WFE-YOLO: A LIGHTWEIGHT PIG BEHAVIOR DETECTION MODEL FOR LIVESTOCK FARMING APPLICATIONS

/

WFE-YOLO: 面向养殖场景的轻量化猪只行为检测模型

Jia LV, Guangjie WANG *), Mengfan ZHANG, Fuzhong LI *) [1]
College of Software, Shanxi Agricultural University, Taigu, Shanxi / China
Tel: +8603546287093; E-mail: lifuzhong@sxau.edu.cn
Corresponding author: Fuzhong LI

DOI: https://doi.org/10.35633/inmateh-78-99

Keywords: *Pig behavior detection; lightweight object detection; WFE-YOLO; fine-grained behavior recognition; smart livestock farming*

ABSTRACT

In a real-life breeding environment, fine-grained pig behavior detection is of great significance for health assessment, welfare monitoring, and intelligent management. However, issues such as dense target distribution, severe occlusion, subtle inter-class differences, class imbalance, and limited computing resources make it difficult to achieve both detection accuracy and computational efficiency. To address these problems, this paper proposes a lightweight pig behavior detection model WFE-YOLO based on YOLOv11, which is used to identify five typical behaviors: standing, lying down, eating, drinking, and chewing. This method conducts collaborative optimization at three levels: training sample distribution, feature representation, and detection head design. Specifically, a weighted sampling strategy is adopted to enhance the learning sufficiency of low-frequency behaviors; a lightweight gated feature extraction module is introduced to improve the fine-grained representation ability; an efficient detection head is designed to reduce structural redundancy and computational overhead. Experimental results show that on the data set of this paper, the Precision, Recall, and mAP@50 of WFE-YOLO reach 0.8154, 0.7803, and 0.8233 respectively; compared with YOLOv11n, the parameter size is reduced from 2.58M to 1.96M, GFLOPs is reduced from 6.3G to 4.4G, and FPS reaches 520.43. Under the experimental settings adopted in this paper, compared with several mainstream YOLO models, WFE-YOLO demonstrates a better balance between detection performance and model complexity, especially in low-frequency and easily confused behaviors such as Drink and Bite, and has a greater advantage. These results indicate that WFE-YOLO provides a lightweight, application-oriented solution for pig behavior monitoring in complex breeding environments, with strong potential for deployment on edge devices.

摘要

在真实养殖环境中，细粒度猪行为检测对于健康评估、福利监测和智能化管理具有重要意义。然而，目标密集分布、严重遮挡、类间差异细微、类别不平衡以及计算资源受限，使得检测精度与计算效率难以兼顾。为解决这些问题，本文提出了一种基于 YOLOv11 的轻量化猪行为检测模型 WFE-YOLO，用于识别五类典型行为：站立、躺卧、进食、饮水和啃咬。该方法从训练样本分布、特征表示和检测头设计三个层面进行协同优化。具体而言，采用加权采样策略增强低频行为的学习充分性；引入轻量化门控特征提取模块以提升细粒度表示能力；设计高效检测头以减少结构冗余和计算开销。实验结果表明，在本文数据集上，WFE-YOLO 的 Precision、Recall 和 mAP@50 分别达到 0.8154、0.7803 和 0.8233；与 YOLOv11n 相比，参数量由 2.58M 降至 1.96M，GFLOPs 由 6.3G 降至 4.4G，FPS 达到 520.43。在本文采用的实验设置下，与多种主流 YOLO 模型相比，WFE-YOLO 在检测性能与模型复杂度之间表现出较优的平衡，尤其在 Drink 和 Bite 等低频且易混淆行为上更具优势。这些结果表明，WFE-YOLO 为复杂养殖环境下的猪行为监测提供了轻量化、面向应用的技术参考，并且具备终端部署的潜力。

[1] *Jia LV Prof. Ph.D. Eng.; Guangjie WANG, M.S. Stud. Agr.; Mengfan ZHANG, M.S. Stud. Eng.*

INTRODUCTION

With the continuous development of large-scale, intensive farming models, the relationship between swine health, welfare, and production management efficiency has garnered increasing attention (*Maes et al., 2020; Hu et al., 2022*).

Swine behavior serves as a key indicator reflecting their activity levels, physiological needs, environmental adaptability, and potential abnormalities, playing a crucial role in health assessment, welfare monitoring, feeding management, and disease early warning (*Pandey et al., 2021; Matthews et al., 2016*). Compared to traditional methods that rely on manual inspections, computer vision-based automatic behavior recognition offers advantages such as non-contact, continuous monitoring, objectivity, and scalability, providing more stable and efficient technical support for modern smart farming. Therefore, developing high-precision, lightweight pig behavior detection models suitable for complex pig barn environments has become a key research direction in the field of precision farming.

In pig behavior monitoring tasks, the selection of target behavior categories not only affects the validity of model training and evaluation but also determines the practical application value of the research results. This study defines five behavioral categories—Stand, Lie down, Eat, Drink, and Bite—as target behavioral categories, primarily because these behaviors are highly representative in pig condition assessment and farming management. Specifically, Stand and Lie down, as fundamental postural behaviors indicating a pig's active and resting states, effectively reflect an individual's behavioral rhythms and basic activity levels; therefore, they are considered the most common and fundamental observation units in pig behavior analysis (*Wemelsfelder et al., 2000*); "Eat" and "Drink" are core intake behaviors essential for maintaining individual growth, metabolism, and physiological homeostasis; their frequency, duration, and abnormal changes are typically closely related to nutrient intake, health status, and environmental adaptation (*Roura et al., 2016*); "Bite" is a negative social behavior in group-housed environments that holds management and early-warning value, often associated with resource competition, stress responses, imbalanced group dynamics, and welfare risks (*Canario et al., 2020*). These five behavioral categories correspond to three practical needs: baseline status monitoring, intake behavior monitoring, and abnormal behavior early warning. They provide a relatively comprehensive reflection of pigs' behavioral states within complex farming environments, thereby holding significant biological relevance and practical monitoring value.

In recent years, advancements in deep learning have significantly driven progress in animal behavior recognition research, particularly with the expanding application of object detection techniques in pig monitoring. Existing research primarily focuses on individual pig detection, identity tracking, pose recognition, and behavioral analysis, and has already achieved some progress in typical behavioral recognition tasks such as standing, lying down, feeding, and aggression. As a representative of single-stage object detection methods, the YOLO series of models has been widely used in pig behavior detection tasks due to its end-to-end architecture, high detection efficiency, and good real-time performance (*Sukkuea et al., 2025*). Relevant studies indicate that YOLO-based methods can, to a certain extent, meet the real-time detection requirements in livestock farming scenarios and maintain good detection capabilities in complex backgrounds (*Ali et al., 2024*). However, existing methods still face multiple challenges in actual pig barn settings: First, in group-housing environments, pigs are densely distributed, and severe occlusion between targets can easily lead to missed or false detections; second, there are minimal visual differences between different behaviors, while the same behavior exhibits significant postural variations across different individuals and life stages, making it difficult to stably characterize fine-grained behavioral features; third, there are relatively few samples of low-frequency behaviors such as drinking and biting/fighting, and the uneven distribution of categories can easily cause the model to be biased toward high-frequency categories during training; fourth, livestock farming scenarios typically impose high demands on model inference speed, parameter scale, and deployment costs, limiting the application of complex detection models on edge devices. Consequently, how to balance detection accuracy, model robustness, and deployment feasibility in complex pig barn environments remains an urgent issue in current pig behavior detection research.

