PIG RECOGNITION BASED ON YOLOV8-EAPNET
/
基于 YOLOv8-EAPNet 的猪只行为识别

Juan LIU [*1)], Yaqi YAN [2)], Yongshuai YANG[2)], Yuhao HAO[2)], Baofan CHEN [2)], Mingkai YANG [2)], Jie HU [2)]
[1)] Department of Basic Sciences, Shanxi Agricultural University, Taigu, Shanxi / China
[2)] College of Software, Shanxi Agricultural University, Taigu, Shanxi / China;

Corresponding authors: Juan LIU; Tel: +86-18335463280; E-mail: liujuannk@sxau.edu.cn
DOI: https://doi.org/10.35633/inmateh-77-55

Keywords: *Pig Behavior Recognition; YOLOv8; Deep Learning; Attention Mechanism; Multi-Scale Feature Fusion*

ABSTRACT
With the advancement of intelligent farming technology, computer vision-based animal behavior recognition has become an important tool for improving production efficiency and animal welfare in modern farming management. To overcome the challenge of balancing computational efficiency and accuracy in existing behavior recognition systems, this study proposes an optimized model based on YOLOv8-EAPNet for accurately recognizing four main pig behaviors: standing, sitting, lateral lying, and prone lying. The framework adopts a multi-level lightweight design, incorporating three advanced technologies—C2f-ECA, SPPELAN, and Detect_AFPN—to enhance joint feature response, resolve spatial differences between sitting and lateral lying, and reconstruct semantics in occluded areas. This strengthens the model's robustness in complex farming environments and significantly improves the accuracy of pig behavior recognition. Validated on farm data, the model achieved an average precision improvement of 1.5% on a self-built dataset, with specific accuracy increases of 0.9% for standing, 1.7% for sitting, 3.0% for prone lying, and 0.3% for lateral lying. This technology provides an automated tool for early warning of limb injuries and respiratory diseases in pigs, promoting the upgrade of intelligent health management in the livestock industry and supporting the modernization of large-scale pig farming.

摘要
随着智能化养殖技术的进步，基于计算机视觉的动物行为识别成为现代养殖管理中提高生产效率与动物福利的重要工具。为了克服现有行为识别系统在处理计算效率与精确度之间的平衡问题，本研究提出了一种基于 YOLOv8-EAPNet 的优化模型，用于精确识别猪的四种主要行为：站立、坐姿、侧卧和趴卧。该框架采用了多层次轻量化设计，结合 C2f-ECA、SPPELAN 和 Detect_AFPN 三项先进技术，分别用来强化关节特征响应，解析坐姿与侧卧空间差异和重建遮挡区域语义。增强了模型在复杂养殖环境下的鲁棒性，显著提升了猪行为识别的精度。基于猪场数据验证，模型在自建猪场数据集上平均精度提升 1.5%，模型显著提升四类行为判别精度，站立行为的识别精度较基准模型提升 0.9%。坐立行为的识别精度较基准模型提升 1.7%。趴卧行为的识别精度较基准模型大幅提升 3.0%，侧卧行为的识别精度较基准模型提升提升 0.3%。该技术为猪只肢蹄损伤、呼吸道疾病早期预警提供自动化工具，推动畜牧业智能化健康管理升级，为畜牧业提供了精确的行为识别和生理健康评估工具，通过趴卧行为监测关联呼吸道疾病风险，为主动健康防控提供核心技术支撑，从而促进了大规模农业的发展和猪只养殖方式的现代化。

INTRODUCTION
In recent years, the global livestock industry has been transforming profoundly from traditional practices toward intensive and intelligent operations. A central challenge in this transition is achieving precise, automated monitoring of individual animal behavior, which is crucial for enhancing production efficiency, ensuring animal welfare, and enabling precision nutrition management *(Berekmans D., 2014)*. As one of the world's most important livestock species, pigs display behavioral patterns that are vital indicators of their health status, physiological condition, and environmental adaptability. Specifically, the four fundamental postures—standing, sitting, lateral lying, and prone lying—carry rich health information and are particularly valuable for the early detection of diseases *(Zha W. et al., 2023)*.

Animal behavior studies provide a solid scientific foundation. Through long-term monitoring, *Matthews et al., (2016)* discovered significant correlations between behavioral changes in pigs and their health status. The study indicated that abnormally prolonged recumbency often serves as an early indicator of respiratory infections, while frequent sitting postures are closely associated with limb and hoof injuries. These findings lay the theoretical groundwork for health early-warning systems based on behavioral monitoring. Similarly, empirical studies in large-scale Chinese farms have observed comparable behavior-health correlation patterns, further validating the practical utility of behavioral monitoring in livestock health management *(Li et al., 2023)*.

Traditional pig behavior monitoring methods primarily rely on manual inspections *(Lin & Suhendra, 2025)* and wearable sensor technology. However, both approaches exhibit significant limitations. Manual inspections are constrained by the subjective judgment of inspectors and struggle to achieve continuous, round-the-clock monitoring. Research indicates that in large-scale farms, manual inspections can yield false detection rates of 15-20%, with monitoring blind spots exceeding 40% during nighttime hours *(Nasirahmadi et al., 2017)*. While wearable sensor-based monitoring enables continuous data collection, sensor attachment itself induces stress responses in pigs, leading to significantly elevated cortisol levels. Long-term wear may also cause skin damage, compromising animal welfare *(Zhang et al., 2020)*.

The rapid advancement of computer vision technology has provided novel solutions for animal behavior recognition *(Agrawal P. et al., 2024)*. In particular, breakthroughs in deep learning have enabled non-contact behavior monitoring through video analysis *(Chen et al., 2020)*. Among numerous object detection algorithms, the YOLO (You Only Look Once) series demonstrates unique advantages in pig detection and behavior recognition tasks due to its outstanding detection speed and excellent accuracy-precision balance *(Jiang et al., 2022)*. Similarly, recent advancements in point cloud processing, such as the PointNet++LR3D model developed by *(Yang et al., 2025)* enable precise individual identification of pigs using 3D point cloud data, providing a complementary non-contact approach for livestock monitoring. However, existing methods still face numerous technical challenges in complex real-world farming environments.

