Computers and Electronics in Agriculture 225 (2024) 109286

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Multi-behavior detection of group-housed pigs based on YOLOX and SCTS-SlowFast

Ran Li [a], Baisheng Dai [a,b,*], Yuhang Hu [a], Xin Dai [a], Junlong Fang [a,b,*], Yanling Yin [a,b], Honggui Liu [b,c], Weizheng Shen [a]

[a] College of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China
[b] Key Laboratory of Pig-breeding Facilities Engineering, Ministry of Agriculture and Rural Affairs, Harbin 150030, China
[c] College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China

ARTICLE INFO

Keywords:
Multi-behavior detection
Group-housed pigs
SCTS-SlowFast
Deep learning

ABSTRACT

Accurately and rapidly recognizing the behaviors of group-housed pigs plays a very important role in pig farm production management. Recognizing multiple behaviors in one scene remains a challenge. This study proposed a multi-behavior detection method of group-housed pigs based on YOLOX and SCTS-SlowFast (SC-Conv-TS-SlowFast) to recognize four behaviors (Eating or Interaction with feeding though, Drinking or Interaction with drinker, Standing and Walking) and locate corresponding pig locations. Firstly, YOLOX object detection module is used to locate the locations of group-housed pigs. Secondly, SCTS-SlowFast behavior recognition module is proposed to classify the behaviors category of pigs in the located regions, in which Self-Calibrated Convolution (SC-Conv) and Temporal-Spatial (TS) attention mechanism are specially introduced to improve behavior feature extraction capability of the model. Finally, the results of two modules are combined to realize the task of multi-behavior detection of group-housed pigs. To evaluate the proposed method, a multi-behavior video dataset of group-housed pigs with 420 video segments is established. This study achieved a mAP value of 80.05% for four behaviors of group-housed pigs, and counted the duration of these behaviors throughout one day from 8:00 to 16:00 as well as their corresponding changing trends. It verifies the potential and feasibility of proposed method in automatically and simultaneously monitoring and analyzing multiple typical behaviors of group-housed pigs. We shared our behavior detection dataset at https://github.com/IPCLab-NEAU/Group-housed-pigs-Multi-Behavior-Detection for precision livestock farming research community.

1. Introduction

Monitoring and analyzing behaviors in group-housed pigs is of great significance in production management of commercial pig farm. Behavioral changes of pigs may reflect welfare condition or environmental stress (Matthews et al., 2017). In a large-scale pig farm, manual observation of pig behaviors not only has the problems of high labor intensity, high labor cost and low efficiency, but also is prone to errors. Therefore, it is necessary to automatically monitor behaviors of group-housed pigs, which is of positive effect in promoting the development of the pig farming industry.

In recent years, the methods based on computer vision have achieved some results in the behavior recognition of pigs. These methods can be divided into two categories: the methods based on traditional image processing and the methods based on deep learning. In the traditional image processing-based methods, Kashiha et al. (2013) used the distance between the pig's nose and drinker to recognize the drinking behavior of pigs. Viazzi et al. (2014) used the Motion History Image (MHI) and Linear Discriminant Analysis (LDA) to recognize aggressive behavior of pigs. Lao et al. (2016) proposed a method of feeding behavioral recognition of sow by the position of feeder and body. Leonard et al. (2019) used the depth sensor to determine a feeding behavior based on the area of the sow's head captured within a region around the feeder. Although traditional image processing-based methods have been used for pig activity monitoring and provide important monitoring data for managers, their application costs are high in large-scale farming facilities, and monitoring performance could easily be affected by the image quality and points detected on the pig's body (Wang et al 2021). Consequently, utilizing computer vision and deep learning technologies for non-contact pig behavior recognition has emerged as an effective solution.

* Corresponding authors at: College of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China.
E-mail addresses: bsdai@neau.edu.cn (B. Dai), jlfang@neau.edu.cn (J. Fang).

https://doi.org/10.1016/j.compag.2024.109286
Received 7 February 2024; Received in revised form 18 July 2024; Accepted 27 July 2024
Available online 31 July 2024

Alameer et al. (2020) proposed an automatic recognition method for feeding behavior in pigs based on grayscale video frames and improved GoogLeNet network. Chen et al. (2020) developed a method based on convolutional neural network (CNN) and long short-term memory (LSTM) to recognize feeding behavior of nursery pigs. Li et al. (2023) used the SlowFast model to extract features of pig care behaviors and introduced the Hidden Markov Model (HMM) to achieve accurate recognition of sow care behaviors. Gao et al. (2023) proposed a recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model. At present, compared with traditional methods, the pig behavior recognition methods based on deep learning have been improved in recognition performance and model generalization. However, most of the above methods focus on recognizing a specific behavior of pigs, this remains a challenge to detect multiple behaviors of pigs in single field of view of surveillance camera. Although, some researchers have carried out related work on domestic animal by object detection methods (Zhang et al., 2019; Wang et al., 2023; Gu et al., 2023), these image-based method fail to fully utilize the sequence information of video, and thus achieves poor robustness (Sun et al., 2023). Therefore, the reliability and practicability of behavior recognition need to be further improved.

This study explores a behavior detection strategy for monitoring behaviors of group-housed pigs by fusing temporal and spatial dimension information. It is designed to not only recognize the behaviors of group-housed pigs in the video but also locate corresponding pig locations (wang et al., 2024). More importantly, it can recognize the multiple behaviors in single field of view. Currently, most of existing behavior detection methods are used for human behavior detection tasks (Faure et al., 2023; Li et al., 2023; Kim et al., 2023). However, due to the complexity of breeding environment, the feature extraction capability of those behavior detection methods still needs to be further improved, which is difficult to be adopted for detecting pig behaviors in a group-housed environment. Therefore, this study designed a multi-behavior detection method of group-housed pigs, which is composed of YOLOX (Ge et al., 2021) object detection module and SCTS-SlowFast behavior recognition module. The proposed method can recognize four behaviors (Eating or Interaction with feeding though, Drinking or Interaction with drinker, Standing and Walking) in a group-housed environment and locate corresponding location of pigs.

The main contributions of this study are as follows:

- A multi-behavior detection method of group-housed pigs based on YOLOX and SCTS-SlowFast was proposed, which can recognize pig multiply behaviors in a rearing environment and locate corresponding pig locations.
- A Temporal-Spatial (TS) attention mechanism is specially designed into behavior recognition module for enhancing detection performance, TS attention mechanism can provide a reference for other behavior detection method.
- The proposed method is proved to achieves an effective performance for detecting four behaviors of group-housed pigs by comparative experiments, which provides a new strategy for automatically and simultaneously monitoring and analyzing multiple typical behaviors of group-housed pigs.

2. Materials

2.1. Data acquisition

The video of group-housed pig behaviors was collected at HongFu Pig Farm, Harbin City, Heilongjiang Province. The length of pig pen is 4.3 m and width is 2.3 m. The pig pen adopts a leaky seam floor and the material is slats. There are two nipple drinkers and two feeding troughs inside the pen. The pigs are piglets of Large white × Landrace of 35–42 days old. The rearing density is about 10 pigs per pen. A Hikvision DS-2CD3345D-I camera was installed in the center of one pig pen to record piglets within 72 h after mixing. The camera was positioned at a height of 2.3 m relative to the ground and pointed downwards. It can capture the entire activity region of piglets, and reduce impact of occlusion between pigs in this study. The camera resolution was 2560 × 1440 pixels with a frame rate of 25 fps, and the video was stored in MP4 format.

2.2. Dataset construction

In this study, the behavior detection datasets of group-housed pigs contained two kinds of datasets, one was for object detection module training and testing, and the other was for behavior recognition module training and testing.

2.2.1. Object detection dataset

This study randomly sampled from the raw 72-hour video data and obtained 2015 images for constructing the object detection dataset. These images were labeled using the LabelImg software to create bounding box annotations around the pig, all annotations were completed by three pig behavior experts with several years of experience. The annotations were stored in the COCO format (Lin et al., 2014). The annotated images are divided into training set (1209 images), validation set (403 images), and testing set (403 images) following a ratio of 6:2:2.

2.2.2. Behavior recognition dataset

The group-housed pig behavior recognition dataset contains a total of 420 video segments, respectively, with a uniform length of 10 s. These video segments were obtained by random cropping from the original 72-hour videos. According to the AVA (Gu et al., 2018) dataset format and the behavior classification in Table 1, the keyframe was selected at per second from these segments for behavior labeling. Three pig behavior experts were invited to annotate four group-housed pig behaviors from these video segments. A total 12,013 behavioral labels were obtained (the examples are shown in yellow box of Fig. 1). Subsequently, these video segments are divided into training set and testing set following a ratio of 8:2. The training set contains 336 video segments, the testing set contains 84 video segments. The illustration of multi-behavior annotating is shown in Fig. 1.

3. Methods

3.1. Multi-behavior detection of group-housed pigs

The purpose of multi-behavior detection is to recognize multiple categories of pig behaviors in single field of view and locate corresponding location of pigs. The proposed behavior detection method is composed of object detection module and behavior recognition module, as illustrated in Fig. 2. YOLOX (Ge et al., 2021) is selected as the object detection module to determine location of group-housed pigs from video keyframes. SCTS-SlowFast is designed as the behavior recognition module to classify behavior categories of pigs in the located regions. Particularly in the recognition module, the SC-Conv (Liu et al., 2020)

Table 1
Definition of group-housed pig behaviors.