To address these issues, existing research typically enhances model performance by focusing on sample reweighting and training bias mitigation, improvements to feature extraction networks, occlusion modeling, and lightweight architecture design. Krasanakis et al. improved the model's learning process through sample reweighting and training bias mitigation by proposing an adaptive sensitive reweighting strategy (*Krasanakis et al., 2018*).

Cai et al. (2023) developed EfficientViT, introducing a lightweight multi-scale attention mechanism to enhance high-resolution dense prediction performance by improving the feature extraction network and controlling computational complexity. These works provide valuable insights for pig behavior detection, but most focus on optimizing a single aspect, with insufficient consideration of the synergistic relationship among training sample distribution, feature modeling capabilities, and computational overhead at the prediction stage.

Consequently, in real-world pig barn environments characterized by dense objects, severe occlusions, subtle behavioral differences, and imbalanced class distributions, existing methods still struggle to achieve a unified balance between high accuracy and low computational complexity. In response, Ma et al. proposed the three-stage PigFRIS framework, which integrates occlusion segmentation, GAN-based restoration, and lightweight recognition modules to enhance the performance of pig face recognition systems by addressing complex occlusion modeling and lightweight recognition (*Ma et al., 2025*). Furthermore, existing research indicates that pig behavior recognition has evolved from early analysis methods based on manual features and traditional image processing to deep learning approaches that integrate object detection, behavior classification, and multi-object tracking. Yang and Xiao conducted a systematic review of video-based pig behavior recognition research, highlighting its significant application value in health monitoring, welfare assessment, and precision feeding management (*Yang et al., 2020*). Chen et al. further summarized the evolutionary path of pig and cattle behavior recognition technology from traditional computer vision to deep learning, noting that complex lighting, individual occlusion, and herd overlap remain key challenges in real-world farming scenarios (*Chen et al., 2021*). To address behavior recognition in group-housing environments, Tu et al. achieved automatic recognition and tracking of pig behaviors such as lying down, feeding, and standing using object detection and an improved DeepSORT method (*Tu et al., 2022*); Li et al. combined YOLOX with a spatiotemporal behavior recognition module to achieve joint detection and recognition of multiple behaviors in group-housed pigs (*Li et al., 2022*). The aforementioned studies indicate that behavior recognition methods designed for real-world pig barn environments must strike a balance between detection accuracy, robustness in complex scenarios, and real-time deployment capabilities. This provides important reference for the research conducted in this paper on lightweight pig behavior detection in complex farming environments.

Based on this, this paper addresses the task of fine-grained pig behavior detection in complex farming environments. YOLOv11 was selected *(Li et al., 2024; Khanam et al., 2024; Jegham et al., 2024; Hidayatullah et al., 2025; He et al., 2025)* as the base detection framework and propose a lightweight improved model, WFE-YOLO, to identify five typical behaviors: Stand, Lie Down, Eat, Drink, and Bite. YOLOv11 was selected as the base model primarily because it strikes a good balance between detection accuracy and inference efficiency in object detection tasks, while its relatively lightweight structure facilitates subsequent lightweight modifications and deployment optimization. Additionally, its multi-scale feature representation capability provides a certain degree of adaptability for identifying targets with varying postures and scales in pig behavior detection. However, the original YOLOv11 still exhibits limitations in this task, including a significant bias toward high-frequency categories, insufficient fine-grained feature representation for complex behaviors, and excessive redundant computations during the detection head prediction stage. To address these issues, this paper performs collaborative optimization across three levels: training data distribution, feature extraction, and detection head design. During the training phase, the YOLOWeightedDataset was introduced to increase the training exposure of low-frequency behavior samples through image-level weighted sampling, thereby mitigating the learning bias caused by class imbalance. In the feature extraction stage, the FasterConvolutionalGatedLinearUnitBlock was designed to enhance the joint modeling capability of spatial and channel information, thereby improving the representation of fine-grained behaviors in complex scenarios. In the prediction stage, an Efficient Detection Head was designed to reduce redundancy and computational overhead through a more compact multi-scale prediction structure, thereby enhancing the model's suitability for lightweight deployment. Through these improvements, this paper attempts to establish a synergistic mechanism among data distribution optimization, feature representation enhancement, and prediction structure simplification to improve the model's overall detection performance in complex pig barn environments.

The main contributions of this paper can be summarized as follows. First, to address the learning bias caused by the class imbalance in complex pig house environments, a weighted sampling strategy was introduced to increase the exposure of low-frequency behavior samples during the training stage. Second, to solve the problem of insufficient fine-grained feature representation under conditions of dense distribution, occlusion, and posture changes, a lightweight feature extraction module integrating spatial and channel interactions was designed.

Third, to better meet the computational constraints in resource-limited livestock monitoring scenarios, a compact detection head was constructed, maintaining strong detection performance while reducing the number of parameters and computational costs. In summary, this paper focuses on improving the accuracy-complexity balance of pig behavior detection in complex farming scenarios and provides a methodological basis for subsequent monitoring systems for practical applications.

The experimental results show that the proposed WFE-YOLO performs well in the pig behavior detection task. Compared with the benchmark model YOLOv11n, the improved model achieves better performance in terms of accuracy, recall rate, and mAP@50, while further reducing the number of parameters and GFLOPs, demonstrating a better balance between accuracy and complexity. Further category analysis reveals that the model shows significant performance improvement in low-frequency and easily confused behavior categories, such as drinking water and chewing. Moreover, under the current dataset and evaluation settings, the proposed method is effective at the model level. The practical significance of this study should be understood as an exploration of model design for applications, rather than the completion of the verification of a complete monitoring or alarm system at the farm end.

MATERIALS AND METHODS

To address challenges in pig behavior detection within large-scale pig barn settings—such as high animal density, mutual occlusion, complex pose variations, imbalanced class distribution, and the need for lightweight deployment—this paper selects YOLOv11 as the base detection framework and proposes a lightweight improved model, WFE-YOLO, based on it. The original YOLOv11 still exhibits limitations in this task, including a significant bias toward high-frequency classes, insufficient fine-grained feature representation of complex behaviors, and excessive computational redundancy in the detection head prediction stage. To address these issues, this paper performs coordinated optimization across three levels: data sampling, feature extraction, and detection head design. Specifically, the YW-Dataset is utilized to improve the distribution of training samples, thereby enhancing the model's ability to learn low-frequency behavioral categories; the FCGB module is introduced to strengthen fine-grained feature representation in complex scenarios; and the EDH module is designed to reduce structural redundancy and computational overhead in the multi-scale prediction stage. These improvements collectively constitute WFE-YOLO, enabling the model to more effectively address issues such as class imbalance, insufficient feature discrimination, and limited prediction efficiency in complex pig barn environments, thereby achieving a better balance between detection accuracy and model complexity.