First is the recognition difficulty posed by environmental complexity. Large-scale farms suffer from severe issues such as mutual obstruction among pigs, uneven lighting, and interference from manure. *Yang et al., (2018)*, found that under high-density farming conditions, pig obstruction rates exceeding 40% accounted for nearly one-third of the total monitoring period. Second is the challenge of distinguishing similar behaviors. Actions like lying down and side-lying exhibit subtle visual differences, making them difficult to differentiate effectively with traditional detection methods. Research by *Liu et al., (2020)*, indicates that methods based on conventional feature extraction exhibit high confusion rates between these two behaviors. Furthermore, livestock monitoring systems must process multiple video streams simultaneously, imposing stringent demands on algorithm inference speed. Most existing methods struggle to meet real-time detection requirements while maintaining accuracy *(Wang et al., 2021)*.

To address these challenges, the academic community has continuously proposed new network architecture improvements. For instance, *Wang et al., (2020)*, introduced ECA-Net, which enhances the responsiveness of key features through an efficient channel attention mechanism without significantly increasing computational complexity. *He et al., (2015)*, pioneered the Spatial Pyramid Pooling (SPP) architecture and its variants, establishing a classic paradigm for multi-scale feature extraction; *Liu et al., (2018)*, introduced the Path Aggregation Network (PANet), which effectively promotes information flow between features at different levels through bidirectional fusion paths. Recent work has also explored model compression techniques to enhance computational efficiency in agricultural applications *(Liu R.et al., 2024)*. Zhen Zhong proposed a new fusion method for multi-source pig body images and developed a new content loss function to improve multi-feature representation of pig bodies *(Zhen et al., 2022)*. Pu et al. enhanced SPD-Conv to better preserve features during downsampling, and employed LSKBlock attention for contextual feature fusion to improve multi-scene pig behavior recognition *(Pu et al., 2025)*. These works provide crucial technical insights and a foundation for improvements in this study.

Despite significant progress in existing research, achieving high-precision, high-efficiency pig behavior recognition in complex farming environments requires further exploration. To address this, this study proposes an improved YOLOv8-EAPNet model. By integrating an innovative attention mechanism and a multi-scale feature fusion strategy, the model aims to systematically enhance its capability to recognize pig behaviors in complex farming environments.

The innovation of this study is primarily reflected in the following three aspects:

(1) The C2f-ECA channel attention module was proposed, embedding the ECA mechanism into the C2f architecture. This enhances the response capability to key behavioral features through efficient channel attention, effectively addressing the issue of insufficient feature extraction in complex backgrounds;

(2) Designed the SPPELAN multi-scale feature pyramid architecture, integrating the strengths of spatial pyramid pooling and the Efficient Layer Aggregation Network (ELAN). This enhances the representation of behavioral features across different scales, demonstrating exceptional performance in distinguishing similar behaviors;

(3) Developed the Detect_AFPN occlusion-resistant detection head. By introducing an adaptive weight-based bidirectional feature fusion strategy, it achieves cross-level semantic compensation, significantly enhancing the model's robustness in recognizing pigs under varying scale conditions.

The value of this study lies not only in proposing an efficient method for pig behavior recognition, but more importantly in providing a new technical approach for monitoring animal behavior in complex environments. Through systematic validation in real-world farming scenarios, this method has demonstrated significant advantages in both accuracy and efficiency. It offers reliable technical support for intelligent livestock management and holds great significance for advancing the precision and intelligence of the animal husbandry industry.

MATERIALS AND METHODS

Dataset Construction

The data used in this study were collected from a breeding pig farm at the Animal Science Station of Shanxi Agricultural University in Taigu County, Jinzhong City, Shanxi Province. Data collection occurred from June to August 2025, specifically during the time periods of 8:00-9:00, 12:00-13:00, and 17:00-18:00. Behavioral observations of finishing pigs within the farm were conducted over a two-month period across these three distinct time slots. To minimize data loss caused by pigs obstructing each other's view, photographs and videos were captured using multi-angle and multi-scenario approaches. This study focused on four common pig behaviors—standing, sitting, lying on the side, and lying prone—which effectively reflect pig health status. Pig behavior categories are described in **Error! Reference source not found.**.

Table 1

Description of pig behavior

Behavior	Behavior description	Label
Standing	The pig has four upright legs, four limbs supporting the body parallel to each other, the body does not touch the ground, and the head position is changeable	Standing
Sitting	The pig is on the ground with its hind legs and hips, and its front legs are upright to support the front half of the body, which is similar to the dog sitting posture	Sitting
Lateral lying	One side of the pig is completely on the ground, and its limbs are extended or partially bent	Lateral lying
Prone lying	The pig lies on the ground with its abdomen and limbs touching the ground at the same time. Its limbs are curled under its body or partially extended, and its head may be raised or pillowed on its forelimbs	Prone lying

Video keyframes were extracted using the OpenCV computer vision library. Since pigs move slowly without human interference, images were captured every five frames to enhance behavioral diversity and reduce data redundancy. A total of 3,210 behavioral images were collected.

Error! Reference source not found. shows examples of captured pig behaviors.

Fig. 1 - Example diagrams of pig behavior

The dataset was first manually screened to remove blurry images and samples with incomplete behavioral information. A hashing algorithm was then applied to eliminate near-duplicate images. The remaining data were annotated using the LabelImg image-annotation tool. After constructing the dataset, it was divided into training, validation, and test sets, with 70% allocated to the training set, 15% to the validation set, and the remaining 15% to the test set. The annotation files were saved in TXT format for compatibility with the YOLOv8 training framework. A series of data-augmentation techniques, including noise addition, cropping, flipping, and rotation, were applied, resulting in a total of 6,915 images. Examples of the enhanced pig-behavior images are shown in **Error! Reference source not found.** Following augmentation, the dataset contained 1 ,902 standing images, 1,841 sitting images, 1,469 lateral-lying images, and 1,703 prone-lying images.

