Postprint: Wheat Grain Phenotypic Identification Method Based on Deep Learning ImCascade R-CNN

Authors: Weiting Pan, Sun Mengli, Yuan Yan, Liu Ping

Date: 2023-12-04T00:00:00+00:00

Abstract

[Objective/Significance] Breeding high-quality and high-yield wheat varieties is the primary objective of wheat breeding, and wheat grain integrity directly affects the wheat breeding process. The relatively small differences in some characteristics between intact and damaged grains are key factors limiting the accuracy of identifying damaged wheat grains based on deep learning. [Method] To address the problem of low detection accuracy for wheat grains, this study establishes an ImCascade R-CNN model and proposes a wheat grain phenotypic identification method to accurately detect wheat grain integrity, segment grains, and obtain phenotypic parameters of intact grains. [Results and Discussion] The ImCascade R-CNN model achieves an average precision of 90.2% in detecting wheat grain integrity. Compared with Cascade Mask R-CNN and Deeplabv3+ models, it can better identify, locate, and segment wheat grains, providing a foundation for obtaining grain phenotypic parameters. The average error rates of this method for measuring grain length and width are 2.15% and 3.74%, respectively, and the standard error for measuring the length-width ratio is 0.15, showing high consistency with manual measurements. [Conclusion] The research results can rapidly and accurately detect grain integrity, obtain phenotypic data of intact grains, and accelerate the breeding of high-quality and high-yield wheat varieties.

Full Text

Abstract

[**Objective/Significance**] Developing high-quality, high-yield wheat varieties represents the primary goal of wheat breeding, and wheat grain integrity directly influences this breeding process. The partial feature differences between intact and damaged grains are minimal, constituting a key factor limiting the accuracy of deep learning-based damaged wheat grain identification. [**Methods**] To address the problem of low detection accuracy for wheat grains, this study established an ImCascade R-CNN model to detect wheat grain integrity, segment grains, and obtain phenotypic parameters of intact grains. [**Results and Discussion**] The ImCascade R-CNN model achieved an average precision of 90.2% for detecting wheat grain integrity. Compared with the Cascade Mask R-CNN model, it better identified, located, and segmented wheat grains, providing a foundation for obtaining grain phenotypic parameters. The average error rates for measuring grain length and width were 2.15% and 3.74%, respectively, and the standard error for the length-width ratio was 0.15, demonstrating high consistency with manual measurements. **Conclusion** The research results enable rapid and precise detection of grain integrity and acquisition of intact grain phenotypic data, accelerating the cultivation of high-quality, high-yield wheat varieties.

Keywords: wheat breeding; ImCascade R-CNN model; grain integrity; semantic segmentation; grain phenotypic parameters; deep learning

Introduction

Wheat serves as the primary source of dietary carbohydrates for the human population, providing 20% of the total caloric intake required by the growing global population. Developing superior wheat varieties and ensuring wheat yield have become key research priorities for breeding professionals. In wheat breeding, improper harvesting and storage can easily cause wheat grain damage, which directly reduces emergence rates and leads to yield loss. Additionally, grain weight is a major determinant of wheat yield and represents a critical breeding target for enhancing production to ensure food security. Wheat grain size and shape show significant positive correlation with grain weight, and attention to phenotypic parameters such as grain length and width also provides an important basis for introducing, screening, and evaluating germplasm resources. Therefore, detecting wheat grain integrity and obtaining phenotypic parameters of intact grains can fundamentally improve breeding efficiency.