Data Collection and Processing

Data collection for this study was conducted in May 2025 in Taigu District, Jinzhong City, Shanxi Province. The subjects of the study were individual pigs and their behavioral activities under group-housing conditions in a real-world pig barn environment. The DJI Action 5 Pro 605C camera was used for image acquisition. All images were saved at the device's native output resolution to fully preserve texture, edge, and structural information within the scenes, thereby minimizing detail loss caused by image compression or downsampling and providing a reliable data foundation for fine-grained feature extraction in subsequent behavior detection.

All images were captured under natural lighting conditions, with collection times spanning various periods throughout the day. Although this study did not precisely quantify light intensity, the naturally distributed shooting times objectively covered a range of lighting conditions from soft light to strong sunlight, ensuring that the dataset included a rich variety of brightness variations, shadow distributions, and local reflections. These naturally occurring variations in lighting help enhance sample diversity and support the model's ability to learn robust features under variable lighting conditions. It should be noted that this dataset does not include nighttime scenes, and no artificial lighting was used during data collection; therefore, this study primarily focuses on the visual detection of pig behavior under natural daylight conditions.

Based on the behavioral characteristics of pigs and the requirements of livestock monitoring, this study defines five behavioral categories: Stand, Lie down, Eat, Drink, and Bite. This study constructed a dataset of 30,197 pig behavior images, which were divided into training, validation, and test sets in an 8:1:1 ratio. The data volumes for these five behavior categories are as follows: Stand (9,391), Lie down (14,519), Eat (4,333), Drink (776), and Bite (1,178). It is evident that there are significant distribution disparities among the different categories in the dataset. While Stand and Lie down have a large number of samples, Drink and Bite have relatively few, exhibiting typical characteristics of class imbalance.

This distribution pattern largely aligns with the frequency of behavior occurrence in real-world farming environments and reflects the objective challenges of pig behavior detection tasks in practical scenarios, providing a realistic basis for subsequent efforts in low-frequency behavior augmentation and class balance optimization.

The dataset constructed in this paper originates from a real pig barn environment. It preserves practical characteristics of group-housing scenarios, such as high target density, mutual occlusion, rich pose variations, and diverse lighting conditions. With its strong scene authenticity, it provides reliable data support for training and evaluating pig behavior detection models in complex farming environments.

Annotation Protocol and Validation Scope

This study constructed the dataset by using instance-level bounding box annotations rather than image-level category annotations. For each identifiable pig individual in the image, a separate bounding box was independently annotated, covering the visible body area of the pig. Each bounding box was assigned a main behavior label from the five predefined categories, namely Stand, Lie down, Eat, Drink, and Bite. This annotation method can simultaneously achieve target localization and behavior classification at the individual level, and is more suitable for fine-grained pig behavior detection tasks in group-rearing scenarios.

Since the data was collected in a real-scale pig farm environment, multiple pigs may appear in the same frame, and different individuals may exhibit different behaviors simultaneously. Therefore, the annotation rules explicitly allow multiple behavior categories to coexist within a single frame. That is, different labels are assigned to each pig based on its actual observable behavior. To improve the consistency of annotations, unified annotation rules were formulated. For transitional behaviors, they are determined based on the dominant observable action; when a certain behavior action accounts for 60% or more in the observable posture or movement pattern, it is labeled as the corresponding behavior. For ambiguous, partially occluded or difficult-to-determine samples, manual cross-review is conducted to further reduce subjective inconsistency.

It should be noted that the current content of this article mainly provides model-level validation under the setting of dataset-level experimental setup. Although the data comes from actual farming scenarios, this research does not constitute a fully verified end-to-end monitoring or alarm system. Therefore, the current validation scope should be understood as: supporting the effectiveness of the proposed model in the pig behavior detection task under more realistic visual conditions, rather than proving that the complete deployment process on the farming end has been completed.

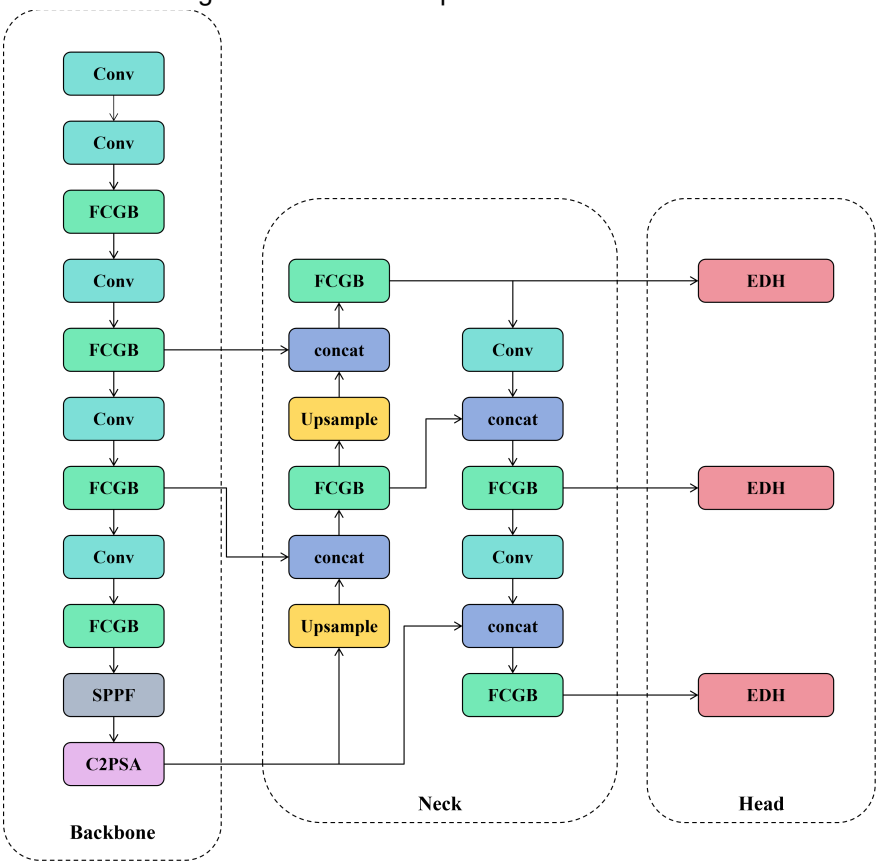

Fig. 1 - Overall Architecture of WFE-YOLO

Model Overview

Based on the YOLOv11 framework, this paper proposes an object detection model named WFE-YOLO that integrates data sampling strategies, feature modeling modules, and detection head structures. The model employs a unified design across three levels—training data distribution, feature representation, and detection head structure—aiming to collaboratively design the model's overall architecture and training process while preserving YOLOv11's original multi-scale detection framework and prediction mechanism. Specifically, WFE-YOLO introduces the YOLOWeightedDataset (YW-Dataset) during the training phase, redistributing training samples through category-frequency-based weighted sampling to adjust the sample distribution during training; At the network architecture level, the original feature extraction module is replaced with the FasterConvolutionalGatedLinearUnitBlock (FCGB) to enhance spatial modeling and channel interaction capabilities; in the detection head, the EfficientDetectionHead (EDH) is introduced to model multi-scale prediction features through a lightweight structural design. Through the synergistic combination of the above modules, WFE-YOLO constructs a unified detection model design framework across three levels: data, features, and prediction. The overall structure is shown in Fig. 1.

