YOLO-LSD: A LIGHTWEIGHT MODEL FOR HIGH-ACCURACY MULTI-BREED SHEEP FACE DETECTION

/

YOLO-LSD：一种轻量级的高精度多品种羊脸检测模型

Xiwen ZHANG[1,2]; Zelin NIU[1]; Yanxin GUO[1,3], Yu CAI[3], Ruiyan SUN[1,3*]
[1] Jiangsu Maritime Institute, College of Marine Electrical and Intelligent Engineering, Nanjing, China
[2] Inner Mongolia Agricultural University, College of Mechanical and Electrical Engineering, Inner Mongolia, China
[3] Industrial Center, Nanjing Institute of Technology, Nanjing, China
*Tel: 0471-4309215; *)Corresponding author E-mail: nit_sry@126.com*
DOI: https://doi.org/10.35633/inmateh-76-97

Keywords: *Sheep face detection; Deep learning; High-accuracy model; YOLOv11; Multi-breed*

ABSTRACT

Sheep face detection is critical for intelligent livestock management and breeding, yet existing models often struggle in complex farm scenarios due to inadequate multi-scale feature utilization and high computational demands. To address these challenges, this study proposes a lightweight multi-breed sheep face detection framework named YOLO-LSD (Lightweight Sheep Face Detection), achieving an optimal balance between detection accuracy and computational efficiency through multi-dimensional optimizations. At the feature enhancement level, the lightweight channel attention mechanism Efficient Channel Attention (ECA) is embedded into the backbone network to dynamically strengthen the channel responses of key facial features through local cross-channel interactions. Concurrently, Ghost convolution is introduced to replace traditional convolutional layers, leveraging feature redundancy mining technology to substantially reduce computational complexity while maintaining the ability to represent diverse facial features across sheep and goat breeds. To address the limited sample diversity in multi-breed datasets, a transfer learning strategy is employed, involving directional fine-tuning of breed-specific facial features based on large-scale pre-trained models to enhance the model's generalization ability across diverse sheep and goat varieties. Experimental results demonstrate that YOLO-LSD achieves a mAP@0.5 of 99.29% on a self-constructed multi-breed sheep face dataset, marking a 0.59% improvement over the baseline YOLOv11. Notably, the parameter count of YOLO-LSD is only 2.4×10^6, while achieving an inference speed of 60 FPS and 6.3 Flops. This study presents a high-precision, lightweight solution for intelligent livestock monitoring systems, offering practical insights for the deployment of multi-breed sheep face detection models in real-world farm applications.

摘要

绵羊面部检测是智能畜牧管理和养殖的关键，但由于多尺度特征利用不足和计算需求高，现有模型在复杂的农场场景中往往难以实现。为了解决这些挑战，本研究提出了一种轻量级的多品种羊人脸检测框架，名为YOLO-LSD (lightweight sheep face detection)，通过多维优化实现了检测精度和计算效率之间的最佳平衡。在特征增强层面，将轻量级通道注意机制ECA嵌入骨干网络，通过局部跨通道交互，动态增强关键面部特征的通道响应。同时，引入Ghost卷积来取代传统的卷积层，利用特征冗余挖掘技术大幅降低计算复杂性，同时保持表示绵羊和山羊品种不同面部特征的能力。为了解决多品种数据集样本多样性有限的问题，采用迁移学习策略，在大规模预训练模型的基础上对特定品种的面部特征进行定向微调，以提高模型在不同绵羊和山羊品种间的泛化能力。实验结果表明，YOLO-LSD在自构建的多品种绵羊面部数据集上的mAP@0.5 达到了 99.29%，比基线YOLOv11提高了0.59%。值得注意的是，YOLO-LSD的参数量仅为 2.4×10^6，同时实现了 60 FPS和6.3 Flops的推理速度。本研究提出了一种高精度、轻量级的智能牲畜监测系统解决方案，为在实际农场应用中部署多品种绵羊面部检测模型提供了实用的见解。

INTRODUCTION

With the increasing advancement of intelligent livestock farming, including intensive breeding, breed improvement, and precision animal health management, the demand for accurate identification and monitoring of sheep, as key livestock resources, has emerged as a critical research focus in agricultural technology (*Sharma et al., 2020*).

Against the backdrop of rapid developments in smart farming systems, high-precision sheep face detection techniques serve not only as a foundational enabler for individual animal tracking and breed classification but also as core technology for optimizing feeding strategies and ensuring animal welfare (*Xue et al., 2024*). The accurate detection of sheep face targets is directly linked to the efficiency of farm management, and its significance becomes particularly pronounced in real-time scenarios such as automated feeding and health monitoring (*Zhang et al., 2024*). However, traditional manual sheep identification methods suffer from notable drawbacks, such as low efficiency, poor real-time performance, susceptibility to false negatives and positives in complex farm environments, and high labor costs (*Hao et al., 2024*).

Traditional manual sheep detection methods have manifested numerous deficiencies in practical applications. The visual screening approach is highly inefficient, rendering it unable to handle the massive volume of monitoring data generated in large-scale farms and falling short of real-time management requirements (*Salama et al., 2019*). In time-sensitive scenarios such as disease outbreak response, the inherent delays in manual identification can lead to untimely intervention, thereby increasing the risk of epidemic spread (*Hitelman et al., 2022*). Moreover, this method is heavily reliant on human resources, with labor costs increasing non-linearly as the farming scale expands. These limitations have become increasingly prominent in the face of the urgent demand for automated and high-precision management in smart agriculture, prompting both academia and industry to explore intelligent detection technologies based on deep learning to overcome the long-standing bottlenecks in efficiency, accuracy, and scalability associated with traditional methods (*Peruzzi et al., 2025*).

In the early stages of sheep face detection, two-stage object detection algorithms, such as Faster R-CNN, were commonly employed (*Zhang et al., 2022*). These algorithms generate candidate bounding boxes through region proposal networks and conduct fine-grained classification for each box, enabling high detection accuracy in complex farm scenarios (*Deng et al., 2022*). Nevertheless, the two-stage cascaded architecture leads to high computational complexity and slow inference speed (*Deng et al., 2021*). Additionally, the manually designed anchor boxes struggle to adapt to the multi-scale facial features across different sheep breeds, often resulting in missed detection of small targets and positioning deviations for large targets. As a result, these algorithms fail to meet the requirements of real-time farm monitoring.

