrated, VGG-UNet exhibits certain misclassifications in the detailed information of some chalky regions, resulting in discrepancies when compared with the ground truth labels. In contrast, our improved VSE-UNet model provides more accurate extraction of chalky regions, effectively capturing the complete edges and preserving subtle features. Consequently, the segmentation results of VSE-UNet are closer to the ground truth, showing significant improvements in boundary clarity, detail restoration, and overall consistency compared with the baseline model.



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

Figure 6. Comparison of 2D segmentation results: (**a**) original micro-CT image; (**b**) ground truth; (**c**) VGG-UNet segmentation results; (**d**) VSE-UNet segmentation results.

3.2.2. Comparison of 3D Surface Model Results

Building upon the 2D segmentation, to further assess the impact of the segmentation results on 3D surface reconstruction, we compare the 3D surface models reconstructed from the segmentation results of VGG-UNet and VSE-UNet. As shown in Figure 7, both models achieve satisfactory reconstruction results for the kernel surface models, showing similar performance in capturing the overall kernel morphology.

(**a**) (**b**) (**c**)

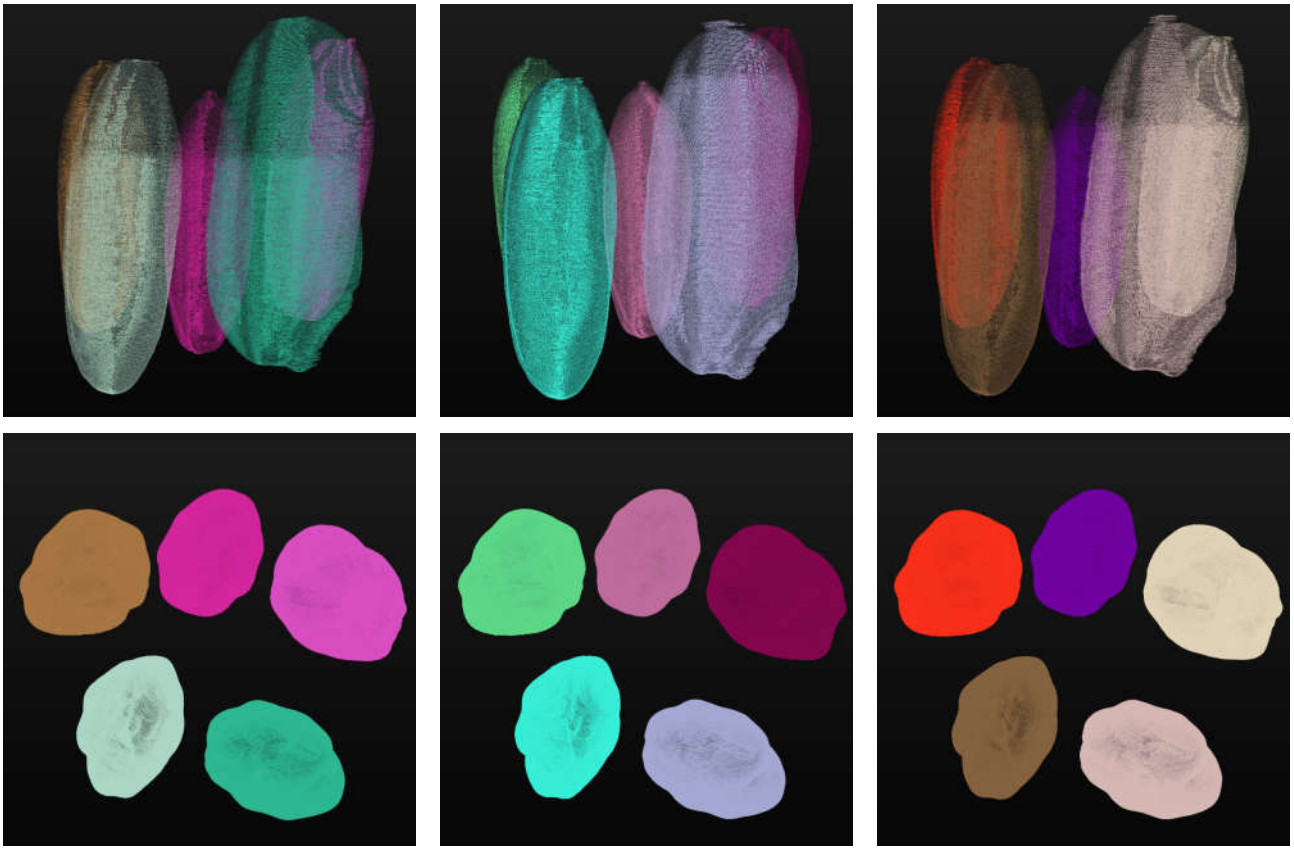
Figure 7. Comparison of 3D surface model results: (**a**) ground truth reconstruction; (**b**) VGG-UNet segmentation reconstruction; (**c**) VSE-UNet segmentation reconstruction.