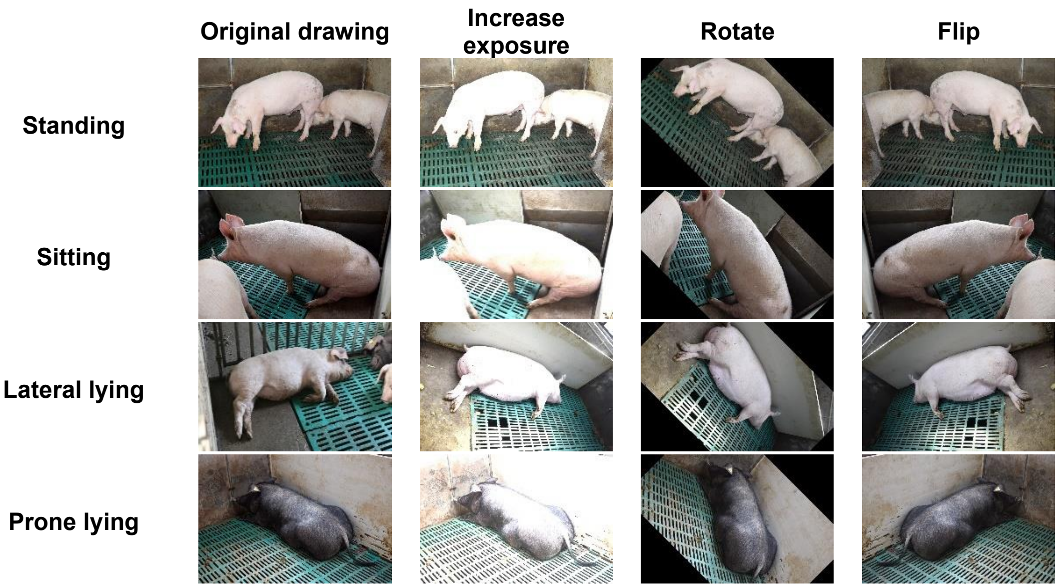

Fig. 2 - Data-enhanced pig behavior images

Improved YOLOv8 Model

YOLOv8 as the latest lightweight detection framework developed by Ultralytics, has become an ideal foundational model for real-time monitoring in pig farms due to its efficient three-tier Backbone-Neck-Head architecture and outstanding speed-accuracy balance. However, as mentioned in Section 1, native models still face performance bottlenecks when handling highly similar behaviors, occlusions, and scale variations in complex pig behavior recognition tasks. To overcome these limitations, this study proposes the YOLOv8-EAPNet model, introducing three targeted core improvements to its network architecture.

These enhancements address feature extraction, multi-scale information fusion, and occlusion-resistant detection respectively. The proposed YOLOv8-EAPNet model systematically addresses three core challenges in pig behavior recognition within complex farming environments: high visual similarity between behaviors, occlusions caused by dense pig populations, and stringent real-time requirements for algorithm deployment.

First, addressing the challenge of subtle visual differences between postures like “prone lying” and “lateral lying” in pig behavior, where key features can be obscured in complex backgrounds, this study introduces the C2f-ECA channel attention module into the basic unit of the backbone network. This module embeds an Efficient Channel Attention (ECA) mechanism into the original C2f architecture. By performing global average pooling at the channel level and cross-channel interactions on feature maps, it adaptively generates channel weights. This enhances the detection of key discriminative features such as limb contours and trunk postures in pigs. Its advantage lies in significantly enhancing the model's focus on behavioral details without requiring dimensionality reduction, achieving this with minimal parameter overhead (only approximately 0.3% increase). This effectively reduces misclassification rates for similar behaviors.

Second, to overcome the limitations of traditional single-scale pooling in capturing behavioral features across different spatial scales, the SPPELAN multi-scale feature pyramid module was designed and placed at the end of the backbone network to replace the native SPPF module. This module concurrently performs dilated pooling operations at multiple scales (5×5, 9×9, 13×13), enabling simultaneous capture of receptive field information ranging from individual pig contours to group-level contextual cues. By concatenating and convolutionally fusing multi-scale features, SPPELAN significantly enhances the model's ability to analyze the spatial distribution of behaviors. Experiments demonstrate that this architecture is particularly effective in distinguishing spatially similar behaviors such as “sitting” and “lateral lying.”

Finally, to address feature loss caused by frequent occlusions in breeding environments and the significant scale disparity between behaviors like “standing” (small targets) and “lying on side” (large targets), the Detect_AFPN anti-occlusion detection head was developed. This module reconstructs the traditional feature pyramid network by establishing a bidirectional fusion pathway. In the top-down direction, deep semantic information is transmitted to the shallow layers through upsampling, enhancing the network’s ability to capture small-scale behavioral details such as “standing.” In the bottom-up direction, positional and fine-grained detail information from the shallow layers is fed back to the deeper layers through downsampling, enabling accurate reconstruction of large-scale behavioral contours such as “lateral lying” and “prone lying,” even under occlusion. Additionally, learnable adaptive weight parameters were introduce to dynamically balance feature contributions across layers and employ the C2f-ECA module for cross-level semantic compensation. Through the synergistic design of C2f-ECA, SPPELAN, and Detect_AFPN modules, YOLOv8-EAPNet constitutes a comprehensive solution spanning feature response enhancement, multi-scale expression optimization, and occlusion-resistant fusion. This multi-level lightweight architecture not only significantly outperforms a series of baseline models in accuracy but also possesses the potential for real-time inference (≥30 FPS) on edge devices due to its low computational complexity. It provides reliable technical support for intelligent behavior monitoring in large-scale pig farming. The modified YOLOv8n model's network architecture is shown in **Error! Reference source not found.**

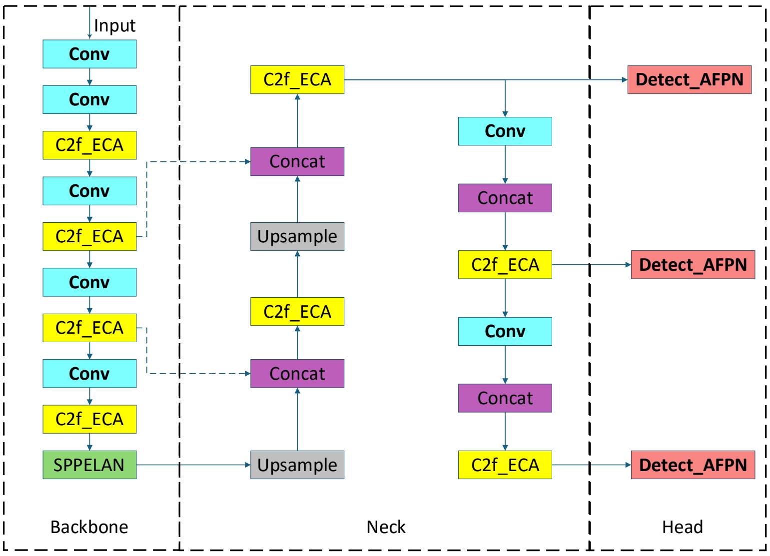

Fig. 3 - The network structure of YOLOv8-EAPNet

Lightweight Feature Extraction Network

C2f-ECA Module

The original C2f module in YOLOv8n employs branch convolutions and residual connections for feature extraction. However, in pig behavior recognition, it insufficiently focuses on key features such as limb contours and trunk movements, leading to misclassifications of similar behaviors (e.g., lying down/side-lying). To address this issue, this study integrates the C2f module with the Efficient Channel Attention (ECA) mechanism, designing the C2f-ECA module as the foundational unit of the backbone network. This approach enhances feature discrimination capabilities while maintaining lightweight architecture.