As a typical representative of single-stage object detection algorithms, the YOLO series transforms object detection into a problem of directly predicting the coordinates of grid cells and class probabilities through an end-to-end regression architecture (*He et al., 2016*). This approach avoids the redundant calculation of candidate bounding box generation in two-stage algorithms, significantly improving inference efficiency while maintaining detection accuracy. On this basis, the YOLOv11 algorithm further optimizes the backbone network structure. By adopting the CSPDarknet lightweight feature extraction module and the efficient Feature Pyramid Network (FPN), it maintains the multi-scale feature representation ability while reducing the number of parameters. It is particularly suitable for complex farm scenarios where small distant sheep faces and large close-up faces coexist. Its improved anchor box adaptation mechanism learns the aspect ratio priors of the sheep face dataset through K-means clustering, enhancing the matching degree between the anchor boxes and the actual facial shapes of different breeds (*Zhang et al., 2023*). This effectively alleviates the adaptation deviation problem of traditional manually designed anchor boxes for varied facial characteristics, such as the rounder faces of some sheep breeds versus the more angular features of others. These characteristics endow YOLOv11 with strong engineering practicality in farm monitoring, making it an ideal baseline model for real-time sheep face detection.

In summary, to overcome the challenges of complex scale variation, cluttered farm backgrounds, and limited sample diversity in multi-breed sheep face detection, this work presents YOLO-LSD, a lightweight enhancement of YOLOv11. Specifically, we embed the ECA module into the backbone to adaptively amplify key facial features and suppress interference from farm backgrounds. Then, we replace standard convolutional layers with Ghost convolution blocks, mining feature redundancy to slash both model parameters and FLOPs without sacrificing multi-scale representation. Finally, the strategy of transfer learning pre-training is adopted. Through pre-training on large-scale animal facial datasets, the learning ability of the model for sheep facial features is further enhanced. Through these multi-dimensional optimizations, YOLO-LSD strikes an optimal balance between detection accuracy and computational efficiency, offering a practical, high-precision solution for real-world farm surveillance.

MATERIALS AND METHODS

Dataset

The dataset constructed in this study integrates two categories: sheep and goats, forming a composite multi-breed sheep face dataset. The sheep samples specifically belong to the Small-tailed Han breed, with a total of 50 individuals selected for data collection. These facial images were acquired in May 2024 at a farm operated by Beiqi Technology Co., Ltd. In Hohhot, Inner Mongolia, China. Image capture was performed using a Canon EOS 600D digital single-lens reflex (DSLR) camera (manufactured by Canon Inc., Tokyo, Japan), with all images stored in JPG format at a resolution of 2736×1824 pixels. For each sheep in the experiment, 50 original facial images were collected.

The goat dataset was sourced from an open-access repository, which contains manually captured images of 10 dairy goats from a farm in China for facial recognition purposes, amounting to 1311 goat face images in total (*Billah et al., 2023*). Data collection took place outdoors over a three-month period, encompassing variations in weather and seasonal conditions. Each goat was photographed individually in three distinct pens and across three-time segments: morning, midday, and afternoon. Each goat collected approximately 80 to 150 raw face images. Both the sheep and goat datasets were collected in real-world outdoor farm environments, encompassing diverse collection conditions such as varying lighting (e.g., morning sunlight, overcast afternoons), seasonal changes, and typical farm backgrounds (e.g., fences, feeding areas). These conditions effectively simulate the complex environments encountered in real-time sheep face detection tasks, ensuring that the proposed method in this study possesses strong practicality and effectiveness in actual application scenarios. Examples of original facial samples from both sheep and goats are presented in Fig.1.

Fig. 1 – Sample images of the sheep and goat faces

To improve the model's adaptability to complex farm environments, data augmentation was applied to the collected sheep face images. Specific operations included: modulating image brightness within a range of -45% to 45%, adjusting contrast by ±45%, rotating images by 45 degrees left and right, and performing vertical flipping. Through these augmentation strategies, augmented images were generated for each sheep in the experiment, effectively expanding the training data volume. Ultimately, each sheep retained 150 facial images (including original and augmented samples), forming the complete multi-breed sheep face image dataset. In addition, to ensure accurate target localization during model training, all images (both original and augmented) were annotated using the Make Sense online tool. For each image, a bounding box was manually drawn to precisely enclose the sheep face region, with the category label uniformly set as "Sheep Face" to standardize the annotation format. The sample annotation diagrams are shown in Fig.2.

Fig. 2 – Sample annotation diagrams

The final multi-breed sheep face dataset was randomly partitioned into training, validation, and test sets at a ratio of 8:1:1. Detailed configuration parameters of the dataset are presented in Table 1.

Table 1

The multi-breed sheep face dataset

Dataset	Images	Proportion
Training	7200	80%
Verification	900	10%
Testing	900	10%
Total	9000	100%

YOLOv11

The state-of-the-art single-stage object detection algorithm YOLOv11 is composed of three main components: backbone network, neck network, and detection head, which are optimized for detection scenarios to achieve efficient sheep face detection (*Li et al., 2025*). The backbone network employs an optimized CSPDarknet architecture, where the basic module CBS consists of a convolution, batch normalization, and activation function (*Jo et al., 2024*). A novel feature preprocessing module is deployed at the front end to process high-resolution images more efficiently, minimizing information loss compared to traditional downsampling. During the feature extraction stage, an improved spatial pyramid pooling (SPP) structure is utilized to finely capture multi-scale spatial features, enhancing adaptability to sheep face targets of different sizes (*Kumar et al., 2023*). The CSPLayer, through optimizing the residual structure, effectively integrates gradient changes of feature maps into the output results, not only reducing the total number of network parameters and computational complexity but also delivering high-quality features to the neck network, thus serving as a core component for efficient feature extraction (*Chen et al., 2025*).

The neck network of YOLOv11 adopts an adaptive feature fusion architecture to dynamically optimize multi-scale feature weights. Through an innovative cross-layer connection mechanism, high-resolution detail features from shallow layers are deeply fused with semantic-rich features from deep layers to construct a more representative feature pyramid. This design enhances the model's detection capability for both distant small targets and nearshore large targets, effectively addressing the challenge of extreme scale distribution of sheep face targets in complex scenarios, and lays a foundation for the accurate prediction of the detection head.

The detection head employs a decoupled design, separating the target classification and localization regression tasks into independent branches. The classification branch outputs multi-class probabilities for sheep face detection, while the localization regression branch optimizes the prediction accuracy of bounding box coordinates and confidence scores by integrating improved loss functions. Multi-scale feature maps are fused to cover sheep face targets of different scales, and anchor boxes adapted to the aspect ratio distribution of sheep faces are generated through K-means++ clustering, replacing manually preset anchors to improve matching efficiency. In the prediction stage, confidence screening and non-maximum suppression (NMS) are used to reduce false positives in complex backgrounds, and dynamic sample matching strategies are employed to balance positive and negative samples, achieving precise detection and localization of multiple sheep face types while maintaining end-to-end efficient inference.