Behavior	Abbreviations	Behavior description
Eating or Interaction with feeding trough	EI	Feed and root food in the feeding trough, or lick, bite and sniff the feeding trough
Drinking or Interaction with drinker	DI	Drink water, or lick and bite the drinker
Standing	ST	The hooves support the body weight, and there is no change in body position.
Walking	WK	The body changes position, and walks and runs in a normal posture

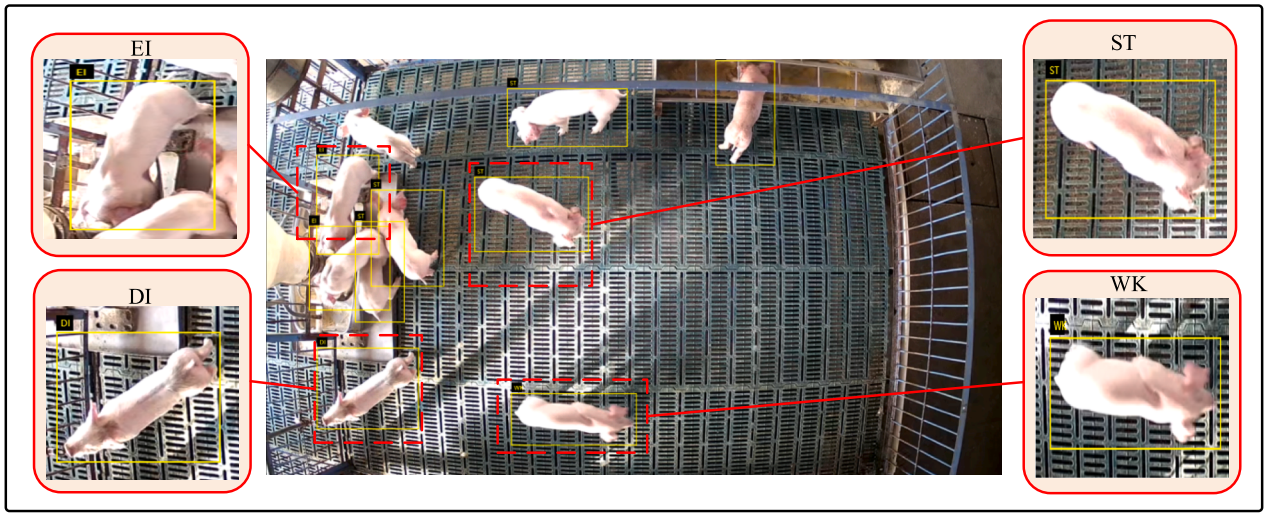

Fig. 1. The illustration of multi-behavior annotating.

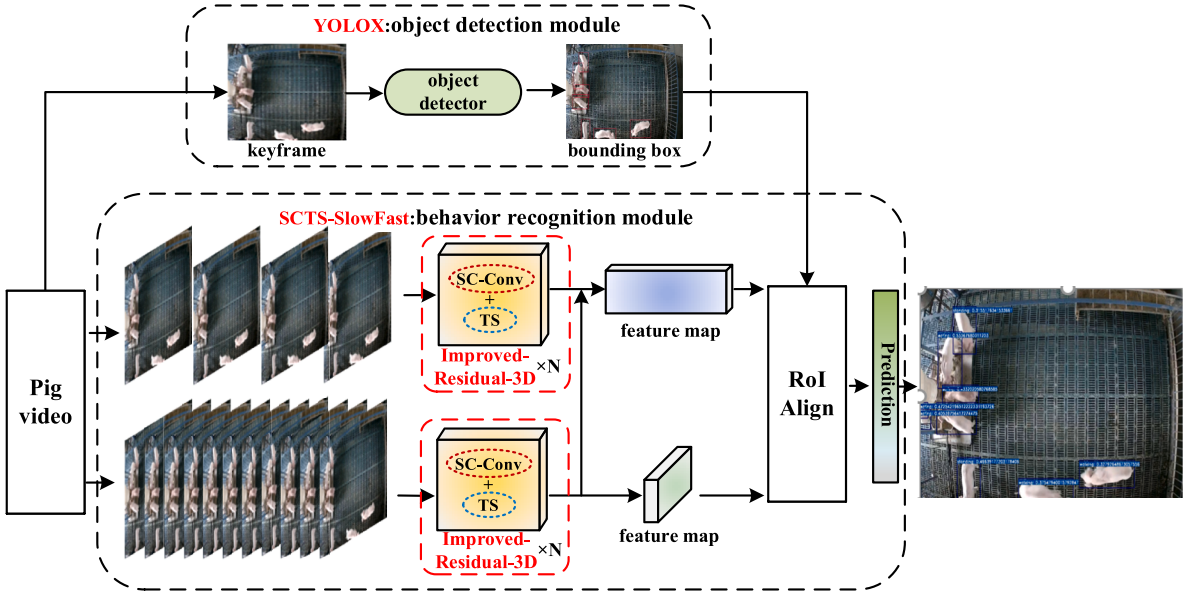

Fig. 2. The schematic of the proposed multi-behavior detection method.

and Temporal-Spatial (TS) attention mechanism are integrated into Residual-3D block (Hara et al., 2018), i.e., Improved-Residual-3D block, to enhance behavior features extraction capability in spatial and temporal dimension, which is described detailly in following sections. Next, the located regions are mapped to the feature map of fixing size with RoI Align (He et al., 2017), and recognizing behavior categories and locating corresponding location of group-housed pigs are achieved by prediction unit.

3.2. YOLOX object detection module

In this study, YOLOX is selected as the object detection module for locating location of group-housed pigs, which is a real-time single-stage object detection algorithm. YOLOX consists of three parts: CSPDarkNet backbone network, PANet, and Decoupled head blocks, which are used for feature extraction, enhanced feature extraction, and multi-scale prediction, respectively (Liu et al., 2023). The structure of YOLOX is shown in Fig. 3.

YOLOX combines the merits of YOLO series (Redmon and Farhadi, 2018; Wu et al., 2021), employing Focus and CSPNet in the backbone network to efficiently extract and analyze image features, and introducing a decoupled structure into the detection head to more accurately and quickly detect objects of different sizes. Unlike traditional anchor-based detector, YOLOX adopts an anchor-free design and incorporates the simOTA dynamic matching mechanism to improve the detection accuracy, convergence rates and reduce parameter redundancy, which achieves an efficient detection performance (He et al., 2023). Particularly, the test accuracy of group-housed pig detection can reach 98.7 % on the object detection dataset established in this study.

3.3. SCTS-SlowFast behavior recognition module

In this study, the behavior categories of group-housed pigs were classified by SCTS-SlowFast, which extended from SlowFast network (Feichtenhofer et al., 2019). SCTS-SlowFast extracts spatiotemporal features through a single stream structure that operates at two different framerates, i.e. Slow pathway and Fast pathway. Slow pathway, operating at low framerate, to capture spatial semantics features, while Fast pathway operating at high framerate, to capture temporal semantics features, as illustrated in Fig. 4, where T represents the temporal

