Novel image segmentation model of multi-view sheep face for identity recognition

Suhui Liu[1,2], Guangpu Wang[1], Chuanzhong Xuan[1,2*], Zhaohui Tang[1,2], Junze Jia[1]

(1. *College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China*;
2. *Inner Mongolia Engineering Research Center for Intelligent Facilities on Prataculture and Aquaculture, Hohhot 010018, China*)

Abstract: Traditional sheep identification is based on ear tags. However, the application of ear tags not only causes stress to the animals but also leads to loss of ear tags, which affects the correct recognition of sheep identity. In contrast, the acquisition of sheep face images offers the advantages of being non-invasive and stress-free for the animals. Nevertheless, the extant convolutional neural network-based sheep face identification model is prone to the issue of inadequate refinement, which renders its implementation on farms challenging. To address this issue, this study presented a novel sheep face recognition model that employs advanced feature fusion techniques and precise image segmentation strategies. The images were preprocessed and accurately segmented using deep learning techniques, with a dataset constructed containing sheep face images from multiple viewpoints (left, front, and right faces). In particular, the model employs a segmentation algorithm to delineate the sheep face region accurately, utilizes the Improved Convolutional Block Attention Module (I-CBAM) to emphasize the salient features of the sheep face, and achieves multi-scale fusion of the features through a Feature Pyramid Network (FPN). This process guarantees that the features captured from disparate viewpoints can be efficiently integrated to enhance recognition accuracy. Furthermore, the model guarantees the precise delineation of sheep facial contours by streamlining the image segmentation procedure, thereby establishing a robust basis for the precise identification of sheep identity. The findings demonstrate that the recognition accuracy of the Sheep Face Mask Region-based Convolutional Neural Network (SFMask R-CNN) model has been enhanced by 9.64% to 98.65% in comparison to the original model. The method offers a novel technological approach to the management of animal identity in the context of sheep husbandry.

Keywords: image segmentation, sheep face, deep learning, multi-view, feature fusion
DOI: 10.25165/j.ijabe.20251806.9678

Citation: Liu S H, Wang G P, Xuan C Z, Tang Z H, Jia J Z. Novel image segmentation model of multi-view sheep face for identity recognition. Int J Agric & Biol Eng, 2025; 18(6): 260–268.

1 Introduction

In China's agricultural economic development, livestock plays a pivotal role, particularly in the critical industry[1,2]. The digital transformation in large-scale sheep farming is gradually unfolding, with a focus on collecting digital information about each sheep to facilitate precision farming, which is becoming the main direction of modern scientific breeding[3-5]. Achieving this goal hinges on the effective identification and management of sheep[6]. Nevertheless, there are issues that require attention. For instance, sheep only exhibit a stress response during the tagging process, and the constraints of conventional techniques that depend on electronic ear tagging must be acknowledged[7]. In order to address these challenges, this study presents a sheep identity management system based on contactless sheep facial image recognition[8,9].

Previous research has extensively explored animal identification using facial information. For instance, Salama et al.[10] and Ning et al.[11] have made significant strides in sheep identification, with the latter achieving an average accuracy of 97.41% using an improved You Only Look Once version 5 small (YOLOv5s) model, surpassing the base model by 2.21%. This provides a robust method for individual livestock identification in smart farming. Zhang et al.[12] proposed a MobileFaceNet-based model for sheep face recognition, enhancing feature extraction capabilities while maintaining a lightweight network architecture. Huang et al.[13] adapted a human face key point detection model, YOLOv5 Face, for pig face key point detection, addressing the challenges posed by pigs' movement and changing facial postures. Marsot et al.[14] employed gradient-weighted class activation mapping (Grad-CAM) for pig face recognition, demonstrating that neural networks primarily classify pigs by extracting facial features. Similar advancements have been made in cow face recognition. Cai et al.[15] used a cascade detector to capture frontal cow face images, while Weng et al.[16] utilized a convolutional neural network to process images from different angles, enhancing robustness and generalization capabilities.

Summarizing the existing methods, animal identification is often approached as an image classification or target detection problem, and implemented using deep learning techniques. However, for intensively reared animals like sheep, models based on classification or detection alone struggle with fine-grained recognition and are prone to misidentification[17,18]. Therefore, a recognition strategy that emphasizes accurate segmentation of facial images was proposed in this study for sheep face identification and introduces an improved parallel hybrid attention mechanism, the

Received date: 2025-01-09 **Accepted date:** 2025-05-07
Biographies: Suhui Liu, PhD, research interest: agricultural engineering and intelligent information processing, Email: suhui0518@126.com; **Guangpu Wang**, MS, research interest: agricultural engineering and intelligent information processing, Email: 1497180397@qq.com; **Zhaohui Tang**, PhD, research interest: precision agriculture, Email: 15204718180@139.com; **Junze Jia**, MS, research interest: agricultural engineering and intelligent information processing, Email: 2687852103@qq.com.
***Corresponding author: Chuanzhong Xuan**, PhD, Professor, research interest: key technologies and equipment for digital agriculture and animal husbandry. College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China. Tel: +86-471-4301062, Email: xcz@imau.edu.cn.

Improved Convolutional Block Attention Module (I-CBAM) . By removing the non-sheep face background region and enhancing feature information, this approach significantly boosts recognition accuracy. Considering the three-dimensional nature of sheep faces, which contrasts with the flatness of human faces, this study constructs a multi-view sheep face dataset. It acquires and fuses sheep face features from the left, front, and right angles using fully convolutional neural networks (FCN), thereby improving recognition accuracy.

2 Materials and methods

2.1 Multi-view sheep face image acquisition device

The multi-view sheep face image acquisition device, shown in Figure 1, is mainly composed of five parts: control board, conveyor belt system, camera system, host computer system, and channel. Sheep enter the conveyor belt module from the channel, and after reaching the camera shooting area, the control box turns off the conveyor belt's power, and the conveyor belt stops running. The cameras, placed on the left, right, and directly in front of the three angles, simultaneously shoot the sheep's face images and transmit them to the upper computer system in real time. The conveyor belt is inclined at an angle of 75°, with a narrow lower and wide upper design to prevent the sheep's legs from falling off during the clamping process. The cameras are Hikvision's standard 3-megapixel B13HV3-IA, with a focal length of 4 mm, a monitoring distance of 6 m, and a resolution of 2304 pixels×1296 pixels, with a product size of 87.1 mm×83.7 mm×171.7 mm. The cameras are manufactured by Hikvision, with the primary production facility located in Hangzhou, China. When a sheep enters the imaging zone, the conveyor stops and the three cameras (left, frontal, right) capture images simultaneously. After imaging, the conveyor restarts to release the animal and the next sheep enters.[19].