Training Configuration for Comparative Experiments

To ensure fairness, all the YOLO models being compared were trained and evaluated under exactly the same experimental settings. The input image resolution was uniformly set to 640×640. The optimization of relevant hyperparameters is as follows: initial learning rate lr0 = 0.01, final learning rate coefficient lrf = 0.01, momentum=0.937, weight decay=0.0005. The warm-up strategy is set to warmup_epochs=3.0, warmup_momentum=0.8, and warmup_bias_lr=0.1. The loss function gain parameters are set to box=7.5, cls=0.5, and dfl=1.5. The nominal batch size is set to nbs=64. Additionally, all the models adopted the same data augmentation strategy, including HSV enhancement (hsv_h=0.015, hsv_s=0.7, hsv_v=0.4), geometric enhancement (translate=0.1, scale=0.5), horizontal flipping (fliplr=0.5), and mosaic enhancement (mosaic=1.0). Therefore, the performance differences observed in the comparison experiments can mainly be attributed to the model structure itself rather than the inconsistent training configurations.

YOLO Weighted Dataset

During YOLO training, the original YOLO dataset employs a uniform sampling strategy, which can easily lead to bias toward high-frequency classes during model training when the class distribution is imbalanced. To mitigate this issue, this paper replaces the YOLO dataset with the YW-Dataset during the training phase. Unlike methods that rely on loss function weighting, the YW-Dataset adjusts the sampling distribution of the training data by introducing an image-level sampling strategy at the dataset level, without altering the network architecture or loss function design.

First, during the dataset initialization phase, the occurrence counts of each target category in the training set are counted, and category-level weights are constructed as the reciprocal of the target count. That is, the category weight is inversely proportional to the number of targets of that category in the training set, thereby giving categories with fewer targets a larger weight. Building on this, for each training image, the corresponding category weights are obtained based on the target categories it contains, and these weights are aggregated using a summation function to assign a unique image-level sampling weight to each image. Subsequently, the sampling weights of all training images are normalized to obtain the corresponding sampling probability distribution. During training, an image is randomly selected from the dataset based on this probability distribution each time data is read, rather than traversing the dataset in a fixed order. This weighted sampling strategy is enabled only during the training phase; the original data loading method is retained during the validation and testing phases to ensure that the model evaluation process adheres to the actual data distribution.

Faster Convolutional Gated Linear Unit Block

In network architecture design, the original C3k2 module performs feature extraction and fusion by stacking multiple Bottleneck submodules, relying primarily on standard convolutional operations to model spatial and channel information. To further enhance the flexibility and efficiency of feature representation, this paper replaces the Bottleneck submodules in the C3k2 module with the proposed FCGB (Faster Convolutional Gated Linear Unit Block) module, whose overall structure is shown in Fig. 2. While preserving the original residual connection structure of C3k2, the FCGB module redesigns the internal feature transformation mechanism, thereby ensuring structural compatibility and network stability.

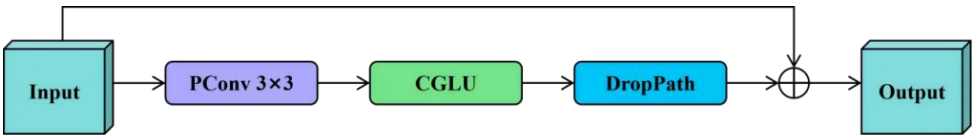

Fig. 2 - FCGB Schematic Diagram

Specifically, the FCGB module first uses partial convolution (PConv) to spatially blend the input features, performing local convolution operations only on a subset of channels. This approach effectively models local spatial relationships while reducing computational overhead, as illustrated in Fig. 3.

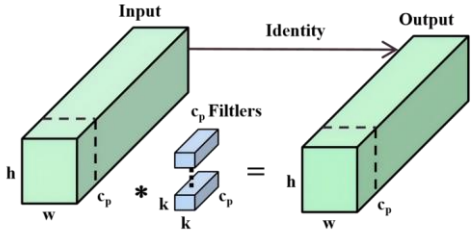

Fig. 3 - PConv Block Diagram

Subsequently, a Convolutional Gated Linear Unit (CGLU) is introduced to perform nonlinear modeling of channel features. Through the synergistic interaction between the feature branch and the gated branch, selective modulation of channel responses is achieved, as shown in Fig. 4.

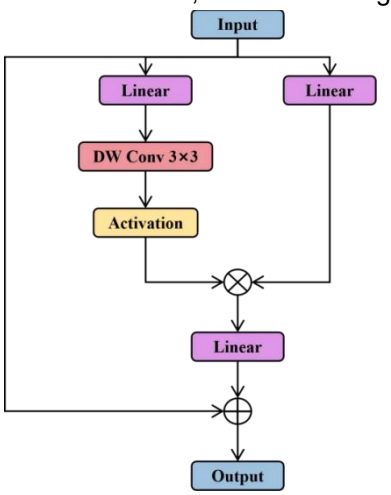

Fig. 4 - CGLU Organizational Chart

Within the residual framework, the FCGB module further introduces a random depth (DropPath) mechanism in the feature transformation branch to enhance stability during the training of deep networks. Through the aforementioned modifications, the C3k2 module achieves collaborative modeling of spatial and channel information while maintaining the overall structural consistency, thereby providing the network with more flexible feature transformation capabilities.