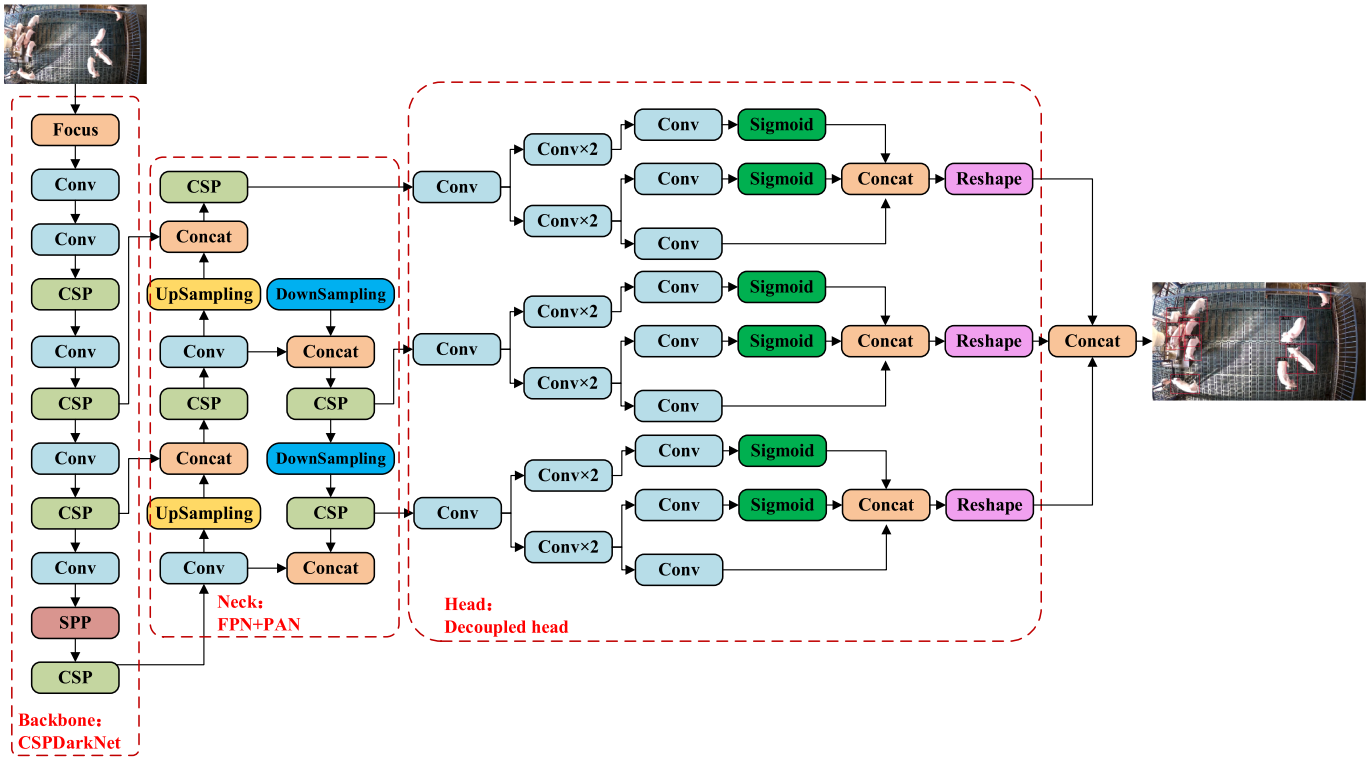

Fig. 3. The structure of YOLOX.

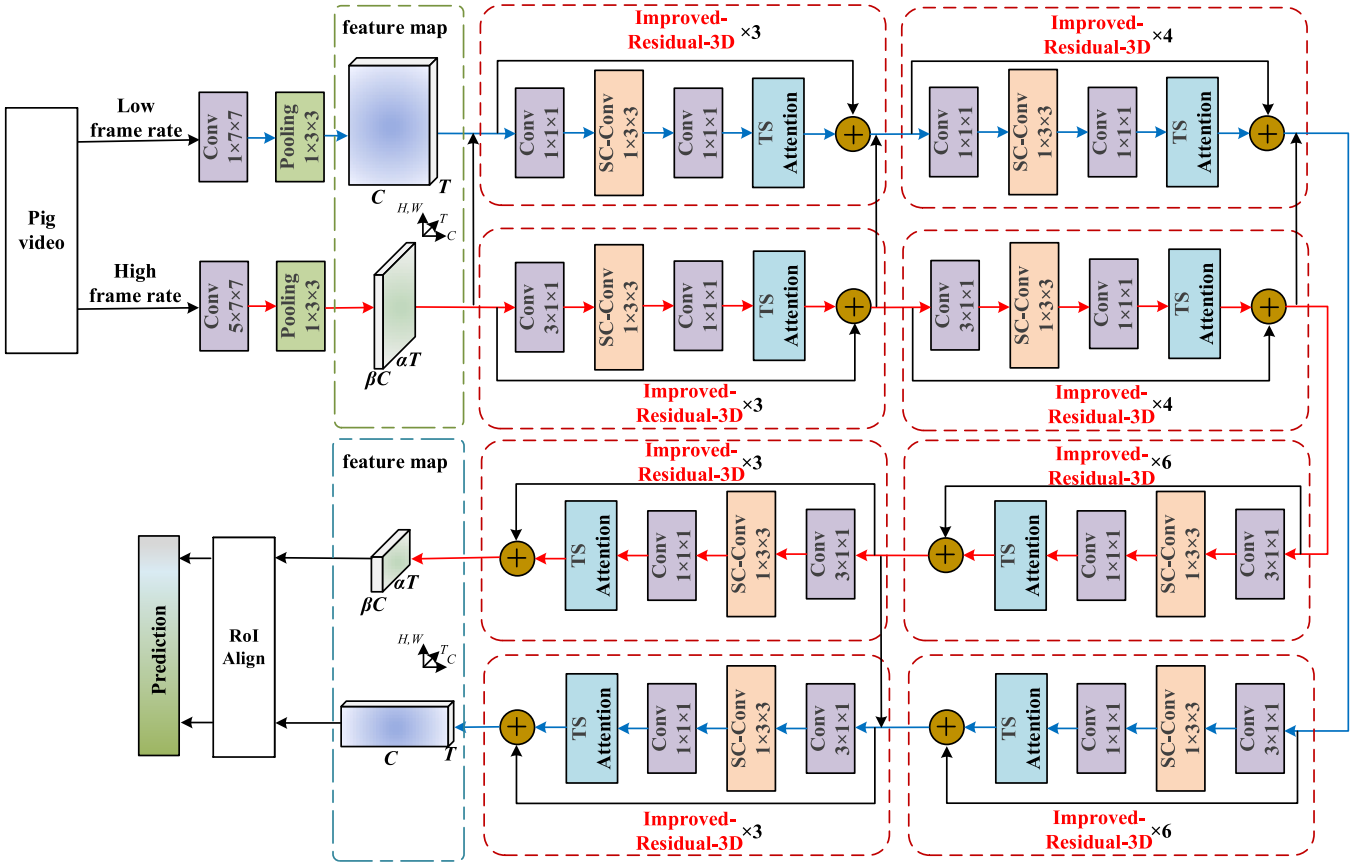

Fig. 4. The structure of SCTS-SlowFast.

resolution, H and W represent the spatial resolution, and C represents the number of channels. α is the frame rate ratio between the Fast and Slow pathways. β is the ratio of channels between the Fast and Slow pathways. The blue line represents the Slow pathway, and the red line represents the Fast pathway.

Firstly, to improve modeling ability of the behavior recognition module, the Self-Calibrated Convolution (SC-Conv) and Temporal-Spatial (TS) attention mechanism are specially integrated into Residual-3D block, i.e., Improved-Residual-3D block. Particularly, SC-Conv can adaptively build long-range spatiotemporal information and inter-channel dependencies, and help recognition module to generate more discriminative representations by incorporating richer information. Meanwhile, in order to improve the temporal and spatial features extraction performance of the recognition module, TS attention mechanism is designed to introduce into the Residual-3D. Next, the lateral connection is used for fusing the extracted feature from Fast pathway to Slow pathway with transforming the temporal dimension. Finally, the features of two pathways combine with the detected features of object detector through RoI Align, and behavior categories are recognized by prediction unit.

3.3.1. SC-Conv

SC-Conv (Self-Calibration Convolution) is different from the standard convolution, which utilizes a small-size convolution kernel to build long-range spatiotemporal information and inter-channel dependencies. Specifically, SC-Conv is conducted convolutional feature transformation on two different scales. One is the same resolution as the original scale space, the other gets a small latent space after down-sampling. The embeddings after transformation in the small latent space are used as references to guide the feature transformation process in the original feature space because of their large fields-of-view (Liu et al., 2020). This study extends the 2D convolution in the original SC-Conv to 3D convolution for effectively retaining the high-dimensional behavioral information (Tran et al., 2015; Guo et al., 2019). The structure of SC-Conv is shown in Fig. 5, "C×T×H×W" is the feature size and "r" is the down-sampling step size.

3.3.2. TS attention mechanism

Temporal-spatial (TS) attention mechanism is composed of temporal attention mechanism and spatial attention mechanism which can help the module automatically focus on the pig behavior characteristics in spatial and temporal dimension, this study adopts the structure of T temporal attention mechanism before S spatial attention mechanism, as shown in Fig. 6.

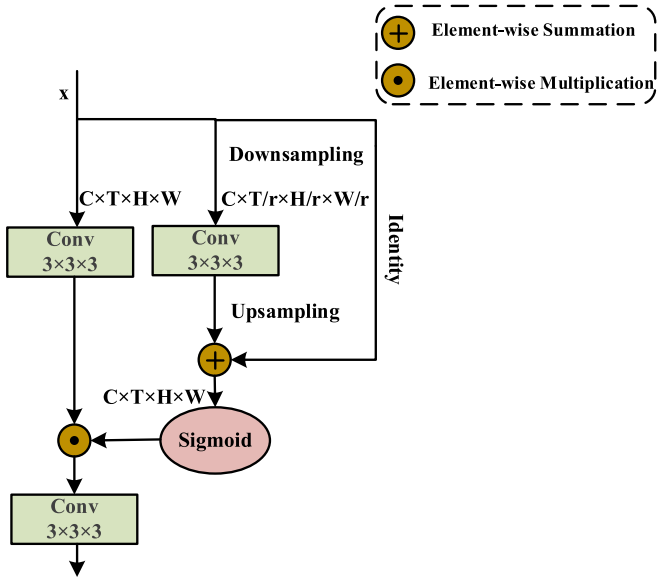

Fig. 5. The structure of SC-Conv.

