Scientific Research Publishing

Open Journal of Applied Sciences, 2026, 16(4), 1386-1399
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

Research on Sheep Face Recognition Based on Deep Learning and YOLOv3

Xu Zhen[1,2,3], Guoqing Chen[1,2,3*], Lerong Wen[2], Haifeng Jia[4]

[1]Inner Mongolia Key Laboratory of Aeolian Physics and Desertification Engineering, Hohhot, China
[2]College of Desert Control Science and Engineering, Inner Mongolia Agricultural University, Hohhot, China
[3]Inner Mongolia Hangjin Desert Ecological Position Research Station, Ordos, China
[4]Inner Mongolia Caodu Grass and Livestock Ecological Technology Co., Ltd., Hohhot, China
Email: *chenguoqing@imau.edu.cn

How to cite this paper: Zhen, X., Chen, G.Q., Wen, L.R. and Jia, H.F. (2026) Research on Sheep Face Recognition Based on Deep Learning and YOLOv3. *Open Journal of Applied Sciences*, **16**, 1386-1399.
https://doi.org/10.4236/ojapps.2026.164080

Received: October 22, 2026
Accepted: April 27, 2026
Published: April 30, 2026

Abstract

To address the issues of low management efficiency, imprecise data tracking, and cumbersome rapid decision-making in traditional grassland animal husbandry, this study proposes a livestock monitoring model based on biometric technology. The study integrates sheep face recognition and in-depth data analysis, and combines them with deep learning algorithms to achieve non-contact, high-precision individual identification of livestock. A total of 400 sheep were tested in typical grassland areas of Inner Mongolia. The results indicated that the system achieved a comprehensive recognition accuracy of 94.1% and a recall rate of 81%, which effectively resolved the problems of traditional ear tags being prone to loss and damage, while also facilitating the monitoring of livestock diseases. This technology enables the full-lifecycle tracking of livestock, encompassing the intelligent monitoring of health status, the precise management of pastures, and the scientific prediction of breeding cycles. The research provides technical support for the digital transformation of grassland animal husbandry, and particularly offers practical assistance for the sustainable development of ecological animal husbandry in degraded grassland regions.

Keywords

YOLOv3, Deep Learning, Biometric Technology, Grassland Livestock Husbandry, Sustainable Development

1. Introduction

Dairy sheep and meat sheep in Inner Mongolia are vital components of China's agricultural sector [1], and smart animal husbandry serves as a key approach to

boosting livestock production efficiency [2]. Currently, the development of smart pastures requires the implementation of intelligent management, meat product traceability, and epidemic monitoring. However, traditional ear-tag identification suffers from issues such as easy detachment and high labor costs; additionally, individual sheep exhibit high similarity, leading to low accuracy in long-distance identification, which fails to meet practical demands [3] [4]. Meanwhile, livestock diseases severely compromise the quality and safety of livestock products [5] [6]. Against this backdrop, biometric detection and recognition solutions based on computer vision technology have become critical. These solutions not only can replace ear tags as essential pasture infrastructure but also record livestock growth data to ensure animal health, facilitate rapid monitoring of herd health status and female livestock reproduction, and provide support for modern livestock production [4]. Therefore, individual identification is the core focus of this study.

Driven by deep learning, biometric technology has achieved breakthroughs in multiple fields, evolving from traditional manual feature extraction to an efficient recognition system centered on convolutional neural networks (CNN) [7] [8] and transfer learning. It has demonstrated remarkable effectiveness in fields such as face recognition, iris recognition, and gait recognition [9]. Specifically, CNN enhance the adaptability of iris recognition to open scenarios [10], while multi-network fusion improves the performance of gait recognition under low-resolution conditions. In the agricultural sector, this technology offers a new pathway for grassland herd management: by combining YOLOv3 with DeepSORT tracking technology [11]-[13], real-time sheep detection and individual identity binding are realized, addressing the problems of traditional ear-tag identification (e.g., easy detachment, high costs, and poor real-time monitoring). Nevertheless, sheep face recognition still has room for optimization, including issues like adaptability to complex grassland environments, insufficient multi-modal feature fusion, and hardware compatibility. In terms of technology integration, multi-modal biometric recognition has emerged as a research hotspot—for instance, multi-modal fusion of hand features enhances recognition robustness [10], and deep learning-based analysis of biometric information improves recognition accuracy.

Looking ahead, deep learning will continue to drive the development of biometric technology [14], providing new ideas for various fields. With technological advancements and expanded application scenarios, biometric technology will achieve higher levels of intelligence and precision, contributing to the development of multiple industries. Modern grassland animal husbandry requires real-time monitoring of individual livestock to update their health status, thereby establishing a real-time IoT-based monitoring big data set. This big data set enables early warning of abnormal herd conditions, prediction of herd productivity, and data-supported health management. However, accurate individual identification of herds remains the most fundamental task to be accomplished [15]. Consequently, this study explores the individual identification of grassland herds based on image recognition technology.