This study utilizes the Open3D library to visualize the kernel surface models after surface reconstruction, providing an intuitive demonstration of the effectiveness of the segmentation and reconstruction methods [33]. The visualization process includes loading the 3D surface models, setting view parameters, and rendering the mesh models, with the final results presented in Figure 7. The comparison shows that both VGG-UNet and VSE-UNet can effectively reconstruct the kernel surface models, with their results closely resembling the models reconstructed from the ground truth. This establishes a solid foundation for the precise calculation of chalkiness-related phenotypic metrics.

In summary, both VGG-UNet and VSE-UNet demonstrate good performance in kernel surface reconstruction, with their 3D visualization results showing comparable quality. These findings indicate that both models possess sufficient robustness and reliability in practical applications, providing strong support for high-throughput automated rice quality assessment and breeding optimization.

3.2.3. Comparison of Three-Dimensional Segmentation Results

After extracting the vertices of the kernel surface model to generate a point cloud model, we apply the DBSCAN clustering algorithm for three-dimensional segmentation. The segmentation results are shown in Figure 8. From these results, it is apparent that different colors represent distinct clusters, corresponding to individual kernels. The DBSCAN-based segmentation method successfully isolates the point cloud models of single kernels.

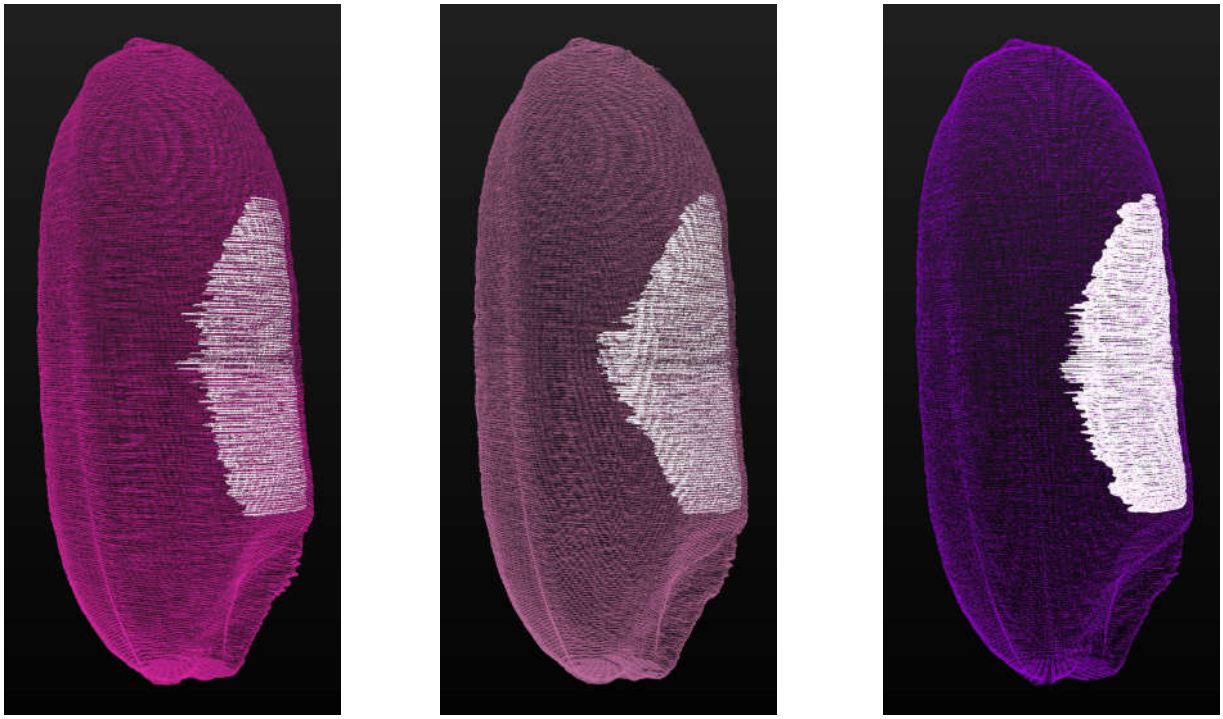


(**a**) (**b**) (**c**)

Figure 8. Comparison of segmentation results of 3D point cloud models: (**a**) segmented point cloud model from ground truth; (**b**) segmented point cloud model using VGG-UNet; (**c**) segmented point cloud model using VSE-UNet.

3.2.4. Comparison of 3D Segmentation Models for Chalkiness Regions

In this section, we employ the minimum oriented bounding boxes of the individually segmented kernels to extract the point cloud models of their corresponding chalkiness regions. The segmentation result for a single kernel is presented in Figure 9. Upon comparison, it is evident that the chalkiness regions segmented by the VSE-UNet model more closely align with the annotations.

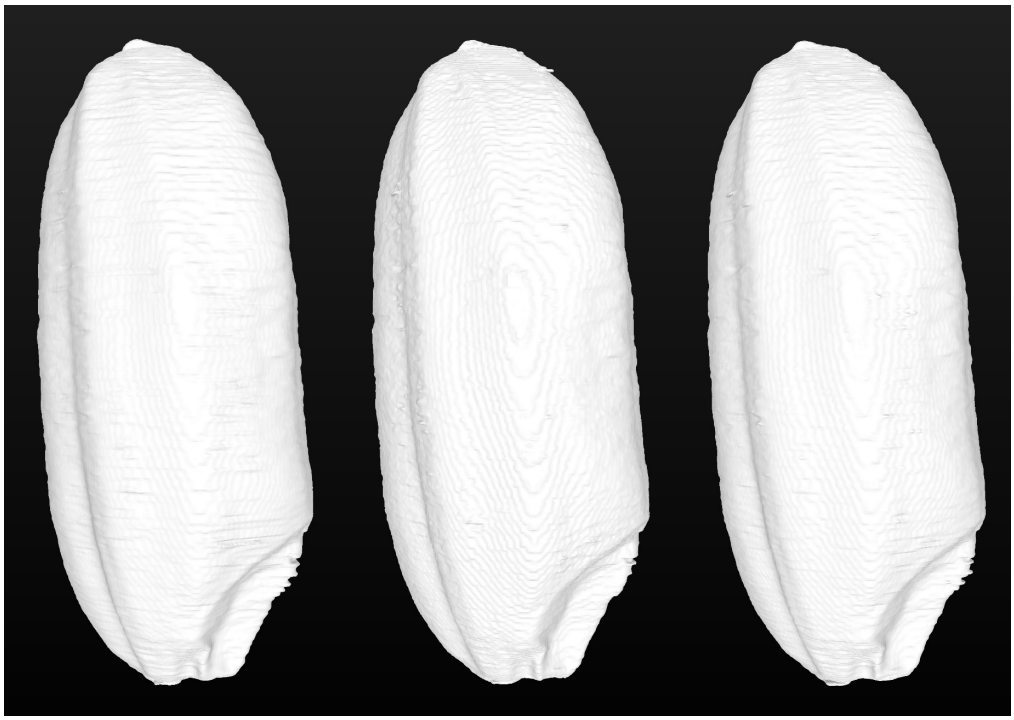


(**a**) (**b**) (**c**)

Figure 9. Comparison of point cloud models of a single kernel and its chalkiness region: (**a**) point cloud model of the kernel and its chalkiness region from ground truth; (**b**) point cloud model of the kernel and its chalkiness region using VGG-UNet; (**c**) point cloud model of the kernel and its chalkiness region using VSE-UNet.