ECA attention

To enhance the model's discriminative ability for sheep face targets in complex backgrounds, this study embeds a lightweight channel attention mechanism, ECA, into the backbone network of YOLOv11. This mechanism adaptively strengthens channel responses related to sheep features, through a local cross-channel interaction strategy, all with almost no additional computational cost (*Peng et al., 2020*).

For the feature map $X \in \mathbb{R}^{H \times W \times C}$ (where $H \times W$ is the spatial dimension and C is the number of input channels) output by the backbone network, spatial dimension information is first aggregated via global average pooling to generate a channel-level descriptor $Z \in \mathbb{R}^{C}$, calculated as:

$$Z_C = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \quad (1)$$

This compresses spatial features into global response values in the channel dimension, providing a basis for channel attention calculation.

Unlike traditional channel attention mechanisms, SE-Net, which rely on fully connected layers and incur high computational costs, ECA achieves lightweight local cross-channel interactions via a 1D convolution with kernel size k. Here, k is adaptively determined by the number of channels C using:

$$k = |\log_2(C) + 1|_{odd} \quad (2)$$

where, $|\cdot|_{odd}$ denotes adjusting the result to the nearest odd integer (e.g., $k = 5$, when $C = 32$) to ensure convolution symmetry and avoid global redundant calculations. After extracting local channel correlations via this 1D convolution, channel attention weights $w_c \in [0,1]$ are generated through a sigmoid function to quantify the importance of each channel for sheep face detection. Among them, weights close to 1 enhance responses in critical feature channels, while weights approaching 0 suppress noise in background interference channels.

Finally, channel attention weights are multiplied element-wise with the original feature map to obtain the enhanced feature map $\hat{X}_c = w_c \cdot X_c$. This design introduces only linear complexity parameters $O(kC)$. While maintaining efficient model inference, it dynamically adjusts the channel interaction range for scale differences of sheep face targets in scenarios: smaller k values focus on high-frequency detail features, whereas larger k values integrate low-frequency semantic features, significantly improving feature discriminability in complex backgrounds and providing more distinguishable inputs for subsequent multi-scale feature fusion and target detection (*Xue et al., 2025*). The structure diagram of the ECA module is shown in Fig. 3.

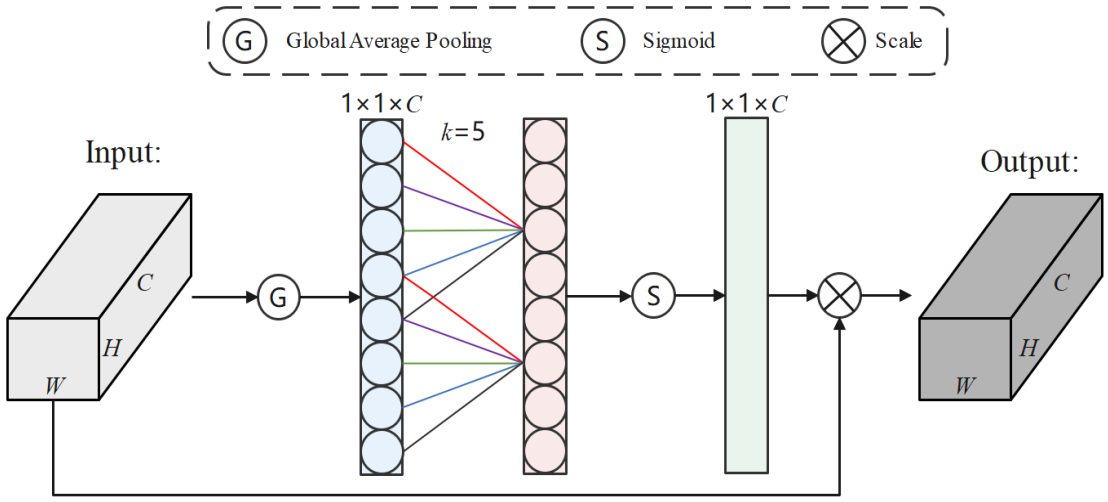

Fig. 3 - The structure diagram of the ECA module

Ghost convolution

To address the issue of low computational efficiency of traditional convolutions in sheep face detection, this study introduces Ghost Convolution into the backbone network of YOLOv11 to achieve lightweight feature extraction by exploiting feature map redundancy. Ghost Convolution decomposes traditional convolution into two stages: intrinsic feature extraction and Ghost feature generation, significantly reducing parameter overhead while preserving critical target semantics (*Dai et al., 2024*).

For an input feature map $X \in \mathbb{R}^{H \times W \times C_{in}}$ (where $H \times W$ is the spatial dimension and C_{in} is the number of input channels), Ghost Convolution first generates a channel-reduced intrinsic feature $F_{int} \in \mathbb{R}^{H \times W \times m} (m \ll C_{in})$, calculated as:

$$F_{int} = Conv(X,\ k,\ m,\ s = 1,\ p = 1) \quad (3)$$

where, k is the convolution kernel size (set to 3×3 in this study) for capturing basic spatial semantic information, such as sheep face contours and textures. Subsequently, lightweight linear transformations are applied to the intrinsic features to generate s-1 groups of Ghost features F_{gh} (s is the expansion factor, set to 2 in this study to balance efficiency and feature diversity).

The output feature map is obtained by concatenating intrinsic and Ghost features:

$$F_{out} = Concat(F_{int}, F_{gh}) \in R^{H \times W \times ms} \quad (4)$$

Compared to traditional convolutions with C_{out} output channels (parameter count: $C_{in} \cdot k^2 \cdot C_{out}$), Ghost Convolution has a parameter count of:

$$Params_{GhostConv} = C_{in} \cdot k^2 \cdot m + m \cdot k^2 \cdot (s - 1) \quad (5)$$

When $s = 2$ and $m = C_{out}/2$, this achieves a significant reduction in computational complexity without loss of feature dimensions.

In this study, replacing traditional convolutions in the backbone with Ghost Convolution is particularly suitable for down sampling high-resolution images. By reusing the spatial correlation of intrinsic features, the generated diverse Ghost features effectively enhance the edge detail representation of distant small sheep faces and the structural semantic expression of nearshore large sheep faces. This lightweight design enables efficient inference on edge devices while retaining the feature discriminability required for sheep face detection in complex scenarios, making it a key technical module for balancing detection accuracy and computational efficiency. The structure diagram of the Ghost module is shown in Fig.4.