2. Basic Principles

Deep learning is a key AI technology with outstanding performance in image classification, speech recognition and other fields but complex to achieve. Built on four advanced technologies, it is implemented in many fields using open-source databases and PaddlePaddle. Convolutional neural networks are important deep learning models, with core modules including convolution and pooling. The convolutional layer extracts image features via operations, serving as feature extraction core; the pooling layer divides images into non-overlapping blocks and samples elements. Additionally, the model relies on ReLU activation function, batch normalization and dropout to jointly ensure recognition performance.

2.1. Object Detection and Recognition

Using convolutional neural networks, two detection algorithms can be chosen for image object detection and classification: one stage detection with fast speed and low accuracy, and two stage detection with slow speed but high accuracy. This time, a two-stage Retina Net algorithm is used to recognize the target image. This algorithm adopts a fusion algorithm of ResNet, FPN, and FCN, and uses Focal loss as the loss function. Focal loss mainly solves the problem of target detection loss being affected by negative samples due to the imbalance of positive and negative sample areas during the object detection process. Firstly, classify the image (**Figure 1(a)**) to identify it as a picture of a sheep. Then, perform object detection (**Figure 1(b)**) to recognize it as a photo of a sheep and mark the position of each sheep in the image.

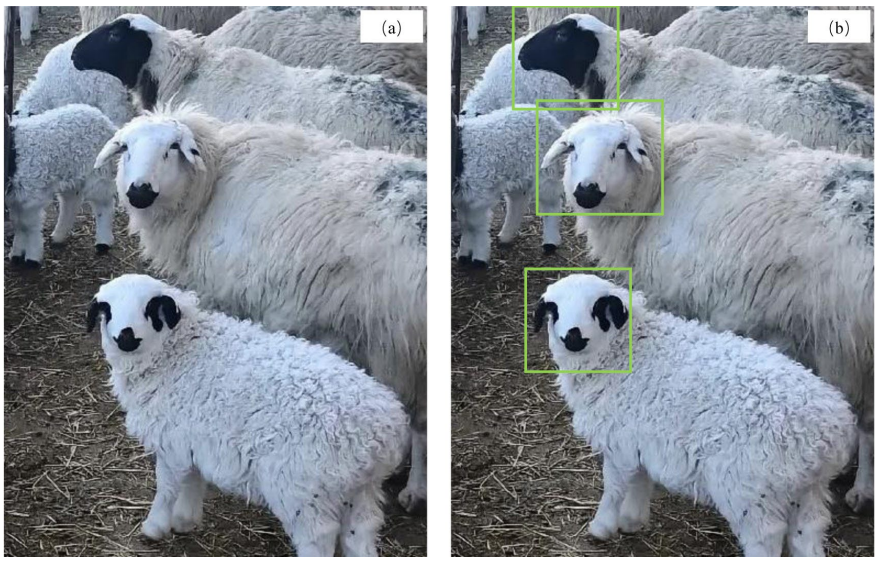

Figure 1. Image classification (a) and object detection (b).

2.2. Detection Object Target Box

In convolutional neural networks, the preferred region needs to be determined, and we use exhaustive search to generate candidate regions. The A pixel on the

image and the B pixel in the lower right corner of A can determine a rectangular box, denoted as AB. A is located in the upper left corner of the image, and B traverses all positions except A to generate rectangular boxes A1 B1, ..., A1 Bn (**Figure 2(a)**). After obtaining a certain position of A in the middle of the image, B traverses all positions in the lower right corner of A and generates rectangular boxes Ak B1, ..., Ak Bn (**Figure 2(b)**). When A traverses all the pixels on the image, B traverses all the pixels in the lower right corner, and finally generates a set of rectangular boxes {AiBj}, which includes all the cocoa selection areas on the image.

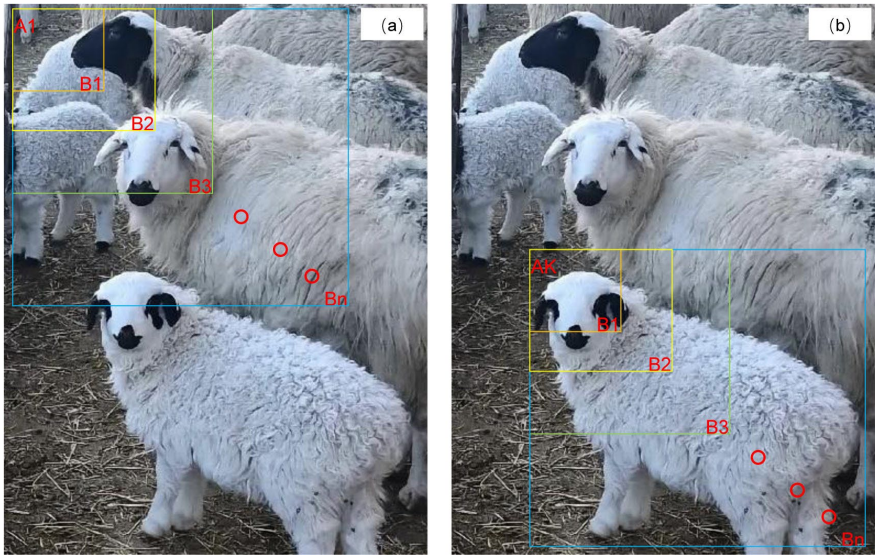

Figure 2. Candidate area confirmation.

