Postprint: Silkworm Recognition and Counting in Industrial Rearing Based on an Improved Mask R-CNN Model

Authors: He Ruimin, Zheng Kefeng, Wei Qinyang, Zhang Xiaobin, Zhang Jun, Zhu Yihang, Zhao Yiying, Gu Qing

Date: 2023-02-17T00:00:00+00:00

Abstract

Precision feeding is one of the core technologies for cost reduction and efficiency improvement in full-age artificial feed factory-scale silkworm rearing, and automated identification and counting of silkworms is a key link in achieving precision feeding. This study obtains digital images of silkworms at the 4th and 5th instar stages during factory-scale rearing based on a machine vision system, and utilizes an improved deep learning model Mask R-CNN to detect silkworm bodies and residual feed. By incorporating a pixel reweighting strategy and a bounding box refinement strategy into the Mask R-CNN model framework, a more robust object detection model is trained from noisy data, achieving optimization of model performance and improving the detection and segmentation capabilities for silkworm and feed boundaries. For the improved Mask R-CNN model, the Average Precision at IoU=0.5 (AP50) for silkworm detection and segmentation is 0.790 and 0.795, respectively, with an identification accuracy of 96.83%; for residual feed detection and segmentation, the AP50 values are 0.641 and 0.653, respectively, with an identification accuracy of 87.71%. The model is deployed on the NVIDIA Jetson AGX Xavier development board, with an average detection time of 1.32 s per image and a maximum detection time of 2.05 s, and its computational speed can meet the requirements for real-time detection of silkworm rearing box units moving on the production line. This study provides a core algorithm for the research and development of precision feeding information systems and feeding devices for factory-scale silkworm rearing, which can improve the utilization rate of artificial feed and enhance the production management level of factory-scale silkworm rearing.

Full Text

Identification and Counting of Silkworms in Factory Farm Using Improved Mask R-CNN Model

HE Ruimin[1], ZHENG Kefeng[2], WEI Qinyang[1], ZHANG Xiaobin[2], ZHANG Jun[1], ZHU Yihang[2], ZHAO Yiying[2], GU Qing[2]*

[1]Shengzhou Mosang High-tech Co., Ltd., Shaoxing, Zhejiang 312400, China
[2]Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China

Abstract: Factory-like rearing of silkworm (*Bombyx mori*) using artificial diet for all instars is a brand-new rearing mode that replaces mulberry leaves with artificial feed and achieves automated large-scale year-round silkworm breeding through environmental control and industrial assembly lines. This approach revolutionizes traditional sericulture and represents an important direction for industry transformation and upgrading. Precise feeding is one of the core technologies for cost reduction and efficiency improvement in factory silkworm rearing, and automatic identification and counting of silkworms is the key to achieving precise feeding. This study obtained digital images of silkworms during the 4th and 5th instars using a machine vision system in factory rearing conditions, and employed an improved deep learning model to detect both silkworms and residual feed. By incorporating a pixel reweighting strategy and a bounding box fine-tuning strategy into the Mask R-CNN framework, a more robust object detection model was trained from noisy data to optimize model performance and improve detection and segmentation capabilities for silkworm and feed boundaries. The improved Mask R-CNN model achieved Average Precision at IoU=0.5 (AP50) of 0.790 and 0.795 for silkworm detection and segmentation, respectively, with an identification accuracy of 96.83%. For residual feed detection and segmentation, the AP50 values were 0.641 and 0.653, respectively, with an identification accuracy of 87.71%. The average detection time per image was 1.32 s, with a maximum detection time of 2.05 s. When deployed on an NVIDIA Jetson AGX Xavier development board, the computational speed meets the requirements for real-time detection of moving silkworm box units on the production line. This research provides a core algorithm for the development of precise feeding information systems and feeding devices in factory silkworm rearing, which can improve artificial feed utilization and enhance production management levels.

Keywords: silkworm; artificial diet; precise feeding; machine vision; deep learning; Mask R-CNN; noisy data

1 Introduction

Factory-like rearing of silkworm using artificial diet for all instars is an innovative breeding mode that replaces mulberry leaves with artificial feed and achieves automated large-scale year-round silkworm breeding through environmental control and industrial assembly lines. This approach revolutionizes traditional sericulture and represents an important direction for industry transformation and upgrading [1-3]. Research on the technology system of factory silkworm rearing has achieved numerous results, with large-scale silkworm factories using artificial diet for all instars having been built and put into operation [4,5]. Artificial feed production constitutes the main input in factory silkworm rearing, and feed processing and feeding are among the most important technical links. Therefore, strict control of feed consumption and improvement of feed utilization are crucial for controlling production costs.