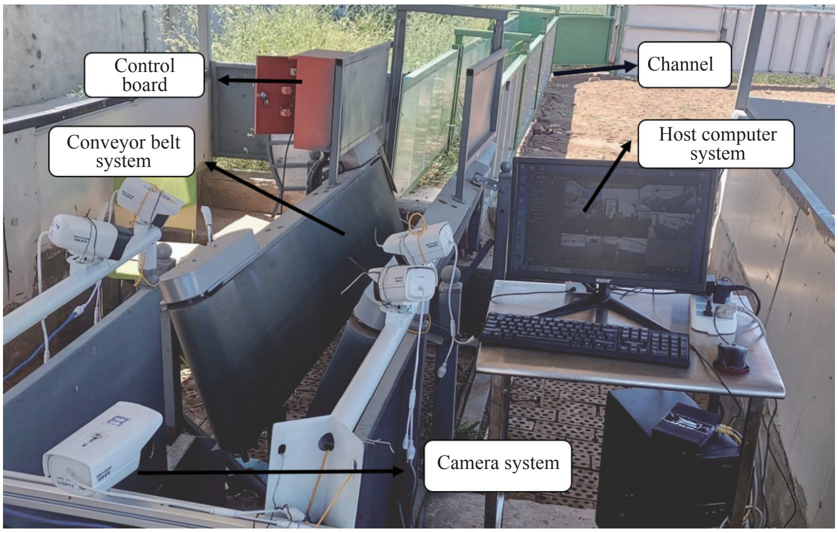


Figure 1 Multi-view sheep face image acquisition device

2.2 Construction of the sheep face dataset

In this study, the small-tailed cold sheep was selected as the subject of research, and the image acquisition site was located at the Heilinger Sheep Farm in Hohhot City. The total number of test sheep was 120, all of which were mature sheep of 1-2 years of age, and the test time was from August 5 to 16, 2022, using the multi-view sheep face image acquisition device. Acquisition of the sheep's left side face, front face, and right side face is obtained from three viewpoint images in order to obtain as much contour information of the sheep's face as possible. The images are stored in JPG format. At the same time, in the channel outside the shooting, a total of 12 430 images were collected for the model validation and testing, with screening out the images with an absence of sheep face, as the sheep face information is incomplete. And 12 000 images from other data were selected for the training of the model, validation and testing. According to the principle of 6:2:2 randomized division of datasets, which is standard in machine learning, the data are divided into 7200 for the training set, 2400 for the validation set, and 2400 for the testing set[20].

Labelme software was used for data annotation, as shown in Figure 2, to accurately punctuate and mark the contours of the sheep face. Each sheep is recorded as a class, and the labeling is divided into 120 classes, which are sheep1, sheep2, sheep3, ..., and sheep120, and the images in the three viewpoints captured are labeled as the left face, the front face, and the right face, respectively.

a. Left side of the face b. Front face c. Right side of the face

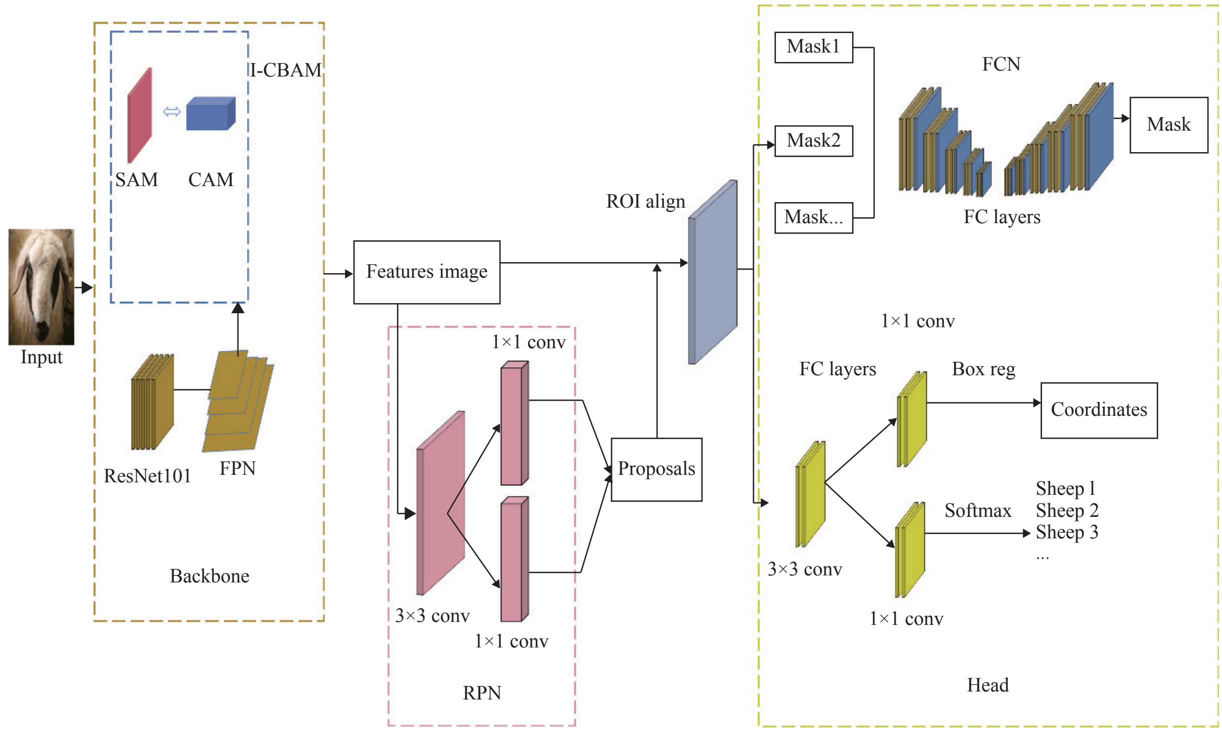
Figure 2 Data labeling during segmentation

Meanwhile, in the process of model training, to reduce model overfitting and improve the robustness and generalization ability of the model, the study used data enhancement to expand the dataset[21]. This is done by randomly reversing the sheep face image, adjusting the noise, darkening by 20%, and brightening by 20%[22].