T temporal attention mechanism. The input feature map is first squeezed the channel dimension and the spatial dimension by the Global Average Pooling operation, which does not reduce the temporal dimension. Next, through one-dimensional convolution, cross-dimensional interaction of squeezed features in temporal information is achieved. Finally, after sigmoid function, the temporal attention map is generated, and perform elementwise multiplication with input feature map to generate features required by the S spatial attention mechanism. The expression of T temporal attention mechanism is shown in Equation (1)-(2).

$$\widehat{F} = \sigma(conv(GAP(F))) \tag{1}$$

$$w_{attT} = \widehat{F} \bullet F \tag{2}$$

where F is the input feature, $F \in R^{C\times T\times H\times W}$, GAP denotes the Global Average Pooling, σ denotes the sigmoid function, $\widehat{F}$ denotes the input feature F after Global Average Pooling, convolutional layer, and sigmoid function. w_{attT} denotes the output feature map.

S spatial attention mechanism. The output feature map of temporal attention mechanism serves as the input of the spatial attention mechanism. In contrast to the squeezing approach of temporal attention mechanism, spatial attention mechanism first squeezes the channel dimension and temporal dimension by Average Pooling and Max Pooling. Next, the two squeezed features are fused, and the spatial attention map is generated by convolution layer and sigmoid function. Finally, Temporal-Spatial attention map is generated by perform elementwise multiplication with temporal attention and spatial attention maps. The expression is shown in Equation (3)-(6).

$$w_{attM} = MaxPool(w_{attT}) \tag{3}$$

$$w_{attA} = AvgPool(w_{attT}) \tag{4}$$

$$w_{attS} = \sigma(conv[w_{attM}; w_{attA}]) \tag{5}$$

$$w_{att} = w_{attS} \bullet w_{attT} \tag{6}$$

where w_{attM} and w_{attA} represent the output features after Max Pooling and Average Pooling of w_{attT}. w_{attS} denotes the output spatial attention map. w_{att} is Temporal-Spatial attention map.

4. Results and analysis

4.1. Experimental settings

The experiment platform of this study is completed under the Windows 11 system and use the programming language of Python 3.10, The batch size and initial learning rate for training are set as 18 and 0.01, the ratio of channels between the Fast and Slow pathways (β) is 0.125, the frame rate ratio (α) is 8, the specific experimental environment are shown in Table 2.

4.2. Evaluation metrics

In this study, the *AP* value of each behavior and the total *mAP* value are adopted to evaluate performance of the proposed group-housed pigs multi-behavior detection method. AP_i is the area under the curve of P_i and R_i, which is an important indicator that reflects the detection performance. The AP_i calculating method is shown in equation (7)-(9). P_i is precision, which measures the correctness of the model for the identified behavior, R_i is recall, which judges the coverage of the model for the detection results of the identified behavior. *TP* is true positive cases, which is the number of correctly detected pig behaviors. *FP* is false positive cases, which is the number of incorrectly detected other

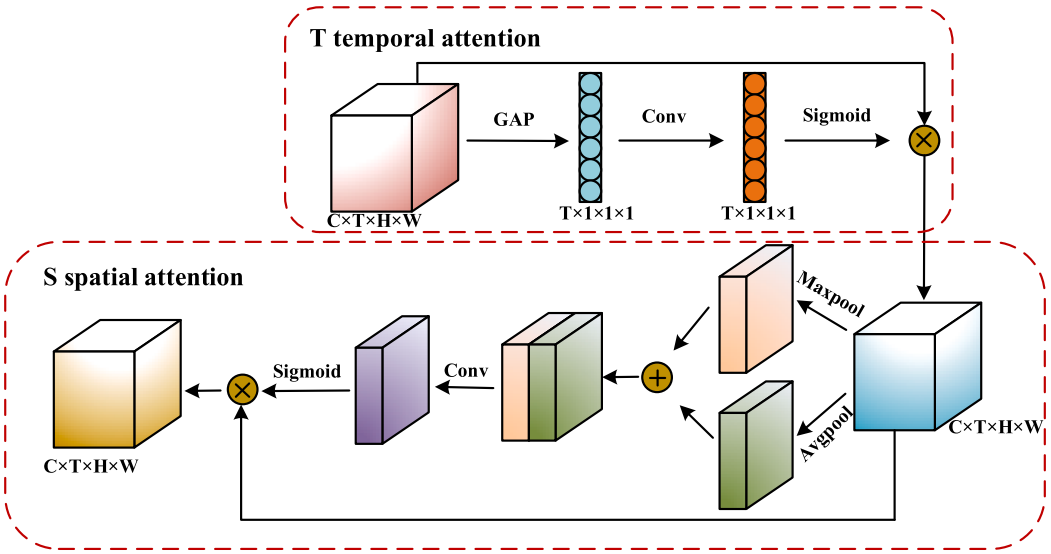

Fig. 6. The structure of TS attention mechanism.

Table 2
Experimental environment parameter.

Parameter settings	Version
CPU	Intel(R) Core (TM) i9-13900KF
GPU	NVIDIA GTX 4090
RAM	DDR5 32G*4
PyTorch	2.0.1

behaviors as positive behaviors. *FN* is false negative cases, which is the number of incorrectly detected pig behaviors. The calculating method of *mAP* is shown in equation (10), which is the average *AP* of k (k = 4) classes of the group-housed pig behaviors.

$$P_i = \frac{TP}{TP + FP} \times 100\% \tag{7}$$

$$R_i = \frac{TP}{TP + FN} \times 100\% \tag{8}$$

$$AP_i = \int_0^1 P_i(R_i) dR_i \tag{9}$$

$$mAP = \frac{1}{k} \sum_{i=1}^{k} AP_i \tag{10}$$

4.3. Experimental results and analysis

In order to verify the effectiveness of behavior detection, this experiment tested performance of the proposed method by the testing set, and the AP values of detecting four behaviors reached 99.10 % (EI), 90.18 % (DI), 75.15 % (ST) and 55.76 % (WK), The *mAP* value of four behavior detections reached 80.05 %, indicating that the proposed method realized the effective detection of group-housed pig behaviors. Fig. 7 is the detection schematic of four behaviors by proposed method which recognize the behavior in single field of view and locate pig locations with bule bounding box. Particularly, the AP values of ST and WK behaviors was significantly lower than other behaviors. Similarly, the confusion matrix (Fig. 8) of the experiment showed a similar phenomenon. Combined with group-housed pig behavior videos, although the proposed method had high performance in the detection of EI and DI behaviors, it could be seen from the confusion matrix that EI, DI, ST and WK behaviors still had errors. Particularly ST and WK behaviors were more prone to misclassification. The reason is that the group-housed pigs video data of this study were filmed at an overhead perspective for reducing negative impact in occluding between group-housed pigs, which led to difficultly in capturing information about the pig leg movements. Because the characteristics of group-housed pig behaviors are similar in spatial dimension, this leads to errors in the detection of group-housed pig behaviors, especially standing and walking behaviors, using the proposed method. In addition, small local movements in some video data are also difficult to effectively distinguish. This error example is described in the following section, and it is also a challenge that we will focus on solving in the future.

In addition, this experiment also explored the detection effect of the method in nighttime case. 20 segments of 10-second nighttime videos were additionally extracted from the original data to test the detection performance of model in nighttime. The results were that the AP values of the four behaviors are 98.73 % (EI), 87.02 % (DI), 74.62 % (ST) and 52.47 % (WK), respectively, with a mAP value of 78.21 %, the detection result were shown in Fig. 9.

Although the proposed method still had strong effectiveness at nighttime, the detection performance for the four behaviors of group-housed pigs was still affected. This is because that behavior videos captured at night appear blurry due to lack of light, and there are errors in the behavior features extraction of pigs. This is a significant challenge for behavior detection tasks. In the future work, nighttime enhancement technology will be introduced to improve the nighttime detection performance and robustness.

4.4. Comparison with object detection-based behavior detection methods

Some researchers have tried to achieve the task of multi-behavior detection by strategy of object detection, and made some progress. Particularly, YOLOv5 has been widely used for pig behavior detection task (Kim et al., 2022; Tu et al., 2024; Luo et al., 2024), In addition, YOLOv5 has also carried out related work on other animal behavior detection tasks (Zeng et al., 2023; Du et al., 2023; Subedi et al., 2023). YOLOv5 is known for its balanced performance between speed and accuracy through coupled head, while YOLOX prioritizes detection accuracy, particularly for smaller objects, through its anchor-free approach and decoupled head design. To verify that the proposed method is more suitable for detecting group-housed pig behaviors, this experiment selects YOLOv5 to compare with the proposed method. Specifically, the same video data used for establishing behavior recognition dataset was selected to construct the dataset for training and testing YOLOv5 detector. The video data was extracted at one frame per second, and a total