Currently, factory silkworm rearing adopts a constant feeding mode, where the amount of artificial feed delivered to each silkworm box unit is the same within the same instar. However, as rearing progresses, the number of silkworms in each box can vary significantly due to multiple factors, including different initial numbers of newly hatched silkworms, silkworm mortality, manual removal of suspected diseased individuals, and silkworms crawling out of the boxes. Consequently, constant feeding leads to uneven feeding, excessive or insufficient feed, ultimately resulting in feed waste or reduced cocoon quality. Therefore, precise feeding of artificial feed is essential for improving feed utilization, reducing rearing costs, and enhancing cocoon quality. Feeding based on silkworm count is an effective method to achieve precise feeding, which requires rapid and accurate detection of silkworm numbers in each box unit, conversion to the required feed amount, and transmission to the feeding device. Upon receiving the signal, the device adjusts the discharge volume in real-time to achieve precise feeding.

Object detection technology in image recognition can be used for silkworm identification and counting. Traditional object detection methods are mostly based on edge-related features of target objects [6,7]. Although they can achieve good detection accuracy and speed in specific scenarios, their adaptability and generalization are weak. In recent years, deep learning technology has been widely applied to object detection. Deep learning-based object detection methods can adaptively extract image features at different levels, and the trained models can be applied to different scenarios, significantly improving model accuracy and generalization capability [8-12]. Mask R-CNN is one of the commonly used deep learning algorithms in recent years and has achieved excellent performance in numerous applications. Mask R-CNN improves upon Faster R-CNN by using RoIAlign instead of RoIPooling and employs bilinear interpolation algorithms to reduce positional errors in bounding box regression [13]. The model not only demonstrates good detection performance but also enables pixel-level segmentation of detected targets, which meets the requirements of this study' s application scenario. However, when detecting silkworms and residual feed, issues such as unclear data annotation, background interference in annotation

regions, and overlapping or 粘连 (adhesion) of target object contours introduce noise into the training data, reducing the accuracy and stability of model detection and mask segmentation. To address this problem, this study improves and adjusts the Mask R-CNN model using noisy data by adding a pixel reweighting strategy and a bounding box fine-tuning strategy to the model framework, thereby enhancing the model' s segmentation capabilities for silkworm and feed boundaries.

This study uses a machine vision system to obtain digital images of 4th and 5th instar silkworms during factory rearing with artificial diet. The improved Mask R-CNN model is employed to detect silkworms and residual feed in the images. The segmentation mask outputs are then used to evaluate silkworm growth development and feed residue conditions, providing algorithmic support for the development of precise feeding equipment and management systems in factory silkworm rearing, enabling precise control of artificial feed delivery, and improving feed utilization efficiency.

---

2 Materials and Methods

2.1 Data Collection

2.1.1 Data Acquisition Data collection was conducted at the artificial diet feeding workshop of Shengzhou Mosang High-tech Co., Ltd. in Zhejiang Province (29°35' N, 120°51' E). The silkworm variety was "Zhong 2016 × Ri 2016," a specialized breed for factory silkworm rearing. The image acquisition environment had a temperature of (25$±$1)°C and humidity of 60%-70%.

Silkworms undergo five instars from hatching to cocooning, with a dormant period (molting) between each instar. After each molt, silkworms grow larger and require more food. Silkworms in the first three instars require minimal feed, accounting for only about 5% of the total feed consumption during the entire growth cycle, while the 4th and 5th instars consume over 95% of the total feed. Therefore, this study focused on identifying silkworms in the 4th and 5th instars only.

The image acquisition equipment was an FLIR Blackfly S USB3 industrial camera with a Changbudao FA3516A lens. The camera specifications were: 20-megapixel resolution, 35 mm fixed-focus lens, F2.8 aperture, C-Mount lens type, maximum resolution of 5472$×$3648, frame rate of 18 f/s, and pixel size of 2.4 m. Silkworm boxes moved horizontally on the assembly line system at 1.5 m/s and stopped for approximately 5 s during feed delivery. The box dimensions were 60 cm × 100 cm. The camera was installed 2 m directly above the stopping position of the silkworm box unit and captured images during the box dwell period. The camera was connected to a laptop computer via a USB 3.0 interface and controlled using the camera' s accompanying application software.