3.2.5. Comparison of Surface Reconstruction Results After 3D Segmentation

In this study, we perform Poisson surface reconstruction on the point cloud models of the individually segmented rice kernels to obtain surface models. The reconstruction results for single kernels are shown in Figure 10. The comparison demonstrates that both VGG-UNet and VSE-UNet achieve comparable performance in reconstructing kernel surface models, with both methods effectively capturing the kernel morphology and producing high-quality surface reconstructions that closely resemble the ground truth model.



(**a**) (**b**) (**c**)

Figure 10. Comparison of surface models of a single kernel reconstructed from point cloud models: (**a**) surface model of a single kernel reconstructed from the ground truth point cloud model; (**b**) surface model of a single kernel reconstructed from the point cloud model segmented using VGG-UNet; (**c**) surface model of a single kernel reconstructed from the point cloud model segmented using VSE-UNet.

3.3. Phenotypic Indicator Calculation Results

The proposed method is applied to a series of consecutive CT images of five rice samples, achieving automated quantitative analysis and 3D reconstruction of chalkiness traits and other related phenotypes. The indicator results are detailed in Table 6.

Table 6. Calculation results of phenotypic indicators for rice samples.

Indicator	Method	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
VK (mm^3)	label	17.17	16.58	15.56	16.53	15.56
	VGG-UNet	17.06	16.47	15.45	16.72	15.45
	VSE-UNet	17.12	16.56	15.53	16.71	15.57
VC (mm^3)	label	1.55	1.40	0.31	0.37	0.36
	VGG-UNet	1.81	1.66	0.42	0.47	0.45
	VSE-UNet	1.68	1.55	0.39	0.41	0.43
CD3D (%)	label	9.00	8.45	1.99	2.24	2.34
	VGG-UNet	10.61	10.10	2.74	2.82	2.88
	VSE-UNet	9.81	9.34	2.52	2.43	2.74
L (mm)	label	6.20	5.98	6.13	6.34	5.86
	VGG-UNet	6.18	6.05	6.14	6.29	6.03
	VSE-UNet	6.18	6.02	6.13	6.31	5.92
W (mm)	label	2.66	2.72	2.64	2.63	2.64
	VGG-UNet	2.72	2.76	2.66	2.69	2.71
	VSE-UNet	2.67	2.73	2.65	2.65	2.65
T (mm)	label	1.99	2.00	1.91	2.02	1.93

	VGG-UNet	1.98	2.01	1.89	2.00	1.90
	VSE-UNet	1.99	2.03	1.89	2.01	1.91
L/W	label	2.33	2.20	2.32	2.41	2.22
	VGG-UNet	2.27	2.19	2.30	2.34	2.23
	VSE-UNet	2.31	2.21	2.32	2.38	2.23

Note: VK (mm^3) and VC (mm^3) represent the volumes of the rice kernels and chalkiness areas, respectively. CD3D (%) represents the 3D chalkiness rate, calculated as the proportion of the chalkiness volume relative to the total rice kernel volume. L (mm), W (mm), and T (mm) represent the length, width, and thickness of the rice kernel, respectively, calculated as the longest, second longest, and shortest edges of the minimum-oriented bounding box. L/W represents the length-to-width ratio, calculated as the ratio of the longest edge to the second longest edge of the minimum-oriented bounding box.

From Table 6, the following observations can be made: The VK measured using VGG-UNet and VSE-UNet generally aligns with the label values across all samples. While VGG-UNet shows slight underestimation in most samples, VSE-UNet maintains values closer to the label, indicating comparable accuracy. For the VC, both models show slight overestimation of the values, with VSE-UNet providing results that are relatively closer to the label values. In terms of the CD3D, VSE-UNet consistently gives values closer to the label compared with VGG-UNet, though both models show some degree of overestimation. Regarding kernel dimensions (L, W, T), both models perform well, with VSE-UNet showing minimal deviations from the label values, indicating that it accurately captures kernel morphology. The L/W values obtained by VSE-UNet closely align with the label values across all samples, with an exact match in Sample 3, suggesting reliable morphological assessments.

