agriculture

MDPI

Article

A Method of Polished Rice Image Segmentation Based on YO-LACTS for Quality Detection

Jinbo Zhou [1], Shan Zeng [1,*], Yulong Chen [2], Zhen Kang [1], Hao Li [1] and Zhongyin Sheng [1]

[1] School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan 430023, China
[2] College of Medicine and Health Science, Wuhan Polytechnic University, Wuhan 430023, China
* Correspondence: zengshan1981@whpu.edu.cn

Abstract: The problem of small and multi-object polished rice image segmentation has always been one of importance and difficulty in the field of image segmentation. In the appearance quality detection of polished rice, image segmentation is a crucial part, directly affecting the results of follow-up physicochemical indicators. To avoid leak detection and inaccuracy in image segmentation qualifying polished rice, this paper proposes a new image segmentation method (YO-LACTS), combining YOLOv5 with YOLACT. We tested the YOLOv5-based object detection network, to extract Regions of Interest (RoI) from the whole image of the polished rice, in order to reduce the image complexity and maximize the target feature difference. We refined the segmentation of the RoI image by establishing the instance segmentation network YOLACT, and we eventually procured the outcome by merging the RoI. Compared to other algorithms based on polished rice datasets, this constructed method was shown to present the image segmentation, enabling researchers to evaluate polished rice satisfactorily.

Keywords: polished rice; RoI; YOLOv5; YOLACT

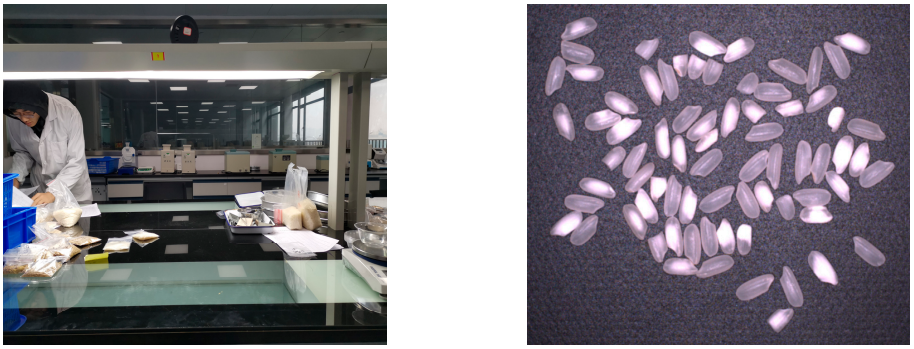
Citation: Zhou, J.; Zeng, S.; Chen, Y.; Kang, Z.; Li, H.; Sheng, Z. A Method of Polished Rice Image Segmentation Based on YO-LACTS for Quality Detection. *Agriculture* **2023**, *13*, 182. https://doi.org/10.3390/agriculture13010182

Academic Editor: Wei Ji

Received: 7 December 2022
Revised: 5 January 2023
Accepted: 9 January 2023
Published: 11 January 2023

1. Introduction

China accounts for more than 28% of global rice production, and the stability of its rice yield plays an important role in world food security. Rice is one of the prominent cereal crops in China, as about 65% of the population feeds on it [1]. The production process of polished rice is mainly composed of three steps: the milling and hulling of grains; the removing of the brown outer bran layer; and the polishing of the bran particles. Polished rice is no less than three-quarters the length of fully milled grains, and is twice or three times the price of broken grains [2]. With the development of the Chinese economy and the improvement of people's living standards, higher-quality rice is required urgently. The traditional manual detection method has been unable to meet the demand, due to slow detection speed, low accuracy and high labor cost, as shown in Figure 1a. Machine-vision-based quality inspection techniques of agricultural products have advanced rapidly in recent years, with their advantages of fast speed, high precision and reproducibility, thus providing the prospect of the quality detection of agricultural products [3]. In order to detect quality appearance in rice accurately, it is necessary to obtain a decomposition diagram of refined rice, by using image processing technology: the attached rice image is partitioned into single grain rice, for further determining its physical and chemical indexes; therefore, image segmentation is a significant stage in rice quality detection. However, in the practical process of rice image segmentation, due to small targets with potential irregular rice particles in the image, the mixture of refined and broken rice and uneven levels of polished rice results in difficulties in the polished rice image segmentation, as shown in Figure 1b.



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

Figure 1. Traditional artificial quality detection and rice sample diagram: (**a**) from Hubei Province Cereals, Oils and Foodstuffs Quality Supervision and Inspection Center; (**b**) irregular rice mixed image.

In the process of image segmentation, different segmentation methods are needed for various types of adherent particles images. More than a thousand segmentation methods have been proposed, after decades of research and development [4–8]. At present, the image segmentation method is generally classified into traditional machine-learning-based and deep-learning-based algorithms.

Traditional machine learning algorithms for image segmentation include different partitioning methods, based on threshold, cluster, edge, morphology, etc. One or more thresholds are used to divide the image grayscale histogram into several types, and then the image with grayscale value at the same class of pixel is regarded as one research object. Although the calculation is simple and fast, object position relationships in space may cause noise sensitivity and have a poor effect on the segmentation of different objects with slight difference in grayscale. The cluster-based image segmentation algorithm uses feature space points to represent the pixels in the image space, according to their positions in the feature space for clustering, and then maps to the image space for the final segmentation result. However, the initial cluster numbers and the initial cluster center positions have a large impact on the segmentation effect. Mariena et al. [9] came up with a hybrid K-mean clustering algorithm with a cluster-centric evaluation technique, to overcome this drawback. The concave point is a feature point formed at the edge of the objects by mutual contact or overlap of objects in the image. The concave point-based image segmentation steps generally include concave point detection, match and pair segmentation. The algorithm effectiveness depends on the correct matching of concave points. YAO et al. [10] applied the Edge Center Mode Proportion (ECMP) method to concave point matching, and then used the Minimum Enclosing Rectangle (MER) to calculate the rice length for identifying whole polished rice. Morphology-based image segmentation algorithms are characterized by simple methods and fast speed, such as the watershed algorithm; however, the traditional watershed algorithm is sensitive to weak edges, resulting in over-segmentation issues, noises in the image and subtle grayscale changes on the surface of the object. In order to solve this problem, some pre-processing methods are introduced before the algorithm, or are combined with other algorithms for improvement. Liang et al. [11] proposed a watershed segmentation algorithm based on generative markers, by combining the optimizing core region strategy and the comprehensive segmentation strategy: this method overcame the problem of over-segmentation and under-segmentation on images with dense and small targets, and the average segmentation accuracy was 98.73%. Gamarra et al. [12] utilized the Marker-Controlled Watershed algorithm (MC-Watershed), combined with the Split and Merge Watershed (SM-Watershed) two-step algorithm, to reach a segmentation balance using the inherent features of cells. Traditional segmentation methods are able to achieve fast and simple segmentation of adherent particles, but their results are easily affected by the external environment, and they have a relatively poor effect on complex adhesive particles segmentation, especially when whole polished rice is mixed with broken rice.