2.1.2 Image Preprocessing The original images were large in size and contained numerous silkworms, posing challenges for object annotation and modeling. To improve annotation efficiency and standardize image dimensions for data processing and subsequent analysis, the original image data were cropped to a uniform size of 2000$×$2000 pixels, from which clearer data were selected for further processing.

Data augmentation can improve image data quality and expand the training dataset scale [14]. This study employed three methods for augmenting original images: rotation and flipping, brightness enhancement, and noise addition. Rotation and flipping are widely used image augmentation methods [14,15], and all images were processed with 90°, 180°, and 270° rotations and horizontal mirror flipping. Due to strict lighting requirements in factory silkworm rearing, the silkworm rooms were relatively dark and could not be supplemented with conventional lighting methods, resulting in low brightness of original images. Therefore, image brightness was increased by 20% to compensate for insufficient ambient light intensity. Additionally, random noise may be generated due to instability of image acquisition equipment. To address this, Gaussian noise with a variance of 0.01 was added to original images following Zhou et al. [14] to enhance model robustness. The processed data were added to the training set for model training.

After image augmentation and numbering, the open-source graphical annotation tool Labelme was used for image annotation. Polygons were drawn to annotate silkworms and residual feed, and annotated images were saved as *.json files. A total of 180 high-quality cropped images were selected for annotation, including 90 images of 4th instar and 90 images of 5th instar silkworms. Each image contained approximately 150-200 silkworm annotations and 30-50 residual feed annotations, with incomplete silkworms at image edges also being annotated. [Figure 1: see original paper] shows examples of cropped original images of 4th and 5th instar silkworms and different preprocessing results.

2.2 Mask R-CNN

Mask R-CNN, proposed by He et al. [13], achieves instance segmentation while performing object detection by adding a mask branch to the Faster R-CNN network. The model follows the Faster R-CNN concept, uses ResNet-FPN architecture for feature extraction, and adds a Mask prediction branch. Mask R-CNN consists of three main modules: Faster R-CNN, RoIAlign, and Fully Convolutional Networks (FCN). It employs the same two-stage detection method as Faster R-CNN. In the first stage, a Region Proposal Network (RPN) is established for feature extraction [16]. In the second stage, Mask R-CNN introduces the RoIAlign method to replace the RoIPooling resampling method of Faster R-CNN [17]. In addition to category detection, Mask R-CNN outputs a binary segmentation mask for each candidate object [18]. The loss function L of Mask R-CNN is expressed as:

$$L = Lcls + Lloc + Lmask$$

where Lcls is the classification loss, Lloc is the bounding box regression loss, and Lmask is the mask loss.

To enable multi-scale silkworm prediction for different sizes, a Feature Pyramid Network (FPN) was adopted for multi-scale feature learning. FPN uses a top-down and bottom-up bidirectional multi-scale bounding box prediction method that fuses features from various levels, enabling them to possess both strong semantic and spatial information [19]. [Figure 2: see original paper] shows the structure of FPN. The Mask R-CNN in this study used a ResNet50-FPN backbone network for feature extraction. During RPN generation, anchor sizes were 32, 64, 128, 256, and 512, with scales of 0.5, 1.0, and 2.0.

2.3 Improved Mask R-CNN Model

2.3.1 Issues with Original Mask R-CNN and Improvements Although the original Mask R-CNN model demonstrates superior performance on natural image segmentation, its segmentation performance for small objects and heavily occluded scenarios requires improvement. Due to mutual overlapping, adhesion, and interference from other background objects, the boundaries of silkworms and residual feed may be ambiguous and difficult to define, leading to errors in annotation data and introducing noise into model training. Under noisy conditions, segmentation models may encounter the following problems: (1) Incorrect category labels corrupt the detector; (2) Incorrect segmentation masks mislead the model into producing imprecise mask predictions. An excessively large annotation region will cover more background area, while an excessively small annotation region cannot cover the complete silkworm or feed. These factors prevent the model from generating accurate masks; (3) Noisy annotations lead to unstable training processes. When training data lack correct category labels or precise mask annotations, the model exhibits instability during training, and an unstable loss function prevents learning parameters from converging to optimal solutions. To address these impacts of annotation noise on the model, this study added a pixel reweighting strategy and a bounding box fine-tuning strategy to the original Mask R-CNN framework to train a more robust object detection model from noisy data, achieving model performance optimization and improvement.