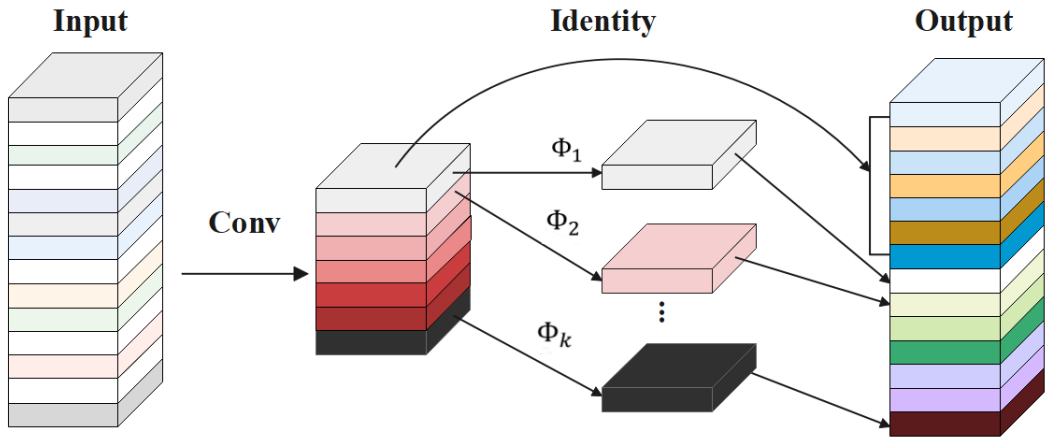

Fig. 4 - The structure diagram of the Ghost module

Transfer learning

To address the challenges of scarce high-quality labeled data and insufficient model generalization in few-shot scenarios for sheep face detection, this study employs a transfer learning strategy to efficiently transfer pre-trained knowledge to the sheep face detection task. Specifically, the constructed multi-breed sheep face dataset was first split into two subsets, A and B, ensuring a balanced distribution of breeds, lighting conditions, and shooting angles in both subsets. Cross-pre-training was then performed: subset A was used as the pre-training dataset, and the model was trained from scratch to obtain pre-trained weights, which were subsequently used as initial weights for training subset B. Conversely, subset B was employed for pre-training to generate weights, which served as the starting point for training subset A. During each pre-training phase, the optimizer's learning rate was adjusted and shallow network layers were frozen to enable the model to gradually learn breed-specific facial features (e.g., muzzle contours, ear shapes) and adapt to variations in farm environments. This cross-pre-training strategy enabled the limited labeled data to be fully leveraged, accelerated convergence in subsequent training, and enhanced generalization ability across different sheep face samples. Finally, the average of the two sets of training results was taken as the final detection performance, effectively mitigating the impact of data imbalance and improving both detection accuracy and robustness.

YOLO-LSD

The YOLO-LSD model builds on YOLOv11 with two targeted modifications, the ECA attention mechanism and Ghost convolutions. Specifically, an ECA module was added immediately before the SPPF block in the backbone and before every C3k2 block in the neck, so channel-wise sheep face features such as hull contours and superstructure lines are adaptively amplified and distinctions between container sheep faces and other classes are enhanced. Meanwhile, the final convolution in the backbone and all convolutions in the neck were replaced with Ghost convolution blocks. Each Ghost layer decomposes a standard convolution into intrinsic feature extraction and lightweight transformation, reusing spatial correlations to preserve multi-scale detail for distant small sheep faces while reducing parameters and computational complexity for fast and resource-efficient inference. The overall structure diagram of YOLO-LSD is shown in Fig. 5.

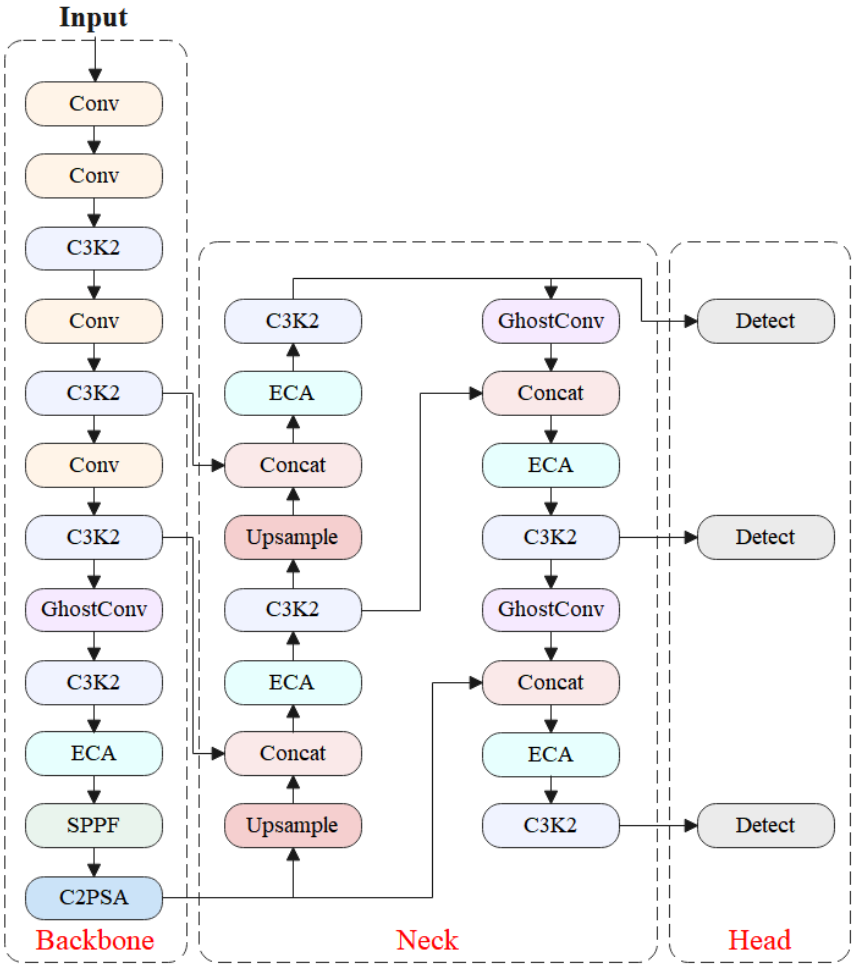

Fig. 5 - The overall structure of YOLO-LSD

Training platform and hyperparameters

The experiment was conducted on a Windows 11 operating system with a hardware platform configured with an i7-9700 processor (3.0 GHz clock speed), 16 GB of RAM, and an NVIDIA GeForce RTX 2080 Ti discrete graphics card, which provided efficient computational support for model training and inference. The algorithm was implemented using the PyTorch 1.12.0 deep learning framework in a Python 3.8 development environment, with GPU computing performance optimized via CUDA 11.6 and cuDNN acceleration technologies to ensure training efficiency for complex models. During training, a batch size of 16, 200 epochs, and a stochastic gradient descent (SGD) optimizer with a momentum parameter of 0.937 were set, balancing model convergence speed and parameter optimization stability while providing a standardized hyperparameter configuration foundation for the multi-module collaborative training of the improved YOLOv11 model.