Overall, the VSE-UNet model demonstrates comparable or slightly improved accuracy and reliability in calculating phenotypic indicators compared with VGG-UNet. The close alignment of VSE-UNet's measurements with the label values across most indicators and samples underscores its effectiveness in automated phenotypic analysis. These precise 3D phenotypic data provide valuable support for evaluating and improving rice chalkiness traits, contributing to a better understanding of chalkiness formation mechanisms and laying a solid foundation for accurate rice quality assessment and optimization.

4. Discussion

The development of high-throughput phenotyping methods for rice quality traits, particularly chalkiness, is crucial for modern breeding and quality assessment. Our study demonstrates the feasibility and advantages of combining micro-CT imaging with deep learning for automated 3D chalkiness detection. Several aspects of this approach warrant further discussion.

First, the improved VSE-UNet model demonstrates consistent performance across different samples (Table 6). The integration of the SE attention mechanism enhances the model's feature extraction capabilities by dynamically adjusting channel-wise feature responses. This adaptive weighting helps capture subtle variations in X-ray attenuation that reflect differences in starch packing density within the grain [15]. These attenuation variations correspond to the distinct starch granule organization observed in chalky versus translucent rice regions, where chalky areas exhibit looser packing and increased air spaces, as confirmed by scanning electron microscopy (SEM) [43]. The enhanced feature extraction, together with the Dice focal loss function that addresses class imbalance between background, kernels, and chalky regions, and dropout for preventing overfitting, contributes to the observed improvements in segmentation accuracy (7.31% improvement in chalkiness IoU, 2.54% in mIoU, and 2.80% in mPA compared with VGG-UNet), particularly in cases with subtle chalkiness variations. However, model performance can be affected by class imbalance in samples with extremely low chalkiness ratios, where the

scarcity of chalky voxels during training can bias the model towards predicting the majority (non-chalky) class.

Second, our unified 3D detection framework significantly improves processing efficiency compared with traditional single-grain methods by enabling simultaneous analysis of multiple grains. The successful integration of isosurface extraction, point cloud conversion, DBSCAN clustering, and Poisson reconstruction provides a robust and automated pipeline for 3D analysis, transforming segmented voxel data into quantifiable morphological traits. The quantitative evaluation of key morphological indicators demonstrates the framework's reliability, with both VGG-UNet and VSE-UNet showing close alignment with ground truth measurements across kernel volume (VK), chalkiness volume (VC), three-dimensional chalkiness degree (CD3D), and kernel dimensions (L, W, T). While both models show slight overestimation in chalkiness-related metrics (VC, CD3D), they maintain high accuracy in kernel morphology measurements, indicating the framework's robustness in capturing diverse phenotypic traits. The consistently high accuracy in kernel morphology measurements, coupled with the reasonable performance in chalkiness detection, demonstrates the framework's capability in comprehensive phenotypic analysis.

Moreover, the modular design of our framework allows for the potential adaptation to other internal grain traits and even different crops. Adapting the framework to other crops will require careful consideration of differences in grain shape, size, and internal structure. Such extensions would require further validation and potential adjustments to the model architecture and parameters. Future research will explore these possibilities, focusing on model generalization across diverse varieties and growth conditions, and developing integrated tools for breeders and agricultural practitioners to facilitate crop improvement.