2.3 Recognition methods for accurate segmentation of sheep face

2.3.1 SFMask R-CNN model structure

This study uses the Mask Region-based Convolutional Neural Networks (Mask R-CNN) model as the baseline, while the latest YOLOv8 was used as the proposed improved plug-and-play experimental analysis. This study added the I-CBAM attention mechanism module to this model. The network parameters are optimized to construct the SFMask R-CNN model for sheep face recognition[23]. The model consists of four parts: Backbone, Region Proposal Network (RPN), Region of Interest (RoI Align), and Head branch, as shown in Figure 3. The backbone network can be combined with various types, including ResNet50, ResNet101, ResNet50+FPN, and ResNet101+FPN. The RPN network generates the proposed regions of interest. The Head branch processes these regions through a series of 4096 1×1 convolution kernels to produce a two-dimensional feature map. This feature map is then subjected to multiple rounds of up-sampling via transposed convolution, also known as inverse convolution, to enhance the resolution of the feature map and refine the final image.[24,25]. The final deconvolution image is classified, and the bounding box is regressed to accurately segment the sheep face image and classify the sheep face targets in the proposed region.



Note: SAM is the spatial attention module; CAM is the channel attention module; I-CBAM is improved convolutional block attention module; ResNet101 is residual network with 101 layers; FPN is feature pyramid network; conv is convolutional layer; RPN is region proposal network; ROI Align is region of interest align; FCN is fully convolutional network; FC layers is fully connected layers.

Figure 3 Model structure of SFMask R-CNN

2.3.2 Hybrid attention mechanism module

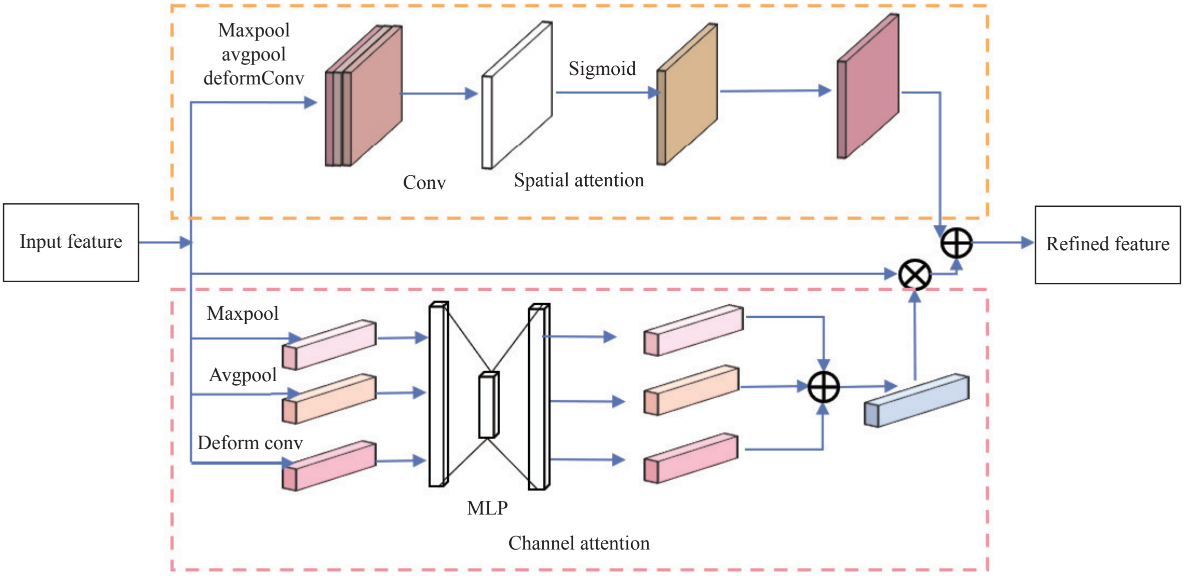
The captured sheep face images suffer from occlusion and background complexity, which leads to high background complexity of the image as well as low contrast between the background and the sheep face target. The segmentation of the sheep face relies on the contour features to differentiate the sheep face, which makes it difficult for the network feature extraction module to capture the complete information of the sheep face contour accurately[26,27]. Therefore, the parallel improved hybrid attention mechanism module, I-CBAM, was proposed with the structure shown in Figure 4. The traditional CBAM module follows the tandem form of channel attention and spatial attention, and the ordered nature of the tandem execution makes the two attention mechanisms possess a dependency on each other[28,29]. Therefore, to fully exploit the performance of channel or spatial attention on the sheep face recognition task scenarios, this study proposed a parallel structure of CBAM and named it I-CBAM. By embedding the I-CBAM module into the features generated by the FPN, the enhanced features are then further sent to the RPN to generate the candidate regions, contributing to a significant improvement in the model performance.

Specifically, in the I-CBAM module, the spatial attention mechanism may enhance the ability to find valid information on the sheep face feature map. Each channel of the feature map in the channel attention mechanism is treated as a feature detector, which enhances the ability to focus on the information in the sheep face image. It can infer the attention weights sequentially along the two dimensions of space and channel for the feature map of the sheep face and then multiply it with the original feature map to adaptively adjust the features, and the I-CBAM adaptively pays more attention to the information in the region of interest to enhance the ability of the model to extract effective features[30].

2.3.3 Experimental environment

The sheep face identification model was performed under the deep learning framework of Python 3.8.13 and PyTorch 1.9.0+ cuda11.1. The computer configuration used was a Windows 10 64-

bit (DirectX12) version-based operating system, a 6-core, 12-thread Intel i7-11600H mobile processor, 24 GB of RAM, an NVIDIA GeForce GPU (Santa Clara, CA, USA), and NVIDIA RTX A5000 graphics card.



Note: Conv is Convolutional layer; MLP is Multilayer Perceptron.

Figure 4 Structure of the I-CBAM model

2.3.4 Evaluation Metrics

The purpose of the evaluation is to test the ability of the algorithm to obtain information about the sheep face, utilizing key metrics such as Accuracy, Precision, Recall, F1-score, Average Precision (AP), etc., as demonstrated in Equations (1)-(5). The higher the metrics, the better the model performance.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{N}+\text{FP}+\text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{3}$$

$$\text{F1-score} = \frac{2(\text{Precision}\times\text{Recall})}{\text{Precision}+\text{Recall}} \tag{4}$$

$$\text{AP} = \frac{1}{\sum_{i=1}^{n} P_i}\sum_{i=1}^{n} P_i\times(R_i - R_{i-1}) \tag{5}$$

where, P_i is the precision at the i recall level; R_i is the i recall level; R_{i-1} is at the i−1 recall level; n is the number of different recall levels.