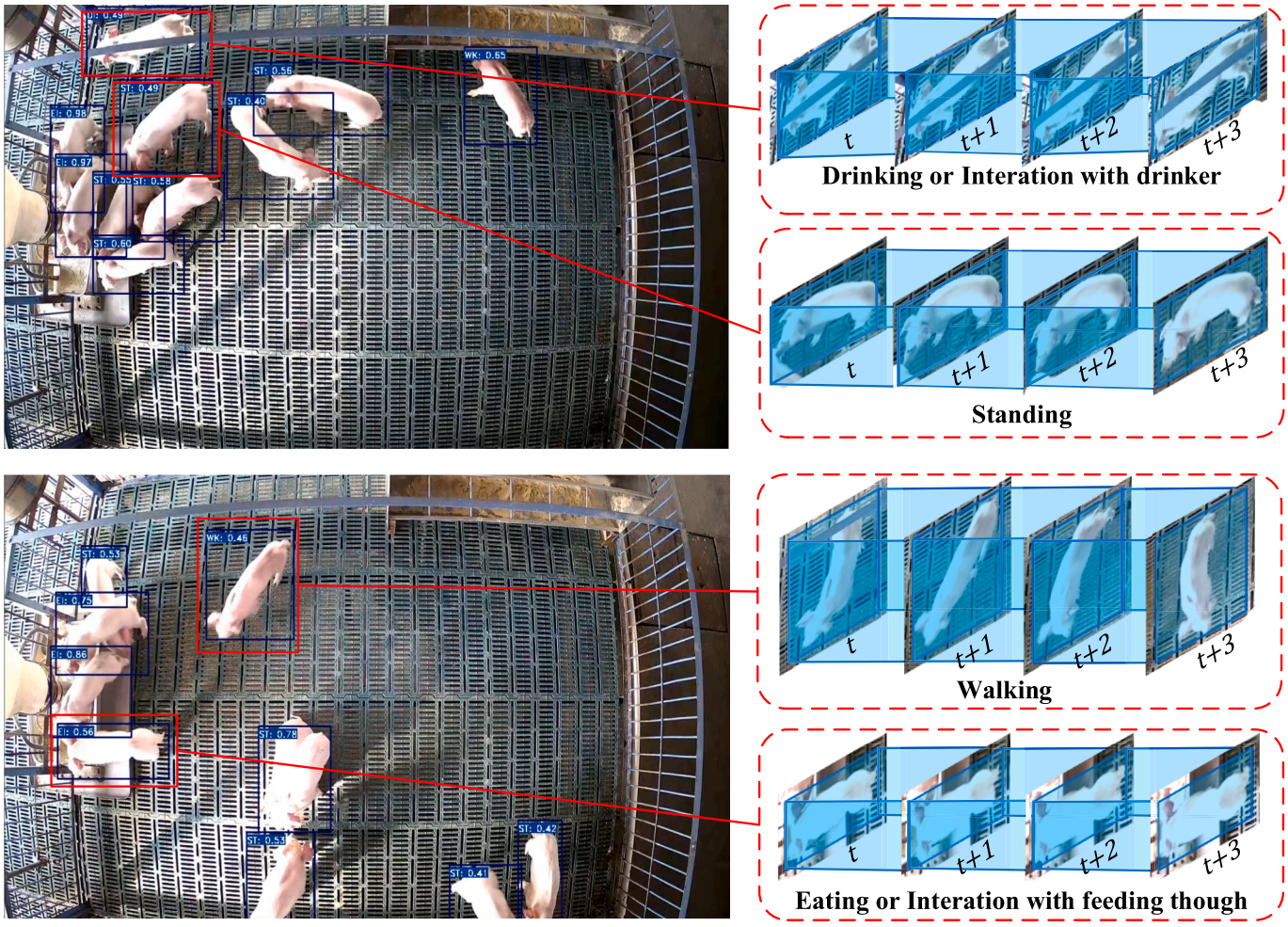

Fig. 7. The detection results of four behaviors.

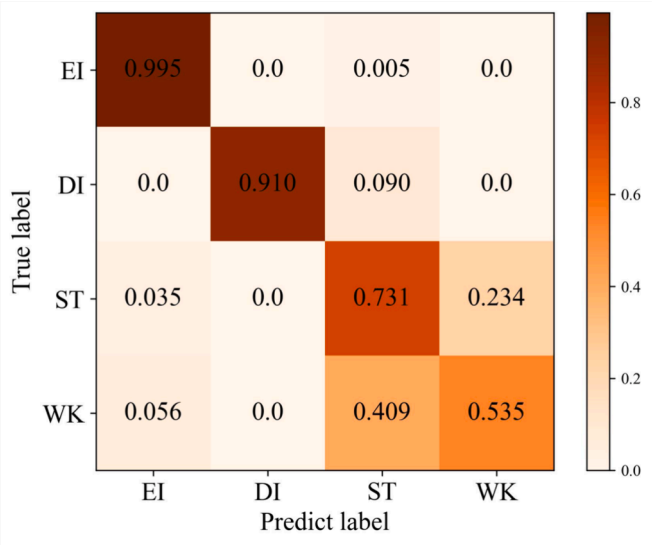

Fig. 8. The confusion matrix of four behaviors detection results.

of 3738 images were obtained. These images were annotated behaviors of group-housed pigs by labelImg tool, and these images were divided into training and testing sets in an 8:2 ratio, comparison results as illustrated in Table 3.

From Table 3, although YOLOv5 has a smaller model size and Params, the proposed method has significant advantages in the detection performance of four behaviors. Compared with YOLOv5, the *mAP* value was improved by 20.45 %, the most obvious advantages were in the detection performance of DI and WK behaviors, which were 35.58 % and 33.96 % better than YOLOv5. There were also increase in the *AP* values of EI and ST which was 4.4 % and 7.75 %. This is because YOLOv5 only focuses on spatial information, lacking the ability to learn temporal information. In the actual rearing environment, the different behaviors of group-housed pigs have similar characteristics. It is difficult to distinguish pig behaviors only by considering spatial information. The proposed method considers both spatial and temporal information, and efficiently recognizes pig behaviors. It is proved that the proposed method is more suitable for the multi-behavior detection of pigs in a group-housed environment.

4.5. Comparison with other behavior detection methods

To further validate the effectiveness of the proposed method in this paper, the ACRN (Sun et al., 2018) and VideoMAE (Tong et al., 2022) were specifically selected for comparative experiments. Among them, both ACRN and the proposed method use 3D convolution for model construction, but ACRN adopts classical S3D-G backbone network, the backbone of proposed method is Improved-Slowfast-R50. VideoMAE is modeled by the transformer which is currently more novel than 3D convolution network. This experiment trained and tested the above models by using the created datasets from this study to verify that the

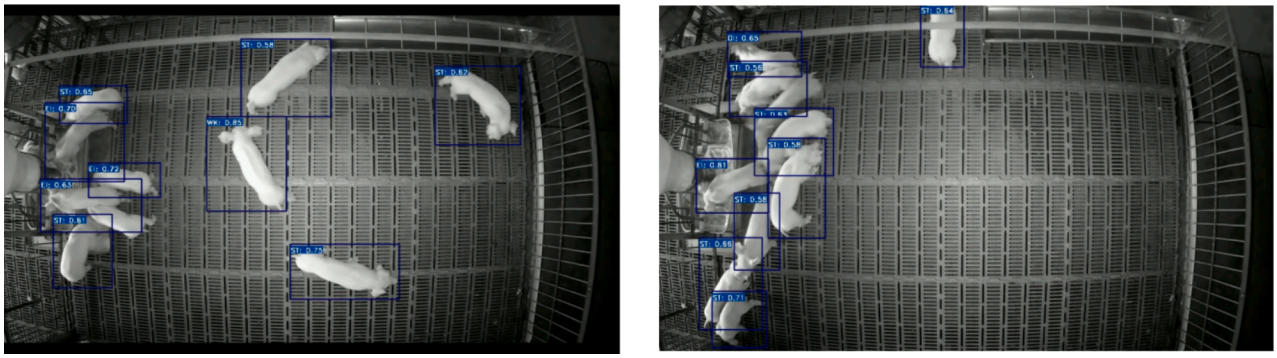

Fig. 9. The detection results of four behaviors in nighttime.

Table 3
Comparison results between YOLOv5 and SCTS-SlowFast.

Model	*mAP*	EI (*AP*)	DI (*AP*)	ST (*AP*)	WK (*AP*)	Params(M)	Model Size (M)
YOLOv5	59.60 %	94.70 %	54.60 %	67.40 %	21.80 %	**7**	**54**
SCTS-SlowFast	**80.05 %**	**99.10 %**	**90.18 %**	**75.15 %**	**55.76 %**	77	304

proposed method has advantages over traditional 3D convolution networks and currently novel transformer models in the task of multi-behavior detection of group-housed pigs, the results are shown in Table 4.

As can be seen from Table 4, the *mAP* value of the proposed method was 80.05 %, which was higher than ACRN and VideoMAE by 4.39 % and 3.27 %. Meanwhile, the model size of proposed method was also smaller than two methods, 367 M and 489 M. And the Params also an advantage of 61 M and 9 M. In the recognition performance of the four behaviors, the proposed method also had an obvious advantage in the recognition of DI, ST and WK behaviors. Compared with the other two methods, the *AP* values of DI were improved by 9.01 % and 6.02 %, the *AP* values of ST were improved by 2.53 % and 2.93 % and the *AP* values of WK were improved by 5.28 % and 3.56 %. There was also a small increase in the *AP* value of EI which was 0.72 % and 0.55 %.

Although ACRN and VideoMAE had an efficient performance in detecting human behavior, due to the complexity of rearing environment, the above two behavior detection methods still have some challenges in detecting pig behaviors. Our method specially introduces SC-Conv and TS attention mechanism to improve the behavior features extraction capability in spatial and temporal dimension, and has a better performance in detecting pig behaviors by comparing other two methods. It also proves that the proposed method has an advantage to detect the multi-behaviors of group-housed pigs. Fig. 10 shows the detection schematic of different behavior detection methods. Where red boxes are examples of false detections and missed detections.