Evaluation Metrics

For the sheep face detection task in complex environments, this study employs Precision, Recall, F1-score, mean Average Precision (mAP@0.5), FPS, GFLOPs, and Parameters as core evaluation metrics to construct a comprehensive evaluation system from three aspects: detection accuracy, efficiency, and model complexity.

Parameters reflect the total number of trainable parameters in the model, used to quantify the complexity of the network structure. Precision is defined as the ratio of the number of correctly detected sheep face samples to the total number of samples detected as sheep face. True negative, true positive, and false negative are the sample numbers of TN, TP, and FN. The calculation formula is:

$$\text{Precision}=\text{TP}/(\text{TP}+\text{FP}) \quad (6)$$

Recall is defined as the ratio of the number of correctly detected sheep face samples to the total number of actual sheep face samples, with the formula:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (7)$$

The F1-score provides a balanced evaluation metric that accounts for both precision and recall through their harmonic mean, with the calculation formula:

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \quad (8)$$

FPS refers to the number of images processed by the model per second, used to measure the algorithm's real-time response capability to dynamic complex scenarios. GFLOPs is used to quantify the model complexity, reflecting the number of floating-point operations required for a single forward pass.

mAP@0.5 calculates the mean average precision for sheep face at an intersection-over-union (IOU) threshold of 0.5, with the formula:

$$\text{AP} = \int_0^1 \text{P(R)d R} \quad (9)$$

where, the average precision (AP) for a single category is obtained by integrating the precision-recall curve:

$$\text{mAP} = \sum_{i=1}^{N}(\text{AP}_i/\text{N}) \quad (10)$$

where, N is the total number of identification types.

RESULTS AND DISCUSSIONS

Comparison experiments of YOLO series algorithms

To further evaluate the detection performance of models for subsequent improvement strategy integration, comparative experiments with YOLO series algorithms were carried out. Representative YOLO models were selected including YOLOv4-tiny, YOLOv5s, YOLOv7-tiny, YOLOv8n, YOLOv10n, and YOLOv11n for comparison. All models were tested under consistent settings, and the results are presented in Table 2.

As Table 2 shows, YOLOv11n demonstrates excellent performance across multiple metrics. In terms of Precision, it achieves 98.56%, outperforming YOLOv4-tiny (89.03%) by 9.53 percentage points, YOLOv5s (90.95%) by 7.61 percentage points, YOLOv7-tiny (92.85%) by 5.71 percentage points, YOLOv8n (95.99%) by 2.57 percentage points, and YOLOv10n (98.28%) by 0.28 percentage points. For Recall, YOLOv11n reaches 96.12%, showing improvements of 5.46 percentage points over YOLOv4-tiny (90.66%), 5.37 percentage points over YOLOv5s (90.75%), 4.15 percentage points over YOLOv7-tiny (91.97%), 0.57 percentage points over YOLOv8n (95.55%), and 0.34 percentage points over YOLOv10n (95.78%).

In terms of the F1-score, which comprehensively reflects the balance between Precision and Recall, YOLOv11n attains 97.32%. It surpasses YOLOv4-tiny (89.82%) by 7.5 percentage points, YOLOv5s (90.84%) by 6.48 percentage points, YOLOv7-tiny (92.40%) by 4.92 percentage points, YOLOv8n (95.78%) by 1.54 percentage points, and YOLOv10n (97.02%) by 0.3 percentage points. For the critical mAP@0.5 metric, YOLOv11n scores 98.70%, outshining YOLOv4-tiny (89.88%) by 8.82 percentage points, YOLOv5s (90.32%) by 8.38 percentage points, YOLOv7-tiny (94.55%) by 4.15 percentage points, YOLOv8n (97.76%) by 0.94 percentage points, and YOLOv10n (98.52%) by 0.18 percentage points. Additionally, regarding model volume (Params), YOLOv11n has only 2.6×10^6 parameters. It is more lightweight compared to YOLOv4-tiny (6.0×10^6), YOLOv5s (7.2×10^6), YOLOv7-tiny (6.3×10^6), YOLOv8n (3.0×10^6), and YOLOv10n (2.7×10^6). These results indicate that YOLOv11n exhibits significant advantages in model performance, detection comprehensiveness, and model volume. Therefore, this study takes YOLOv11n as the benchmark model, and subsequent improvement strategies will be added based on it to further enhance detection capabilities while maintaining its lightweight characteristics.

Table 2

The training results of the YOLO series algorithms

Model	Precision (%)	Recall (%)	F1-score (%)	mAP@0.5 (%)	Params (10^6)
YOLOv4-tiny	89.03	90.66	89.82	89.88	6.0
YOLOv5s	90.95	90.75	90.84	90.32	7.2
YOLOv7-tiny	92.85	91.97	92.40	94.55	6.3
YOLOv8n	95.99	95.55	95.78	97.76	3.0
YOLOv10n	98.28	95.78	97.02	98.52	2.7
YOLOv11n	98.56	96.12	97.32	98.70	2.6

Ablation Experiment

To assess the impact of individual and combined improvements on model performance, ablation experiments were carried out, with results presented in Table 3.

Table 3

The ablation results of YOLO-LSD

Method	ECA	Ghost convolution	F1-score (%)	mAP@0.5 (%)	Params (10^6)	FPS	GFLOPs
1	×	×	97.32	98.70	2.6	56	6.4
2	√	×	97.90	98.88	2.6	50	6.6
3	×	√	97.12	98.31	2.4	62	6.2
4	√	√	97.98	98.96	2.4	60	6.3

The effects of integrating the ECA attention mechanism and Ghost module into the baseline YOLOv11n (Method 1) were analyzed, to evaluate their contributions to key metrics: F1-score, mAP@0.5, model complexity (Params), inference speed (FPS), and computational cost (GFLOPs).