3 Results

3.1 Feature extraction network ablation test

The feature extraction network of the backbone network of the SFMask R-CNN model was used for sheep face identification using ResNet50, ResNet101, ResNet50+FPN, and ResNet101+FPN, respectively[31]. The results obtained from the experiments are listed in Table 1, which shows that the recognition performance of the ResNet network combined with the FPN network is better[32]. Because the FPN can effectively enhance the accuracy of the network for single target detection of sheep faces, the model performs better after combining the FPN network. In contrast, the ResNet101+FPN network improves the F1, P, R, and ACC values by 1.45%, 1.88%, 1.01%, and 1.77%, respectively, compared to the ResNet50+FPN network. However, due to the deeper network of ResNet101, using the image pyramid to construct the feature pyramid, it is slower to compute the image features independently on each image scale, which increases the detection time in this dataset by 2.6 ms[33]. Therefore, ResNet101+FPN is used as the feature extraction network for SFMask R-CNN to improve the accuracy of sheep face identity recognition.

Table 1 Detection effect of different backbone networks on test set

Model	F1/%	P/%	R/%	ACC/%	Time/ms
ResNet50	83.62	83.01	84.24	85.89	109.5
ResNet101	87.07	86.89	87.25	87.79	111.3
ResNet50+FPN	89.11	88.01	90.24	90.24	115.3
ResNet101+FPN	90.56	89.89	91.25	92.01	117.9

3.2 Parametric analysis of the RPN module

In order to improve the segmentation performance of sheep faces, the model region of SFMask R-CNN suggests the network RPN module, designing a total of five kinds of anchor points: 32×32, 64×64, 128×128, 256×256, 512×512. By comparing the different sizes of the anchor points, the anchor point multiscale transformation is completed[34]. From this, the region most likely to have a sheep face target is selected, the region is detected and segmented, and the performance test results of the model are listed in Table 2.

Table 2 Recognition accuracy for different anchor sizes

Model	Anchor size	Anchor value	ACC/%	Time/ms
	Original model	(32, 64, 128, 256, 512)	82.80	110.9
ResNet101+FPN	Double the size	(64, 128, 256, 512, 1024)	71.50	106.8
	Double the size	(16, 32, 64, 128, 256)	90.00	115.6

Comparing the anchor sizes of different RPN modules, it can be found that expanding the anchor size by a factor of 1, i.e., anchor sizes of (64, 128, 256, 512, 1024), the feature extraction network has the lowest detection results. Reducing the anchor sizes of each layer by a factor of 1, i.e., anchor sizes of (16, 32, 64, 128, and

256), the ResNet101+FPN as the feature extraction network of the SFMask R-CNN achieves the highest *AP* value and can learn complex features better. In the case of choosing ResNet101+FPN as the feature extraction network, the *AP* value of reducing the anchor size by a factor of 1 is 6.2% higher than the original model, with an increase in time of 4.7 ms. It is 11.4% higher than increasing the anchor size by a factor of 1, with an increase in time of 8.8 ms. Considering all the factors, the RPN module chooses the anchors with a factor of 1 reduction in the size of the anchors[35].

3.3 SFMask R-CNN sheep face segmentation performance analysis

Figure 5 presents a comparative analysis of the results obtained from two distinct models in the context of sheep face segmentation. The original sheep face image is displayed on the left, the segmentation result of the Mask R-CNN model is shown in the middle, and the segmentation result of the SFMask R-CNN model is displayed on the right. The Mask R-CNN model produces a purple mask on the face of the sheep, indicating basic segmentation capability, although the edge details may not be sufficiently precise. In contrast, the SFMask R-CNN model generates a considerably more refined mask, and the model's hybrid attention mechanism enables the localization and capture of the local features of the face of the sheep, thereby facilitating the complete segmentation of the face of the sheep. The segmentation lines of the facial contour edges and neck of the sheep are smoother and closely fit the contours of the face of the sheep, including the complex curves of the ears and face. This results in a more accurate image of the face of the sheep being obtained. This demonstrates that SFMask R-CNN exhibits superior accuracy in detail processing, which is a crucial aspect for the model to attain high-precision recognition[36].

a. Original image b. Mask R-CNN c. SFMask R-CNN

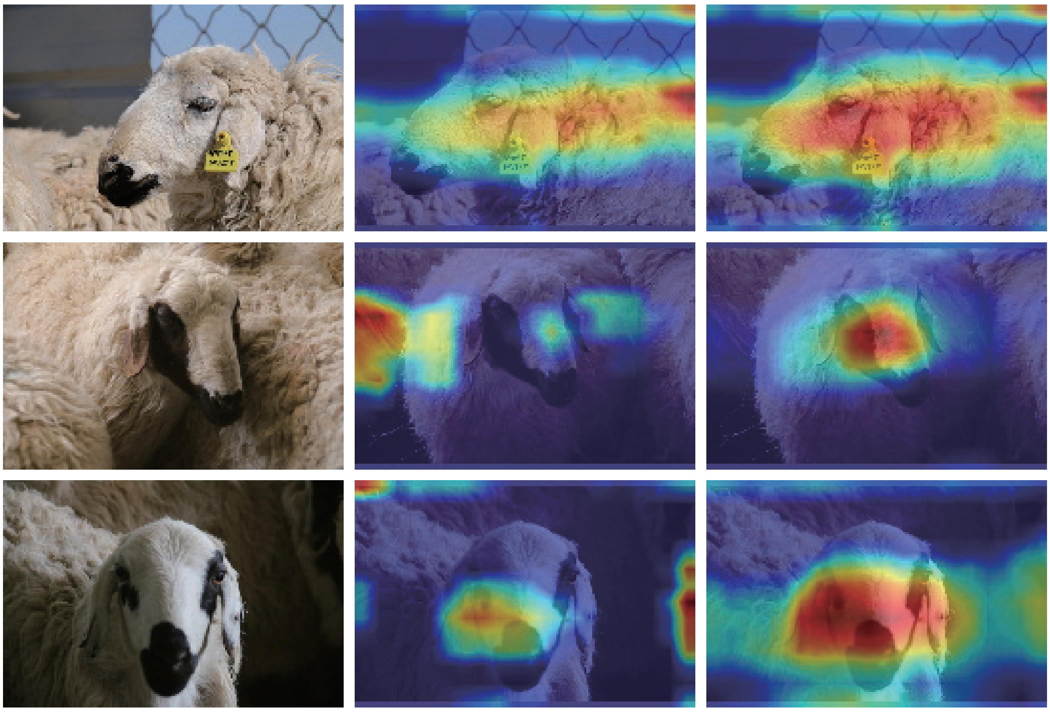
Figure 5 Segmentation of sheep face for different models

The SFMask R-CNN model is further tested to segment the sheep face images with blurred, low-light, and sheep face masked, respectively, and the results of sheep face segmentation are shown in Figure 6. Figure 6a is the sheep face segmentation map with a blurred face image due to sheep movement, Figure 6b is the sheep face segmentation map with low light, and Figure 6c is the sheep face segmentation map with part of the face masked. It is known from Figure 5 that the sheep face mask features can be extracted. However, it is easy to need clarification on the background interference region to extract irrelevant features, such as the SFMask R-CNN model segmentation. However, background clutter can cause models to extract irrelevant features. Even under blur, low light, or partial occlusion, SFMask R-CNN still extracts valid face-mask features and achieves accurate segmentation.