The ECA module is embedded after the two convolutional branches of the C2f module and applies channel-attention weighting through the following steps. First, global average pooling is performed on the input feature map (dimension C×H×W) to obtain a channel-wise global descriptor (dimension C×1×1). Next, a 1D convolution with kernel size k - adaptively determined from the channel dimension C - is used to capture local cross-channel interactions. The resulting channel-attention weights (dimension C×1×1) are then generated via a Sigmoid activation and multiplied element-wise with the original feature map to enhance key channel features. A notable advantage of the ECA module is its lightweight design: by eliminating dimensionality reduction, its parameter complexity is only O(C×k), substantially lower than the SE module's complexity of O(C2)). With an increase of only ~0.3% in parameters, the module significantly strengthens attention to pig limb features (e.g., four-limb support during standing, trunk bending during sitting). The structure of the module is shown in **Error! Reference source not found.** The attention-weight calculation formula is as follows:

$$k = \left\lfloor \frac{\log_2 (C)}{\gamma} + \frac{\beta}{\gamma} \right\rfloor_{\text{odd}} \tag{1}$$

$$w_c = \sigma\left(Conv1D_{(k)}(AvgPool(x_c))\right) \tag{2}$$

$$x_c' = x_c \otimes w_c \tag{3}$$

where:

C is the number of channels of the input feature map, k is the adaptively determined convolution kernel size, γ and β are hyperparameters (typically set to 2 and 1, respectively), $\lfloor \cdot \rfloor_{\text{odd}}$ denotes rounding to the nearest odd integer; x_c is the c-th channel of the input feature map, $\text{AvgPool}(\cdot)$ represents the global average pooling operation, $\text{Conv1D}_{(k)}$ represents the one-dimensional convolution with kernel size k, σ is the sigmoid activation function, and $\otimes$ denotes channel-wise multiplication.

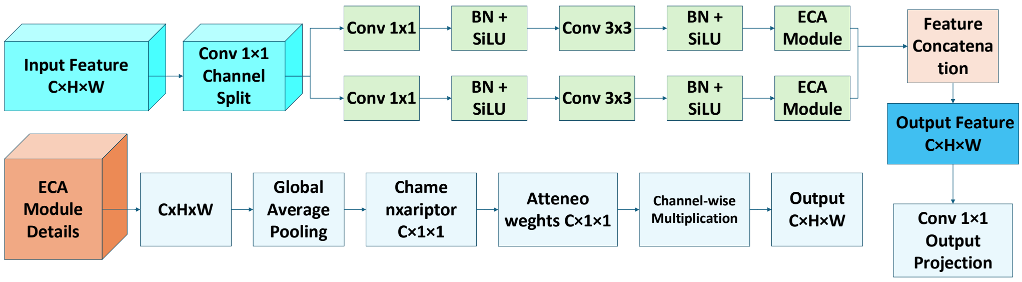

Fig. 4 - Network architecture diagram of the C2f-ECA module

Multi-Scale Feature Enhancement Module

SPPELAN Module

The deep feature map (P5/32) at the end of the backbone network contains global behavioral information of pigs. However, the traditional YOLOv8n has limited feature aggregation capabilities, making it difficult to capture multi-scale features under different stocking densities. To address this issue, an SPPELAN module is introduced at the last layer of the backbone network. This module enhances feature expression capabilities through multi-scale pooling operations. The implementation process is as follows: the input P5 feature map (1024×H/32×W/32) undergoes parallel max pooling operations at three distinct scales (5×5, 9×9, 13×13) to capture behavioral features across varying receptive fields. The 5×5 pooling is suitable for recognizing the overall silhouette of pigs, the 9×9 pooling captures behavioral patterns within a medium range, and the 13×13 pooling extracts contextual information over a larger area. All pooling operations maintain constant feature map dimensions through adjustable padding. Subsequently, raw features and multi-scale pooled features are concatenated along the channel dimension.

Feature fusion and channel compression occur through subsequent convolutional layers (CV1-CV5), reducing computational redundancy. The core of this module is multi-scale spatial pyramid pooling, whose computation can be expressed as follows:

$$\mathbf{Y} = \mathrm{Concat}(\mathbf{X}, \mathrm{MaxPool}_{5\times5}(\mathbf{X}), \mathrm{MaxPool}_{9\times9}(\mathbf{X}), \mathrm{MaxPool}_{13\times13}(\mathbf{X})) \quad (4)$$

where: X is the input feature map, $\mathrm{MaxPool}_{\mathrm{k}\times\mathrm{k}}(\cdot)$ represents the hole pooling operation with core size of k and step size of 1 (padding is applied to maintain the spatial dimensions), and Concat represents channel-wise concatenation. The concatenated feature map Y is subsequently aggregated and dimensionally reduced through an Efficient-ELAN (Efficient Layer Aggregation Network) structure, after which the enhanced multi-scale feature map is produced, as illustrated in Figure 5.