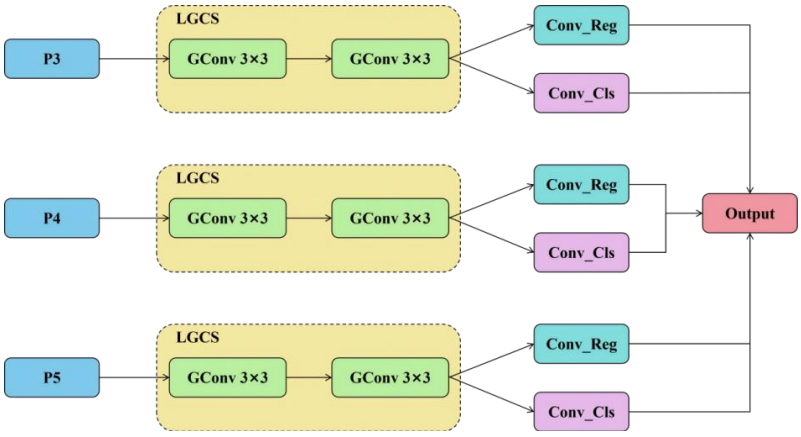

Fig. 5 - EDH Block Diagram

Efficient Detection Head

In the original detection head of YOLOv11, multi-scale features from the neck network (corresponding to layers P3, P4, and P5) are fed into separate regression and classification prediction branches for prediction. The original detection head constructs a prediction subnetwork at each scale, consisting of multi-layer convolutional and deep separable convolutional modules, to perform step-by-step transformation and modeling of the features. Under the multi-scale detection setup based on P3, P4, and P5, this design requires the construction of independent prediction subnetworks at different scales, thereby increasing the structural complexity and parameter size of the detection head. Based on the aforementioned structural characteristics, this paper proposes an EDH detection head aimed at simplifying and restructuring the detection head while preserving the original YOLOv11 multi-scale detection framework and the distributed regression (DFL)-based prediction mechanism. The overall structure is shown in Fig. 5.

Specifically, for feature maps from P3, P4, and P5, EDH introduces a Lightweight Group-Convolution Stem (LGCS) module at each detection scale. This module consists of two 3×3 group convolutions (GConv) connected in series, performing stepwise spatial blending and nonlinear transformations on the features while maintaining the number of feature channels at each scale. Through group convolution, the LGCS reduces the connection density between channels while forming a local spatial receptive field. Building on this, EDH directly generates bounding box regression distributions and class prediction results at the corresponding scale via 1×1 convolution, thereby avoiding the need to repeatedly construct multi-layer convolutional stacked prediction sub-networks at each scale. Through the above design, EDH achieves a more compact implementation of a multi-scale detection head.

Evaluation Criteria

To quantitatively evaluate the performance of the proposed model in the pig behavior detection task, this paper adopts Precision, Recall, and mAP@50 as the primary evaluation metrics. The definitions of Precision and Recall are shown in Eq.1 and Eq.2, respectively. Precision measures the proportion of targets predicted as positive samples by the model that are actually positive samples, i.e., the accuracy of the detection results. As shown in Eq.1:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

In this context, TP (True Positive) refers to the number of targets correctly detected by the model, while FP (False Positive) refers to the number of targets incorrectly identified by the model. A higher Precision indicates fewer false positives and more reliable prediction results.

Recall measures the proportion of true targets in the dataset that are successfully detected by the model, i.e., the model's ability to detect targets. As shown in Eq. 2:

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

In this context, FN (False Negative) refers to the number of objects that actually exist but were not detected by the model. A higher Recall indicates that the model misses fewer objects and has stronger coverage of the targets.

In object detection tasks, relying solely on Precision or Recall does not fully reflect model performance; therefore, this paper further adopts mAP@50 (mean Average Precision at IoU = 0.5) as a comprehensive evaluation metric. Specifically, with an IoU threshold set to 0.5, the average precision (AP) for each behavior category is first calculated, and then the AP values for all categories are averaged to obtain mAP@50. This metric comprehensively reflects the model's overall detection performance across different pig behavior categories and serves as a key standard for evaluating the performance of detection models.

In addition to Precision, Recall, and mAP@50, FPS (frames per second) was also used as a supplementary runtime indicator in the comparative analysis to reflect the inference efficiency of different models under the same test setting. It should be noted that FPS in this study is intended as a preliminary measure of runtime efficiency, rather than a full hardware-side deployment evaluation.

RESULTS

Ablation Experiment Section

To analyze the roles of each component module in the overall design of WFE-YOLO, this paper designs a series of ablation experiments. Specifically, the ablation experiments focus on three key design elements: the weighted data sampling strategy YW-Dataset during the training phase, the feature modeling module

FCGB, and the detection head architecture EDH. Based on the YOLOv11 baseline model, each of the aforementioned modules was introduced individually, and multiple configuration combinations were constructed to systematically evaluate their contributions to the model architecture and training process. This analysis provides a clearer understanding of the overall design of WFE-YOLO. The experimental results are presented in Table 1.

When only YW-Dataset was introduced, without altering the network architecture or computational complexity, the model achieved mAP@50from 0.7933 to 0.804, with P and R improving from 0.7712 and 0.7508 to 0.7809 and 0.7651, respectively. These results indicate that by adjusting the sample selection strategy during the training phase, the model exhibits a certain degree of balance in the response distribution across different behavior categories during the prediction phase.

When FCGB is introduced alone, the model adopts a convolutional structure with a gating mechanism during the feature extraction phase and reconstructs the feature modeling path by reducing redundant channel computations. The corresponding results show that the number of parameters decreased from 2.58 M to 2.23 M, GFLOPs decreased from 6.3G to 5.6G, 0.8029, and P reached 0.792. This indicates that by introducing a gating mechanism and optimizing channel computations during the feature extraction stage, the model is able to maintain stable representation of key visual features of pig behavior while reducing computational complexity.

When using EDH alone, the model's detection head architecture was redesigned for the prediction phase. Experimental results show that the number of parameters and GFLOPs were reduced to 2.31 million and 5.1 GFLOPs, respectively, while the P, R, and mAP@50 metrics also improved compared to the baseline model. This is because EDH reduces the stacking of multiple convolutional layers in the prediction head and adopts a more direct feature transformation path, enabling multiscale features to complete their spatial modeling before entering the regression and classification prediction stages. Consequently, it reduces computational overhead in the prediction phase while maintaining the stability of prediction results.

Under multi-module configurations, structural adjustments at each stage form a synergistic relationship within the overall detection process. For example, when FCGB and EDH are introduced simultaneously, the model adopts a more simplified design in both the feature extraction and prediction stages, reducing parameters and GFLOPs to 1.96M and 4.4G, respectively, while demonstrating a certain degree of improvement in metrics such as mAP@50. After incorporating the YW-Dataset, FCGB, and EDH into the final model, a synergistic workflow was established across three levels: training data distribution, feature modeling, and prediction architecture. This enabled the model to maintain relatively low computational complexity while demonstrating stable detection performance in the pig behavior detection task.

In summary, the ablation experiment results indicate that the YW-Dataset, FCGB, and EDH collaboratively optimize the model's overall workflow across three layers—training sample sampling, feature extraction architecture, and prediction head design—achieving a reasonable balance between detection performance and model complexity in the pig behavior detection task.

Table 1

Ablation Experiment Results

YOLOv11n	YW-Dataset	FCGB	EDH	P	R	mAP@50	Params(M)	GFLOPs(G)
√				0.7712	0.7508	0.7933	2.58	6.3
√	√			0.7809	0.7651	0.804	2.58	6.3
√		√		0.792	0.7552	0.8029	2.23	5.6
√			√	0.7866	0.76	0.8115	2.31	5.1
√	√	√		0.8038	0.776	0.8149	2.23	5.6
√	√		√	0.8062	0.7654	0.8199	2.31	5.1
√		√	√	0.7981	0.7771	0.8155	1.96	4.4
√	√	√	√	0.8154	0.7803	0.8233	1.96	4.4

Fig. 6 shows a comparison of the mAP@50 metrics between the baseline model YOLOv11n and the proposed WFE-YOLO for different pig behavior categories (Stand, Lie down, Eat, Drink, and Bite). It can be observed that WFE-YOLO achieves a certain degree of performance improvement across all behavior categories, indicating that the proposed model improvements demonstrate good adaptability under various behavioral patterns.