As long as the classification of each candidate region is accurate enough, it is certain that a region that is close enough to the actual object can be found. The exhaustive search method may yield accurate prediction results, but its computational complexity is also enormous. Assuming H = W = 100, the total number will reach 2.5 × 107 points. Assuming the classification is fine enough, exhaustive search can theoretically complete the detection task, but it requires designing bounding boxes, anchor boxes, and intersection to union ratios for object detection to accurately provide candidate regions.

$$\frac{\mathrm{W}^2 H^2}{4} \tag{1}$$

Boundary box: Boundary boxes are often used to represent the specific position of objects and can be inserted into rectangular boxes of objects. It can be inferred that boundary boxes are a detection task that simultaneously predicts and detects the position and category of objects. There are usually two forms of bounding boxes: xyxy (x_1, y_1, x_2, y_2) and xywh (x, y, w, h).

In detection tasks, bounding boxes are called real boxes because the labels on the training dataset provide the coordinates of the real bounding box of the target

object, which are (x_1, y_1, x_2, y_2). The model accurately predicts the possible positions of the target object, and the bounding boxes predicted by the model are called prediction boxes [16].

Anchor box: Anchor box is a type of target box imagined by people. Draw an anchor box of a specific size, find a center point, and draw a rectangular box.

In the detection task, an anchor box is formed on the image, and it is determined whether the target object is included before proceeding to the next step of object detection. Due to the mismatch between the object and the anchor box, it is necessary to make fine adjustments to the original anchor box to form an anchor box that can accurately obtain the position of the object. The model needs to predict the magnitude of the fine adjustment, and different models have different methods for generating anchor boxes and different micro adjustment amplitudes, because the anchor box position is fixed at the beginning.

Intersection to union ratio: Intersection to union ratio is a measurement indicator in object detection. In object detection tasks, the intersection to union ratio is used as a measurement indicator in object detection, and the specific calculation formula is as follows:

$$I_O U = \frac{A \cap B}{A \cup B} \tag{2}$$

The green area in the "Intersection" section of the figure represents the overlapping area of two boxes, while the green area in the "Merge" section represents the merged area of two boxes. Dividing these two areas yields the intersection to union ratio between them, which is also known as the intersection to union ratio (**Figure 3**).

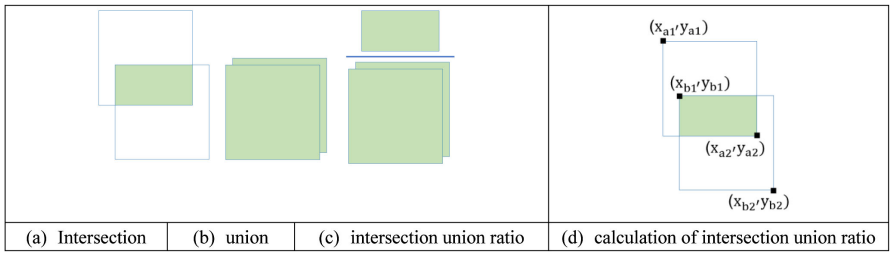

Figure 3. Comparison of Intersection and Intersection Ratio Operations.

If the positions of these two rectangular boxes A and B are A (x_{a1}, y_{a1}, x_{a2}, y_{a2}) and B (x_{b1}, y_{b1}, x_{b2}, y_{b2}), respectively, as shown in **Figure 3(d)**: If there is an intersection between the two, the coordinates of the upper left corner of the intersection are:

$$x_1 = \max\left(x_{a1}, x_{b1}\right), y_2 = \max\left(y_{a1}, y_{b1}\right) \tag{3}$$

The coordinates of the lower right corner of the intersection are:

$$x_2 = \min\left(x_{a2}, x_{b2}\right), y_2 = \min\left(y_{a2}, y_{b2}\right) \tag{4}$$

Calculate the area of the first part to be delivered:

$$\text{interection} = \max\left(x_2 - x_1 + 1.0, 0\right) \cdot \max\left(y_2 - y_1 + 1.0, 0\right) \tag{5}$$

The area of rectangular boxes A and B is:

$$\begin{aligned} s_A &= \left(x_{a2} - x_{a1} + 1.0\right) \cdot \left(y_{a1} - y_{a2} + 1.0\right) \\ s_B &= \left(x_{b2} - x_{b1} + 1.0\right) \cdot \left(y_{b1} - y_{b2} + 1.0\right) \end{aligned} \tag{6}$$

Calculate the combined area:

$$\text{union} = s_A + s_B - \text{intersection} \tag{7}$$

Calculate the intersection to union ratio:

$$I_O U = \frac{\text{intersection}}{\text{union}} \tag{8}$$

In order to clearly demonstrate the relationship between the size of the intersection ratio and the degree of overlap, Figure 4 shows the corresponding positional relationship between two target boxes under different intersection ratios, ranging from $I_o U = 0.95$ to $I_o U = 0.00$.