While the use of micro-CT technology involves higher initial costs compared with traditional 2D imaging methods, several factors justify this investment in the context of rice chalkiness research. First, micro-CT provides unique advantages, including non-destructive 3D analysis, high-precision internal structure visualization, and complete morphological information, which is essential for accurate chalkiness phenotyping [44]. These capabilities are particularly valuable for breeding programs, where precise phenotypic data can significantly accelerate the selection process. Second, our proposed high-throughput framework substantially improves the efficiency of micro-CT analysis through simultaneous multi-grain detection and automated processing, effectively reducing the per-grain analysis cost. The framework's automation capabilities minimize manual intervention, further enhancing operational efficiency. Additionally, the comprehensive and accurate phenotypic data obtained through this method provide valuable insights for breeding programs and quality assessment that could lead to long-term economic benefits through improved variety development.

Although our high-throughput micro-CT framework demonstrates significant advantages for rice chalkiness phenotyping, several limitations should be acknowledged. First, the current method requires specialized equipment and computational resources, which may present accessibility challenges for smaller research institutions or programs with limited budgets. Second, while our framework reduces per-grain analysis costs, the initial time investment for system optimization and protocol standardization remains substantial. Future work should focus on improving edge-case detection algorithms and exploring cost-sharing models for equipment utilization.

5. Conclusions

This study presents a novel high-throughput, automated, and non-destructive 3D method for detecting rice chalkiness using micro-CT imaging and an enhanced deep learning framework (VSE-UNet). By integrating these technologies, we achieved the effective 3D segmentation of kernels and chalky regions, enabling the simultaneous analysis

of multiple grains. Experimental results show that VSE-UNet significantly improves segmentation accuracy for chalky regions, while maintaining similar computational efficiency. The unified 3D detection framework overcomes the limitations of traditional single-grain methods, offering reliable measurements of key 3D morphological indicators (VK, VC, CD3D, L, W, T, L/W). Our method, particularly with the VSE-UNet model, demonstrates a strong potential in characterizing the 3D features of chalkiness. This work establishes a practical high-throughput approach for the automated 3D phenotyping of rice chalkiness, which contributes both to the fundamental research on chalkiness formation and to practical applications in rice breeding and quality assessment.

Author Contributions: Conceptualization, D.L. and Z.C.; methodology, Z.C.; software, Y.D. and Z.C.; validation, Z.C., B.L. and C.X.; formal analysis, Z.C.; investigation, Z.C.; resources, X.Z. and D.L.; data curation, D.L.; writing—original draft preparation, Z.C.; writing—review and editing, Z.C., Y.D., B.L. and D.L.; visualization, Z.C.; supervision, D.L.; project administration, X.Z. and D.L.; funding acquisition, D.L., B.L., X.Z., C.X. and Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 62401203, and in part by the Scientific Research Fund of the Hunan Provincial Education Department under Grants 24B0194 and 24A0498.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lu, Y.; Tang, Y.; Zhang, J.; Liu, S.; Liang, X.; Li, M.; Li, R. Variations and trends in rice quality across different types of approved varieties in China, 1978–2022. *Agronomy* **2024**, *14*, 1234. https://doi.org/10.3390/agronomy14061234.
2. Zhou, H.; Xia, D.; He, Y. Rice grain quality-traditional traits for high quality rice and Health-Plus substances. *Mol. Breed.* **2020**, *40*, 1. https://doi.org/10.1007/s11032-019-1080-6.
3. Wada, H.; Hatakeyama, Y.; Onda, Y.; Nonami, H.; Nakashima, T.; Erra-Balsells, R.; Morita, S.; Hiraoka, K.; Tanaka, F.; Nakano, H. Multiple strategies for heat adaptation to prevent chalkiness in the rice endosperm. *J. Exp. Bot.* **2019**, *70*, 1299–1311. https://doi.org/10.1093/jxb/ery427.
4. Cheng, X.; Shi, W.; Zhang, M.; Yue, H.; Dai, J.; Hu, L.; Zhu, G. Research progress on chalkiness formation mechanism in rice. *Chin. Agric. Sci. Bull.* **2024**, *40*, 1–7. https://doi.org/10.11924/j.issn.1000-6850.casb2023-0104.
5. Qiu, Y.X.; Dong, H.; Li, Y.X.; Wang, T.K.; Song, S.F.; Li, L. Research progress on genes related to chalkiness in rice. *Hybrid Rice* **2023**, *38*, 12–20. https://doi.org/10.16267/j.cnki.1005-3956.20220628.268.
6. Fan, P.; Xu, J.; Wei, H.; Liu, G.; Zhang, Z.; Tian, J.; Zhang, H. Recent research advances in the development of chalkiness and transparency in rice. *Agriculture* **2022**, *12*, 1123. https://doi.org/10.3390/agriculture12081123.
7. Wang, C.; Caragea, D.; Kodadinne Narayana, N.; Hein, N.T.; Bheemanahalli, R.; Somayanda, I.M.; Jagadish, S.V.K. Deep learning based high-throughput phenotyping of chalkiness in rice exposed to high night temperature. *Plant Methods* **2022**, *18*, 9. https://doi.org/10.1186/s13007-022-00839-5.
8. Saha, K.K.; Al Riza, D.F.; Ogawa, Y.; Suzuki, T.; Sugimoto, T.; Kondo, N. Assessment of chalkiness index of sake rice using transmission imaging. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *275*, 121149. https://doi.org/10.1016/j.saa.2022.121149.
9. Zhou, J.; Zeng, S.; Chen, Y.; Kang, Z.; Li, H.; Sheng, Z. A method of polished rice image segmentation based on YO-LACTS for quality detection. *Agriculture* **2023**, *13*, 182. https://doi.org/10.3390/agriculture13010182.
10. Shao, H.; Pu, J.; Mu, J. Pig posture recognition based on computer vision: Dataset and exploration. *Animals* **2021**, *11*, 1295. https://doi.org/10.3390/ani11051295.
11. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-time instance segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9156–9165.
12. Zia, H.; Fatima, H.S.; Khurram, M.; Hassan, I.U.; Ghazal, M. Rapid testing system for rice quality control through comprehensive feature and kernel-type detection. *Foods* **2022**, *11*, 2723. https://doi.org/10.3390/foods11182723.