As shown in Table 3, Method 2 introduces only the ECA attention mechanism. Compared to the baseline (Method 1), it improves the F1-score from 97.32% to 97.90% and mAP@0.5 from 98.70% to 98.88%. However, this comes with a trade-off: FPS decreases from 56 to 50 and GFLOPs increase from 6.4 to 6.6, while model parameters (Params) remain unchanged at 2.6×10^6. This indicates ECA enhances feature attention for better accuracy but adds marginal computational overhead.

Method 3 incorporates only the Ghost module. Relative to the baseline, it reduces model parameters to 2.4×10^6 and increases FPS to 62. However, detection performance slightly degrades: F1-score drops to 97.12% and mAP@0.5 to 98.31%. The Ghost module effectively streamlines model complexity for efficiency but sacrifices minor accuracy, highlighting a classic accuracy-efficiency trade-off.

Method 4 combines both ECA and Ghost modules. It achieves a balanced improvement: F1-score reaches 97.98%, mAP@0.5 rises to 98.96%. Meanwhile, model parameters were reduced to 2.4×10^6, and GFLOPs to 6.3. FPS is 60, which is faster than Method 1 and balances the speed loss from ECA alone. This synergy shows ECA compensates for Ghost's accuracy loss, while Ghost offsets ECA's computational overhead, resulting in a more robust model.

In summary, the ECA module boosts accuracy but adds computational cost, while the Ghost module enhances efficiency but slightly reduces accuracy. Their combination (Method 4) achieves the best balance: improved F1-score (+0.66%), mAP@0.5 (+0.26%), lighter parameters (2.4×10^6), and reasonable speed (60 FPS). Thus, integrating both ECA and Ghost modules optimizes the trade-off between performance and efficiency, validating their synergistic value for enhancing YOLOv11n in subsequent refinements.

Comparative experiment of transfer learning

To evaluate the influence of transfer learning on model performance, comparative experiments were conducted, and the results are presented in Table 4. As shown in Table 4, the baseline model (YOLOv11n) achieves an F1-score of 97.98% and an mAP@0.5 of 98.96%, with 2.4×10^6 parameters, 60 FPS, and 6.3 GFLOPs. When transfer learning is introduced, the F1-score improves from 97.98% to 98.12%, and the mAP@0.5 rises from 98.96% to 99.29%.

Importantly, transfer learning does not bring any negative impacts on model complexity and efficiency. The number of parameters remains unchanged at 2.4×10^6, and both FPS and GFLOPs stay the same as the baseline. This indicates that transfer learning can effectively leverage pre-trained knowledge to enhance the model's feature representation and generalization ability, thereby improving detection accuracy without increasing computational overhead or model size.

Table 4

The training results of transfer learning

Model	F1-score (%)	mAP@0.5 (%)	Params (10^6)	FPS	GFLOPs
YOLO-LSD	97.98	98.96	2.4	60	6.3
+ Transfer learning	98.12	99.29	2.4	60	6.3

The training curves of YOLO-LSD are shown in Fig.6. From the training curves, it can be seen that as the number of iterations progresses, train/box_loss, train/cls_loss, and train/dfi_loss all show a significant and stable downward trend. The corresponding losses on the validation set (val/box_loss, val/cls_loss, val/dfi_loss) also decrease synchronously, indicating that the model's learning of sheep face features is continuously deepening and the gradient update is stable. Meanwhile, metrics/precision and metrics/recall rise rapidly and then tend to be stable. Metrics/mAP50 is close to 1.0, and metrics/mAP50-95 also increase steadily, reflecting the gradual maturity of the model's classification and positioning capabilities and good generalization.

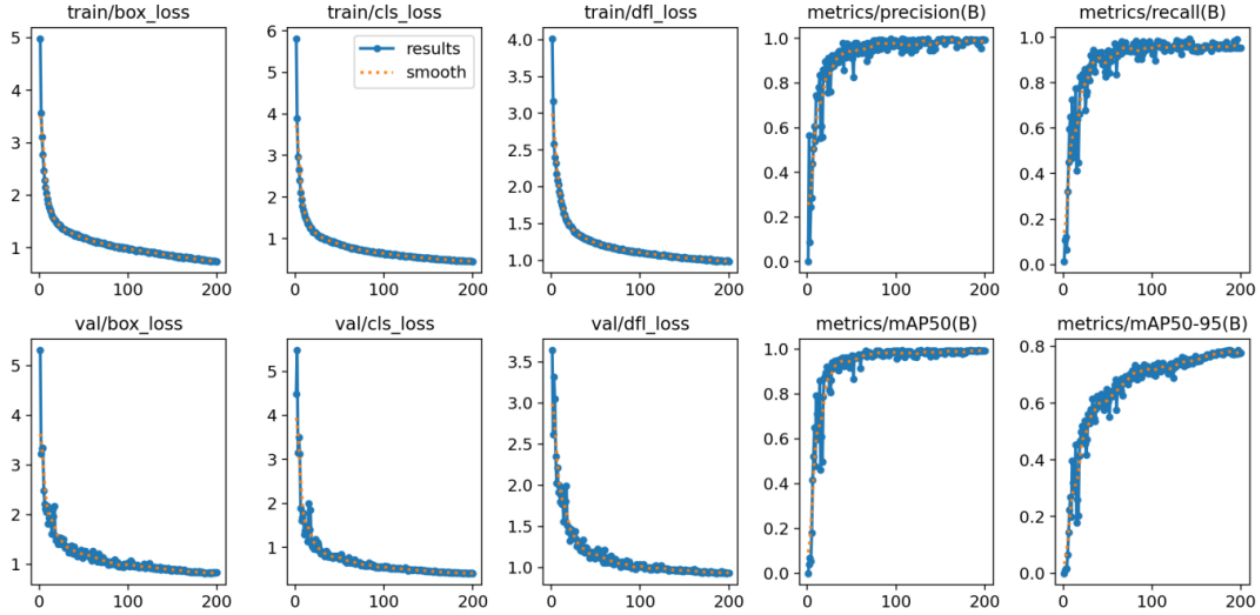

Fig. 6 - The training curves of YOLO-LSD

The detection effect of YOLO-LSD on individual sheep is shown in Fig. 7, and the effect on groups of sheep is shown in Fig. 8. By observing the specific detection results, it can be found that in different time periods and farm environments, the model can accurately identify the 'Sheep Face' category. For individual sheep detection, whether in close-up or long-distance shots, the bounding boxes show a high degree of fitting, and the confidence of category prediction is mostly above 0.80. When testing flocks of sheep, the confidence of each sheep face is above 0.75, and all sheep faces are effectively detected. The test results demonstrate that the proposed YOLO-LSD model has strong adaptability to complex farm environments.