In recent years, Convolutional Neural Networks (CNNs) have been widely used as an advanced algorithm in image classification, object detection and instance segmentation, which avoids complex pre-processing of images, and directly inputs original images. The CNNs consist of a series of connected convolutional layers. The output of the upper layer after convolution is used as the input of the next layer in the forward propagation: each layer calculates the activation value based on its output value and activation function, and then passes the value to the next layer. In backward propagation, the error generated by forward propagation is used to update the parameters of each convolutional layer by the gradient descent method and chain rule: thus, the CNNs are able to learn useful features from image data automatically. The research has confirmed that improving network performance by adjusting the structure of CNNs can be better applied in the field of machine vision [13–17].

Deep-learning-based segmentation algorithms are mainly implemented via CNNs, and are independent of artificial feature extraction. Additionally, their networks can learn features during self-training, further reduce the influence of the external environment on the segmentation results, and improve robustness by simulating particle images under different conditions. Wang et al. [18] combined CNNs and Gradient-weighted Class Activation Mapping (Grad-CAM) to achieve automatic detection of chalkiness in grain images, accurately capturing chalkiness caused by high night temperature in rice. The method trained a CNN model to distinguish between chalky and non-chalky grains, using Grad-CAM to identify the area of a grain that was indicative of the chalky class, and then using a smooth heat map to quantify the degree of chalkiness. Based on EfficientNet-B3, Li et al. [19] introduced a Dual Attention Network (DAN) to sum up the output of two channels for changing feature representation and further focusing on feature extraction: their method realized the classification of rice germ integrity with an accuracy of 94.17%, providing guidance for the rice and grain processing industry. Xiong et al. [20] proposed a rice panicle segmentation algorithm called Panicle-SEG, based on simple linear iterative clustering superpixel regions generation, CNNs classification and entropy rate superpixel optimization: the algorithm was a robust method for panicle segmentation, creating new opportunities for non-destructive yield estimation. Ni et al. [21] developed a web browser-based application (Web App), by training two deep learning models, including MobileNet SSD and MobileNet-UNet, and achieved accurate and fast determination of blueberry scrapes along with online user access. This Web App provided a basis for blueberry breeders, farmers and packers to assess berry bruises. Jia et al. [22] proposed a new instance segmentation method named FoveaMask, which firstly extracted the features of input images by ResNet, fused by Feature Pyramid Networks (FPN), and carried out the classification and bounding-box regression of each spatial position on feature maps directly by full convolution method. RoI Align layer was then applied, to fix the size of the feature region, and at the same time to maintain accurate spatial locations. Finally, instance-level fruit segmentation was completed by using embedded mask branches on each proposal of pixel-level classification, which showed strong generalization ability on different shapes of fruits, and balanced the contradiction of accuracy and efficiency simultaneously. Pérez-Borrero et al. [23] proposed a strawberry instance segmentation method based on improved Mask R-CNN, which designed a new architecture for backbone and mask networks, removed the object classifier and the bounding-box regressor, and replaced the non-maximum suppression algorithm with a new region grouping and filtering algorithm, without increasing the complexity of the calculation. Lu et al. [24] proposed applicable segmentation methods for an intelligent Sichuan pepper-picking robot that could identify the fruit in images from various growing environments. This method not only showed high accuracy for the recognition and segmentation of Sichuan peppers but also provided support for the visual recognition of pepper-picking robots in the field.

At present, deep-learning-based segmentation algorithms are widely applied in the image segmentation of adhesive particles, but the high leakage rate and inaccurate mask-quality segmentation should be considered, especially in the image segmentation of polished rice with small and multiple targets: hence, precise decomposition of complex images

is required in the appearance quality detection of polished rice. As object detection network YOLO [25–27] is characterized by fast speed, low leakage rate and high accuracy, and YOLACT [28] with fast detection and high portability, producing a high quality and dynamic stability mask for simple target images, a new combined method of image segmentation on adherent rice is proposed in this paper. In particular, the model regards irregular mixed sticky polished rice as segmentation objects based on both the YOLOv5 and YOLACT methods, and collects the photos of different sticky rice grains on the conveyor belt to produce the dataset. The training model is used to achieve the refined segmentation of polished rice, and the main contributions of this constructed algorithm are as follows: (1) the complex image segmentation with small and multiple objects is decomposed into the segmentation of multiple single large objects, which enables the YOLACT model to learn the features of the polished rice, and to acquire segmentation results with low leakage detection rate and a high-accuracy mask; (2) a better YOLACT model can be trained via fewer samples, due to decreased image complexity; (3) the object-level labeling in the object detection network reduces the workload of the pixel-level labeling required for the instance segmentation network.

2. Methods and Materials

2.1. Overall Workflow

The workflow of this study is shown in Figure 2. In Step 1, the weights of YOLOv5 and YOLACT were trained by the self-built dataset. In order to ensure the accuracy of the training weights, the training accuracy of Mean Average Precision (mAP) was set to 95% and 75% in YOLOv5s and YOLACT, respectively. The weights greater than the training accuracy were selected, and the weight with the best training accuracy was further selected. The performance of this weight on the test set was used as the measurement criterion for the accuracy of the weights in Step 1. In Step 2, the object detection network YOLOv5 was then used to predict the location information of the RoI, and to extract the RoI from the whole polished rice image, to reduce image complexity and maximize object feature differences. Eventually, the mask of the RoI images was obtained by the instance segmentation network YOLACT in Step 3, and the final results were obtained by merging and restoring the RoI.

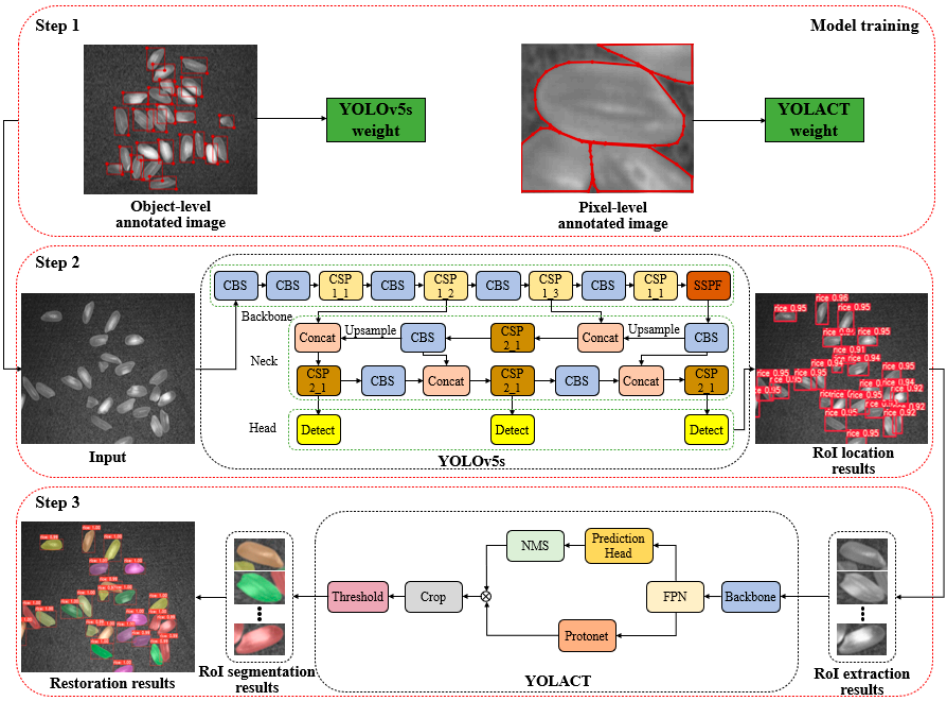


Figure 2. Flowchart of polished rice image segmentation.