Figure 7 shows the feature heat map of different models for sheep face segmentation. The SFMask R-CNN model in Figure 7c has a darker color, higher color saturation, and more obvious critical

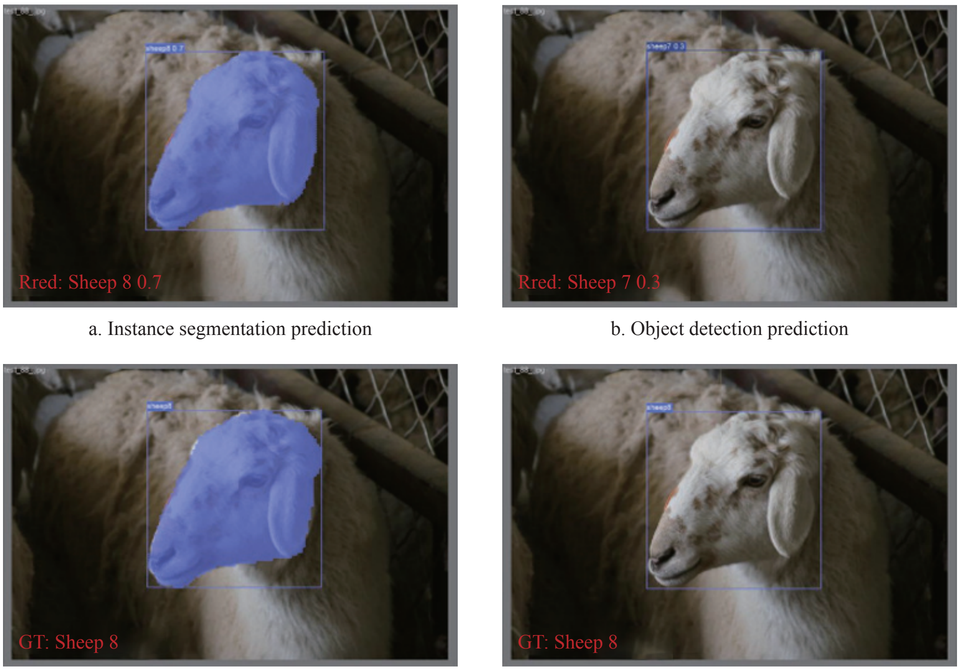
a. Original image b. Mask R-CNN c. SFMask R-CNN

Figure 6 Sheep face segmentation of different models in specific cases features compared to the Mask R-CNN model in Figure 7b, which is conducive to the accurate segmentation of sheep face[37]. The spatial attention of SFMask R-CNN involves localizing the sheep's face feature information. It locates the region of interest more accurately, and the channel attention of the model plays the role of capture, capturing the critical information to a greater extent and effectively suppressing the irrelevant background information so that the target features of the sheep face play a more dominant role. The spatial attention of SFMask R-CNN plays the role of localization of sheep face feature information, locating the region of interest more accurately. The channel attention of the model captures the critical information to a greater extent, effectively suppressing irrelevant background information to let the target features of the sheep face play a more critical role.



a. Original image b. Mask R-CNN c. SFMask R-CNN

Figure 7 Heat map of the sheep face with different models



a. Instance segmentation prediction b. Object detection prediction

c. Ground truth for instance segmentation d. Ground truth for object detection

Figure 8 Instance segmentation and object detection model recognition result display

3.4 Analysis of experimental results of different sheep face recognition models

Table 3 presents a comparative and analytical evaluation of the performance of sheep face recognition with the inclusion of various improved modules, with the objective of assessing their efficacy and efficiency in practical applications. Mask R-CNN is employed as the benchmark model and evaluated in accordance with key performance indicators.

Table 3 Performance comparison of sheep face recognition models

Model	ACC/%	*P*/%	*R*/%	F1/%	Time/ms
YOLOv5	90.57	89.32	88.21	89.76	101.7
YOLOv8	91.73	90.26	91.44	91.07	90.1
Mask R-CNN	91.01	91.89	90.69	91.79	109.4
Mask R-CNN+CBAM(S-C)	93.25	93.25	92.64	92.94	119.2
Mask R-CNN+CBAM(C-S)	94.21	94.25	93.23	93.74	118.9
SFMask R-CNN	97.56	97.04	97.22	97.64	117.9

In terms of accuracy, the SFMask R-CNN model achieves an accuracy of 97.56%, thereby demonstrating its superiority in correctly recognizing sheep faces. This performance is superior to that of the YOLOv5 model, which achieved an accuracy of 90.57%, and even exceeds that of the newer YOLOv8 model, which attained an accuracy of 93.56%[38]. The accuracy of the SFMask R-CNN model validates its robust feature extraction and segmentation capabilities. The precision and recall of the SFMask R-CNN model are 97.04% and 97.22%, respectively, which demonstrate the accuracy and completeness of the model predictions, which are crucial for reliable sheep identification in real-world applications. Notwithstanding the commendable performance exhibited by the SFMask R-CNN model, its processing time of 117.9 ms was found to be longer than that of the enhanced YOLOv8 model, which had a processing time of 90.1 ms. The elevated processing time can be attributed to the more intricate architectural design of the model and the supplementary computational demands of the I-CBAM module.

The performance metrics presented in Table 3 clearly demonstrate the superiority of the SFMask R-CNN model in the sheep face recognition task. Notwithstanding the augmented processing time, its elevated accuracy rate renders it a potent instrument for precision animal husbandry. Further work could concentrate on optimizing the model architecture with a view to reducing processing time while maintaining high accuracy rates. This would serve to reinforce the role of the model in advancing agricultural engineering and animal husbandry practices.