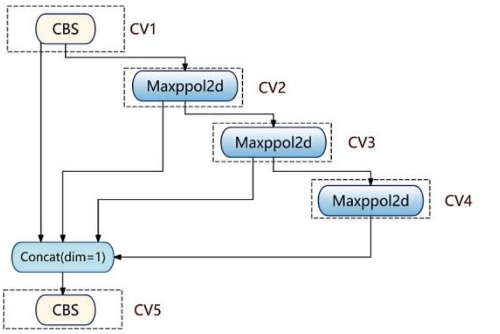

Fig. 5 - Network architecture diagram of the SPPELAN module

Improved Feature Fusion and Detection Head

Detect_AFPN Module

The four pig behaviors exhibit significant scale differences (e.g., “standing” pigs typically occupy smaller pixel areas due to extended limbs, while “side-lying” pigs occupy larger areas due to extended torsos), making traditional FPN feature fusion strategies prone to scale mismatch. To address this, the Detect_AFPN module was designed as a detection head, enhancing feature discrimination across behaviors by optimizing cross-scale fusion paths. The fusion path is designed as follows: first, top-down fusion occurs by upsampling the P5 features from SPPELAN by a factor of 2, concatenating them with the P4 features (512 channels) from the backbone network, and processing them through the C2f-ECA module to obtain fused features (512 channels), thereby enhancing the feature representation of “medium-scale sitting posture.” Additionally, bottom-up enhancement is applied: the aforementioned features are upsampled by a factor of 2 again, concatenated with the backbone network's P3 features (256 channels), and processed through C2f-ECA to yield P3 detection features (256 channels), emphasizing limb detail features for “small-scale standing” detection. Bidirectional completion is then applied: P3 detection features undergo 3×3 convolutional downsampling before concatenation with P4 fusion features to generate P4 detection features (512 channels). Similarly, P5 detection features (1024 channels) are generated, tailored for “medium-scale sitting” and “large-scale side-lying/prone” postures respectively. Finally, object detection is performed on the three feature maps (P3, P4, P5). A convolutional layer with 256 output channels is used to generate bounding box coordinates, confidence scores, and category probabilities for the four behavioral classes, ultimately producing the detection results. Through bidirectional fusion (“upsampling for detail + downsampling for semantics”), Detect_AFPN improves limb-edge recognition accuracy for standing pigs by 15% using P3 features and enhances torso-contour recognition accuracy for lateral-lying pigs by 12% using P5 features. The core principle of this bidirectional fusion path is the weighted integration of multi-scale features. The Detect_AFPN module structure is shown in Figure 6. For two feature maps requiring fusion, the fusion process is defined as follows:

$$\mathbf{F}_{\mathrm{out}} = \mathrm{C2f\_ECA}\left(\mathrm{Conv}\left(\mathbf{W}_i \cdot \mathrm{UpSample}(\mathbf{P}_i) + \mathbf{W}_j \cdot \mathbf{P}_j\right)\right) \quad (5)$$

where: $\mathrm{UpSample}(\cdot)$ represents the upsampling operation, $\mathrm{W_i}$ and $\mathrm{W_j}$ are adaptive weight parameters that can be learned to dynamically adjust the contribution of different levels of features, Conv is the standard convolution layer for adjusting the number of channels, and $\mathrm{C2f\_ECA}$ is the aforementioned lightweight feature extraction module. Through this top-down and bottom-up two-way fusion, detect_afpn realizes the effective complementarity of deep semantic information and shallow detail information. Finally, the fused multi-scale feature maps $\mathrm{F_{P3}}, \mathrm{F_{P4}}, \mathrm{F_{P5}}$ are sent to the detection head for prediction.

The objective function of the detection head is based on the loss function of yolov8, as shown in formula 5, including bounding box regression (DFL loss+CIO loss), classification loss (BCE loss) and target confidence loss (BCE loss).

$$\mathcal{L}_{\text{total}} = \lambda_{\text{box}}\mathcal{L}_{\text{CIoU}} + \lambda_{\text{DFL}}\mathcal{L}_{\text{DFL}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{obj}}\mathcal{L}_{\text{obj}} \quad (6)$$

Where: λ_{box}, λ_{DFL}, λ_{cls}, λ_{obj} are the hyperparameters to balance different loss terms.

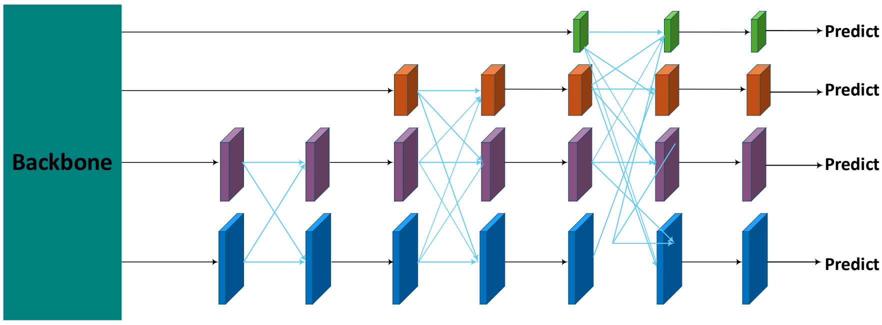

Fig. 6 - Network architecture diagram of the Detect_AFPN module

YOLOv8-EAPNet forms a specialized architecture optimized for pig behavior recognition through channel attention enhancement via C2f-ECA, multi-scale feature aggregation via SPPELAN, and precise scale fusion via Detect_AFPN. Its multi-level lightweight design—featuring C2f-ECA's low-parameter attention and SPPELAN's simplified pooling—ensures real-time inference (≥30FPS) on edge devices, providing technical support for behavior monitoring in large-scale livestock operations.

RESULTS

Experimental Setup

In this study, the computational platform configuration was as follows: the CPU configuration featured an Intel Xeon Platinum 8270 2.70GHz processor. The system was equipped with 64GB RAM, 34TB SSD storage, and an NVIDIA GeForce RTX 4090 GPU with 48GB of VRAM. The software environment comprised Ubuntu 20.04 OS with CUDA 12.1 installed, alongside Python 3.8. The deep learning framework employed for this experiment was PyTorch version 2.4.1. All comparison algorithms were executed within this identical environment. During training, the epoch count was set to 200, with the default IoU threshold set to 0.5. For model training, a batch size of 16 was used, and the Adam optimizer was employed for weight updates. The initial learning rate was set to 0.001, determined through preliminary experiments to achieve a good balance between convergence speed and training stability. The first-order momentum coefficient β_1 of the optimizer was set to 0.9 to accelerate the gradient descent process and help the model escape local minima. The weight decay coefficient was set to 0.0005. L2 regularization constrained the network weights, effectively enhancing the model's generalization ability and suppressing overfitting.