Traditional methods for wheat grain appearance and quality identification rely primarily on manual inspection, which is subjective, time-consuming, and labor-intensive, failing to meet the requirements for rapid and accurate large-scale wheat grain detection. In recent years, deep learning technology represented by Convolutional Neural Networks (CNN) has developed rapidly and been widely applied in image recognition. Compared with traditional machine learning techniques, CNN's feature learning capability can significantly improve detection performance. Song et al. designed a YOLOv5 MDC lightweight network for detecting severely adhered wheat grains, which reduced storage space while maintaining comparable accuracy to the YOLOv5s model. For grain integrity identification, Huang et al. introduced CNN and transfer learning for grain defect detection, achieving a 15.8% accuracy improvement compared with Support Vector Machine (SVM) algorithms. Zhu et al. developed a CNN-based wheat grain integrity image detection system that improved identification accuracy by up to 5.77% compared with SVM and Back Propagation (BP) neural network models. Gao et al. proposed a wheat grain integrity detection method based on Res24_D_{{{CBAM}}_{{Atrous}}}, improving detection precision by 3%-4% over the original Res34 baseline.

However, these studies focused on wheat grain integrity identification but failed to precisely segment wheat grain contours to obtain grain length and width parameters. Comprehensive, large-scale, and multi-scale wheat grain appearance and quality identification platforms are rapidly developing. Zhao et al. developed a high-throughput phenotyping system for wheat grains based on image analysis that can measure parameters such as grain length, short diameter, and length-width ratio. Existing commercial devices such as the SC-M grain appearance quality classifier and SC-G automatic seed analyzer cannot identify wheat grain integrity, and their method of determining length and width parameters using grain bounding rectangles is susceptible to grain arrangement orientation.

With the emergence of Region-CNN (R-CNN) structures, combining candidate box detectors and region classifiers to solve object detection problems has become mainstream. To reduce redundant computation in R-CNN, SPP-Net (Spatial Pyramid Pooling-Network) and Fast R-CNN were introduced for regional feature extraction, significantly accelerating detection speed, though training steps remained cumbersome. Later, Faster R-CNN introduced the Region Proposal Network (RPN), further reducing runtime and becoming an optimal solution for object detection, though it suffered from mismatches between RPN receptive fields and actual object scales. Cascade R-CNN, composed of a series of detectors trained with progressively increasing Intersection over Union (IoU) thresholds, avoids overfitting during training and quality mismatch during inference.

2.1 Data Collection

Experimental data were collected at the National Key Laboratory of Wheat Breeding at Shandong Agricultural University using the Fielder wheat variety. The acquisition device is shown in [Figure 1: see original paper]. The platform surface consisted of black cardboard, with a Jierui Weitong DW800_2.9mm camera (800$×$600 pixel resolution) fixed at a distance of 21 cm from the platform. A total of 300 images were captured, with each image containing 60-120 wheat grains.

Using LabelMe, intact and damaged wheat grains were annotated. To enhance model robustness, images and labels were simultaneously augmented through random flipping, random rotation, random cropping, brightness adjustment, contrast adjustment, and Gaussian noise addition, as shown in [Figure 2: see original paper]. Thirty original images were randomly selected for testing, while the remaining images were augmented to 1,620 images and split into training and validation sets at a 2:1 ratio, ensuring that each original image and its augmented versions belonged to the same set.

2.2 Wheat Grain Phenotypic Identification Method

To address the problems of low identification accuracy for wheat grain integrity, inability to segment grain contours, and limited acquisition of grain phenotypic parameters, this study improved the Cascade Mask R-CNN model by establishing the ImCascade R-CNN model and proposing a novel wheat grain phenotypic identification method. This method enables rapid and precise detection of wheat grain integrity while simultaneously segmenting grain contours and obtaining phenotypic parameters of intact grains, reducing the missed detection rate for adhered grains. The complete workflow is shown in [Figure 3: see original paper].

Wheat grain images are first input into the ImCascade R-CNN model to extract key features. Through classification and bounding box regression branches, the model identifies wheat grain integrity. The mask generation branch produces mask maps for individual wheat grains, segmenting grain contours and ultimately determining grain count in the image to calculate length and width parameters of intact wheat grains.