2.2. Image Acquisition Device

An image acquisition device was designed, to acquire the randomly distributed image data of the polished rice. The acquisition equipment was mainly divided into hardware and software parts. The hardware part consisted of hopper, roller, conveyor, cameras, lighting devices and computing equipment. The schematic diagram of this device is shown in Figure 3. A pipeline operation was conducted in the acquisition process, and randomly distributed and non-overlapping polished rice was obtained by passing through the roller under the operation of the conveyor after the initial storage of the hopper. Vimba Viewer software was used to control the camera to acquire the original image data of the polished rice, and the captured images were transmitted to the computer for further analysis. The software part mainly included image acquisition, deduplication, image segmentation of rice grains, and quality detection of rice grains after accurate segmentation, such as the identification analysis of head, broken and chalky rice. The device acquired images at about 25 frames per second, and the conveyor belt speed was about 224 pixels per frame. There were approximately 43 rice grains in each image. After deduplication, the actual detection speed of the rice grains was about 30 instances per second. To prevent external light from damaging the image quality, the illumination device was composed of an LED light source and a black box. Considering economy, portability and image quality of acquisition, a Prosilica GC1600CH industrial camera from Allied Vision, Germany, with a resolution of 1620 × 1220 1620 × 1220 and a frame rate of 25 frames per second, using a macro lens, was used in this study. To meet the performance of different networks, the operating system in the used computing device was Ubuntu 18.04, with Intel core i9-10900K CPU, 32 GB RAM and NVIDIA GeForce RTX 3090 GPU, 24 GB video memory.

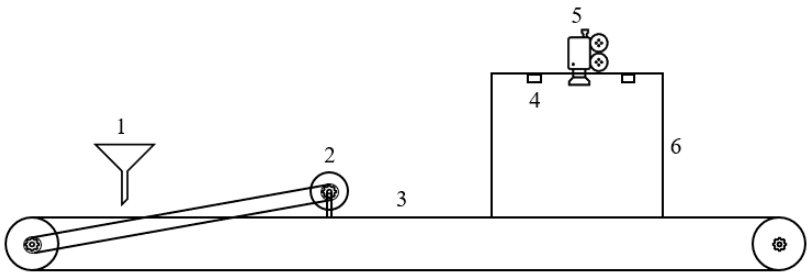
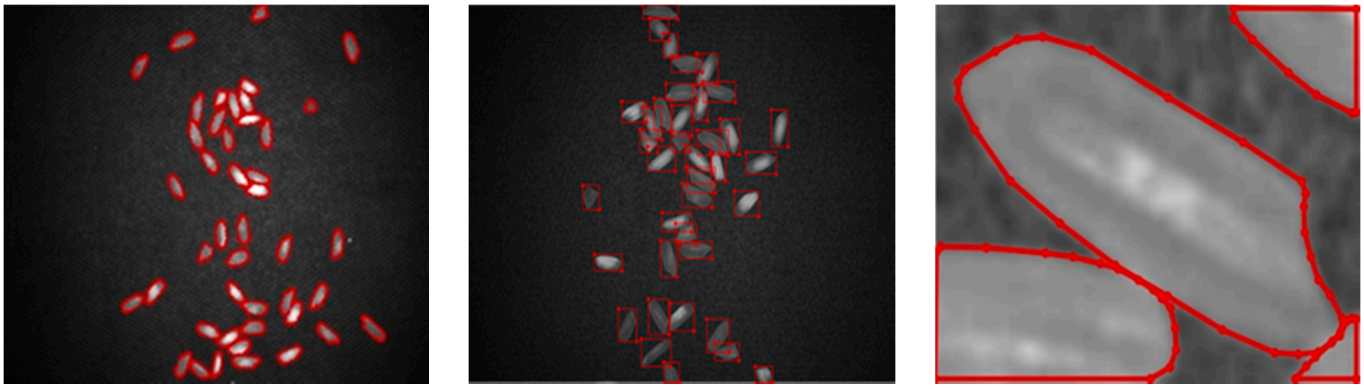


Figure 3. Schematic diagram of the polished rice image acquisition device: 1. hopper; 2. rollers; 3. belt conveyor; 4. toroidal light source; 5. CCD camera; 6. black box.

2.3. Polished Rice Dataset Production

A CCD camera with a resolution of 1620 × 1220 was used for the collection of the polished rice images, manually labeled using Labelme, to treat different rice grains as different instances of the same class, as shown in Figure 4a,b. Labelme is an open source annotation tool for labeling data required in target detection and image segmentation. In particular, Labelme can generate rectangular or polygon boxes in images, and at the same time produce corresponding json files containing all the boundary coordinate information and labeled object categories in the image. In the labeling process, the image segmentation needs to accurately label the boundary points of each instance, and the target detection only needs to label the approximate positions of each instance (the upper-left vertex and the lower-right vertex of the targets), as shown in Figure 4b,c. Thus, the pixel-level annotation of the image segmentation was more difficult than the object-level annotation of the target detection. In the annotation of the whole polished rice images and small polished rice images, the workload of the former was much greater than the latter, due to the different number of instances, as shown in Figure 4a (45 instances) and 4b (4 instances). For experimental requirements, this proposed method divided the YOLOv5 format data into 800 training sets, 100 validation sets and 100 testing sets, and the YOLACT format data into 1172 training sets, 147 validation sets and 147 testing sets, both after annotation, while the data in the other compared algorithms were divided into 727 training sets, 91 validation sets and 91 testing sets after labeling. From the total number of images, the data quantity of

this constructed method was greater than that of other algorithms; however, the workload of this method was much smaller than that of other algorithms, in terms of labeling. Moreover, the number of training segmentation instances required by this method was much less than others.



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

Figure 4. Sample labeling of polished rice images: (**a**) whole image labeling style of other algorithms; (**b**) whole image labeling style of YO-LACTS; (**c**) small icon labeling style of YO-LACTS.

2.4. Model Structure

In 2015, Joseph Redmon presented YOLO [25], a fast and accurate convolutional neural network for object detection, to treat the object detection task as a regression method of target region and category prediction, doing away with the extraction of candidate boxes in the Faster R-CNN [29], and completing fast and accurate end-to-end detection. YOLOv5, one type of network that improves on YOLOv4 [30], reserves the performance benefit of the YOLO series, and increases its speed and flexibility.