13. He, Y.; Fan, B.; Sun, L.; Fan, X.; Zhang, J.; Li, Y.; Suo, X. Rapid appearance quality of rice based on machine vision and convolutional neural network research on automatic detection system. *Front. Plant Sci.* **2023**, *14*, 1190591. https://doi.org/10.3389/fpls.2023.1190591.
14. Yu, H. Key technologies for rice chalkiness rate measurement based on machine vision. Master's Thesis, Hunan Agricultural University, Changsha, China, 2018.
15. Su, Y.; Xiao, L.-T. 3D visualization and volume-based quantification of rice chalkiness in vivo by using high resolution micro-CT. *Rice* **2020**, *13*, 69. https://doi.org/10.1186/s12284-020-00429-w.
16. Cai, Z.; Liu, X.; Ning, Z.; Wang, S.; Li, D. Automated three-dimensional non-destructive detection of rice chalkiness driven by VGG-UNet. In Proceedings of the 9th International Conference on Intelligent Computing and Signal Processing (ICSP), Xian, China, 19–21 April 2024; pp. 1722–1726. https://doi.org/10.1109/ICSP62122.2024.10743996.
17. Lu, Y.; Wang, R.; Hu, T.; He, Q.; Chen, Z.S.; Wang, J.; Liu, L.; Fang, C.; Luo, J.; Fu, L.; et al. Nondestructive 3D phenotyping method of passion fruit based on X-ray micro-computed tomography and deep learning. *Front. Plant Sci.* **2023**, *13*, 1087904. https://doi.org/10.3389/fpls.2022.1087904.
18. Yu, L.; Liu, L.; Yang, W.; Wu, D.; Wang, J.; He, Q.; Chen, Z.; Liu, Q. A non-destructive coconut fruit and seed traits extraction method based on micro-CT and DeepLabV3+ model. *Front. Plant Sci.* **2022**, *13*, 1069849. https://doi.org/10.3389/fpls.2022.1069849.
19. Miao, Y.; Wang, R.; Jing, Z.; Wang, K.; Tan, M.; Li, F.; Zhang, W.; Han, J.; Han, Y. CT image segmentation of foxtail millet seeds based on semantic segmentation model VGG16-UNet. *Plant Methods* **2024**, *20*, 169. https://doi.org/10.1186/s13007-024-01288-y.
20. Zhu, X.; Huang, Z.; Li, B. Three-dimensional phenotyping pipeline of potted plants based on neural radiation fields and path segmentation. *Plants* **2024**, *13*, 3368. https://doi.org/10.3390/plants13233368.
21. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
22. Zhu, W.; Huang, Y.; Zeng, L.; Chen, X.; Liu, Y.; Qian, Z.; Du, N.; Fan, W.; Xie, X. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med. Phys.* **2019**, *46*, 576–589. https://doi.org/10.1002/mp.13300.
23. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
24. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
25. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
26. Li, X.; Fang, J.; Zhao, Y. A multi-target identification and positioning system method for tomato plants based on VGG16-UNet model. *Appl. Sci.* **2024**, *14*, 2804. https://doi.org/10.3390/app14072804.
27. Mukasheva, A.; Koishiyeva, D.; Sergazin, G.; Sydybayeva, M.; Mukhammejanova, D.; Seidazimov, S. Modification of U-net with pre-trained ResNet-50 and atrous block for polyp segmentation: Model TASPP-UNet. *Eng. Proc.* **2024**, *70*, 16. https://doi.org/10.3390/engproc2024070016.
28. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
29. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
30. Guo, M.-H.; Xu, T.-X.; Liu, J.-J.; Liu, Z.-N.; Jiang, P.-T.; Mu, T.-J.; Zhang, S.-H.; Martin, R.R.; Cheng, M.-M.; Hu, S.-M. Attention mechanisms in computer vision: A survey. *Comp. Vis. Media* **2022**, *8*, 331–368. https://doi.org/10.1007/s41095-022-0271-y.
31. Xiao, P.C.; Xu, W.G.; Zhang, Y.; Zhu, L.G.; Zhu, R.; Xu, Y.F. Research on scrap classification and rating method based on SE attention mechanism. *Chin. J. Eng.* **2023**, *45*, 1342–1352. https://doi.org/10.13374/j.issn2095-9389.2022.06.10.002.
32. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. https://doi.org/10.1109/3DV.2016.79.
33. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. https://doi.org/10.1109/TPAMI.2018.2858826.