2.2.1 Wheat Grain Integrity Detection Model

To improve detection accuracy for damaged wheat grains, an ImCascade R-CNN model suitable for wheat grain integrity detection was developed, as shown in [Figure 4: see original paper]. Based on the Cascade Mask R-CNN model, the following improvement strategies were implemented: (1) The backbone network was changed to ResNeXt to prevent gradient vanishing and reduce parameters; (2) The activation function was changed to Mish to improve detection efficiency and generalization; (3) A multilayer convolutional structure was introduced in the detector to fully exploit hidden wheat grain features; (4) The Soft-Non Maximum Suppression (Soft-NMS) algorithm was adopted to determine candidate boxes, achieving precise segmentation of adhered wheat grain regions.

(1) Backbone Network Improvement. The Cascade Mask R-CNN model uses the ResNet architecture as its backbone, with basic modules shown in Figure 5: see original paper. While deepening or widening network layers typically improves accuracy, ResNet networks with greater depth, larger convolutional kernels, or more parameters increase design difficulty and computational costs while suffering from gradient vanishing. To improve feature extraction accuracy and prevent gradient vanishing while reducing parameters, the ImCascade R-CNN model employs a ResNeXt network with group convolution for feature extraction. The basic structure is shown in Figure 5: see original paper, where each group performs the same tensor calculation as ResNet on the input, then concatenates the transformed feature vectors as output.

ResNeXt performs both residual calculations and split-transform-merge operations. The internal structure adopts a branching strategy with consistent topology for hyperparameter sharing, changing the number of branches through group numbers to greatly enhance model scalability and accuracy. The split-transform-merge structure is expressed as:

$$\sum_{i=1}^{c} T_i(x)$$

where F_y represents output features, F_x represents input features, T_i represents identical branch structures, and c represents the number of branches (32 in this model). Compared with ResNet, ResNeXt has fewer parameters while maintaining the same topology, offering better portability without increased model complexity.

(2) Residual Module Improvement. In deep neural networks, each layer's input is the previous layer's output, with lower nodes obtained through activation functions acting on upper nodes. Linear models often underfit when detecting objects with minor differences. To address this, the ImCascade R-CNN model incorporates nonlinear factors into the activation function linear model. Activation functions enhance network expression capability, with nonlinear activation functions significantly improving feature expression. ResNeXt's basic residual block includes two weight layers, as shown in Figure 6: see original paper. However, the ReLU function on the main path eliminates negative signals, causing information loss. Due to ReLU's poor backpropagation performance, the residual module was redesigned to focus on identity mapping operations in ResNeXt for information propagation throughout the network.

The residual module consists of residual and identity branches. The identity branch uses ResNeXt's standard block, which updates slowly during backpropagation. Therefore, the residual branch of the ResNeXt structure was modified to create a direct path for information propagation and signal control. The improved residual module is shown in Figure 6: see original paper. Compared with the original ResNeXt structure, the improved version achieves higher accuracy without increased parameters or model complexity while preventing feature overfitting and demonstrating stronger adaptability.

(3) Detector Structure Improvement. In the Cascade Mask R-CNN model, features are first processed through the ROI Align layer, then classified and regressed in the detector. Its structure, shown in Figure 7: see original paper, contains only two fully connected layers. The extracted image features are directly input to fully connected layers for classification and regression, making the detector highly sensitive to input features such as object size, position, and orientation, resulting in low classification/regression performance and poor detection accuracy. To address these issues, a multilayer convolutional structure was introduced to improve the detector. The enhanced detector, shown in Figure 7: see original paper, contains five convolutional layers and one fully connected layer, adding a series of convolutional operations between the ROI Align layer and the fully connected layer. The front convolutional layers extract shallow features from local regions, while deeper layers extract more abstract features. Multilayer convolutions automatically extract features at different levels, fully mining hidden features and significantly improving detection performance to solve inaccurate localization during wheat grain detection.