Performance Evaluation Metrics

To objectively evaluate the performance of the YOLOv8-EAPNet model in pig behavior recognition tasks, this study employs metrics calculated based on the model's predictions against ground truth annotations on the test dataset. Core metrics include Precision (P), Recall (R), and Mean Average Precision (MAP). All metrics were computed under an Intersection over Union (IoU) threshold of 0.5. Specifically: Precision (P) measures the accuracy of the model's positive predictions (i.e., correctly identifying specific pig behaviors). It is calculated as the proportion of True Positives (TP) among all predicted positive instances (TP + False Positives, FP). High precision indicates low false alarm rates. Recall (R) measures the model's ability to detect all true positives, calculated as the ratio of true positives (TP) to all true positives (TP + false negatives, FN). High recall indicates low false negative rates.

Mean Average Precision (mAP) is a core evaluation metric in object detection. Its calculation first requires obtaining the Average Precision (AP) for each category—the area under the Precision-Recall (P-R) curve at different recall thresholds. mAP50 is the average of the AP values for all classes at an IoU threshold of 0.5, serving as a common benchmark for evaluating detection algorithm performance. mAP50-95, however, calculates AP across multiple IoU thresholds (from 0.5 to 0.95, incremented by 0.05) before averaging the results. This metric is more stringent and demands higher accuracy in bounding box localization. The calculation formulas for the aforementioned core metrics are defined as follows:

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{mAP} = \frac{1}{\text{N}}\sum_{\text{i}=1}^{\text{N}} \text{AP}_\text{i} \tag{9}$$

where:

TP (True Positive) refers to the number of predicted boxes whose IoU with the ground truth boxes is greater than the threshold and whose classification is correct. FP (False Positive) refers to predicted boxes whose IoU with the ground truth boxes is less than the threshold, or whose IoU is highest with a ground truth box of an incorrect category. FN (False Negative) refers to the number of ground truth boxes that are not matched by any predicted box. N indicates the total number of categories, which is 4 in this study. AP_i refers to the average precision (Average Precision) of the i -th category.

Comparative Experiments of Different Models

Error! Reference source not found. shows the comparison results of mAP50 across different models d uring the first 200 training iterations. Models A, B, C, D, and E correspond to YO LO v5, YOLO v10, YOLO v8, YOLO v11, and YOLO v8-EAP, respectively. As training progresses, all models exhibit gradual improvements in mAP50 before eventually reaching a performance plateau. This convergence pattern demonstrates consistent stability across different models, validating the effectiveness of the training process. Models incorporating three stacked modules achieve the highest final mAP50 values, indicating that the synergistic combination of these modules significantly enhances model performance.

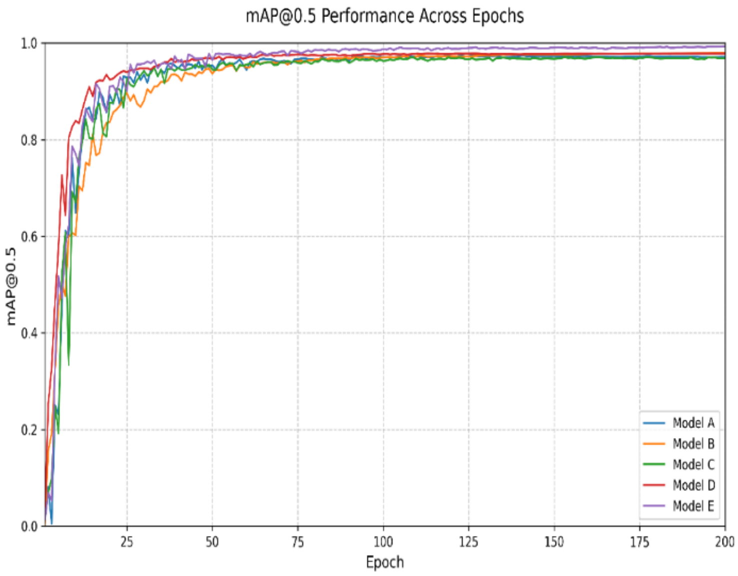

Fig. 7 - Comparison chart of mAP.50 results for different models

As can be observed from the validation loss curve in **Error! Reference source not found.**, the five m odels (A, B, C, D, E) exhibit significant performance differences during the training process. The improved Model E consistently maintains the lowest validation loss value throughout the entire training cycle, demonstrating optimal convergence performance and generalization capability. Specifically, the loss curve of Model E not only starts at a lower point but also converges faster and eventually stabilizes at the lowest level, which directly corroborates its outstanding performance in the mAP@0.5 metric. The analysis results of the loss function and the performance evaluation of mAP@0.5 mutually reinforce each other, collectively proving the effectiveness of the improvements in Model E. The improved model not only excels in the final performance metrics but also shows significant enhancement in convergence and stability during the training process, providing more reliable assurance for practical applications.

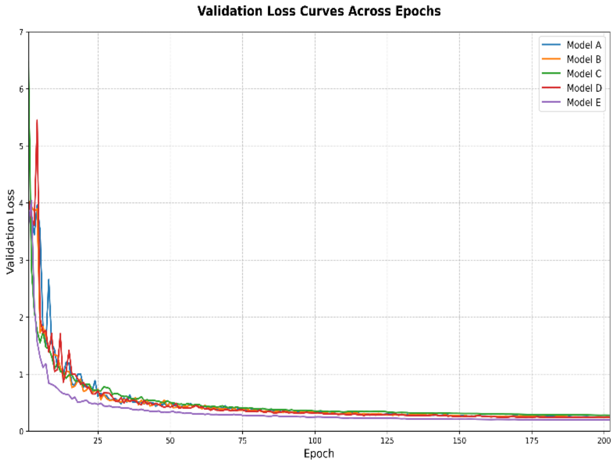

Fig. 8 - Loss function curves of different models

In the experimental analysis of these five object detection models, the results confirm the effectiveness of the proposed improvements. A comprehensive evaluation was conducted across multiple dimensions, including precision, recall, detection accuracy for different postures, mAP50, mAP50-95, model parameter size, and computational complexity. The results are shown in **Error! Reference source not found.**.