4.6. Effectiveness analysis of different improvement strategies

This experiment evaluates the effect of different improvement strategies on multi-behavior detection performance of group-housed pigs by the created datasets from this study, the result was shown in Table 5.

The improvement of introducing SC-Conv was shown in the second row of Table 5, compared with the original SlowFast, the *AP* values of detecting DI and ST are more significantly improved which increased 5.75 % and 2.1 %, the detection performance of EI behavior was also improved by 0.11 %. Although the *AP* value of WK behavior detection decreased by 0.9 %, there was a significant increase in the *mAP* value for the four behaviors detection, which increases 1.76 %. This was because SC-Conv focus on the dependencies between spatiotemporal and channel information, and reduces the performance in temporal feature extraction, in which it leads to degrade detection performance in WK behavior, but improves the detection performance in EI, DI and ST behaviors. The third row of Table 5 was the result of introducing the TS attention mechanism. It was shown that the detection performance four behaviors were improved by comparing with the SlowFast, in which the *AP* value increase 0.31 %, 6.91 %, 0.73 % and 3.07 %. This is because TS attention mechanism enhanced the ability of behavior features extraction, which helps the SCTS-SlowFast behavior recognition module automatically focus on the pig behavior characteristics in spatial and temporal dimension. The fourth row of Table 5 was the result of the proposed method which integrates the SC-Conv and the TS attention mechanism. From the table, it can be seen that the *AP* values of detecting four behaviors reached 99.10 % (EI), 90.18 % (DI), 75.15 % (ST) and 55.76 % (WK). The *mAP* value of four behavioral detections reached 80.05 %. Compared with the original SlowFast, the *AP* values of the four behaviors are improved by 0.38 %, 6.97 %, 4.27 %, 3.79 %, and the *mAP* value was also improved by 3.84 %.

4.7. Comparison with different attention mechanisms

In this experiment, the proposed attention mechanisms of different combinations were compared with TS (Temporal-Spatial) attention mechanism through the created datasets from this study. In addition, this experiment also compared the proposed TS attention mechanism with CBAM (Woo et al., 2018) and ECA (Wang et al., 2020) attention mechanisms, which have similar structures to TS, to validate the effectiveness of the TS attention mechanism proposed in this study, the comparison results were shown in Table 6. Where T (Temporal attention) denotes only temporal attention mechanism, S (Spatial attention)

Table 4
The comparative results of different behavior detection methods.

Model	*mAP*	EI (*AP*)	DI (*AP*)	ST (*AP*)	WK (*AP*)	Params(M)	Model Size (M)
ACRN	75.66 %	98.38 %	81.17 %	72.62 %	50.48 %	138	518
VideoMAE	76.78 %	98.55 %	84.16 %	72.22 %	52.20 %	86	333
SCTS-SlowFast	**80.05 %**	**99.10 %**	**90.18 %**	**75.15 %**	**55.76 %**	**77**	**304**

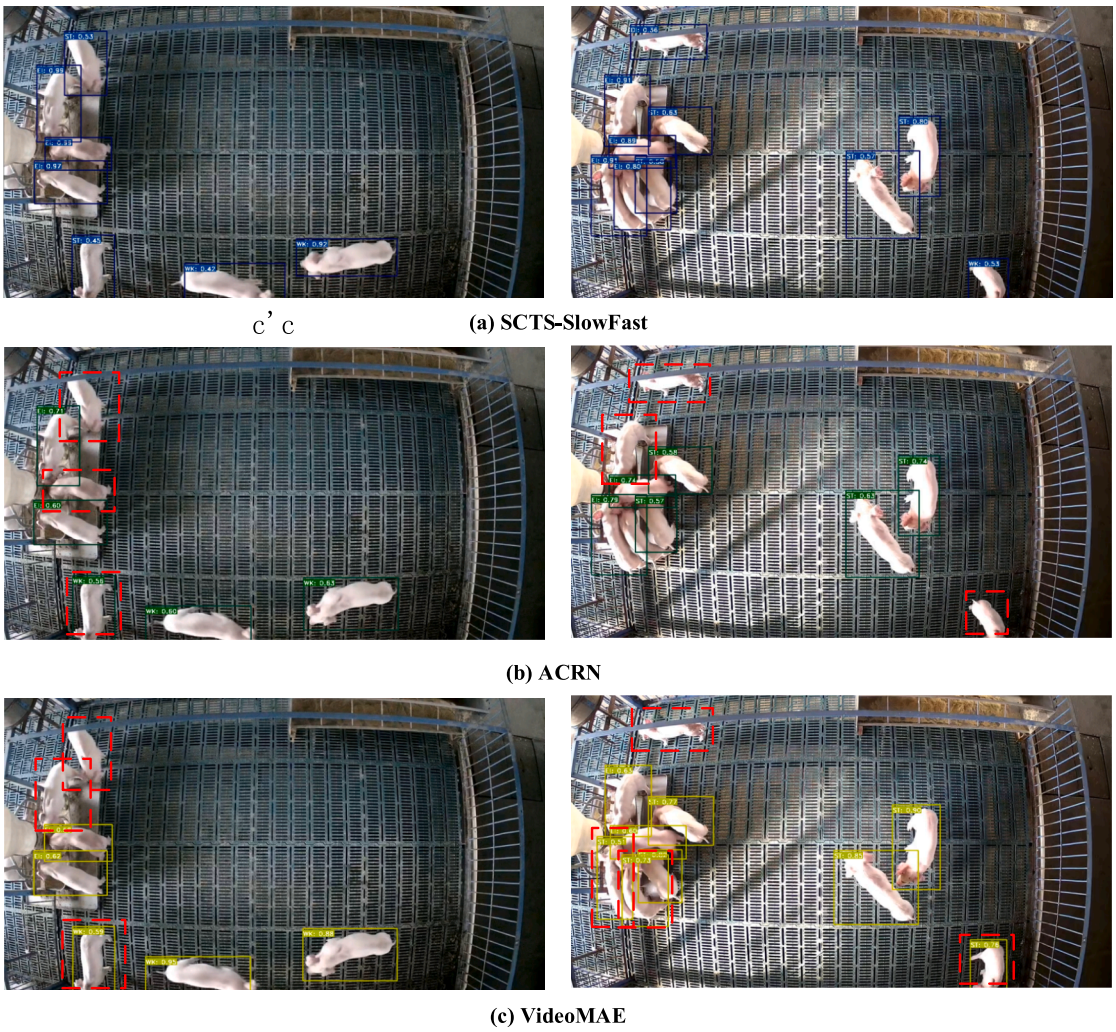

Fig. 10. Detection result of different behavior detection methods: (a) the detection result of SCTS-SlowFast, (b) the detection result of ACRN, (c) the detection result of VideoMAE.

Table 5
Comparison results of different improvement strategies.

Model	*mAP*	EI (*AP*)	DI (*AP*)	ST (*AP*)	WK (*AP*)
SlowFast	76.20 %	98.72 %	83.21 %	70.88 %	51.97 %
SlowFast + SC-Conv	77.96 %	98.83 %	88.96 %	72.98 %	51.07 %
SlowFast + TS	78.95 %	99.03 %	90.12 %	71.61 %	55.04 %
SlowFast + SC-Conv + TS	**80.05 %**	**99.10 %**	**90.18 %**	**75.15 %**	**55.76 %**

Table 6
Comparative results for different attention mechanisms.

Model	*mAP*	EI (*AP*)	DI (*AP*)	ST (*AP*)	WK (*AP*)
SlowFast + CBAM	77.14 %	98.48 %	84.61 %	71.48 %	53.98 %
SlowFast + ECA	77.47 %	98.14 %	85.72 %	71.43 %	54.61 %
SlowFast + T	77.27 %	98.61 %	84.53 %	71.77 %	54.18 %
SlowFast + S	78.10 %	98.90 %	89.42 %	71.95 %	52.12 %
SlowFast + ST	78.74 %	98.77 %	89.38 %	**73.11 %**	53.71 %
SlowFast + TS	**78.95 %**	**99.03 %**	**90.12 %**	71.61 %	**55.04 %**

denotes only spatial attention mechanism, ST (Spatial-Temporal attention) denotes spatial attention mechanism followed by temporal attention mechanism, and TS (Temporal-Spatial attention) denotes temporal attention mechanism followed by spatial attention mechanism.

From Table 6, it can be seen that there was the best performance by using the structure of temporal attention mechanism before the spatial one, in which the *mAP* value was higher than the detection results by using CBAM, ECA and other structures attention mechanism, it was reached 78.95 %. The *AP* values of the four behaviors through using TS attention mechanism were higher than those using CBAM by 0.55 %, 5.51 %, 0.13 % and 1.06 %, it also was improved by 0.89 %, 4.40 %, 0.18 % and 0.43 % through comparing with ECA attention mechanism. This is mainly because CBAM and ECA focus on channel or spatial features, while TS attention mechanism focus on features in temporal and spatial dimension. In the behavior recognition task of group-housed pigs, it is difficult to accurately distinguish complex behaviors in lacking of temporal dimension feature. Therefore, TS attention mechanism is more advantageous than CBAM and ECA. Compared with other structure of the proposed attention mechanism, the *mAP* value was improved by 1.68 %, 0.85 % and 0.21 %. Although, in the detection of standing behavior, the *AP* value was 1.5 % lower than the detecting result by using the ST structure, the recognition performance of the other three behaviors had obvious advantages. Overall, the proposed TS attention mechanism outperforms the other attention mechanisms, it can more effectively improve the detection performance of pig behaviors in a rearing environment.