2.3.2 Main Processing Flow The training process of the improved Mask R-CNN model based on noisy data is as follows: (1) The dataset is divided into a training set and a meta-test set, where images in the meta-test set have been manually verified as completely correct annotations, while the training set annotations contain noise. (2) Transfer learning strategy is used to fine-tune and retrain the pre-trained CNN model. The pre-trained model was trained on the COCO (Common Objects in Context) dataset. The original Mask R-CNN framework is used to train the training set data to obtain initial classification, bounding box regression, and mask segmentation results. (3) The

pixel reweighting strategy takes classification loss values as input, with 1 representing parameters in the pixel reweighting strategy. After passing through a two-layer perceptron, it outputs weight β as the noise level of pixels. When pixel labels are incorrect, the pixel reweighting module outputs smaller weights, while for correct pixel labels, it outputs larger weights. (4) The bounding box fine-tuning strategy takes bounding box regression loss as input, with 2 representing parameters in the bounding box fine-tuning strategy. After passing through a two-layer perceptron, it outputs bounding box displacement transformation parameters to adaptively modify imprecise annotation boundaries and produce more accurate estimates. After bounding box fine-tuning, the detector is optimized through more accurate regression loss.

The specific algorithms for the pixel reweighting strategy and bounding box fine-tuning strategy are detailed in Xu et al. [20]. [Figure 3: see original paper] shows the framework flow of the improved Mask R-CNN model for silkworm and residual feed detection. All 180 images were divided into four datasets: training set, validation set, meta-test set, and test set. One hundred images were randomly selected as the training set for model training. Twenty images were selected as the validation set for hyperparameter tuning and selection. The meta-test set contained manually verified completely correct annotation data for pixel reweighting and bounding box fine-tuning, consisting of 20 images. Forty images served as the test set for model performance evaluation. Each dataset contained equal numbers of 4th and 5th instar silkworm images.

2.3.3 Detection Speed Detection speed is an important evaluation metric for object detection algorithms. In this study, the feed delivery process occurs on an assembly line where silkworm box units continuously move and only stop briefly during feed dispensing. Therefore, to achieve uninterrupted detection, the model' s computational speed must meet certain requirements. Two metrics were used to evaluate the computational speed of the deep learning model: maximum running time (Tmax) and average running time (ART) [14,16]. Under specific hardware configurations, Tmax represents the longest time required for model detection on a single test set image, while ART represents the average time required for model detection on test images, expressed in seconds per image.

2.4 Model Performance Evaluation

Prediction results can be divided into four categories: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). TP represents positive samples predicted as positive by the model, TN represents negative samples predicted as negative, FP represents negative samples predicted as positive, and FN represents positive samples predicted as negative. Before determining the classification of these four prediction types, the Intersection over Union (IoU) threshold must be defined. IoU measures the overlap rate between detected boundaries and ground truth boundaries (annotation boundaries), expressed as

the ratio of the overlapping area to the union area of the two regions:

IoU = (CandidateBox GroundTruth) / (CandidateBox GroundTruth)

where CandidateBox is the detected result boundary and GroundTruth is the annotation boundary. If IoU equals 1, the prediction completely coincides with the annotation; the closer IoU is to 1, the better the prediction. In this study, IoU = 0.5 was defined as the threshold for determining prediction results. If IoU is greater than 0.5, the prediction is considered successful and classified into one of the four categories; if less than 0.5, it is considered a failed prediction.

2.4.1 Accuracy Accuracy is the proportion of correctly classified samples to the total number of samples [8]:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

where TP+TN is the number of correctly classified samples, and TP+TN+FP+FN is the total number of samples. The accuracy for the entire dataset is the average of all image prediction results.