YOLOv5 consists of four parts: Input; Backbone network; Neck; and Head, as shown in Figure 5. In the Input part, Mosaic data augmentation, adaptive image scaling and auto-learning bounding box anchors are used in YOLOv5, which enriches the background and the small target of the detected object, reduces the information redundancy due to scaling filling, and greatly improves the network robustness. In the Backbone part, YOLOv5 adopts Focus and CSPNet [31] structures. Focus is a structure similar to down-sampling, reducing image size and increasing feature channel by slicing images; it also maintains effective information when reducing feature dimension. CSPNet, as a structure similar to a residual network, can effectively enhance the learning ability of CNNs, and reduce calculation amounts. In the Neck part, YOLOv5 retains the FPN [32] and Path Aggregation Network (PAN) [33] structures of YOLOv4, and adopts the CSPNet-designed CSP2 structure, to enhance the fusion ability of the network features. In the Head part, YOLOv5 utilizes the CIoU loss as the loss function of the bounding box that is derived from the Intersection over Union (IoU) throughout a series of improvements, focusing on both overlapping and non-overlapping areas, and achieving fast convergence.

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{1}$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{3}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{4}$$

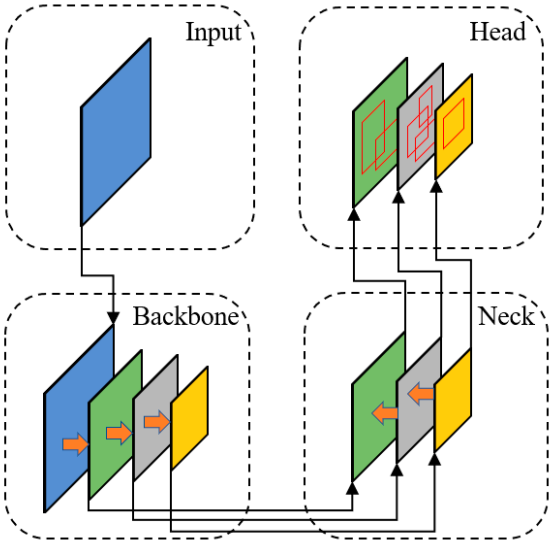


Figure 5. The four parts of the YOLOv5 detection model.

In Formula (1), α is a parameter used to balance the ratio, and v is a parameter used to measure the difference of width and height between prediction box and ground truth. The values of α and v are shown in Formulas (3) and (4).

YOLACT was designed by RetinaNet [34] via a series of improvements. Compared to two-stage instance segmentation networks, YOLACT is a one-stage network independent of the RoI concept, repooling operation and quantization errors, and thus significantly improves the inference speed and predicted mask quality.

YOLACT mainly consists of four parts: the Backbone network; FPN; the Prediction Head branch; and the Protonet branch, as shown in Figure 6. The Backbone network firstly extracts the feature from the image information. When FPN fuses the multi-scale feature information, it also provides a basis for the generation of the Prediction Head branch and the Protonet branch. The Prediction Head branch generates category confidence, position regression parameters and mask coefficients on each anchor, and the Protonet branch generates a set of prototype masks. Meanwhile, YOLACT linearly combines prototype mask with mask coefficients, to obtain the instance mask. These operations can be efficiently implemented by using single matrix multiplication and sigmoid.

$$M = \sigma\left(PC^T\right) \tag{5}$$

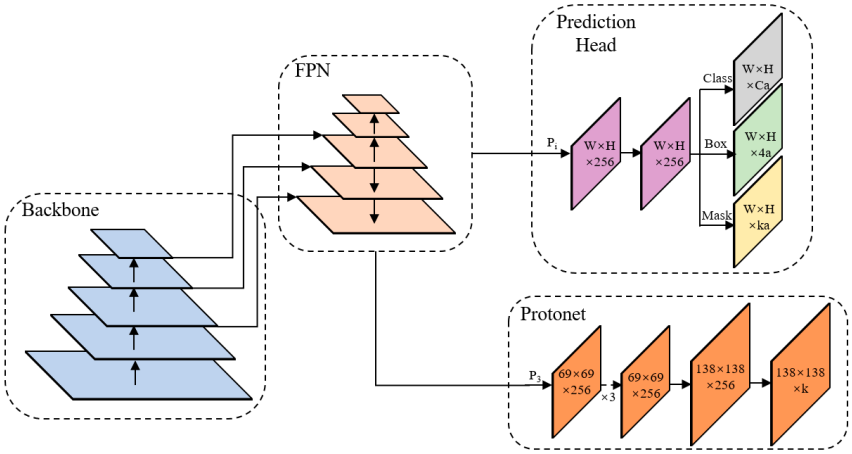


Figure 6. The four parts of the YOLACT segmentation model.

In Formula (5), P is a prototype mask matrix of $w * h * k$, and C is the product of instance number n and k of the mask coefficient matrices.

In addition, YOLACT also uses Crop to clear the mask boundary, and uses Threshold to binarize the mask, making it adaptable to small and multiple targets. A fast NMS algorithm is used to improve segmentation speed with only a slight loss of accuracy.

YOLOv5 achieves high accuracy compared to the instance segmentation network, topping out at nearly 55.0 mAP on the COCO2017 dataset, while the YOLACT network only reaches 31.2 mAP. On the self-built dataset of polished rice, YOLOv5 had a lower miss detection rate compared to the YOLACT network, as shown in Figure 7b. Nevertheless, YOLOv5 as an object detection network only detects the position of each polished rice without the generation of its mask. Although the YOLACT network attained an important balance of speed and performance in instance segmentation, the YOLACT-predicted results showed the poor performance of leakage detection and the inaccurate mask quality in the whole polished rice image (Figure 7c), due to the segmentation of small and multi-object adhesive rice particles. Therefore, the use of YOLOv5 in the decomposition of complex images, and the use of YOLACT in the segmentation of decomposed images, are regarded as effective methods to solve this issue.

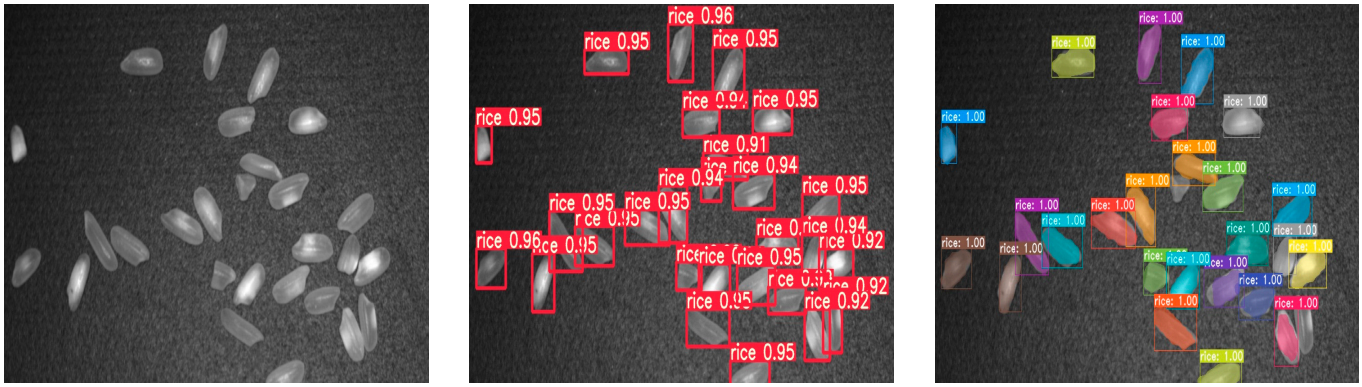


Figure 7. Comparison of direct prediction results of different networks on the whole polished rice map: (**a**) original graph; (**b**) YOLOv5 network prediction; (**c**) YOLACT network prediction.