3.5 Comparative experiments of instance segmentation and object detection for sheep face recognition

Table 4 presents a comparative analysis of the base Mask R-CNN and YOLOv8 models, which have demonstrated robust performance in accurate segmentation and target detection tasks. The Mask R-CNN model exhibits notable proficiency in instance segmentation, and its enhanced variant, SFMask R-CNN, is capable of providing more granular pixel-level segmentation. The process of sheep identification is facilitated by the accurate extraction of regions of interest (RoI), which helps to overcome the high degree of similarity between the faces of different sheep. This approach has been shown to achieve an accuracy rate of 97.56%. This illustrates that precise segmentation facilitates a more comprehensive comprehension of the image content and interactions between objects. In contrast, the YOLOv8 model and its enhanced iteration demonstrate commendable efficacy in target detection, particularly the YOLO v8+I-CBAM variant, which attains an accuracy of 95.93%. The YOLO v8 model's key advantage lies in its rapid detection speed and high efficiency, rendering it more suitable for real-time feedback applications[39].

In conclusion, instance segmentation modeling is the most effective approach for target detection in sheep face recognition, particularly in terms of accuracy. However, it does impose a burden on the annotation process. Conversely, CBAM (S-C) demonstrates superior performance in terms of accuracy, suggesting that spatial information plays a pivotal role in modeling the sheep face recognition problem with target detection frame annotation. This finding lends further support to our assertion that accurate sheep face region features are of paramount importance for sheep identification. Therefore, for the purpose of precise identification and localization of sheep faces, the SFMask R-CNN recognition model, which is based on accurate segmentation, represents a more suitable choice for the task of high-precision sheep face recognition.

Table 4 Comparison of instance segmentation and object detection

Typology	Model	ACC/%	*P*/%	*R*/%	F1/%
Instance segmentation	Mask R-CNN	91.01	91.89	90.69	91.79
	Mask R-CNN +CBAM(S-C)	93.25	93.25	92.64	92.94
	Mask R-CNN +CBAM(C-S)	94.21	94.25	93.23	93.74
	Mask R-CNN +I-CBAM	97.56	97.04	97.22	97.64
Target detection	YOLO v8	91.73	90.26	91.44	91.07
	YOLO v8+CBAM(S-C)	90.51	93.46	92.82	93.50
	YOLO v8+CBAM(C-S)	92.88	93.55	93.53	93.10
	YOLO v8+I-CBAM	95.93	94.98	96.24	95.60

Figure 8 illustrates the recognition outcomes of the same sheep, contrasting the two methodologies of target detection and instance segmentation. In the instance segmentation image, the sheep face is accurately covered by a blue mask, and the model exhibits a high degree of confidence in recognizing the sheep's number[39]. Conversely, in the target detection image, although the model recognizes the sheep's face, its confidence level is comparatively lower. The results demonstrate that instance segmentation has a beneficial effect on the accuracy of recognition when mask region labeling is introduced.

3.6 Loss function

Figure 9 depicts five loss curves, each representing the training loss of a distinct model engaged in the sheep face recognition task. The overall trend demonstrates a gradual decrease in loss values across all models, indicating that they are all learning and improving their predictive capabilities. Nevertheless, the rate of loss decline and the final loss value attained vary between models[40]. The SFMask R-CNN model demonstrates superior learning and generalization abilities, converging to lower loss values more rapidly.

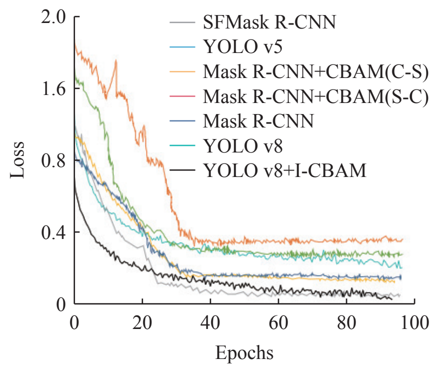


Figure 9 Loss profile for each model

3.7 Sheep face recognition with feature fusion in multiple perspectives

Due to the strong sense of three-dimensional features of sheep faces, it is very easy for sheep to be photographed with only the left, front, or right side of the face, while the left, front, and right side features are very different but all correspond to the same sheep, which makes the generalization ability of the network poor and easy to fall into a local optimum. The SFMask R-CNN model performs accurate instance segmentation of the sheep faces from different perspectives. The sheep faces from each perspective are segmented and labeled with different masks, which helps to extract and fuse the features from each perspective. Therefore, in order to improve the accuracy of sheep face identity recognition, the identity recognition method using fusion of sheep face feature data from different viewpoints is shown in Figure 10, where SFMask R-CNN sheep face segmentation is performed on the sheep face image data from different viewpoints, and the segmented features are subsequently fed into a fully convolutional neural network (FCN) for classification to identify the sheep[41,42].

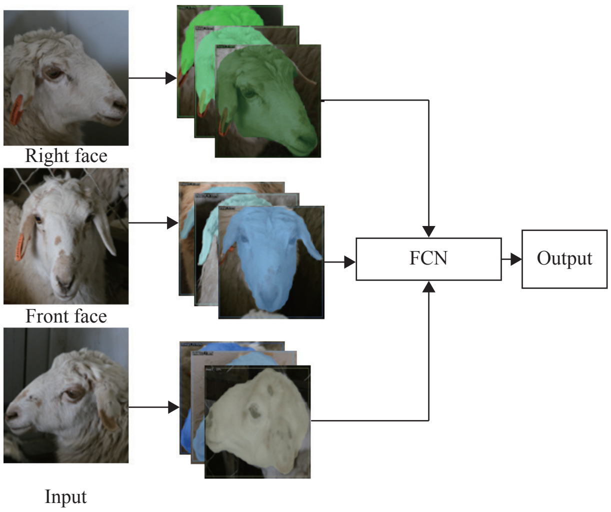


Figure 10 Sheep face recognition with multi-view feature fusion

As listed in Table 5, the sheep face identity recognition results under the SFMask R-CNN model for sheep face datasets with different views are listed, from which it can be seen that the model's F1 value for the left face, the front face, the right face, and the fused features are above 96.12%, the AP value is above 95.31%, and the accuracy rate is above 95.23%. By fusing the feature data under multiple viewpoints, the model generates more comprehensive feature information about the sheep's face, and the recognition accuracy is improved, which is 1.13% higher than the accuracy of the front face.