34. Lorensen, W.E.; Cline, H.E. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Comput. Graph.* **1987**, *21*, 163–169. https://doi.org/10.1145/37402.37422.
35. Schroeder, W.; Martin, K.; Lorensen, B. *The Visualization Toolkit*, 4th ed.; Kitware: New York, NY, USA, 2006. ISBN 978-1-930934-19-1.
36. Zhou, Q.Y.; Park, J.; Koltun, V. Open3D: A modern library for 3D data processing. *arXiv* **2018**, arXiv:1801.09847.
37. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland, OR, USA, 2–4 August 1996; Volume 96, pp. 226–231.
38. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing, Cagliari, Sardinia, 26–28 June 2006*; Eurographics Association: Goslar, Germany, 2006; Volume *7*.
39. Lin, J.; Chen, H.; Wu, R.; Wang, X.; Liu, X.; Wang, H.; Wu, Z.; Cai, G.; Yin, L.; Lin, R.; et al. Calculating volume of pig point cloud based on improved Poisson reconstruction. *Animals* **2024**, *14*, 1210. https://doi.org/10.3390/ani14081210.
40. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
41. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
42. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 5693–5703.
43. Dwiningsih, Y.; Kumar, A.; Thomas, J.; Ruiz, C.; Alkahtani, J.; Al-hashimi, A.; Pereira, A. Identification of genomic regions controlling chalkiness and grain characteristics in a recombinant inbred line rice population based on high-throughput SNP markers. *Genes* **2021**, *12*, 1690. https://doi.org/10.3390/genes12111690.
44. Su, Y.; Zhou, Z.; Ma, L. Micro-CT technology and its application in plant research. *Exp. Sci. Technol.* **2022**, *20*, 7–12. https://doi.org/10.12179/1672-4550.20210096.