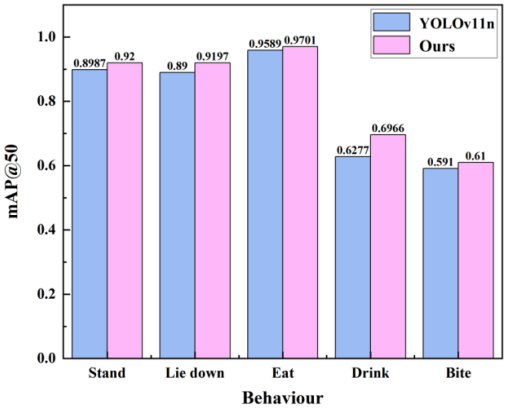

Fig. 6 - Comparison of mAP@50 between YOLOv11n and WFE-YOLO across different pig behavior categories

For the three behavior categories—Stand, Lie Down, and Eat—where action features are relatively stable and the proportion of samples is high, WFE-YOLO demonstrates a significant improvement even though the baseline model already possesses high detection accuracy. This indicates that while maintaining the original multi-scale feature representation capability, the introduced data sampling strategy and feature modeling module do not adversely affect the discrimination ability for common behaviors and further enhance the discriminative power of the feature representation. Particularly for the Eat behavior, both models achieved high mAP@50 scores, yet WFE-YOLO still delivered relatively superior detection results.

For the Drink and Bite categories—which have relatively few samples, short behavior durations, and significant visual variations—the performance improvement of WFE-YOLO was even more pronounced. In particular, the improvement in mAP@50 for the “Drink” behavior was greater than for other behavior categories, indicating that by adjusting the sample distribution during the training phase and enhancing the expressive capabilities of feature modeling and prediction within the network architecture, the model possesses stronger discriminative power when handling complex and easily confused behaviors.

In summary, WFE-YOLO demonstrates a consistent trend of performance improvement across different behavior categories and exhibits a more pronounced advantage in categories with relatively higher detection difficulty. This further validates the effectiveness and stability of the proposed method in pig behavior detection tasks.

As shown in Fig. 7, (a) is the original image, (b) is the image from the YOLOv11n baseline model, and (c) is the result from the improved WFE-YOLO model. By comparing the detection results of the original image, YOLOv11n, and WFE-YOLO, it can be observed that the improved model demonstrates more stable detection performance in scenarios with a high density of pigs, occlusion between individuals, and complex pose variations. When there are a large number of pigs and the spatial distance between individuals is relatively close, WFE-YOLO can detect more complete individuals, and the distribution of prediction boxes is more reasonable, whereas YOLOv11n exhibits low confidence in the prediction of detection boxes at certain locations. Furthermore, when pigs are located at the edges of the image or parts of their bodies are occluded, WFE-YOLO still maintains good detection consistency.

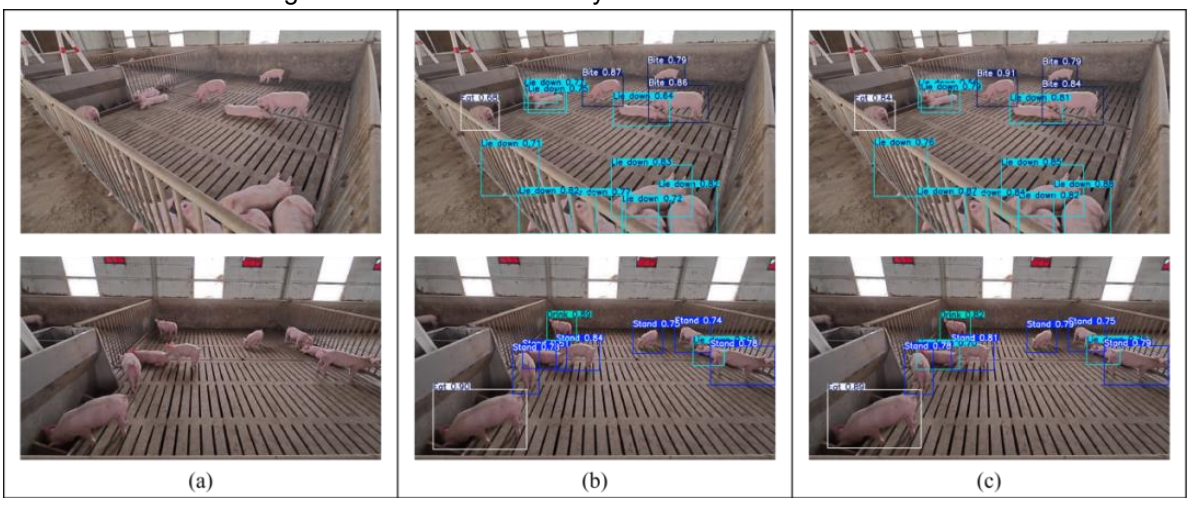

Fig. 7 - Comparison of detection results before and after improvement

Furthermore, when multiple behaviors coexist in a scene or individuals are in transitional states, WFE-YOLO demonstrates higher consistency in behavior category prediction, whereas YOLOv11n exhibits confusion in behavior classification in such scenarios. WFE-YOLO also exhibits greater robustness in complex scenarios, particularly in situations with dense multi-objects, severe occlusion, and diverse behavioral patterns, where detection results are more stable and reliable.

As shown in Fig. 8, (a) depicts the original image, (b) shows the image from the YOLOv11n baseline model, and (c) displays the results from the improved WFE-YOLO model. A comparison of the heatmap visualizations between YOLOv11n and WFE-YOLO reveals significant differences in the spatial distribution characteristics of the regions of interest between the two models. Compared to the baseline model, the attention distribution of WFE-YOLO is more concentrated within the target region, exhibiting a clearer attention pattern.