3. Network Selection

3.1. Evaluation Indexes

mAP is usually used as an important indicator in instance segmentation evaluation metrics. The Average Precision (AP) was defined as the area formed by the Precision–Recall curve and the horizontal axis, and a higher AP value meant better model performance. mAP was calculated as the average of the APs of each category. The AP was equivalent to mAP due to the polished rice being labeled as one class in this paper. The value comparison between the pixel-level IoU of the predicted mask and the true mask, and the threshold value, was used to determine whether the mask prediction was correct. The confusion matrix of classification results calculated by the predicted data and the true situation is shown in Table 1.

Table 1. Confusion matrix.

Predicted	**Actual**	
	Positive	Negative
Positive	TP (True positive)	FP (False positive)
Negative	FN (False negative)	TN (True negative)

The precision P and recall R are defined as:

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

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

The object detection task was calculated in a similar way, only computing out the intersection ratio of detected and real boxes. The experimental results selected the IoU with different thresholds (namely IoU_{all}, IoU_{50} and IoU_{75}), to comprehensively evaluate the performance of the network model on the polished rice dataset, in which AP_{all} was the average AP from IoU_{50} to IoU_{95} with an interval of 5.

3.2. Experimental Verification

YOLOv5 includes five versions: YOLOv5n; YOLOv5s; YOLOv5m; YOLOv5l; and YOLOv5x, which can be selected according to different requirements. To select an appropriate YOLOv5 network, we compared the performance of different versions of the YOLOv5 network on the whole image, as shown in Table 2.

Table 2. Performance of different versions of YOLOv5 on the polished rice dataset.

	Box AP_{all} (%)	**Params (M)**	**Speed/Image (ms)**
YOLOv5n	88.4	1.9	11.3
YOLOv5s	**92.1**	**7.2**	**11.4**
YOLOv5m	92.8	21.2	15.3
YOLOv5l	94.4	46.5	17.4
YOLOv5x	95.6	86.7	27.0

AP means the accuracy and miss detection rate of network prediction; model parameter means the network portability and equipment configuration requirement; and detection speed means the efficiency of the network at image processing. A good lightweight network is shown to possess both accurate prediction and efficient processing capabilities. YOLOv5s has a high 92% of AP, while its model parameters are much lower than YOLOv5m, YOLOv5l and YOLOv5x, and its inference speed is only 0.1 ms lower than YOLOv5n: therefore, YOLOv5s was selected as the initial segmentation network, considering AP, inference speed and model parameter quantity.

As the YOLACT network was used for further re-segmentation of the results of the YOLOv5 network segmentation, it did not need to learn complex features: this enabled the preservation of the accuracy of the YOLACT model as much as possible, while streamlining the network structure and reducing the parameter quantity; therefore, the number of FPN layers was adjusted in Figure 8. Meanwhile, the performance of YOLACT with different FPN layers on the self-built small resolution dataset of the polished rice was compared, in order to verify the performance of the improved algorithm. During training, the network input image sizes were all set to 200, the batch sizes were all set to 16, and the other parameters were kept consistent. Table 3 indicates the best training results obtained after sufficient rounds of training.

From the experimental results in Table 3, it can be seen that the Mask AP first increased and then decreased as the FPN layer decreased, while the Box AP continued to increase under the approximately the same conditions. It is essential to notice that the Mask AP remained higher than the Box AP in all cases. The FPN layer was designed to process the multi-scale computation, and to enhance the performance of small object detection. The image size after the YOLOv5 segmentation was smaller, and the percentage of rice grains needed for segmentation was larger compared to the image, with fixed position and simple background. As the mask needed to be more precisely extracted at the end stage, YOLACT with backbone network of resnet101 and FPN layer of 4 was selected in the processing method, considering the experimental results for Mask AP.

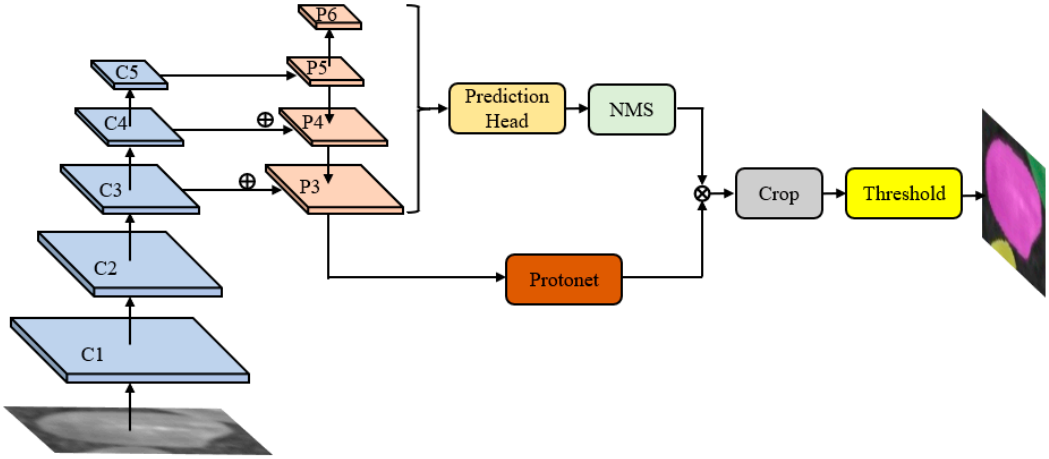


Figure 8. The YOLACT network structure after removing the P7 layer from the FPN structure.

Table 3. Performance of YOLACT network with different parameters on the polished rice dataset.

	FPN Num	Params (M)	Speed (Frame/s)	Box (%)			Mask (%)		
				AP_{all}	AP_{50}	AP_{75}	AP_{all}	AP_{50}	AP_{75}
ResNet101	3	194.4	3.76	79.34	97.91	89.84	79.99	97.85	91.44
	4	196.7	3.74	77.55	97.86	89.43	**81.28**	98.83	91.58
	5	199.0	3.73	71.78	95.59	85.17	77.13	96.74	89.65
ResNet50	3	118.0	4.54	78.57	98.76	90.50	80.17	97.88	90.99
	4	120.4	4.52	77.53	97.83	89.52	**80.55**	97.86	91.51
	5	122.7	4.51	75.06	97.86	90.15	78.96	98.52	91.30

4. Experimental Results

YO-LACTS was compared to YOLACT [28], Mask R-CNN [35], Mask Scoring R-CNN [36], SOLO [37] and SOLOv2 [38], based on the results of whole image rice segmentation. In order to compare the algorithm performance more accurately, the input image size was set to 550 for YOLACT training and prediction on the whole image. As the Mask R-CNN input size had to be a multiple of 2, it was set to 512 when training and testing on the whole image. The best input size of 1333 × 800 was selected for SOLO, SOLOv2 and Mask Scoring R-CNN. YO-LACTS set the YOLOv5s input size to 550, and set the YOLACT input size to 256 when predicting the YOLOv5s results. All the models mentioned in the paper underwent sufficient rounds of training on the same polished rice dataset, and the optimal results were recorded in the final stage.