R-CNN generates candidate regions through selective search, segments the image, and merges similar regions. The candidate boxes are uniformly adjusted to 227 × 227 and input into a CNN (such as AlexNet) to extract 4096 dimensional features. SVM is used for classification and non-maximum suppression is used to remove overlapping detection boxes, but the process is complex and time-consuming. YOLO divides the image into N × N grids, and directly predicts the bounding box position and category probability for each grid, achieving end-to-end detection and greatly improving speed. SSD is improved based on VGG16 by adding multi-scale convolutional layers, using default boxes for detection on different feature layers, and combining multiple aspect ratio presets to achieve adaptability to targets of different sizes. The three represent the technological evolution paths from regional proposal, single-stage detection to multi-scale prediction.

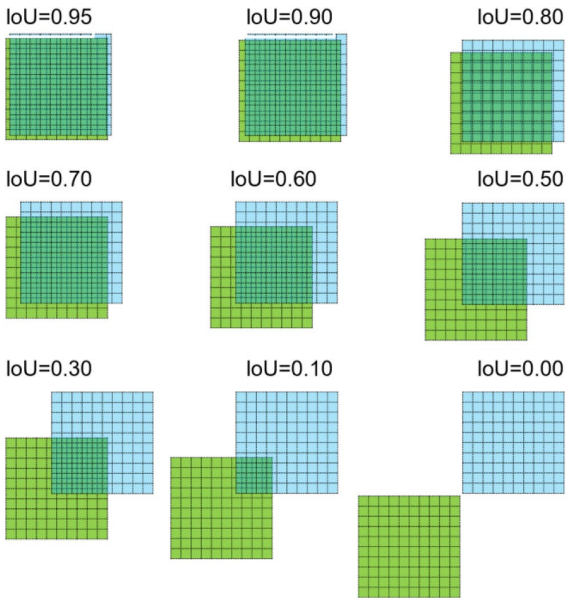

Figure 4. Relative position diagram between two boxes under different intersection and merger ratios.

2.3. Target Detection Evaluation

In order to further evaluate the performance of the sheep detection model and select evaluation indicators for the experimental results, including recall rate R, accuracy rate P, mean average precision mAP, and average precision AP. The recall rate R is an evaluation criterion that reflects the incompleteness of the target detected by a model; Accuracy P is an evaluation criterion that reflects the accuracy of a model's predictions; AP is the area under the Precision call curve, and generally speaking, the better the classifier, the higher the AP value. MAP is the average of multiple categories of AP. Mean is the average of the AP for each class, resulting in mAP. The size of mAP must be in the range of [0, 1], the larger the better. The calculation formulas for P, R, AP, and mAP are shown in equations (9), (10), (11), and (12), respectively.

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$AP = \int_0^1 P(R)\, d(R) \tag{11}$$

$$\text{mAP} = \frac{1}{m} \sum_{i=1}^{m} AP_i \tag{12}$$

2.4. Object Detection YOLOv3 Algorithm

Data preprocessing is necessary before the YOLOv3 algorithm, and it is a crucial and primary step in training convolutional neural network structures. Firstly, selecting appropriate data preprocessing methods can prevent overfitting, and then implementing data reading and preprocessing can help accelerate processing. The biometric technology based on YOLOv3 first normalizes the input biometric image with a size of 416 × 416, and then extracts multi-scale features through the Darknet-53 backbone network, generating feature maps at three scales: 13 × 13, 26 × 26, and 52 × 52. Using Feature Pyramid Network (FPN) to fuse deep and shallow features, three prior boxes are pre-set at each grid point to predict boundary coordinates and category probabilities. Finally, the non maximum suppression (NMS) algorithm is used to optimize the detection results, eliminate redundant boxes, and retain the optimal prediction, forming an end-to-end recognition pipeline from feature extraction to target localization, providing reliable technical support for individual identity recognition.

Data retrieval is the process of storing all descriptive information of an image in records, where each element contains a description of the image. The subsequent program demonstrates how to retrieve the image and annotate it based on the description in the records. Data preprocessing is the random processing of images, but with minimal changes. The main purpose is to suppress overfitting and improve the generalization ability of the model by increasing the training dataset. Common methods include randomly changing brightness, contrast, and

color, random filling, random cropping, random scaling, random flipping, random scrambling of the real box arrangement order, and using numpy to implement these data augmentation methods.