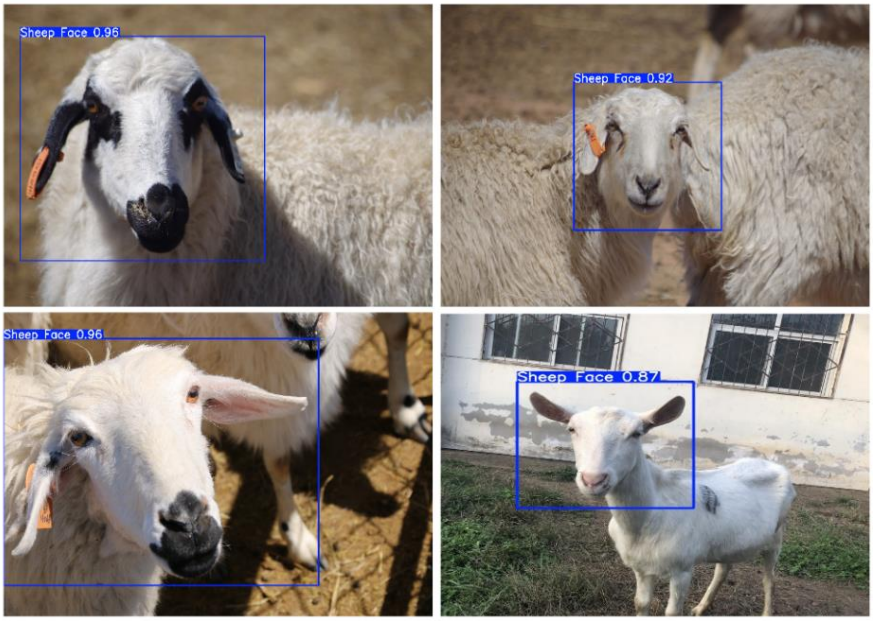

Fig. 7 - Sample images of individual sheep face detection results

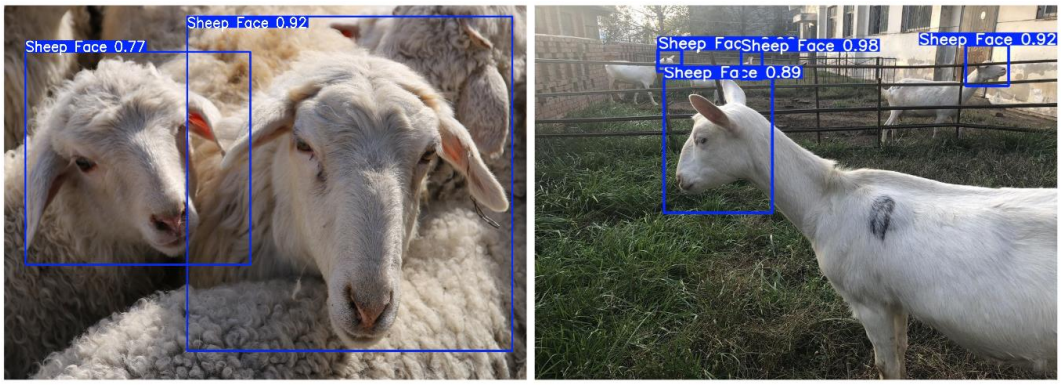

Fig. 8 - Sample images of detection results for groups of sheep

Comprehensively considering the convergence of the training curves and the recognition effect of the detection diagrams, the YOLO-LSD model proposed in this study has effectively completed the training task, constructed a stable and generalizable feature learning system, and can efficiently perform sheep face detection tasks. It provides reliable technical support for scenarios such as intelligent livestock management and farm monitoring, and fully verifies the practicality and effectiveness of the model design in sheep face detection tasks.

To demonstrate the practical deployment of the YOLO-LSD model, a real-time sheep face detection system was developed using LabVIEW 2018, as shown in Fig. 9. This system was tested in an actual farm breeding environment, specifically capturing the scenario where sheep pass through a detection channel, which naturally simulates the complex backgrounds (e.g., fences, vegetation) and variable lighting conditions encountered in real-time monitoring. The target detection interface within the system visualizes results on high-resolution RGB images, where two sheep faces are accurately identified, with confidence scores of 0.82 and 0.71, and bounding boxes fitting tightly to the facial regions even in a semi-overlapping situation. The parameter setting module allows for the configuration of the image path, processing type, and automatic timestamping, while the identification result module records detection quantities, staff information, and identification dates, and provides functions for data clearing, screenshot capturing, and result saving. Overall, this system not only validates the YOLO-LSD model's strong adaptability to complex farm environments for multi-target detection but also supports practical livestock management tasks like rapid flock counting and individual health monitoring, bridging the gap between algorithm development and on-site application.