2.4.2 Average Precision Precision and Recall are commonly used metrics for evaluating deep learning model performance, calculated as:

Precision = TP / (TP + FP)
Recall = TP / (TP + FN)

where TP+FP represents the number of predicted objects, and TP+FN represents the number of actual objects. Precision indicates the proportion of correct predictions among all positive predictions, while Recall indicates the proportion of correctly predicted target samples among all target samples. Higher Precision and Recall values indicate better model performance. Plotting Precision on the vertical axis and Recall on the horizontal axis produces a Precision-Recall (P-R) curve. Average Precision (AP) is defined as the average Precision value at different Recall values, calculated as the integral of the P-R curve [21]:

AP = $\int_0^1$ p(r) dr

where p represents Precision and r represents Recall. The AP value is the area under the P-R curve. AP is one of the most commonly used metrics for object detection model performance evaluation. In this study, AP was calculated at an IoU threshold of 0.5, denoted as AP50.

---

3 Results and Analysis

3.1 Experimental Setup and Model Training Parameters

Transfer learning strategy was used to fine-tune and retrain the pre-trained model. The pre-trained model was trained on the COCO dataset. The initial

learning rate was 0.02, reduced to 0.0001 at 50,000 steps and to 0.00001 at 70,000 steps. FPN was executed on the outputs of residual block units 2, 3, 4, and 5. The open-source deep learning framework PyTorch was used for model training with Python as the programming language. Experiments were conducted on the Ubuntu operating system with a computer configuration of 32.0 GB RAM and an Intel® Core™ i7-9700K CPU @ 3.60 GHz × 8 processor. Training was performed in parallel on four NVIDIA Tesla V100 Graphics Processing Units (GPUs). The validation set data were used for model evaluation and hyperparameter tuning to select optimal parameters for building the model, which was then evaluated using the test set data. Other initial model parameters are shown in .

3.2 Detection and Segmentation Results

Prediction results were compared with annotation data to evaluate model performance. Detection results of the improved Mask R-CNN model are shown in [Figure 4: see original paper], where bounding boxes represent detection results for silkworms and residual feed, and masks represent segmentation results. The figure demonstrates that the improved Mask R-CNN model performed well in locating target objects and could accurately identify silkworms. Despite similarities between molted skin and silkworm bodies, the model could accurately distinguish between them. The model also showed good identification performance for overlapping silkworms.

shows the performance of the Mask R-CNN model and the improved Mask R-CNN model. The original Mask R-CNN achieved detection and segmentation AP50 values of 0.764 and 0.768 for silkworms, respectively, with an identification accuracy of 95.23%. Residual feed identification performed worse than silkworm identification, with detection and segmentation AP50 values of 0.602 and 0.611, respectively, and an identification accuracy of 85.35%. The improved Mask R-CNN model achieved silkworm detection and segmentation AP50 values of 0.790 and 0.795, respectively, with an identification accuracy of 96.83%. For residual feed, the detection and segmentation AP50 values were 0.641 and 0.653, respectively, with an identification accuracy of 87.71%. These results indicate that the improved Mask R-CNN model offers enhanced performance compared to the original Mask R-CNN and demonstrates good performance in detecting and segmenting both silkworms and residual feed, making it suitable as a core algorithm for developing precise feeding control systems and hardware devices in factory silkworm rearing.

In terms of detection speed, the trained improved Mask R-CNN model achieved an average running time (ART) of 0.075 s and a maximum running time (Tmax) of 0.142 s for detecting test set images on a computer configured with an NVIDIA Tesla V100 GPU and i7-9700K CPU. When the model was deployed on the NVIDIA Jetson AGX Xavier development board for testing, the ART was 1.32 s and Tmax was 2.05 s. This computational speed meets the real-time detection requirements for silkworm bodies and residual feed based on machine vision

systems on the feeding assembly line.

3.3 Model Performance for Different Instars

Due to different morphological characteristics of silkworms at different instars, models trained with image data from different growth stages may exhibit varying detection performance. To compare model performance trained with data from different instars, the entire dataset was divided into two categories: 4th instar and 5th instar, and silkworm detection models were trained and tested separately. As shown in , the 5th instar model performed better than the complete data model shown in , while the 4th instar model performed worse than the complete data model. Compared with each other, the 5th instar dataset-trained model demonstrated better performance on test data than the 4th instar model. This indicates that the instar of training images affects model detection performance, with models for older silkworms performing better than those for younger ones. This is because 4th instar silkworms are relatively small, dark yellow in color, and densely distributed, while 5th instar silkworms have clearer contour features, larger individuals, and less overlapping.