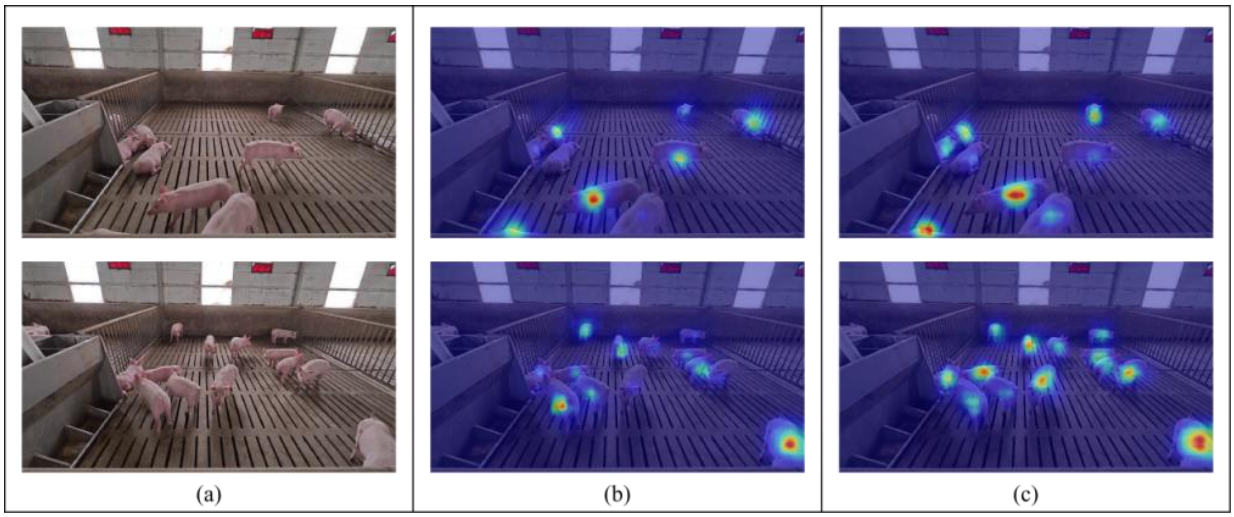

Fig. 8 - Before-and-After Heatmap Comparison

In scenarios with multiple coexisting targets and complex pose variations, the heatmap distribution of WFE-YOLO maintains focused attention on target instances, with relatively distinct boundaries for the high-response regions corresponding to different individuals, demonstrating a clearer spatial attention structure.

Overall, the advantages of WFE-YOLO not only lie in its detection performance, but also in the lightweight design under the current experimental setup. These results indicate that the proposed method has certain potential for practical applications in the task of pig behavior analysis in complex breeding environments. Similar trends can also be seen in recent studies on improving YOLO for pig behavior detection and recognition, and related IoT-based pig house monitoring work further highlights the value of integrating behavior perception with environmental perception in smart breeding (*Li et al., 2023; Liu et al., 2025; Rong et al., 2025*).

Comparative experiment

For comparative evaluation, all models were assessed on the same dataset split and under the same evaluation criteria used in this study. The comparison is therefore intended as a controlled relative evaluation within the current experimental setting, rather than as a universal benchmark across all pig behavior detection scenarios.

To systematically evaluate the performance of the proposed WFE-YOLO model in the pig behavior detection task, this paper selected several representative models from the YOLO series for comparative experiments, including YOLOv5n, YOLOv6n, YOLOv8n, YOLOv8s, YOLOv9t, YOLOv10n, YOLOv11s, YOLOv12n, and YOLOv13n.

These models encompass network architectures of different versions and scales, reflecting variations in structural design and computational complexity among current mainstream single-stage object detection methods. By conducting comparative evaluations on a unified dataset under consistent training configurations, the relative performance of WFE-YOLO in terms of detection accuracy and model complexity can be systematically analyzed, thereby providing a more robust basis for performance evaluation.

Table 2 presents the comparative experimental results of WFE-YOLO against various YOLO series models on the pig behavior detection task. Under the unified dataset, training strategy, and evaluation criteria of this paper, WFE-YOLO achieved an mAP@50 of 0.8233, with Precision and Recall reaching 0.8154 and 0.7803, respectively, while having only 1.96M parameters and 4.4G GFLOPs. This indicates that the model possesses good overall detection performance and low computational complexity under the current experimental settings. In addition to detection accuracy and model complexity, FPS was further introduced as a runtime efficiency indicator to provide a preliminary assessment of real-time applicability. As shown in Table 2, WFE-YOLO achieved 520.43 FPS under the current test setting, which was the highest among the compared models, indicating that the proposed model has favorable inference efficiency while maintaining competitive detection performance.

Further comparison of models of different scales reveals that, compared to medium-scale models such as YOLOv8s and YOLOv11s, WFE-YOLO achieves a superior mAP@50 result while significantly reducing the number of parameters and computational load; compared to lightweight models such as YOLOv5n, YOLOv8n, and YOLOv10n, WFE-YOLO also demonstrates strong overall competitiveness. This indicates that the method proposed in this paper achieves a good balance between accuracy and computational complexity, making it more suitable for pig behavior detection tasks under resource-constrained conditions.

It should be noted that when comparing the experimental results of this paper with those of other studies, metrics across different literature are often not strictly comparable due to potential differences in data sources, behavior classification, image acquisition conditions, annotation standards, and evaluation protocols. Therefore, this paper focuses more on relative comparisons between models under standardized experimental conditions. Furthermore, our method emphasizes adaptability to complex real-world pig barn environments, lightweight deployment capabilities, and superior recognition of low-frequency, easily confused behaviors. Combining ablation experiments with category-level results reveals that WFE-YOLO not only enhances overall detection performance but also demonstrates more significant improvements in behavior categories with higher detection difficulty, such as “Drink” and “Bite.”

In summary, the advantages of WFE-YOLO are not merely reflected in the improvement of a single accuracy metric, but are further demonstrated in multiple aspects such as lightweight design, robustness in complex scenarios, and the ability to recognize key low-frequency behaviors, highlighting its potential value in practical smart livestock farming applications.

Table 2

Experimental Results Comparing Different Models

Model	P	R	mAP@50	Params(M)	GFLOPs(G)	FPS
YOLOv5n	0.7956	0.7643	0.7989	2.5	7.1	383.12
YOLOv6n	0.7688	0.7313	0.8	4.23	11.7	370.76
YOLOv8n	0.7881	0.7555	0.8043	3.01	8.1	486.87
YOLOv8s	0.8076	0.7661	0.8138	11.13	28.4	412.84
YOLOv9t	0.7492	0.7673	0.7936	1.97	7.6	501.53
YOLOv10n	0.7985	0.7302	0.7985	2.7	8.2	490.31
YOLOv11s	0.8097	0.7769	0.8112	9.41	21.3	427.09
YOLOv12n	0.7833	0.7331	0.7952	2.51	5.8	483.9
YOLOv13n	0.7858	0.7456	0.8012	2.45	6.1	289.87
Ours	0.8154	0.7803	0.8233	1.96	4.4	520.43

CONCLUSIONS

This paper proposes WFE-YOLO, a lightweight pig behavior detection model for complex pig barn environments characterized by dense targets, occlusion, subtle behavioral differences, class imbalance, and limited computing resources. The main value of this work lies in showing that coordinated improvements in sample distribution, feature representation, and detection head design can improve the accuracy–complexity trade-off of pig behavior detection under the current experimental setting.

At the same time, the scope of the present study should be clearly defined. Although the proposed method is motivated by practical livestock monitoring needs, this work does not include a full end-use validation pipeline such as long-term continuous surveillance, automatic farm-side alerting, or farmer-oriented decision-support integration. Therefore, the application significance of this study should be understood as providing a model-level and application-oriented technical basis, rather than as demonstrating a fully validated deployment system.