4.8. Statistical analysis of daily behavior of group-housed pigs

To verify the effectiveness of the proposed method, this experiment selects a period of 08:00–16:00 in one day to monitor the behaviors of 35–42 days age piglet. Based on the behavior characteristics of pigs, that the duration of pig feeding varies from one minute to 28.3 min (Fornós et al., 2022), this experiment adopts 30 min intervals to count the duration of behaviors. The variation trends of four behaviors of group-housed pigs are shown in Fig. 11, the horizontal coordinate is the time, and the vertical coordinate is the behavior duration of all group-housed pigs during this period.

From Fig. 11, it can be seen that the group-housed pigs were dominated by EI behavior and ST behavior in the most of the time and the duration of DI behavior was the least among the one-day. Among them, the duration of EI behavior was highest between 11:30 and 12:00. In behavior study of pigs, group-housed pigs can exhibit relatively synchronized feeding behavior, which tends to be particularly pronounced when pigs are hungry or interested in feed. Combining video clips, when some individuals in the group start feeding, others are stimulated to join the feeding activity (Zwicker et al., 2015). It led to a peak in the duration of feeding behavior during this period. The duration of WK behavior was relatively longer in the 10:00–11:30 and 15:00–16:00, and we found from this period video that there was a period of aggressive and chasing behaviors between group pigs. Since the test segment was the mixed groups period video, there was a high probability of aggressive and chasing behaviors after mixing (Buettner et al., 2015). When there was aggressive or chasing behavior among group-housed pigs, WK behavior was recognized in this time segment. This is also the reason for the peak in walking behaviort. In the future, we will further research this type of interaction and fast behaviors.

Furthermore, to validate the effectiveness of this method in monitoring the behavior of group-housed pigs. Three experts were invited to conduct manual observations and statistical analysis of the monitoring videos, which were cross-referenced with the predictions by the proposed method. Fig. 12 shown the comparison between the human observation results and the predicted results of four behaviors, the horizontal coordinate is the time, and the vertical coordinate is the behavior duration of group-housed pigs during this period.

From the Fig. 12, the predicted curves of EI and DI exhibited relatively small errors compared to the manually observed true curves. Within each 30-minute, the errors between the predicted values of EI, DI and the true values observed manually were −0.18 min and −0.08 min. The errors between the predicted values of ST and WK and their manually observed true values were + 3.25 min and −1.23 min. Although the relatively larger monitoring errors for ST and WK behaviors compared to EI and DI, the monitoring errors for each behavior within every 30-minute were all less than 4 min. The above results verify that the proposed method can efficiently monitor and analyze the pig daily behaviors in rearing environment, it also provides a data basis for promoting the development of the pig farming industry.

5. Error analysis and discussion

Fig. 13 shows two examples of false detection by the proposed methods. In Fig. 13 (a), a standing pig was incorrectly detected as walking behavior. In this video segment, there was a slight change in the head and legs of pig, but the overall spatial position was not significantly displaced, it did not meet the criteria for walking behavior. Therefore, when the body of pig changes but the position did not change, the proposed method still had difficult to distinguish the behaviors between walking and standing. This was the main reason that the *AP* value of recognizing walking and standing behaviors was relatively low. In the future research, we will adopt the multi-view technique to collect leg and body information, and combine with leg movement features and body spatial features to distinguish walking and standing behaviors.

In addition, Fig. 13 (b) shows a drinking pig which was incorrectly detected as standing behavior. This was because that the problem of occluding between the pig barn equipment was difficult to avoid due to collect video data in overhead perspective. In the Fig. 13 (b), the steel beams of the pig barn just covered drinker and the head of the recognized pig, and the pig body did not obviously move in spatial position. Therefore, the proposed method cannot capture information about the drinking behavior of pig and lead to be misclassified as standing behavior. In the future research, we will adopt the way with the least external interference for video data acquisition as much as possible.

In the rearing environment, it easily poses a great difficult for video data acquisition because of the pig occlusion. Although this study used an overhead perspective video data acquisition method to avoid those

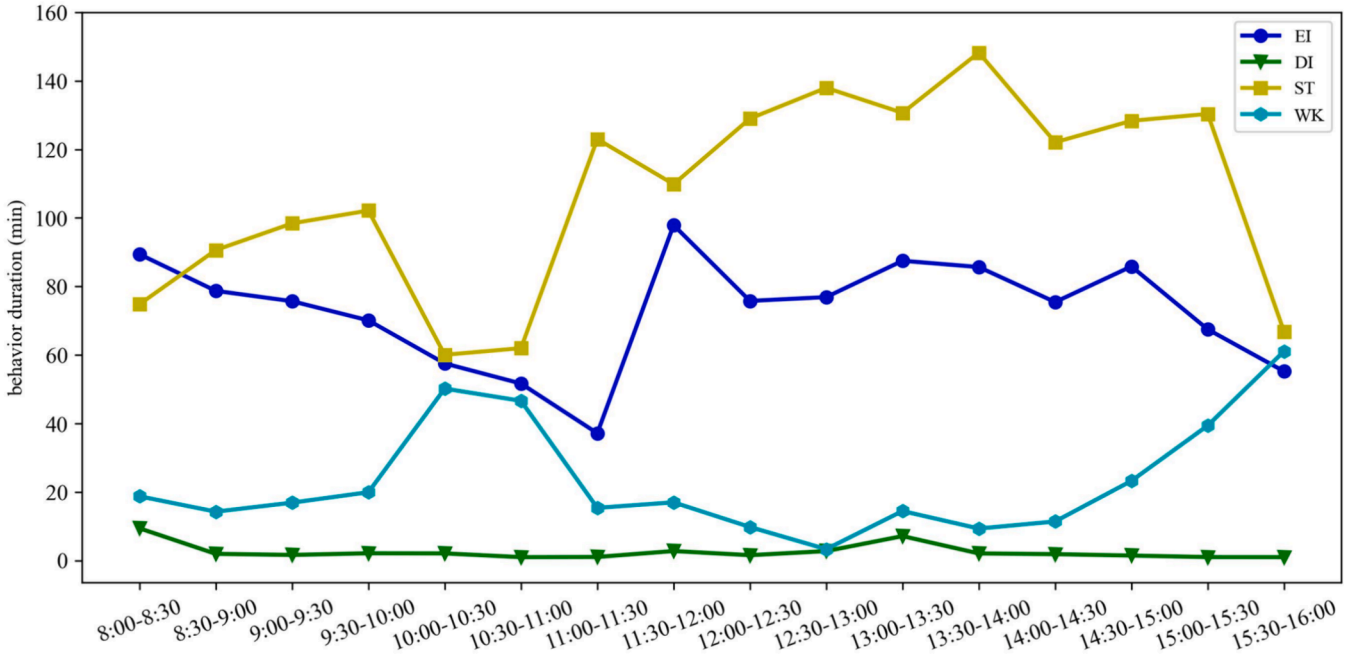

Fig. 11. The variation trends of four behaviors detection results during one day.

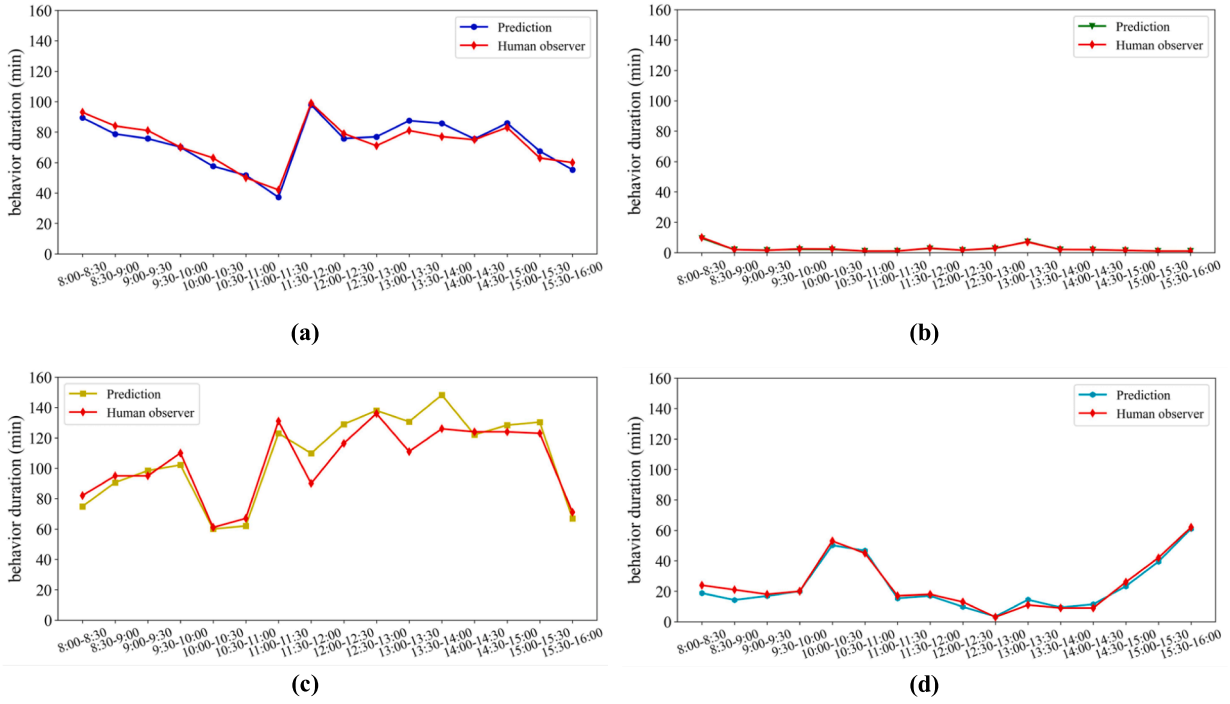

Fig. 12. Comparison between the human observation results and the predicted results: (a) the curve of EI behavior, (b) the curve of DI behavior, (c) the curve of ST behavior, (d) the curve of WK behavior.