This article standardizes the animal and plant dataset based on the Pascal VOC 2007 set format and uses the LableImg tool for animal and plant image annotation. Use the VOC_ label. py script to navigate this path to your own dataset, with the class selected as "face". Finally, running the script will generate files such as train.txt, val.txt, test.txt, etc., and generate a labels file in devkit. Change the model parameters of YOLOv3 and adjust the batch. When batch equals 64, it's enough, then adjust other parameters to end. The parameter settings are shown in Table 1. If the average loss of the result is less than 0.0607, the operation will be stopped at this time.

Table 1. Parameter settings for YOLOv3.

Parameter	Value	Parameter	Value
Batch	64	Exposure	1.5
Subdivisions	2	Hue	0.1
Width	416	learning_rate	0.001
Height	416	Burn_in	400
Channels	3	Max_batches	5200
Momentum	0.9	Policy	Steps
Decay	0.0005	Steps	3800
Angle	0	Scales	0.1
Saturation	1.5		

Adopting the Triplet loss proposed in FaceNet as the loss function. The Triplet Loss function consists of three images: the reference image (anchor), the positive example image (positive), and the negative example image (negative). Through continuous training, there is a small gap between the positive example image and the negative example image, but a large gap between them. The selected VGG-Face16 and resNet50 represent two types of convolutional network structures with different network layers (16 layers and 50 layers), respectively, combined with two embeddings (2096 and 128) to construct a sheep face recognition model.

3. Result Analysis

3.1. Dataset Preprocessing Results

3.1.1. Individual Labeling of Livestock Herds

The dataset of this article mainly comes from photos of livestock herds in Yangchang Village, Zhaojun Town, Dalate Banner, Ordos. LabelImg software was used to label the above data as sheep, including pictures of sheep in various scenarios, totaling 400 images. Divide the annotated dataset into training, testing,

and validation sets in an 8:1:1 ratio, resulting in 320 training sets, 40 testing sets, and 40 validation sets, respectively. Finally, the dataset was organized to obtain its label situation as shown in **Figure 5**: (a) The y-axis of the figure represents the number of labels, and the x-axis represents the type of labels. (b) In the figure, the horizontal axis represents the ratio of label width to image width, and the vertical axis represents the ratio of label height to image height.

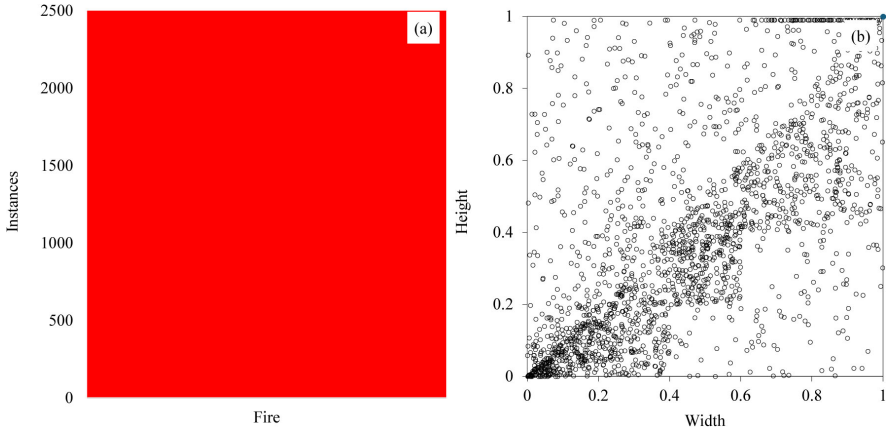

Figure 5. Distribution of sheep face labeling labels (a) and aspect ratio (b).

3.1.2. Individual Tracking of Livestock Herds

The Deepsort object tracking algorithm is based on detectors, such as the YOLO object detector. Firstly, the object detector detects sheep in the image and passes the detection coordinates of the sheep to the Deepsort algorithm. The Deepsort algorithm predicts the target position of the sheep at the next moment based on the Kalman filter, and finally completes the target tracking based on the Hungarian algorithm (**Figure 6**), and assigns a unique ID number to each target.

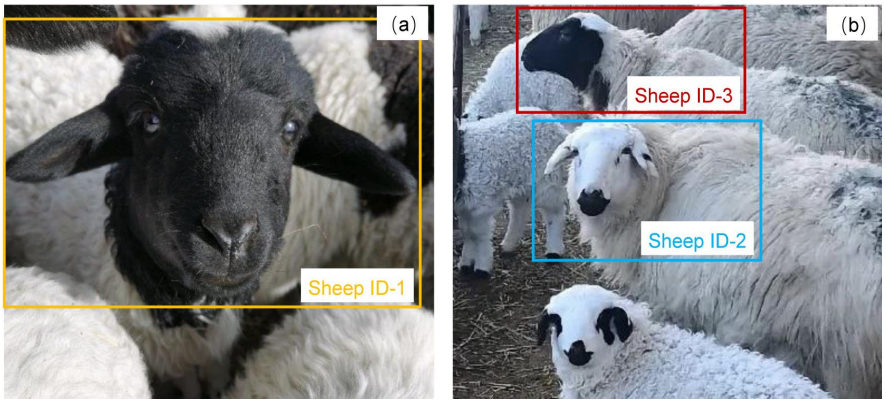

Figure 6. Individual ID tracking of livestock herd.