(4) Non-Maximum Suppression Improvement. The Cascade Mask R-CNN model uses Non-Maximum Suppression (NMS), which selects candidate boxes with the highest local detection scores based on IoU values to determine the optimal detection location. However, most wheat grains in images are close together or adhered. When candidate boxes overlap significantly, grains are missed. To address this, NMS was optimized. The Soft-NMS algorithm sets a weight function for adjacent candidate boxes' detection scores based on overlap size. When the IoU between candidate box b_i and the highest-scoring box M exceeds threshold μ_t, the candidate box's detection score decays linearly, with greater overlap causing more severe decay. The confidence reset function for Soft-NMS is:

$$s_i = s_i e^{-\frac{\text{iou}(M,b_i)^2}{\sigma}}$$

where s_i represents the detection score of candidate box b_i, and $\text{iou}(M, b_i)$ represents the overlap between M and b_i. Candidate boxes are retained when $s_i \geq \mu_t$ and removed when $s_i < \mu_t$. The Soft-NMS algorithm does not operate on all candidate boxes, so computational load does not increase.

2.2.2 Method for Obtaining Phenotypic Parameters of Intact Wheat Grains

Wheat grain phenotypic parameters are important evaluation indicators reflecting grain appearance quality and morphological characteristics. This study focuses on thousand-grain weight, grain length, grain width, and length-width ratio. Through the ImCascade R-CNN model, wheat grain integrity is detected and segmented to obtain mask maps as shown in Figure 8: see original paper. Connected component labeling is performed on the image as shown in Figure 8: see original paper. Each individual wheat grain connected component is processed to extract phenotypic parameters based on connectivity and boundary characteristics.

When capturing wheat grain images, a 20 mm × 30 mm reference object was included to obtain actual phenotypic parameters and establish the correspondence between target length and pixel count at a fixed shooting distance. Wheat grain area is calculated through proportional relationships with the reference object. The calculation methods for each phenotypic parameter are as follows:

Thousand-Grain Weight (TGW) represents the weight of one thousand wheat grains. Wheat grain quantity is obtained through the counting process shown in [Figure 8: see original paper]. Sample weight is measured and proportionally converted, with multiple measurements averaged:

$$\text{TGW} = \frac{W_{\text{grain}} \times 1000}{M}$$

where W_{grain} represents wheat grain weight (g) and M represents wheat grain count.

Wheat grains approximate an oval shape as shown in [Figure 9: see original paper]. The grain edge contour is obtained from the mask, and the farthest distance between points is iteratively determined as grain length. Actual grain length calculation follows:

$$d'_h = \max_{i,j=1,2,\ldots,n} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

$$d_h = \frac{c_h}{c'_h} \times d'_h$$

where d'_h and d_h represent grain length within the connected component and actual grain length (mm), respectively; x_i, x_j, y_i, y_j are coordinates of arbitrary points on the grain mask edge curve; c'_h and c_h are reference object length in the image and actual reference length (mm), respectively.

Actual wheat grain area and width are calculated as:

$$S = \frac{S_p \times h_w}{h_p}$$

$$d_w = 4 \times \frac{S}{\pi \times d_h}$$

where S represents actual wheat grain area (mm^2), S_p represents actual reference object area (600 mm^2), h_w represents pixel count within a single wheat grain connected component, h_p represents pixel count within the reference object connected component, d_w represents actual wheat grain width (mm), and π is 3.14.

3 Results and Analysis

3.1 Grain Integrity Detection Results and Analysis

3.1.1 Evaluation Metrics Mean Average Precision (mAP) is the evaluation metric for object detection models. This study uses it to measure ImCascade R-CNN's ability to detect wheat grain integrity. Higher mAP values indicate stronger identification and localization performance. mAP is estimated from Precision (P) and Recall (R), calculated as:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{mAP} = \int_0^1 P(R)\, dR$$

$$\mathrm{mAP}_{50} = \frac{\sum_{i=1}^{n} \mathrm{AP}_i}{n}$$

where TP represents correctly predicted wheat grains, FP represents background objects incorrectly predicted as wheat grains, FN represents wheat grains incorrectly predicted as background, and n represents the number of categories.