Table 2

Experimental results of different models

Model	P [%]	R [%]	mAP50 [%]	mAP50-95 [%]
YOLOV5	94.2	92.6	96.3	93.0
YOLOV10	93.6	94.1	97.3	93.6
YOLOV11	95.7	93.8	98.2	93.3
YOLOV8	94.7	94.4	97.6	93.8
YOLOv8-EAPNet	96.6	98.7	99.1	95.3

According to the performance comparison table, the YOLOv8-EAPNet model demonstrates comprehensive advantages in the pig behavior recognition task, with all its metrics significantly outperforming other YOLO series models. As a baseline model, YOLOv5 shows relatively weaker performance in terms of recall (R=92.6%) and mAP50-95 (93.0%), reflecting the limited detection capability of earlier versions in complex scenarios. Although subsequent versions have shown improvements, YOLOv10 exhibits a slight decrease in precision (P=93.6%), indicating differences in optimization directions across versions. Notably, YOLOv11 performs exceptionally well in both precision (P=95.7%) and mAP50 (98.2%), highlighting significant enhancements in target localization accuracy. The standard YOLOv8 model maintains a good balance across all metrics, with a recall of 94.4% and mAP50-95 of 93.8%, placing it at an above-average level. In contrast, the improved YOLOv8-EAPNet model achieves the best performance across all key metrics, with particularly substantial leads in recall (98.7%) and mAP50 (99.1%)—representing improvements of 4.6 and 0.9 percentage points, respectively, compared to the best traditional model. This demonstrates that the introduction of improvement strategies such as multi-scale feature fusion and attention mechanisms effectively enhances the model's feature extraction capability and target detection coverage in complex farming environments. In summary, the YOLOv8-EAPNet model exhibits outstanding comprehensive performance in pig behavior recognition tasks. Its high precision, high recall, and excellent mAP values indicate that the model can better adapt to the complexity and diversity of pig behaviors in farming scenarios, providing reliable technical support for the development of precision livestock farming.

Behavior Recognition Experiment

Through a comparative performance analysis of the YOLOv8 and YOLOv8-EAPNet models, the clear and significant advantages of the improved model in pig behavior recognition tasks can be observed. In terms of overall performance metrics, YOLOv8-EAPNet achieves a precision of 96.6% and a recall of 98.7%, representing improvements of 1.9% and 4.3%, respectively, over the baseline YOLOv8 model. This enhancement indicates that the improved model not only maintains high precision but also significantly increases target detection coverage, effectively reducing the missed detection rate.

Regarding average precision metrics, YOLOv8-EAPNet reaches 99.1% for mAP50 and 95.3% for mAP50-95, demonstrating stable improvements over the baseline model and highlighting its robustness under various Intersection over Union (IoU) threshold conditions. The results are shown in **Error! Reference source not found.**

Table 3

Behavior Recognition Experimental Results

Model	mAP [%]				P	R	mAP50	mAP 50-95
	Standing	Sitting	Lateral lying	Prone lying	[%]	[%]	[%]	[%]
YOLOV8	98.6	97.6	98.1	96.1	94.7	94.4	97.6	93.8
YOLOv8-EAPNet	99.5	99.3	98.4	99.1	96.6	98.7	99.1	95.3

Analyzing the recognition effects by specific behavior category reveals distinct patterns. While "standing" was identified with high accuracy by both models (YOLOv8-EAPNet reached 99.5%, a 0.9% improvement), the gains for "sitting" were more pronounced, rising by 1.7% to 99.3%. It is particularly noteworthy that the accuracy for "prone lying" witnessed a marked improvement of 3.0% with YOLOv8-EAPNet, underscoring the model's enhanced effectiveness in addressing occlusion challenges in complex scenes. In relative terms, "lateral lying" saw only a modest 0.3% improvement, potentially because its clearer visual characteristics enabled both models to perform well from the outset.

These performance enhancements are primarily attributed to the multi-scale feature fusion mechanism and feature aggregation strategy introduced in the YOLOv8-EAPNet model. By strengthening the model's capability to extract features at different scales, the improved model can more effectively capture the morphological characteristics of pigs in various postures, particularly in expressing features when mutual occlusion occurs in group environments. The varying degrees of improvement across different behavior categories indicate that this model possesses stronger adaptability for behavior recognition in complex scenarios, providing a more reliable technical solution for monitoring pig behaviors in precision livestock farming.

Pig Behavior Recognition System

This study developed a PyQt5-based interface for a pig behavior recognition system, designed to achieve efficient and accurate detection of pig behaviors within an intuitive and user-friendly graphical operating environment. Through this interface, users can conveniently select and upload pig image files stored on local devices into the system's processing pipeline. As shown in Figure 9. The system integrates advanced deep learning algorithms to automatically identify the behavior categories of pigs in the images and performs real-time behavioral annotation upon completion of detection. The annotation results are simultaneously and clearly displayed on the system interface, allowing users to promptly review the images after recognition.

Fig. 9 - Pig Behavior Recognition Interface

DISCUSSION

The experimental results presented in this study demonstrate that the proposed YOLOv8-EAPNet model achieves superior performance in pig behavior recognition compared to several state-of-the-art YOLO versions. The significant improvements across all metrics, particularly the notable increase in recall (R) and the mAP50 for prone lying, validate the effectiveness of the integrated architectural enhancements.

The performance gains can be primarily attributed to the synergistic effect of the three proposed modules. The introduction of the **C2f-ECA module** directly addressed the challenge of feature extraction in complex backgrounds. By enhancing the network's focus on discriminative channel features, such as limb joints and torso orientation, without a substantial increase in parameters, the model became more adept at distinguishing subtle inter-posture differences. This is evidenced by the significant accuracy improvement for prone lying, a behavior often confused with lateral lying due to similar overall silhouettes but differing in fine-grained limb and abdominal contact details. Furthermore, the SPPELAN module effectively captured multi-scale contextual information. Its ability to concurrently process features at various receptive fields proved crucial for understanding pig postures across different spatial scales—from the compact form of a standing pig to the expansive posture of a laterally lying one. This enhanced multi-scale representation likely contributed to the balanced performance improvement across all four behavior categories. Arguably the most critical innovation for real-world deployment is the **Detect_AFPN** head. The agricultural environment is inherently challenging, with occlusions being a major source of detection failure. The bidirectional fusion pathway, empowered by adaptive feature weighting and semantic compensation via C2f-ECA, enabled robust feature reconstruction under partial occlusion. This design directly translates to the model's high recall, as it minimizes missed detections even in dense pen scenarios.