3.4 Impact of Silkworm Overlap on Detection Performance

In production, overlapping among silkworms is common and may affect detection results. This section analyzes the detection performance of the improved Mask R-CNN model for overlapping silkworms. [Figure 5: see original paper] shows detection and segmentation output examples for overlapping silkworms. In test images, overlapping silkworms were counted individually. The test set contained 823 overlapped silkworms, accounting for 13.4% of the total silkworm count. Silkworm bodies that were segmented into two or three parts due to overlapping but were accurately identified as the same silkworm were considered correct identifications. The overall detection accuracy for these overlapped silkworms was 95.06%, slightly lower than the detection accuracy for the entire test dataset (96.83%). This indicates that overlapping situations affect detection performance to some extent. Nevertheless, the improved Mask R-CNN model still demonstrates satisfactory detection performance for overlapping silkworms, with accuracy close to that of the complete dataset, indicating strong detection capability for silkworms and the ability to handle complex overlapping situations.

3.5 Impact of Data Augmentation on Detection Performance

To evaluate the impact of data augmentation on model performance, models built using the complete dataset were compared with those built using datasets with different processing images removed. As shown in , the three different image augmentation methods affected model accuracy to varying degrees. The brightness enhancement method contributed most significantly to model performance. Removing brightness-enhanced images reduced model accuracy by 3.49%, with noticeable decreases in detection AP50 and segmentation AP50 as

well. Rotation and flipping processing contributed less to model performance, with model accuracy decreasing by 2.04% after removing rotated and flipped images. Gaussian noise addition had no significant effect on model performance.

4 Discussion and Conclusion

4.1 Discussion

Factory silkworm rearing with artificial feed represents a technological innovation in sericulture, providing a new direction for China's sericulture development. Feeding artificial feed based on silkworm count can effectively improve feed utilization, reduce costs, and enhance overall cocoon quality. Accurate silkworm counting also provides reference data for predicting cocoon yield, estimating feed consumption, and calculating silkworm losses, supporting production decision-making and improving production management levels in factory silkworm rearing. In recent years, deep learning technology has been increasingly applied to object detection. Object detection results based on digital images depend on multiple factors, including target object size, distribution, overlap degree, image quality, and training sample size. In this study, data categories (different instars, overlapping silkworms) affected detection results, consistent with findings from Tian et al. [22]. Data augmentation processing can improve the detection capability of deep learning models, which aligns with other research results [14,15]. The brightness enhancement method provided the greatest performance improvement, which is consistent with the dark lighting conditions in silkworm rooms that cannot be supplemented with conventional lighting methods.

As the entire technology system continues to be updated and improved, large-scale industrialized factory silkworm rearing using artificial feed for all instars is becoming increasingly mature. This study confirms the feasibility of deep learning technology for detecting silkworms and feed residue in factory silkworm rearing. However, several issues remain to be addressed in future work. For example, excessive overlapping among silkworms reduces model detection performance, and future efforts need to analyze its impact further and increase such training samples to improve model detection and segmentation capabilities. For residual artificial feed detection, identification is challenging due to irregular shapes, diverse and uneven surface textures, and partial coverage by silkworm bodies. Additionally, silkworm feces have similar color and surface texture to residual feed, making data annotation difficult and error-prone, while incorrect annotations lead to identification errors. Therefore, the model still has considerable room for improvement in residual feed detection performance.

The segmentation mask outputs from the improved Mask R-CNN model can be further analyzed to estimate silkworm size and uniformity, as well as residual artificial feed weight. This information can be used for feed consumption management and cocoon size and uniformity prediction. Moreover, this technology has significant application potential in phenotypic analysis for silkworm

breeding, including evaluation of silkworm size, uniformity, mortality rate, and growth rate calculation.

4.2 Conclusion

To achieve automatic identification and counting of silkworms in rearing boxes, this study proposed an improved Mask R-CNN model using noisy data for detecting silkworms and residual artificial feed, providing a core algorithm for developing precise feeding management systems and feeding devices in factory silkworm rearing.