It can be seen from Table 4 that YO-LACTS was close to YOLACT, in terms of parameter quantity and testing speed, and that its prediction accuracy for the entire image in Mask AP_{all} was higher than that of the other algorithms. Compared to the test results of the algorithms on the polished rice image, the algorithm segmentation results are shown in Figure 9. In brief, YO-LACTS possessed a lower leakage detection rate in whole images, and showed higher accuracy in masking on amplified details.

However, some conditions included uneven illumination, mixed impurities, large aggregation density of adhered particles, and even overlap in practical rice quality detection, usually posing a huge challenge to algorithm stability. Therefore, this paper tested the model performance on the complex images of adhesive rice, shown in Figure 10. Based on the segmentation results, YO-LACTS exhibited greater robustness compared to other algorithms.

Table 4. Comparison between YO-LACTS and other methods.

	Image Size	Backbone	Params (M)	Speed (Frame/s)	Mask (%)		
					AP_{all}	AP_{50}	AP_{75}
YO-LACTS	550 × 550, 256 × 256	ResNet50 ResNet101	134.8 211.1	4.31 3.52	75.49 **83.90**	89.72 98.83	86.09 94.91
YOLACT	550 × 550	ResNet50 ResNet101	122.7 199.0	4.51 3.73	**73.69** 73.66	98.00 98.99	96.74 97.52
Mask R-CNN	512 × 512	ResNet50 ResNet101	170.0 244.0	2.38 1.76	**78.00** 76.80	97.00 98.00	93.80 95.80
Mask Scoring R-CNN	1333 × 800	ResNet50 ResNet101	481.4 630.9	1.63 1.15	75.49 **81.90**	89.72 96.00	86.09 93.80
SOLO	1333 × 800	ResNet50 ResNet101	318.3 470.6	1.72 1.26	82.50 **82.60**	95.00 95.00	93.70 93.50
SOLOv2	1333 × 800	ResNet50 ResNet101	369.4 546.9	1.70 1.21	**82.20** 80.70	95.00 94.50	93.90 90.90

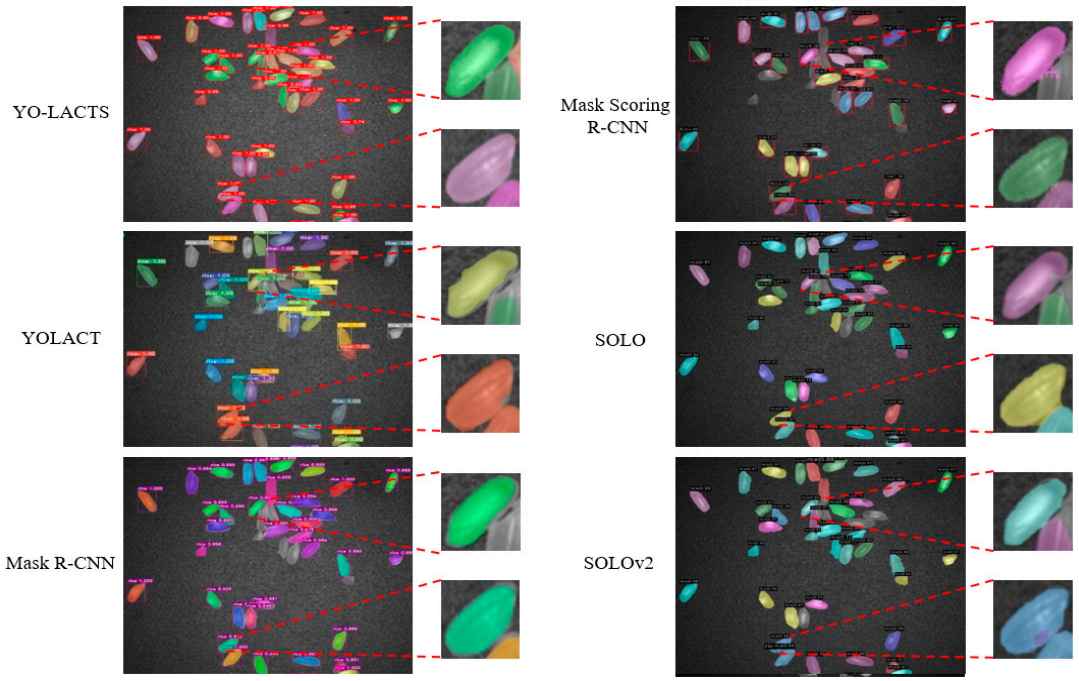


Figure 9. Comparison of YO-LACTS and other algorithms on polished rice image segmentation.

To further evaluate the quality of the rice grains effectively, the type of rice grains was classified, the quantity predicted, the actual instances of rice grains were compared, and the corresponding error rate was calculated in Table 5. Meanwhile, the related confusion matrix is shown in Figure 11. Then, a single rice grain in the whole image was numbered, and some key evaluation parameters, such as chalkiness, broken rice rate and chalky grain percentage, were determined, and are displayed in the upper left corner of Figure 12. Eventually, the type, chalk rate, aspect ratio and chalky area (marked as red) of each polished rice were visualized and compared with the original images, as shown in Figure 13.

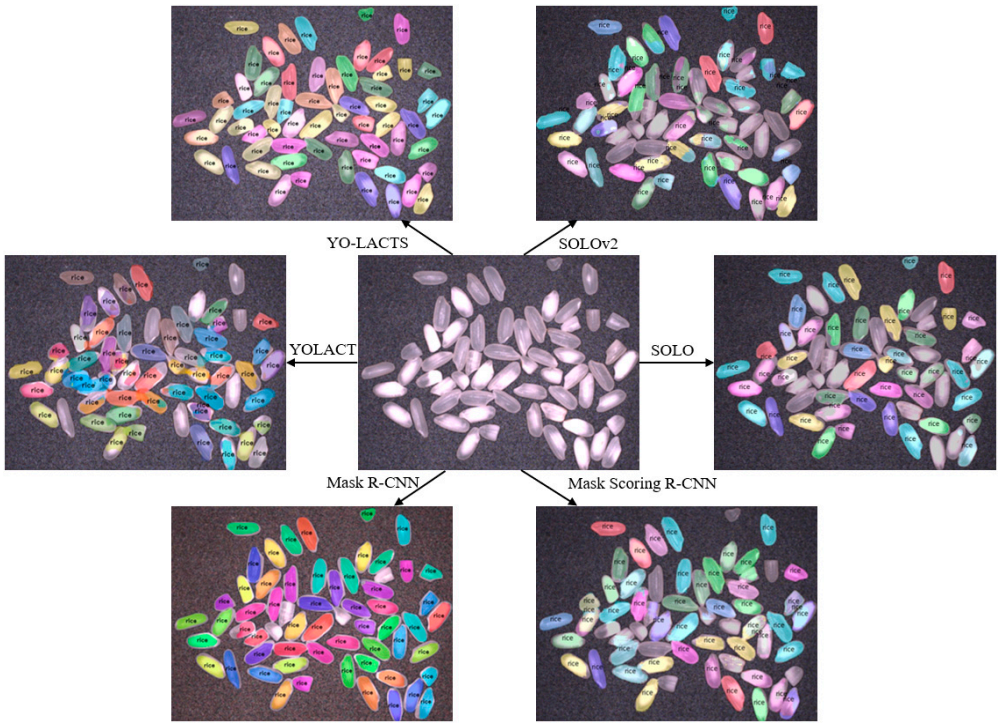


Figure 10. Comparison of YO-LACTS and other algorithms on complex polished rice image segmentation.