Table 5 Comparison of assessment indicators under multiple perspectives

Multi-perspective	F1/%	*AP*/%	*ACC*/%	Time/ms
Left face	96.12	95.49	96.25	133.9
Front face	97.56	96.89	97.79	131.6
Right face	96.51	95.31	95.23	133.4
Integration features	97.07	97.89	98.65	132.9

4 Discussion

The present study introduces the SFMask R-CNN model, a significant advancement in the field of livestock face recognition, specifically for sheep. This model addresses the limitations of traditional ear tag methods, which are not only stressful for the animals but also prone to failure due to tag loss. The approach of this study leverages precise facial segmentation for identity management in livestock farming, demonstrating significant improvements over existing models such as Mask R-CNN and YOLO v8 in terms of accuracy, precision, recall, and F1-score.

Advantages of the SFMask R-CNN Model: Our SFMask R-CNN model has demonstrated significant improvements over traditional methods and existing models in terms of accuracy, precision, recall, and F1-score. This enhancement is largely due to the model's ability to perform precise segmentation of the face of the sheep, which is a critical step in identity recognition. The integration of the I-CBAM module has been particularly effective in focusing on the critical features of the sheep face while suppressing background interference, thereby optimizing the extraction and utilization of feature information.

Innovative Aspects: The construction of a multi-view sheep face dataset and the application of FCN for multi-feature fusion represent innovative aspects of our study. This approach addresses limitations of traditional ear tag methods and provides a more comprehensive understanding of the three-dimensional structure of sheep faces. The ability to fuse features from different angles not only improves recognition accuracy but also offers insights into the variability of facial features across different views.

Challenges and Limitations: Despite the promising results, our model faces challenges in diverse environmental conditions and with different sheep breeds. The performance of the model in low-light conditions or with sheep of varying ages and breeds requires further investigation. Additionally, the computational complexity and processing time of the model should be considered, especially for real-time applications in large-scale farming operations.

Comparative Analysis with Existing Models: This study compared the SFMask R-CNN model with the Mask R-CNN and YOLO v8 models, revealing substantial improvements. This comparative analysis highlights the effectiveness of our model's segmentation capabilities and the I-CBAM module in enhancing feature extraction. However, it also prompts us to consider the trade-offs between accuracy and computational efficiency, as well as the potential for overfitting due to the complexity of the model.

Implications for Animal Welfare and Precision Farming: The non-invasive nature of our model aligns with the growing emphasis on animal welfare in modern agriculture. By eliminating the stress associated with ear tagging, our model contributes to a more humane approach to livestock management. Furthermore, the precision offered by our model supports the growing trend towards precision farming, where individual animal data can be used to optimize breeding practices and health management.

Future Research Directions: For future research, expanding the dataset to include a more diverse range of sheep breeds and ages is essential. Additionally, exploring the model's performance under varying environmental conditions, such as different lighting and weather conditions, will enhance its applicability in real-world scenarios. Investigating the potential integration of our model with other sensor technologies, such as thermal imaging or Radio Frequency Identification (RFID), could also provide a more comprehensive identification system.

5 Conclusions

In conclusion, the SFMask R-CNN model presents a significant advancement in sheep face recognition technology. Its ability to accurately segment and recognize sheep faces, coupled with its innovative multi-view feature fusion, positions it as a powerful tool for improving livestock management practices. While challenges remain, particularly regarding environmental robustness and computational efficiency, the potential benefits of our model for animal welfare and precision farming are substantial. Future work will focus on addressing these challenges and further enhancing the model's capabilities to meet the demands of modern agricultural practices.

Acknowledgements

This work was supported by the Fundamental Research Funds for Inner Mongolia Directly Affiliated Universities (Grant No. BR221032) and the First Class Disciplines Research Special Project (Grant No. YLXKZX-NND-009).

[References]

[1] Jie C, Cao X K, Wang S X, Zhang J Q, Yue B L, Zhang X Y, et al. 3D genome organization and its study in livestock breeding. Journal of Integrative Agriculture, 2024; 23(1): 39–58.

[2] Huang L, Qian B, Guan F, Hou Z, Zhang Q. Goat face recognition model based on wavelet transform and convolutional neural networks. Transactions of the CSAM, 2023; 54(5): 278–287.

[3] Dhanya V G, Subeesh A, Kushwaha N L, Vishwakarma D K, Nagesh Kumar T, Ritika G, et al. Deep learning based computer vision approaches for smart agricultural applications. Artificial Intelligence in Agriculture, 2022; 6: 211–229.

[4] Muhammad F A, Giovanni P B, Antonio V, Maria F S, Fahad A, Giovanni L, et al. The groundbreaking impact of digitalization and artificial intelligence in sheep farming. Research in Veterinary Science, 2024; 170: 105197.

[5] Shamshiri R R, Weltzien C, Hameed I A, Yule I J, Grift T E, Balasundram S K, et al. Research and development in agricultural robotics: A perspective of digital farming. Int J Agric & Biol Eng, 2018; 11(4): 1–14.

[6] Madkour K, Dawood Mahmoud A O, Sewilam H. The use of artemia for aquaculture industry: An updated overview. Annals of Animal Science, 2023; 23(1): 3–10.

[7] Dong S S, Sui B, Shen Y J, Meng H B, Zhao L X, Ding J T, et al. Investigation and analysis of the linkage mechanism and whole process cost of livestock manure organic fertilizer. Int J Agric & Biol Eng, 2020; 13(2): 223–227.

[8] Wang D W, Deng L M, Ni J G, Gao J Y, Zhu H F, Han Z Z. Recognition pest by image-based transfer learning. Journal of the Science of Food and Agriculture, 2019; 99(10): 4524–4531.

[9] Song C Y, Zhang F, Li J S, Xie J Y, Yang C, Zhou H, et al. Detection of maize tassels for UAV remote sensing image with an improved YOLOX Model. Journal of Integrative Agriculture, 2023; 22(6): 1671–1683.

[10] Salama A, Hassanien A E, Fahmy A. Sheep identification using a hybrid deep learning and Bayesian optimization approach. IEEE Access, 2019; 7: 31681–31687.

[11] Ning J F, Lin J Y, Yang S Q, Wang Y S, Lan X Y. Face recognition method of dairy goat based on improved YOLO v5s. Transactions of the CSAM, 2023; 54(4): 331–337.

[12] Zhang H M, Zhou L X, Li Y H, Hao J Y, Sun Y, Li S Q. Sheep face recognition method based on improved MobileFaceNet. Transactions of the CSAM, 2022; 53(5): 267–274.