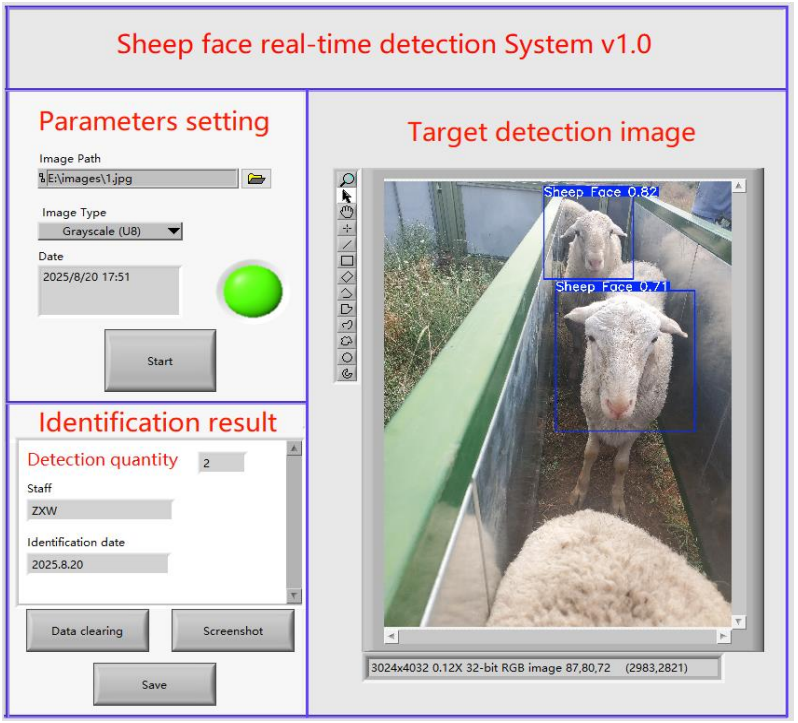

Fig. 9 - Operation diagram of the real-time sheep face detection system

CONCLUSIONS

This study proposes YOLO-LSD, a lightweight sheep face detection model developed based on YOLOv11. Through multi-dimensional optimizations-including feature enhancement, lightweight inference optimization, and few-shot generalization enhancement—the model effectively addresses the issues of insufficient multi-scale feature utilization and low computational efficiency in complex farm scenarios. Experimental results on the self-constructed multi-breed sheep face dataset show that YOLO-LSD achieves 99.29% mAP@0.5, with optimized model parameters and computational overhead, and a competitive inference speed on edge devices (e.g., farm-mounted embedded systems). Beyond technical performance, the model's practical deployment demonstrates tangible value for livestock management: by enabling automated, real-time monitoring of individual and group sheep, YOLO-LSD alleviates labor-intensive manual inspection, enhances the timeliness and precision of health and behavior assessment, and supports data-driven decision-making in farm operations. By synergizing these optimization strategies, YOLO-LSD breaks through the performance bottlenecks of traditional models in farm scenarios, providing a solution that balances performance and computational efficiency for intelligent livestock management systems—ultimately advancing operational efficiency, animal welfare, and data-informed practices in modern agriculture. Future research will focus on further lightweight optimization for low-power devices and multi-modal detection, integrating infrared data to enhance detection robustness in low-light or nighttime conditions, promoting the application of sheep face detection models in broader agricultural engineering scenarios.

ACKNOWLEDGEMENTS

We acknowledge that this work was supported by the Doctoral Research Start-up Fund Project of Jiangsu Maritime College (2024BSKY18).

REFERENCES

[1] Billah M., Wang X., Jiang Y., (2022), Real-time goat face recognition using convolutional neural network. *Computers and Electronics in Agriculture*, Vol 194, pp. 106730.

[2] Chen J., Yu R., Yang M., Che W., Ning Y., Zhan Y. (2025). SN-YOLO: A Rotation Detection Method for Tomato Harvest in Greenhouses. *Electronics*, Vol 14, pp. 3243.

[3] Dai D., Wu H., Wang Y., Ji P. (2024). LHSDNet: A Lightweight and High-Accuracy SAR Ship Object Detection Algorithm. *Remote Sens*., Vol 16, pp. 4527.

[4] Deng X., Yan X., Hou Y., Wu H., Feng C., Chen L., Bi M., Shao Y. (2021). Detection of behaviour and posture of sheep based on YOLOv3. *INMATEH-Agricultural Engineering*, Vol 64, Issue 2, pp. 457-466. DOI: https://doi.org/10.35633/inmateh-64-45

[5] Deng X., Zhang S., Shao Y., Yan X., (2022). A real-time sheep counting detection system based on machine learning. *INMATEH-Agricultural Engineering*, Vol 67, Issue 2, pp. 85-94. DOI: https://doi.org/10.35633/inmateh-67-08

[6] Hao M., Sun Q., Xuan C., Zhang X., Zhao M., Song S. (2024). Lightweight Small-Tailed Han Sheep Facial Recognition Based on Improved SSD Algorithm. *Agriculture*, Vol 14, pp. 468.

[7] He K., Zhang X., Ren S., Sun J., (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778.

[8] Hitelman A., Edan Y., Godo A., Berenstein R., Lepar J., Halachmi I. (2022). Biometric identification of sheep via a machine-vision system. *Computers and Electronics in Agriculture*, Vol 194, pp. 106713. https://doi.org/10.1016/j.compag.2022.106713

[9] Jo S., Woo J., Kang C.H., & Kim S.Y. (2024) Damage detection and segmentation in disaster environments using combined YOLO and Deeplab [J]. *Remote sensing*, 16(22), 4267.

[10] Kumar D., & Muhammad N. (2023) Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8 [J]. *Sensors*, 23(20), 8471.

[11] Li S., Wang R., Wang S., Yue P., Guo G. (2025). YOLO-FFRD: Dynamic Small-Scale Pedestrian Detection Algorithm Based on Feature Fusion and Rediffusion Structure. *Sensors*, Vol 25, pp. 5106, China.

[12] Peng, Z. H., Li, Q., & Liang, S. (2020) Multi-Scale Dense Selective Kernel Spatial Attention Network for Single Image De-raining [C]. 2020 *IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, *IEEE*, 341-347.

[13] Peruzzi, G., Galli, A., Giorgi G., & Pozzebon, A. (2025) Sleep posture detection via embedded machine learning on reduced set of pressure sensors [J]. *Sensors*, 25(2), 458.

[14] Salama A., Hassanien A.E., Fahmy A.A., (2019). Sheep identification using a hybrid deep learning and bayesian optimization approach. *IEEE Access*, Vol 7, pp. 31681-31687.
[15] Sharma A., Jain A., Gupta P., (2020). Machine learning applications for precision agriculture: A comprehensive review. IEEE Access, Vol 9, pp. 4843-4873.
[16] Xue H., Liu L., Wu Q., He J., Fan Y. (2025). Defect Detection Algorithm for Photovoltaic Cells Based on SEC-YOLOv8. Processes, Vol 13, pp. 2425.
[17] Xue J., Hou Z., Xuan C., Ma Y., Sun Q., Zhang X., Zhong L., (2024), A Sheep Identification Method Based on Three-Dimensional Sheep Face Reconstruction and Feature Point Matching. Animals, Vol 14, Issue 13, pp. 1923.
[18] Zhang X., Hou Z., Xuan C., (2022). Design and experiment of recognition system for coated red clover seeds based on machine vision. *INMATEH-Agricultural Engineering*, Vol 66, Issue 1, pp. 62-72. DOI: https://doi.org/10.35633/inmateh-66-06
[19] Zhang X., Xuan C., Ma Y., Su H., (2023). A high-precision facial recognition method for small-tailed Han sheep based on an optimised Vision Transformer, *Animal*, Vol 17, pp. 100886.
[20] Zhang X., Xuan C., Ma Y., Tang Z., Gao X., (2024), An efficient method for multi-view sheep face recognition. *Engineering Applications of Artificial Intelligence*, Vol 134, pp.108697.