3.1.2 Comparison of Results Before and After ImCascade R-CNN Model Improvement To validate ImCascade R-CNN's detection performance, results were compared with the Cascade Mask R-CNN model. Detection results are shown in [Figure 10: see original paper], where both models use candidate boxes to categorize wheat grains and segment contours via masks, though some missed detections are visible. Statistical results from 10 randomly selected test images are presented in .

The Cascade Mask R-CNN model adequately detects wheat grain integrity and distinguishes intact from damaged grains but exhibits missed detections with an average miss rate of 15.4%. The primary reasons are: (1) The detector with only two fully connected layers inaccurately localizes low-resolution wheat grains; (2) The NMS algorithm forces scores of adjacent candidate boxes from adhered grains to zero, causing detection failure. The ImCascade R-CNN model addresses these issues by adding convolutional layers and replacing NMS with Soft-NMS, which retains adjacent boxes using a decay function rather than zeroing them. As shown in , ImCascade R-CNN accurately detects wheat grain integrity, segments adhered grains, effectively avoids missed detections, and produces more precise contour segmentation.

Loss curves during model training are shown in [Figure 11: see original paper]. Under the same iteration count, ImCascade R-CNN converges faster and achieves lower final loss values, demonstrating superior training effectiveness compared with Cascade Mask R-CNN.

P-R curves, important indicators for model detection performance, reflect computational precision and recall. Greater convexity in P-R curves indicates the model maintains high recall and precision simultaneously. P-R curves for both models are shown in [Figure 12: see original paper]. As detection precision increases, ImCascade R-CNN's recall decreases less noticeably, indicating better localization performance. As recall increases, ImCascade R-CNN's precision decreases less noticeably, indicating better classification performance. The mAP_{50} curves in [Figure 13: see original paper] show ImCascade R-CNN's mAP_{50} value is significantly higher.

To analyze ImCascade R-CNN's detection performance, ablation experiments were conducted using four modification modes: changing the backbone to ResNeXt, introducing multilayer convolution (CONV) structure in the detector, adding Mish as the activation function, and adopting Soft-NMS. Results in show each modification significantly improves wheat grain integrity detection performance. Compared with Cascade Mask R-CNN, ImCascade R-CNN improves precision, recall, and mAP_{50} by 16.3%, 17.4%, and 14.5%, respectively.

3.1.4 Comparative Experiments with Different Models To further analyze detection performance, ImCascade R-CNN was compared with Cascade Mask R-CNN and Deeplabv3+ for semantic segmentation. Results in show ImCascade R-CNN achieves superior precision, recall, and mAP values. The ImCascade R-CNN model demonstrates clear advantages in wheat grain integrity identification, localization, and segmentation, with an average detection precision of 90.2%.

3.2 Phenotypic Parameter Acquisition Results and Analysis

The proposed method's results were compared with manual measurements, yielding average error rates of 2.15% for grain length and 3.74% for grain width, with a length-width ratio standard error of 0.15. Statistical fitting of grain length and width obtained through the proposed method produced determination coefficients (R^2) of 0.9351 and 0.8217, respectively, as shown in [Figure 14: see original paper]. These results demonstrate high consistency with manual measurements, confirming the method meets wheat seed testing requirements.

Conclusion

Manual identification of wheat grain phenotypic parameters is inefficient and subjective. To achieve large-scale rapid and precise detection and accelerate wheat breeding, this study developed the ImCascade R-CNN model and proposed a wheat grain phenotypic identification method. To address missed detections and low localization precision in Cascade Mask R-CNN, ImCascade R-CNN improved the backbone network to ResNeXt, added Mish as the activation function, introduced multilayer convolution structure in the detector, and adopted Soft-NMS. These improvements increased precision, recall, and mAP_{50} by 16.3%, 17.4%, and 14.5%, respectively. Compared with Deeplabv3+, ImCascade R-CNN achieved 90.2% average detection precision for wheat grain integrity, better identifying, localizing, and segmenting wheat grains to provide a foundation for phenotypic parameter acquisition.