Table 5. Statistic results of rice grain types before improvement.

	Predicted (Instance)	Actual (Instance)	Error Rate (%)
Head rice	14	19	26.3
Chalky rice	31	26	19.2
Broken rice	16	16	0.0

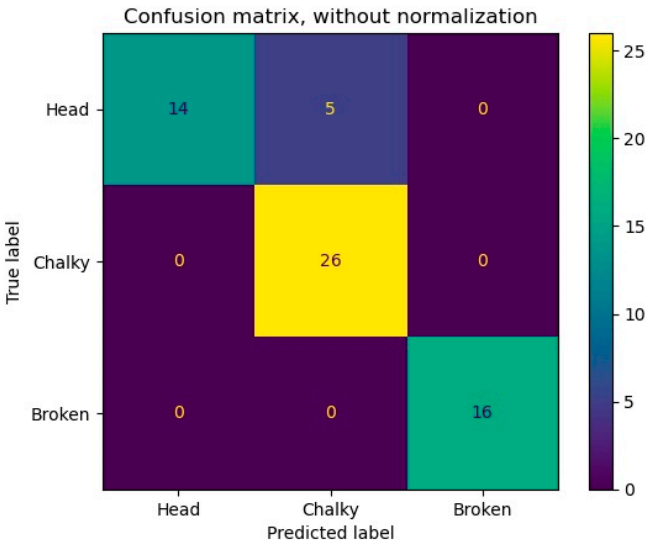


Figure 11. Confusion matrix of rice grain types before improvement.

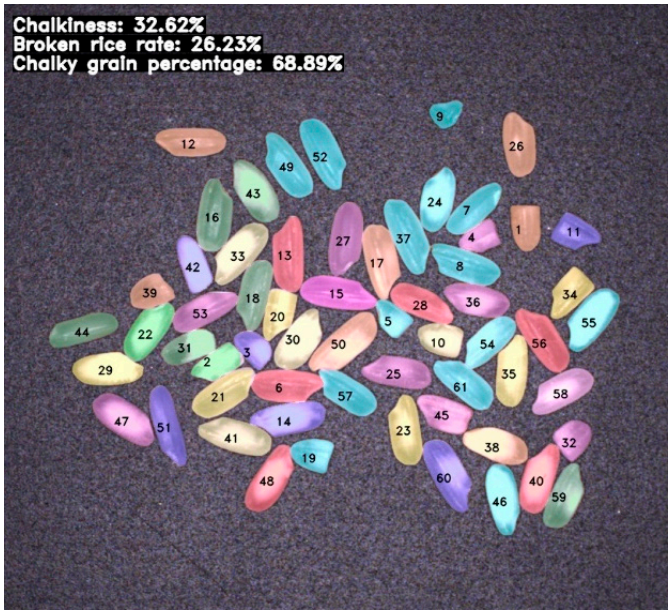


Figure 12. Image visualization of multiple rice grains.

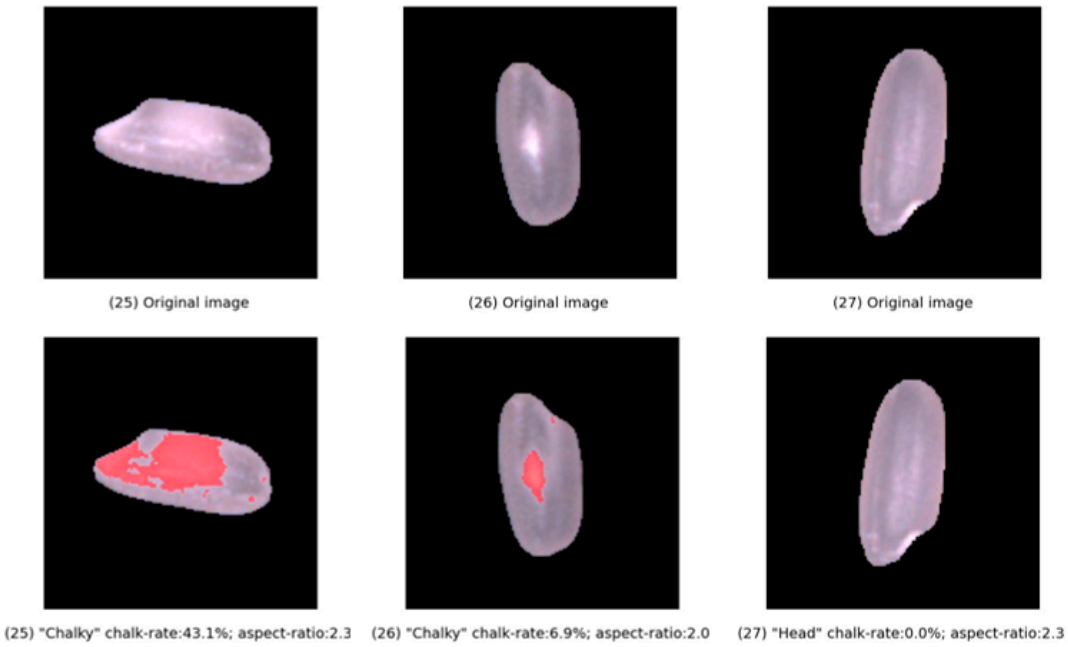


Figure 13. Image visualization of a single polished rice grain.

From the analysis of the model test results, YO-LACTS with strong robustness completed the accurate segmentation of the rice grains under complex conditions, and further assessed their quality. By comparing the original image and the visual analysis image of the numbered rice grain (26) in Figure 13, it can be seen that overexposure in some areas of rice grain image caused the misjudgment of the rice grain type and chalky area: in order to reduce this error rate, the judgment of the exposure degree was added, and then tested on a new test set of 500 images. The identification results and confusion matrix are shown in Table 6 and Figure 14. The test results demonstrate that YO-LACTS achieves the accurate segmentation of rice grains, enabling researchers to evaluate their quality satisfactorily.

Table 6. Statistical results of rice grain types after improvement.

	Predicted (Instance)	Actual (Instance)	Error Rate (%)
Head rice	15,917	15,949	0.2
Chalky rice	5478	5446	0.6
Broken rice	392	392	0.0

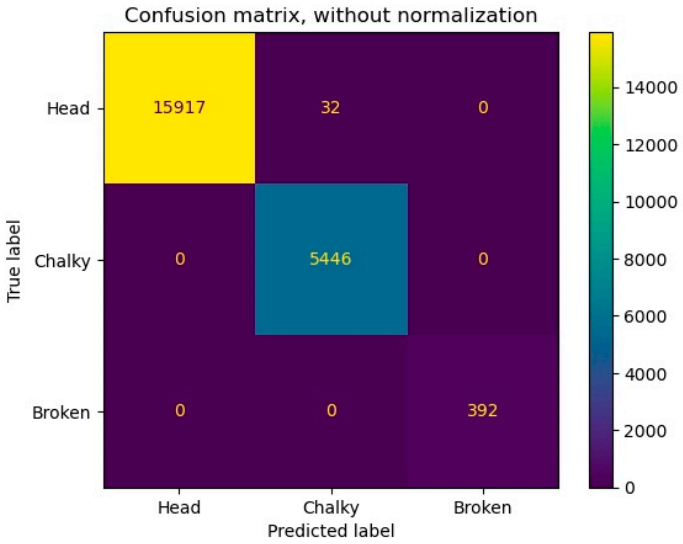


Figure 14. Confusion matrix of rice grain types after improvement.