3.2. Sheep Face Recognition Evaluation

3.2.1. Target Detection Model Training

The experiment used a self-built sheep dataset image, with sheep as the sample. The YOLOv3 network was trained using end-to-end stochastic gradient descent

method, and the input image size was 640 × 640. The parameter values are set as follows: batch size is 64, momentum value is 0.937, decay value is 0.0005, total rounds are 100, and initial learning rate is 0.01. When one round of training is completed, the model training situation is validated using the validation set. The loss curves of the training set and validation set (**Figure 7**) show that from the 0th to the 50th batch on the training set, the loss of the training set decreases during training, and from the 50th to the 100th batch, the loss of the training set tends to stabilize, and the model gradually converges. Around the 25th batch on the validation set, the model gradually converges, and around the 50th batch, the Object loss shows an upward trend, indicating that the model may be overfitting.

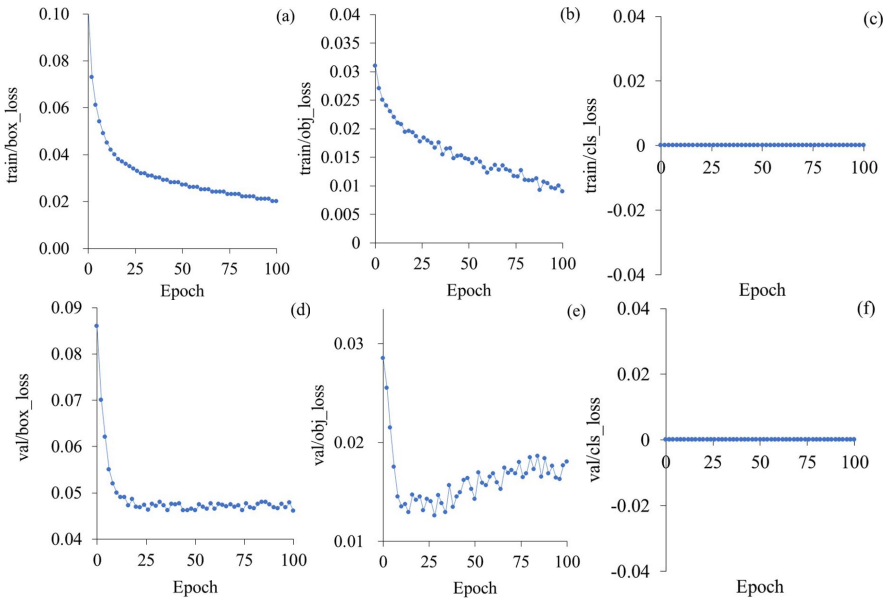

Figure 7. Loss curves of training and validation sets.

The performance curve of the validation set shows that mAP is the area enclosed by the precision Pr and recall Re plotted as two axes, and mAP and mAP are the average detection accuracy under general threshold and high threshold, respectively (**Figure 8**). When the values corresponding to the curve gradually stabilize, the optimal training model can be determined. In Figure 8, the top and bottom are the loss curves of the training set and validation set, respectively. It can be seen that the three types of loss functions have a significant decrease in the model training from 0 to 50 times, and remain stable at 50 to 100 times, indicating that training improves the detection performance of the model.

The mAP curve fluctuates significantly from 0 to 50 times at the beginning of training, indicating that the convergence speed of the model in the early stage of training is fast and meets the requirements of model training. After 50 iterations, the model remained stable with minimal changes, indicating good training and no overfitting. The curve around 90 to 100 tends to stabilize, indicating that the training of the sheep detection model is basically completed at this time (**Figure**

8). The selection of the optimal model is calculated based on mAP-U accounting for 10% and mAP-H accounting for 90%. The accuracy and recall were 94.1% and 81%, respectively, and the mAP reached 63.3%.