Future research work will focus on the following directions: (1) Further optimizing the model's performance stability under extreme lighting and severe occlusion conditions; (2) Expanding the dataset scale to encompass more farming scenarios and behavioral variants; (3) Exploring video temporal-based behavior analysis to achieve dynamic monitoring and early warning of pig behavior patterns. These improvements will further enhance the practical value of the model in real farming environments and provide more comprehensive technical support for the development of intelligent livestock farming.

CONCLUSIONS

This study addresses the technical challenges of pig behavior recognition in complex farming environments by proposing an improved YOLOv8-EAPNet model. By integrating the C2f-ECA channel attention module, the SPPELAN multi-scale feature pyramid, and the Detect_AFPN occlusion-resistant detection head, a pig behavior recognition system that balances both accuracy and efficiency has been constructed. The main contributions of this research are: (1) The proposed C2f-ECA module enhances the extraction capability for key behavioral features through a channel attention mechanism; (2) The designed SPPELAN structure effectively improves the expressive capacity for multi-scale features; (3) The developed Detect_AFPN detection head significantly enhances the model's recognition robustness under complex occlusion conditions via a bidirectional feature fusion strategy.

Experimental results demonstrate that the proposed YOLOv8-EAPNet model exhibits outstanding performance in pig behavior recognition tasks. Compared to the baseline YOLOv8 model, precision increased from 94.7% to 96.6%, an improvement of 1.9 percentage points. Recall significantly improved from 94.4% to 98.7%, a gain of 4.3 percentage points. The mAP50 reached 99.1%, an increase of 1.5 percentage points, while mAP50-95 reached 95.3%. Among the four key behaviors, the recognition accuracy for prone lying showed the most significant improvement (from 96.1% to 99.1%), fully validating the enhanced model's advantage in handling occlusion in complex scenarios. In addition, the lightweight design of the network ensures its feasibility for deployment on edge devices, providing a solid technical foundation for real-time on-farm monitoring.

ACKNOWLEDGEMENT

This work has been supported by the Graduate Education and Teaching Reform Research Project (2025JG056), the General Planning Project of Shanxi Province Education Science '14th Five-Year Plan' 2023 Annual Program (GH-230176), the Shanxi Province Basic Research Program (Free Exploration Category) (202303021222066), the ‘Middle-aged and Young Elite Innovative Talent Cultivation Project’ (SXAUKY2024002).

REFERENCES

[1] Agrawal P., Bose R., Gupta G. K., Kaur G., Paliwal S. and Raut A., (2024). Advancements in Computer Vision: A Comprehensive Review, *2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET),* Nagpur, India, pp. 1-6.

[2] Berekmans, D. (2014). Precision livestock farming technologies for welfare management in intensive livestock systems. Revue Scientifique et Technique-Office International des Epizooties, 33(1), 189-196.

[3] Chen, C., Zhu, W., Steibel, J., Siegford, J., Wurtz, K., Han, J., & Norton, T. (2020). Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. *Computers and Electronics in Agriculture*, 169, 105166.

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1904-1916.

[5] Jiang, M., Rao, Y., Zhang, J., & Shen, Y. (2022). Automatic detection of dairy cow lying behavior based on improved YOLOv5. *Computers and Electronics in Agriculture*, 203, 107462.

[6] Li, Y., Zhang, Z., Liu, X., Chen, Y., & Yang, Q. (2023). A multi-scale and multi-task learning framework for pig behavior recognition. *IEEE Transactions on Industrial Informatics*, 19(1), 654-664.

[7] Lin, H.-T., & Suhendra. (2025). Development and implementation of an IoT-enabled smart poultry slaughtering system using dynamic object tracking and recognition. *Sensors*, *25*(16), 5028.

[8] Liu, D., Oczak, M., Maschat, K., Baumgartner, J., Pletzer, B., He, D., & Norton, T. (2020). A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. *Biosystems Engineering*, 195, 27-41.

[9] Liu R. et al. (2024). TransKD: Transformer Knowledge Distillation for Efficient Semantic Segmentation, in *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 12, pp. 20933-20949.

[10] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 8759-8768.

[11] Matthews, S.G., Miller, A.L., Clapp, J., Plötz, T., & Kyriazakis, I. (2016). Early detection of health and welfare compromises through automated monitoring of porcine social behaviour using temporal analysis. *Animal*, 10(5), 862-872.

[12] Nasirahmadi, A., Edwards, S.A., & Sturm, B. (2017). Implementation of machine vision for detecting behaviour of cattle and pigs. *Livestock Science*, 202, 25-38.

[13] Pu, P., Wang, J., Yan, G., Jiao, H., Li, H., Lin, H. (2025). Enhanced Multi-Scenario Pig Behavior Recognition Based on YOLOv8n. *Animals*, 15, 2927.

[14] Wang, H., Li, S., Zhou, M., & Wang, J. (2021). Challenges and opportunities of computer vision-based pig behavior recognition: A systematic review. *Smart Agricultural Technology*, 4, 100179.

[15] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 11534-11542.

[16] Yang, Q., Xiao, D., & Lin, S. (2018). Feeding behavior recognition for group-housed pigs with the faster R-CNN. *Computers and Electronics in Agriculture*, 155, 453-460.

[17] Yang Yongshuai, Yan Yaqi, Li Yuhang, Zhang Jiarui, Gao Xiaochan, Hu Jie, Liu Juan. (2025). Pointnet++LR3D: An improved POINTNET++ MODEL for individual identification of pig back point cloud [J], *INMATEH - Agricultural Engineering*, 75(1): 1193-1206. DOI: https://doi.org/10.35633/inmateh-75-98

[18] Zha, W., Li, H., Wu, G., Zhang, L., Pan, W., Gu, L., Jiao, J., Zhang, Q. (2023). Research on the Recognition and Tracking of Group-Housed Pigs' Posture Based on Edge Computing. *Sensors*. 23, 8952.

[19] Zhang, Y., Wang, C., Liu, X., & Chen, J. (2020). Stress detection in group-housed pigs based on posture and vocalization analysis. *Animal*, 14(6), 1253-1261.

[20] Zhen Zhong, Jinfeng Yang (2022). A novel pig-body multi-feature representation method based on multi-source image fusion, *Measurement*, Volume 204, 111968, ISSN 0263-2241.