5. Conclusions

The problem of sticking rice particles has been a difficult point that hinders industrialized and intelligent development of rice quality inspection. In this paper, a new YO-LACTS algorithm combining YOLOv5 and YOLACT is proposed for the refined segmentation of adhesive rice grains in images, which lays the foundation for rice grain counting, grain shape detection and other physicochemical index detection. In addition, YO-LACTS can also be applied in other fields of small and multi-object segmentation, such as cells and widgets. Compared to other algorithms, YO-LACTS is more stable, and shows better results on mask quality. However, there are some flaws in this method. Although the detection boxes predicted by YOLOv5s had high accuracy compared to other networks, and the prediction boxes deviation of the YOLOv5s network could be corrected by its expansion, the detection boxes predicted by YOLOv5s still had a minute quantity of leakage detection and repetition. It is necessary to further calibrate the predicted results of YOLOv5s, to improve the segmentation accuracy of the polished rice image.

Author Contributions: All authors contributed significantly to the work. Conceptualization, J.Z. and S.Z.; software, J.Z.; validation, J.Z., Z.S., S.Z. and Y.C.; formal analysis, J.Z., Y.C. and H.L.; investigation, Z.K.; writing—original draft preparation, J.Z. and S.Z.; writing—review and editing, J.Z., Z.S., Z.K. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Hubei province Natural Science Foundation for Distinguished Young Scholars, grant NO. 2020CFA063, and funded by excellent young and middle-aged scientific and technological innovation teams in the colleges and universities of Hubei Province, grant NO. T2021009.

Data Availability Statement: Not applicable.

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

References

1. Nie, L.; Peng, S. Rice production in China. In *Rice Production Worldwide*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 33–52. [CrossRef]
2. Yadav, B.; Jindal, V. Changes in head rice yield and whiteness during milling of rough rice (*Oryza sativa* L.). *J. Food Eng.* **2008**, *86*, 113–121. [CrossRef]
3. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [CrossRef]
4. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef]
5. Pal, N.R.; Pal, S.K. A review on image segmentation techniques. *Pattern Recognit.* **1993**, *26*, 1277–1294. [CrossRef]
6. Bali, A.; Singh, S.N. A review on the strategies and techniques of image segmentation. In Proceedings of the 2015 Fifth International Conference on Advanced Computing & Communication Technologies, Haryana, India, 21–22 February 2015; IEEE: New York, NY, USA, 2015; pp. 113–120. [CrossRef]
7. Pham, D.L.; Xu, C.; Prince, J.L. Current Methods in Medical Image Segmentation. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 315–337. [CrossRef]
8. Ghosh, S.; Das, N.; Das, I.; Maulik, U. Understanding Deep Learning Techniques for Image Segmentation. *ACM Comput. Surv.* **2019**, *52*, 1–35. [CrossRef]
9. Muda, T.Z.T.; Salam, R.A. Blood cell image segmentation using hybrid K-means and median-cut algorithms. In Proceedings of the IEEE International Conference on Control System, Penang, Malaysia, 25–27 November 2011; IEEE: New York, NY, USA, 2011; pp. 237–243. [CrossRef]
10. Yao, Y.; Wu, W.; Yang, T.; Liu, T.; Chen, W.; Chen, C.; Li, R.; Zhou, T.; Sun, C.; Zhou, Y.; et al. Head rice rate measurement based on concave point matching. *Sci. Rep.* **2017**, *7*, 41353. [CrossRef]
11. Liang, J.; Li, H.; Xu, F.; Chen, J.; Zhou, M.; Yin, L.; Zhai, Z.; Chai, X. A Fast Deployable Instance Elimination Segmentation Algorithm Based on Watershed Transform for Dense Cereal Grain Images. *Agriculture* **2022**, *12*, 1486. [CrossRef]
12. Gamarra, M.; Zurek, E.; Escalante, H.J.; Hurtado, L.; San-Juan-Vergara, H. Split and merge watershed: A two-step method for cell segmentation in fluorescence microscopy images. *Biomed. Signal Process. Control* **2019**, *53*, 101575. [CrossRef]
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
14. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
15. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
17. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
18. Wang, C.; Caragea, D.; Narayana, N.K.; 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. [CrossRef]
19. Li, B.; Liu, B.; Li, S.; Liu, H. An Improved EfficientNet for Rice Germ Integrity Classification and Recognition. *Agriculture* **2022**, *12*, 863. [CrossRef]
20. Xiong, X.; Duan, L.; Liu, L.; Tu, H.; Yang, P.; Wu, D.; Chen, G.; Xiong, L.; Yang, W.; Liu, Q. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* **2017**, *13*, 104. [CrossRef]
21. Ni, X.; Takeda, F.; Jiang, H.; Yang, W.Q.; Saito, S.; Li, C. A deep learning-based web application for segmentation and quantification of blueberry internal bruising. *Comput. Electron. Agric.* **2022**, *201*, 107200. [CrossRef]
22. Jia, W.; Zhang, Z.; Shao, W.; Hou, S.; Ji, Z.; Liu, G.; Yin, X. FoveaMask: A fast and accurate deep learning model for green fruit instance segmentation. *Comput. Electron. Agric.* **2021**, *191*, 106488. [CrossRef]
23. Pérez-Borrero, I.; Marín-Santos, D.; Gegúndez-Arias, M.E.; Cortés-Ancos, E. A fast and accurate deep learning method for strawberry instance segmentation. *Comput. Electron. Agric.* **2020**, *178*, 105736. [CrossRef]
24. Lu, J.; Xiang, J.; Liu, T.; Gao, Z.; Liao, M. Sichuan Pepper Recognition in Complex Environments: A Comparison Study of Traditional Segmentation versus Deep Learning Methods. *Agriculture* **2022**, *12*, 1631. [CrossRef]
25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
26. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]
27. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.

28. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-Time Instance Segmentation. In Proceedings of the 2019 IEEE/CVF In-ternational Conference on Computer Vision (ICCV), Seoul, South Korea, 27 October–2 November 2019; pp. 9156–9165. [CrossRef]
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
30. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
31. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
32. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125. [CrossRef]
33. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768. [CrossRef]
34. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
35. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969. [CrossRef]
36. Huang, Z.; Huang, L.; Gong, Y.; Huang, C.; Wang, X. Mask Scoring R-CNN. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 6409–6418. [CrossRef]
37. Wang, X.; Kong, T.; Shen, C.; Jiang, Y.; Li, L. SOLO: Segmenting Objects by Locations. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 649–665.
38. Wang, X.; Zhang, R.; Kong, T.; Li, L.; Shen, C. Solov2: Dynamic and fast instance segmentation. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17721–17732.