[13] Huang Z J, Xu A J, Zhou S Y, Ye J H, Weng X X, Xiang Y. Key point detection method for pig face incorporating reparameterization and attention mechanism. Transactions of the CSAE, 2023; 39(12): 141–149.

[14] Mathieu M, Mei J Q, Shan X S, Ye L Y, Feng P, Yan X J, et al. An adaptive pig face recognition approach using convolutional neural networks. Computers and Electronics in Agriculture, 2020; 173: 105386.

[15] Cai C, Song X X, He J R. Algorithm and realization for cattle face contour extraction based on computer vision. Transactions of the CSAE, 2017; 33(11): 171–177.

[16] Weng Z, Meng F H, Liu S Q, Zhang Y, Zheng Z Q, Gong C L. Cattle face recognition based on a Two-Branch convolutional neural network. Computers and Electronics in Agriculture, 2022; 196: 106871.

[17] Cui Q N, Pan H W, Li X K, Zhang K J, Chen W P. CMSuG: Competitive mechanism-based superpixel generation method for image segmentation. Journal of Intelligent & Fuzzy Systems, 2022; 43(4): 4409–4430.

[18] Li K, Yu L Q, Heng P. Domain-incremental cardiac image segmentation with style-oriented replay and domain-sensitive feature whitening. IEEE Transactions on Medical Imaging, 2023; 42(3): 570–581.

[19] Gu J P, Wang Z H, Gao R H, Wu H R. Cow behavior recognition based on image analysis and activities. Int J Agric & Biol Eng, 2017; 10(3): 165–174.

[20] Zhu H, Yang H T, Guo L W, Zhang Y D, Wang Y R, Huang M K, et al. FaceScape: 3D facial dataset and benchmark for single-view 3D face reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023; 45(12): 14528–14545.

[21] Boone L, Biparva M, Mojiri F P, Ramirez J, Masellis M, Bartha R, et al. ROOD-MRI: Benchmarking the robustness of deep learning segmentation models to out-of-distribution and corrupted data in MRI. NeuroImage, 2023; 278: 120289.

[22] Zhou H R, Huang D, Wu B M. Deep recognition of rice disease images: how many training samples do we really need? Journal of the Science of Food and Agriculture, 2024; 104(13): 8070–8078.

[23] Zhu N Y, Liu X, Liu Z Q, Hu K, Wang Y K, Tan J L, et al. Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. Int J Agric & Biol Eng, 2018; 11(4): 21–28.

[24] Xiong H T, Li J, Wang T W, Zhang F, Wang Z Y. EResNet-SVM: An overfitting-relieved deep learning model for recognition of plant diseases and pests. Journal of the Science of Food and Agriculture, 2024; 104(10): 6018–6034.

[25] Liu W, Yan Z W, Wang J W, Ning Z M, Min Z. Ultrawideband real time monitoring system based on electro-optical under-sampling and data acquisition for near-field measurement. IEEE Transactions on Instrumentation and Measurement, 2020; 69(9): 6603–6612.

[26] Dong X H, Kim W Y, Zheng Y, Oh J Y, Ehsani R, Lee K H. Three-dimensional quantification of apple phenotypic traits based on deep learning instance segmentation. Computers and Electronics in Agriculture, 2023; 212: 108156.

[27] Chen G J, Yue X Y, Zhu Y N, Fu Y P. Study on Building Extraction Algorithm of Remote Sensing Image Based on Multi-scale Feature Fusion. Computer Science, 2023; 50(9): 202–209.

[28] Chen X S, Dang P N, Zhang E Z, Chen Y X, Tang C Y, Qi L. Accurate recognition of rice plants based on visual and tactile sensing. Journal of the Science of Food and Agriculture, 2024; 104(7): 4268–4277.

[29] Shi Y Q, Gao W J, Shen T T, Li W T, Li Z H, Huang X W. Calorie detection in dishes based on deep learning and 3D reconstruction. Computers and Electronics in Agriculture, 2025; 229: 109832.

[30] Hui F, Xie Z, Li H, Guo Y, Li B, Liu Y, et al. Image-based root phenotyping for field-grown crops: An example under maize/soybean intercropping. Journal of Integrative Agriculture, 2022; 21(6): 1606–1619.

[31] Mohammad R, Samira M, Hamidreza K. ResNet deep models and transfer learning technique for classification and quality detection of rice cultivars. Expert Systems With Applications, 2024; 247: 123276.

[32] Rehman Y, Hafsa A, Bhatti D M, Toor W T, Ahmad M, Mazzara M. Detection of small size traffic signs using regressive anchor box selection and DBL layer tweaking in YOLOv3. Applied Sciences, 2021; 11(23): 11555.

[33] Zhang Y, Ma Y L, Li Y L, Wen L F. A deep learning approach of RQD analysis for rock core images via Cascade Mask R-CNN-Based model. Rock Mechanics and Rock Engineering, 2024; 57(12): 11381–11398.

[34] Singh V. Sunflower leaf diseases detection using image segmentation based on particle swarm optimization. Artificial Intelligence in Agriculture, 2019; 3: 62–68.

[35] Li H, Tian Z N, Li X H, Arebi P. Identifying effective nodes in term of the influence maximization on the social IoT networks using deep learning-based node embedding. Computing, 2025; 107(1): 38.

[36] Ma B L, Hua Z X, Wen Y C, Deng H X, Zhao Y J, Pu L R, et al. Using an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments. Artificial Intelligence in Agriculture, 2024; 11: 70–82.

[37] Sapkota R, Ahmed D, Karkee M. Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. Artificial Intelligence in Agriculture, 2024; 13: 84–99.

[38] Zhang W, Cheng Z S, Chen J, Gao W. Hierarchical collaboration for referring image segmentation. Neurocomputing, 2025; 613: 128632.

[39] Li F Y, Peng H, Yu S H, Zeng F, Chen F, Yang F. Revisiting the loss functions in sequential recommendation. Engineering Applications of Artificial Intelligence, 2024; 138: 109366.

[40] Valentin S, Lancelot R, Roche M. Fusion of spatiotemporal and thematic features of textual data for animal disease surveillance. Information Processing in Agriculture, 2023; 10(3): 347–360.

[41] Ge L Z, Zou K L, Zhou H, Yu X W, Tan Y Z, Zhang C L, et al. Three dimensional apple tree organs classification and yield estimation algorithm based on multi-features fusion and support vector machine. Information Processing in Agriculture, 2022; 9(3): 431–442.

[42] Praneet S, Amy R R. Task-aware image quality estimators for face detection. EURASIP Journal on Image and Video Processing, 2024; 2024: 44.