The improved Mask R-CNN model demonstrated good detection capability for silkworms and artificial feed residue, achieving overall detection accuracies of 96.83% and 87.71%, respectively. The AP50 values for silkworm detection and segmentation were 0.790 and 0.795, respectively, while those for residual feed detection and segmentation were 0.641 and 0.653, respectively. When tested on the NVIDIA Jetson AGX Xavier development board, the model achieved an ART of 1.32 s and Tmax of 2.05 s, enabling real-time detection of silkworm bodies and residual feed on industrial assembly lines.

The model established in this study meets the requirements for rapid and accurate detection of silkworms in rearing box units on industrial assembly lines in terms of both accuracy and computational speed. Therefore, it can serve as the core algorithm for developing precise feeding control information systems and feeding devices in factory silkworm rearing.

Future work will focus on further improving model accuracy, robustness, and stability. To enhance model performance, larger-scale training image datasets will be introduced, particularly samples with overlapping and adhered silkworms and more diverse residual feed samples. Additionally, different model structures will be selected for comparative analysis to conduct more in-depth research on image feature extraction, detection, and contour segmentation for silkworms and residual feed.

References

[1] TANAKA Y, SUDO M. Studies on the technology of artifical diet rearing for parental strains of the silkworm, 4: The relationship between the water content of artificial diets for the fifth larval instar and egg laying results[J]. Journal of Dainippon Silk Foundation, 2006, 53: 1-5.

[2] WU Y, ZHANG S, WANG H, et al. Inheritance pattern of feeding habit on artificial diet in different bombyx mori varieties[J]. Science of Sericulture, 2017, 43(4): 595-600.

[3] QIAN Q, CHEN W. Research and application progress of artificial diet for silkworm[J]. Bulletin of Sericulture, 2016, 47(2): 11-14.

[4] DONG J, PAN M, WU H. Thinking on speeding up the transformation and development of sericulture Industry—The exploration and enlightenment based on the BABEI' s silkworm rearing in the factory[J]. Bulletin of Sericulture, 2018, 49(2): 14-16.

[5] WANG L, HU S. Babe model of promoting industrial sericulture by feeding artificial diet of full larval stage[J]. Bulletin of Sericulture, 2020, 51(1): 37-45.

[6] DOU J, LI J. Robust object detection based on deformable part model and improved scale invariant feature transform[J]. Optik-International Journal for Light and Electron Optics, 2013, 124(24): 6485-6492.

[7] HONG G S, KIM B G, HWANG Y S, et al. Fast multi-feature pedestrian detection algorithm based on histogram of oriented gradient using discrete wavelet transform[J]. Multimedia Tools and Applications, 2015, 75(23): 1-17.

[8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2014.

[10] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(9): 1904-1916.

[11] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2015.

[12] REN S, HE K, 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, 2015, 39(6): 1137-1149.

[13] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]// IEEE International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2017.

[14] ZHOU C, HU J, XU Z, et al. A Novel greenhouse-based system for the detection and plumpness assessment of strawberry using an improved deep learning technique[J]. Frontiers in Plant Science, 2020, 11: ID 559.

[15] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417-426.

[16] ZHANG Y, XIAO D, CHEN H, et al. Rice panicle detection method based on improved Faster R-CNN[J]. Transactions of the CSAM, 52(8): 231-240.

[17] WEN Q, LUO Z, CHEN R, et al. Deep learning approaches on defect detection in high resolution aerial images of insulators[J]. Sensors, 2021, 21(4): ID 1033.

[18] WJPD A, YT A, RONG L.B, et al. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot[J]. Computers and Electronics in Agriculture, 2020, 172(6): ID 105380.

[19] M.AKHAN, ZHANG Y D, SHARIF M, et al. Pixels to classes: Intelligent learning framework for multiclass skin lesion localization and classification[J]. Computers and Electrical Engineering, 2021, 90: 1-20.

[20] XU Y, ZHU L, YANG Y, et al. Training robust object detectors from noisy category labels and imprecise bounding boxes[J]. IEEE Transactions on Image Processing, 2021, 30: 5782-5792.

[21] ZHANG Y, CHU J, LENG L, et al. Mask-refined R-CNN: A network for refining object details in instance segmentation[J]. Sensors, 2020, 20(4): ID 1010.

[22] TIAN Y, YANG G, WANG Z, et al. Instance segmentation of apple flowers using the improved mask R-CNN model[J]. Biosystems Engineering, 2020, 193: 264-278.

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