Several limitations should also be acknowledged. First, the current study focuses on frame-level detection of five observable behaviors, and does not model longer temporal dependencies or continuous behavioral processes. Second, ambiguous and transitional behaviors remain challenging both for annotation and for visual recognition. Third, the present validation is limited to the current dataset and scenario conditions, and cross-scenario generalization still requires further study.

In future work, more comprehensive application-level validation should be conducted by integrating continuous video monitoring, temporal behavior analysis, and alert-oriented system design. In summary, this study provides a lightweight methodological framework for pig behavior detection in complex farming environments, while full deployment and end-use validation are left for future research.

ACKNOWLEDGEMENT

This work was supported by the Scientific Research Startup Project for Introduced Talents of Shanxi Agricultural University, No.2023BQ28; Shanxi Provincial Department of Education Science and Technology Innovation, No.2024L058.

REFERENCES

[1] Ali, M. L., Zhang, Z., (2024). The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection. Computers, Vol. 13, pp. 336, Basel/Switzerland.

[2] Cai, H., Li, J., Hu, M., Gan, C., Han, S., (2023). EfficientViT: Lightweight multi-scale attention for high-resolution dense prediction. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17302-17313, Paris/France.

[3] Canario, L., Bijma, P., David, I., Camerlink, I., Martin, A., Rauw, W. M., Flatres-Grall, L., van der Zande, L., Turner, S. P., Larzul, C., Rydhmer, L., (2020). Prospects for the analysis and reduction of damaging behaviour in group-housed livestock, with application to pig breeding. *Frontiers in Genetics*, Vol. 11, pp.611073, Lausanne/Switzerland.

[4] Chen, C., Zhu, W., Norton, T., (2021). Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. *Computers and Electronics in Agriculture*, Vol. 187, pp.106255, Netherlands.

[5] He, L., Zhou, Y., Liu, L., Cao, W., Ma, J., (2025). Research on object detection and recognition in remote sensing images based on YOLOv11. *Scientific Reports*, Vol. 15, pp. 14032, London/United Kingdom.

[6] Hidayatullah, P., Syakrani, N., Sholahuddin, M. R., Gelar, T., Tubagus, R., (2025). YOLOv8 to YOLO11: A comprehensive architecture in-depth comparative review. *arXiv Preprint*, arXiv:2501.13400, United States.

[7] Hu, Y., Yu, Y., (2022). Scale difference from the impact of disease control on pig production efficiency. *Animals*, Vol. 12, pp. 2647, Basel/Switzerland.

[8] Jegham, N., Koh, C. Y., Abdelatti, M., Hendawi, A., (2024). YOLO evolution: A comprehensive benchmark and architectural review of YOLOv12, YOLO11, and their previous versions. *arXiv Preprint*, arXiv:2411.00201, United States.

[9] Khanam, R., Hussain, M., (2024). YOLOv11: An overview of the key architectural enhancements. *arXiv Preprint*, arXiv:2410.17725, United States.

[10] Krasanakis, E., Spyromitros-Xioufis, E., Papadopoulos, S., Kompatsiaris, Y., (2018). Adaptive sensitive reweighting to mitigate bias in fairness-aware classification. *Proceedings of the 2018 World Wide Web Conference*, pp. 853-862, Lyon/France.

[11] Li, R., Dai, B., Hu, Y., Dai, X., Fang, J., Yin, Y., Liu, H., Shen, W., (2024). Multi-behavior detection of group-housed pigs based on YOLOX and SCTS-SlowFast. *Computers and Electronics in Agriculture*, Vol. 225, pp. 109286, Netherlands.

[12] Li, Y., Li, J., Duan, L., Na, T., Zhang, P., Zhi, Q., (2023). Detection of eating behaviour in pigs based on modified YOLOX. INMATEH - Agricultural Engineering, Vol. 71, pp. 44-52, Bucharest/Romania.

[13] Liu, J., Yan, Y., Yang, Y., Hao, Y., Chen, B., Yang, M., Hu, J., (2025). Pig recognition based on YOLOv8-EAPNet. *INMATEH - Agricultural Engineering*, Vol. 77, pp. 676-688, Bucharest/Romania.

[14] Ma, R., Chung, S., Kim, S., Kim, H., (2025). PigFRIS: A three-stage pipeline for fence occlusion segmentation, GAN-based pig face inpainting, and efficient pig face recognition. *Animals*, Vol. 15, pp. 978, Basel/Switzerland.

[15] Maes, D. G. D., Dewulf, J., Piñeiro, C., Edwards, S., Kyriazakis, I., (2020). A critical reflection on intensive pork production with an emphasis on animal health and welfare. *Journal of Animal Science*, Vol. 98, pp. S15-S26, Oxford/United Kingdom.

[16] Matthews, S. G., Miller, A. L., Clapp, J., Plötz, T., Kyriazakis, I., (2016). Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. *Veterinary Journal*, Vol. 217, pp. 43-51, Netherlands.

[17] Pandey, S., Kalwa, U., Kong, T., Guo, B., Gauger, P. C., Peters, D. J., Yoon, K.-J., (2021). Behavioral monitoring tool for pig farmers: Ear tag sensors, machine intelligence, and technology adoption roadmap. *Animals*, Vol. 11, pp. 2665, Basel/Switzerland.

[18] Rong, L., Fan, J., Guo, X., Tong, Z., Xu, W., Pan, Y., Li, S., Zhang, W., Sun, F., (2025). Research on environmental monitoring and comprehensive evaluation system of pig house based on Internet of Things technology. *INMATEH - Agricultural Engineering*, Vol. 75, pp. 501-514, Bucharest/Romania.

[19] Roura, E., Koopmans, S.-J., Lallès, J.-P., Le Huerou-Luron, I., de Jager, N., Schuurman, T., Val-Laillet, D., (2016). Critical review evaluating the pig as a model for human nutritional physiology. *Nutrition Research Reviews*, Vol. 29, pp. 60-90, United Kingdom.

[20] Sukkuea, A., Akkajit, P., (2025). Improved detection and classification of precise behaviors in group-housed pigs using deep learning models. *Engineered Science*, Vol. 37, pp. 1808, Singapore.

[21] Tu, S., Zeng, Q., Liang, Y., Liu, X., Huang, L., Weng, S., Huang, Q., (2022). Automated behavior recognition and tracking of group-housed pigs with an improved DeepSORT method. *Agriculture*, Vol.12, pp. 1907, Basel/Switzerland.

[22] Wemelsfelder, F., Hunter, E. A., Mendl, M. T., Lawrence, A. B., (2000). The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science*, Vol. 67, pp. 193-215, Netherlands.

[23] Yang, Q., Xiao, D., (2020). A review of video-based pig behavior recognition. *Applied Animal Behaviour Science*, Vol. 233, pp. 105146, Netherlands.