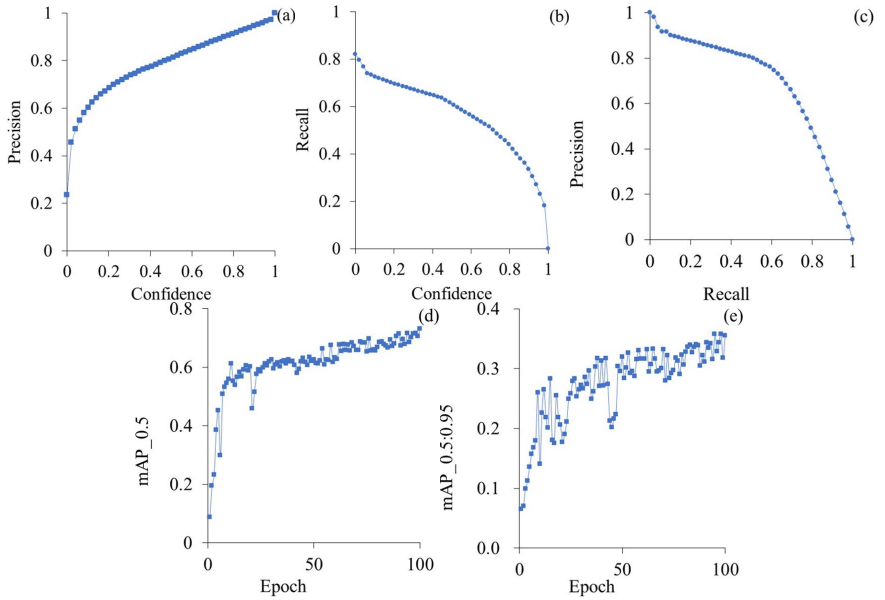

Figure 8. Verification set performance curve.

3.2.2. Accuracy Evaluation of Sheep Face Recognition

After training the sheep face detection model with YOLO v3 object detection, computer vision technology is used to read the video stream, perform sheep face recognition and detection, and save the sheep face information in the detection box. After a series of tests, the accuracy is 94.1% and the recall rate is 81%. Real time detection of sheep face data has been achieved. The training data for sheep face detection includes sheep with different postures (Figure 9), different lighting conditions, and camera shake, indicating that the model can accurately detect sheep under different movements and angles, meeting the needs of sheep face detection.

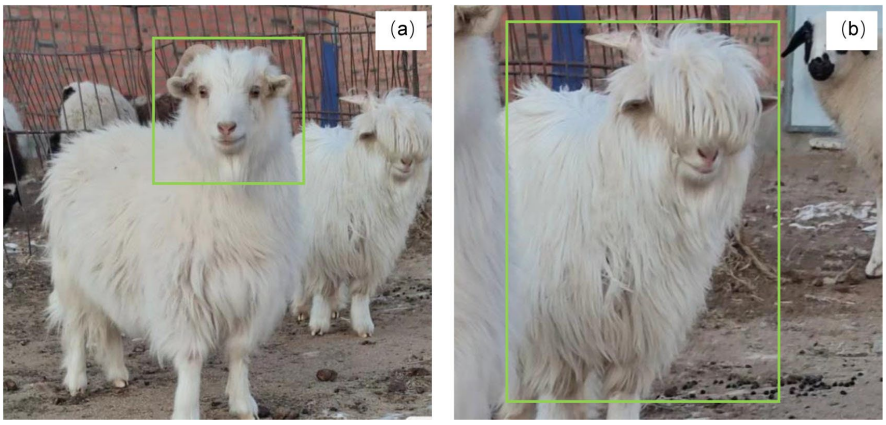

Figure 9. Sheep face detection results ((a) front view; (b) side view).

Divide the sheep image dataset of different scenes into training set, testing set, and validation set in an 8:1:1 ratio, and conduct model training based on this division. When the performance test results of the model using VGGFace neural network and embedding dimension 2096 are obtained, the performance of this model is relatively high. Using a specific model to recognize sheep faces in different situations within the same sheepfold. The use of transfer method to train models with a small number of samples can achieve high accuracy and recall, and can better complete the task of recognizing frontal sheep faces in a small dataset.

4. Discussion

In this study on sheep face recognition models, nose stripe recognition and comparative analysis between the two should be included in order to better and more accurately recognize sheep, improve accuracy, and achieve the ideal state. In addition to methods such as data collection, data cleaning, model construction, model training, and deployment, methods to improve recognition accuracy should also consider lighting conditions to make recognition more accurate. Generally speaking, conventional recognition systems require uniform illumination of light in the recognition area, without scattered light such as shadows and flashes. Some high-end products have reduced lighting requirements, but their costs are relatively high. Therefore, in order to improve the accuracy of sheep face recognition, it is necessary to add supplementary lighting equipment in situations with poor lighting conditions. There is also a hardware factor, which refers to the performance of the camera and control motherboard in the recognition system. The commonly used recognition camera pixels are between 2 million and 4 million pixels, not necessarily the higher the pixel, the better.

5. Conclusion

The sheep face detection dataset for training scenarios is divided into training set, testing set, and validation set in an 8:1:1 ratio, and the results meet the requirements. YOLOv3 accuracy and recall display can achieve real-time monitoring of sheep faces under natural light. In smaller datasets and with lower threshold for devices (hardware camera pixels of 2 - 4 million), it saves hardware configuration costs and meets the requirements for use in pastoral areas.