The method obtained grain length and width with average error rates of 2.15% and 3.74%, respectively, and length-width ratio standard error of 0.15, with determination coefficients of 0.9351 and 0.8217 compared with manual measurements. These results demonstrate high consistency with manual identification, meeting wheat seed testing requirements and accelerating the cultivation of high-quality, high-yield wheat varieties.

Conflict of Interest Statement: This study has no conflicts of interest among researchers or with publicly disclosed research findings.

References

[1] NAGY É, LEHOCZKI-KRSJAK S, LANTOS C, et al. Phenotyping for testing drought tolerance on wheat varieties of different origins[J]. South African journal of botany, 2018, 116: 216-221.

[2] CHEN J, LIAN Y, ZOU R, et al. Real-time grain breakage sensing for rice combine harvesters using machine vision technology[J]. Agricultural engineering technology, 2020, 40(30): ID 94.

[3] GAO L, YANG J, SONG S J, et al. Genome-wide association study of grain morphology in wheat[J]. Euphytica, 2021, 217(8): 1-12.

[4] LI H Y, YIN G H. On development trend of China's wheat breeding from perspective of national grain security[J]. Jiangsu agricultural sciences, 2022, 50(18): 36-41.

[5] GAO H, ZHEN T, LI Z H. Detection of wheat unsound kernels based on improved ResNet[J]. IEEE access, 2022, 10: 20092-20101.

[6] SHAHIN M A, SYMONS S J. Detection of fusarium damage in Canadian wheat using visible/near-infrared hyperspectral imaging[J]. Journal of food measurement & characterization, 2012, 6(1/2/3/4): 3-11.

[7] FENG J K, ZHENG Y, LI P, et al. Research on the identification of wheat grain varieties based on feature selection[J]. Journal of Chinese agricultural mechanization, 2022, 43(7): 116-123.

[8] ZHAO W Y, LIU S Y, LI X Y, et al. Fast and accurate wheat grain quality detection based on improved YOLOv5[J]. Computers and electronics in agriculture, 2022, 202(11): ID 107426.

[9] WANG Y, LI Y, WU T T, et al. Fast counting method of soybean seeds based on density estimation and VGG-two[J]. Smart agriculture, 2021, 3(4): 111-122.

[10] LIU H, WANG Y Q, WANG X M, et al. Study on detection method of wheat unsound kernel based on near-infrared hyperspectral imaging technology[J]. Spectroscopy and spectral analysis, 2019, 39(1): 223-229.

[11] SONG H B, WANG Y F, DUAN Y C, et al. Detection method of severe adhesive wheat grain based on YOLO v5-MDC model[J]. Transactions of the Chinese society for agricultural machinery, 2022, 53(4): 245-253.

[12] HUANG S, FAN X F, SUN L, et al. Research on classification method of maize seed defect based on machine vision[J]. Journal of sensors, 2019, 2019: 1-9.

[13] ZHU S P, ZHUO J X, HUANG H, et al. Wheat grain integrity image detection system based on CNN[J]. Transactions of the Chinese society for agricultural machinery, 2020, 51(5): 36-42.

[14] XU L X, CHEN J W, DING G H, et al. Indoor phenotyping platforms and associated trait measurement: Progress and prospects[J]. Smart agriculture, 2020, 2(1): 23-42.

[15] ZHAO H M, GE C J, JIA J Q, et al. High-throughput phenotyping system of wheat grains based on image analysis[J]. Shandong agricultural sciences, 2021, 53(6): 113-120.

[16] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 6154-6162.

[17] PURKAIT P, ZHAO C, ZACH C. SPP-net: Deep absolute pose regression with synthetic views[EB/OL]. arXiv:1712.03452, 2017.

[18] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2016: 1440-1448.

[19] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[20] WU W Q, YIN Y J, WANG X G, et al. Face detection with different scales based on faster R-CNN[J]. IEEE transactions on cybernetics, 2019, 49(11): 4017-4028.

Visit www.smartag.net.cn for free access to the full electronic version.

Note: Figure translations are in progress. See original paper for figures.