Acknowledgments

We are grateful to the reviewers and editors for their valuable comments and suggestions, which have significantly improved the quality of this manuscript. This work was supported by the Science and Technology Plan Project of Inner Mongolia Autonomous Region (Grant No. 2022YFYZ0009 and 2023YFDZ0078).

Authors' Contributions

Guoqing Chen, Lerong Wen, and Xu Zheng designed the experiments and interpreted the results; Xu Zheng wrote the draft; Zheng Xu and Guoqing Chen pre-

pared the figures and reviewed the manuscript; all authors revised the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zhu, M.F. and Cheng, G.Q. (2025) The Realistic Challenges, Key Issues, and Promotion Strategies for the High-Quality Development of China's Herbivorous Animal Husbandry Industry. *Journal of Social Sciences*, No. 6, 164-175. (In Chinese)

[2] Yang, C.D., Qi, J.L., Yang, T.H. and Zhang, L.J. (2025) The Spatial Correlation Network and Driving Mechanism for the Coupling and Coordination of Digital Rural Construction and Green Transformation of Animal Husbandry in China. *Research on Agricultural Modernization*, **47**, 14-25. (In Chinese)

[3] Zhang, W.Y. (2020) Research and Implementation of Smart Ranch Management System. Master's Thesis, Harbin University of Science and Technology.

[4] Alomair, R., Al-Amoudi, A., Javaid, A., Alnaser, M. and Al binali, S. (2024) Enhancing Precision Livestock Farming Management with AI-Driven Ear Tag Detection and OCR Recognition. 2024 *IEEE International Conference on Technology Management, Operations and Decisions* (*ICTMOD*), Sharjah, 4-6 November 2024, 1-6. https://doi.org/10.1109/ictmod63116.2024.10878196

[5] Tomley, F.M. and Shirley, M.W. (2009) Livestock Infectious Diseases and Zoonoses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 2637-2642. https://doi.org/10.1098/rstb.2009.0133

[6] Ouali, M., Belhouadjeb, F.A., Soufan, W. and Rihan, H.Z. (2023) Sustainability Evaluation of Pastoral Livestock Systems. *Animals*, **13**, Article 1335. https://doi.org/10.3390/ani13081335

[7] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/iccv.2015.169

[8] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/iccv.2017.322

[9] Sun, Z., Tian, L., Du, Q., Bhutto, J.A. and Wang, Z. (2022) Feature Learning via Multi-Action Forms Supervising Force for Face Recognition. *Neural Computing and Applications*, **34**, 4425-4436. https://doi.org/10.1007/s00521-021-06598-z

[10] Zhang, X., Song, Y., McCurdy, W., Wang, X. and Zuo, F. (2024) Revising the Problem of Partial Labels from the Perspective of CNNs' Robustness. 2024 *IEEE/ACIS 22nd International Conference on Software Engineering Research, Management and Applications* (*SERA*), Honolulu, 30 May-1 June 2024, 88-93. https://doi.org/10.1109/sera61261.2024.10685603

[11] Zou, X., Yin, Z., Li, Y., Gong, F., Bai, Y., Zhao, Z., *et al*. (2023) Novel Multiple Object Tracking Method for Yellow Feather Broilers in a Flat Breeding Chamber Based on Improved YOLOv3 and Deep SORT. *International Journal of Agricultural and Biological Engineering*, **16**, 44-55. https://doi.org/10.25165/j.ijabe.20231605.7836

[12] Zhang, C., Liu, X., Shang, Q., Wu, C., Wang, Y., Wang, L., *et al*. (2025) Personnel Target Detection in Infrared Environment Based on YOLOv3-Tinier Network and Its FPGA Implementation. *Infrared Physics & Technology*, **150**, Article ID: 106015. https://doi.org/10.1016/j.infrared.2025.106015

[13] Nigam, N., Singh, D.P. and Choudhary, J. (2025) Real-Time Traffic Spatial Occupancy Calculation with Modified YOLOv3 in Complex Environments. *Procedia Computer Science*, **260**, 717-724. https://doi.org/10.1016/j.procs.2025.03.251

[14] Chuong, V.H., Cuong, V.H., Dat, V.N., Thanh, N.T.C., Thanh, P.T. and Le Quan, N. (2024) Facial Expression Recognition: A Lite Deep Learning-Based Approach. In: Yang, XS., Sherratt, S., Dey, N. and Joshi, A., Eds., *Proceedings of Ninth International Congress on Information and Communication Technology*, Springer, 125-135. https://doi.org/10.1007/978-981-97-3559-4_10

[15] Wei, B. (2020) Sheep Face Detection and Recognition Based on Deep Learning. Master's Thesis, North-West A&F University.

[16] Kolsrud, D. (2015) A Time-Simultaneous Prediction Box for a Multivariate Time Series. *Journal of Forecasting*, **34**, 675-693. https://doi.org/10.1002/for.2366