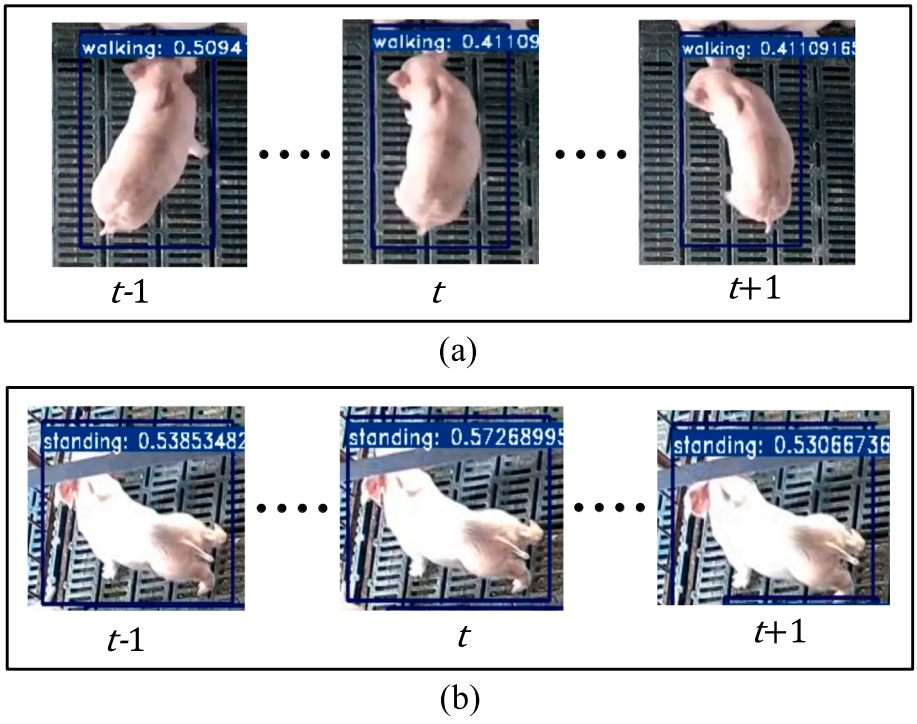

Fig. 13. Examples of false detection:(a) standing behavior is falsely detected as walking behavior, (b) drinking behavior is falsely detected as standing behavior.

issue, it also resulted in loss the leg movement information of pigs. This is the reason why it is challenging to accurately distinguish similar behaviors of group-housed pigs. Therefore, in the future, the effective occlusion-resistant module will be introduced in the proposed method to avoid the situations of false detection because of pigs and equipment occlusion. Additionally, several improvements will be considered to enhance the usability of the proposed method in pig farms. Specifically, this study will design cross-age detection models according to different body weights and sizes of group-housed pigs, and introduce object tracking strategies to realize behavior monitoring and tracking of pigs at different periods. Furthermore, a lightweight structure will be designed to reduce the parameters in the behavior detection model, while ensuring the practicality of the proposed method in actual pig farms.

6. Conclusion

To automatically monitor multiple typical behaviors of group-housed pigs in one scene, this study proposed a multi-behavior detection method for group-housed pigs based on YOLOX object detection module and SCTS-SlowFast behavior recognition module, it can effectively recognize four behaviors (Eating or Interaction with feeding though, Drinking or Interaction with drinker, Standing and Walking) of group-housed pigs and locate the corresponding locations of pigs. YOLOX can accurately extract the locations of group-housed pigs. SCTS-SlowFast is able to efficiently recognize multiple behaviors of group-housed pigs. Based on multi-behavior dataset of group-housed pigs, the proposed method achieved a mAP value of 80.05 %, and the AP value in four behaviors is 99.10 % (EI), 90.18 % (DI), 75.15 % (ST) and 55.76 % (WK). This study also verifies the effectiveness of the proposed improvement and proves that the detection performance of proposed method is better than existing behavior detection methods through the comparison experiment. In addition, the duration of group-housed pig behaviors and their corresponding changing trends during a day is statistically analyzed by the proposed mothed. It verifies the feasibility of proposed method in automatically monitoring multiple pig behaviors in the group-housed environment, and provides a scientific basis for improving the production efficiency of pig farms. In the future, the occlusion-resistant module will be added to improve the generalization and accuracy of the method, while cross-age detection model, object tracking strategy and lightweight structure will be designed to enhance the practicality of the proposed method in pig farms.

CRediT authorship contribution statement

Ran Li: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Baisheng Dai:** Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yuhang Hu:** Validation, Resources. **Xin Dai:** Software, Funding acquisition. **Junlong Fang:** Validation, Resources. **Yanling Yin:** Validation, Software, Resources. **Honggui Liu:** Funding acquisition, Conceptualization. **Weizheng Shen:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared my dataset on github

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 31902210 and 32172784, in part by the Natural Science Foundation of Heilongjiang Province under Grant QC2018074 and YQ2023C012.

References

Alameer, A., Kyriazakis, I., Dalton, H.A., et al., 2020. Automatic recognition of feeding and foraging behaviour in pigs using deep learning[J]. Biosyst. Eng. 197, 91–104.
Buettner, K., Scheffler, K., Czycholl, I., et al., 2015. Network characteristics and development of social structure of agonistic behaviour in pigs across three repeated rehousing and mixing events[J]. Appl. Anim. Behav. Sci. 168, 24–30.
Chen, C., Zhu, W., Steibel, J., et al., 2020. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory[J]. Comput. Electron. Agric. 169, 105166.
Du, L., Lu, Z., Li, D., 2023. A novel automatic detection method for breeding behavior of broodstock based on improved YOLOv5[J]. Comput. Electron. Agric. 206, 107639.
Faure G J, Chen M H, Lai S H., (2023). Holistic interaction transformer network for action detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 3340-3350.
Feichtenhofer, C., Fan, H., Malik, J., et al., 2019. Slowfast networks for video recognition [C]. Proceedings of the IEEE/CVF international conference on computer vision 6202–6211.
Fornós, M., Sanz-Fernández, S., Jiménez-Moreno, E., et al., 2022. The feeding behaviour habits of growing-finishing pigs and its effects on growth performance and carcass quality, a review[J]. Animals 12 (9), 1128.
Gao, Y., Yan, K., Dai, B., et al., 2023. Recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model with spatiotemporal attention mechanism[J]. Comput. Electron. Agric. 205, 107606.
Ge Z, Liu S, Wang F, et al., (2021). Yolox, Exceeding yolo series in 2021[J]. arXiv preprint arXiv,2107.08430.
Gu C, Sun C, Ross D A, et al., 2018. Ava, A video dataset of spatiotemporally localized atomic visual actions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 6047-6056.
Gu, Z., Zhang, H., He, Z., et al., 2023. A two-stage recognition method based on deep learning for sheep behavior[J]. Comput. Electron. Agric. 212, 108143.
Guo, S., Lin, Y., Li, S., et al., 2019. Deep spatial–temporal 3D convolutional neural networks for traffic data forecasting[J]. IEEE Trans. Intell. Transp. Syst. 20 (10), 3913–3926.
Hara, K., Kataoka, H., Satoh, Y., 2018. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 6546–6555.
He, K., Gkioxari, G., Dollár, P., et al., 2017. Mask r-cnn[C]. Proceedings of the IEEE international conference on computer vision 2961–2969.
He, Q., Xu, A., Ye, Z., et al., 2023. Object detection based on lightweight YOLOX for autonomous driving[J]. Sensors 23 (17), 7596.
Kashiha, M., Bahr, C., Haredasht, S.A., et al., 2013. The automatic monitoring of pigs water use by cameras[J]. Comput. Electron. Agric. 90, 164–169.
Kim, T., Kim, Y., Kim, S., et al., 2022. Estimation of number of pigs taking in feed using posture filtration[J]. Sensors 23 (1), 238.
Kim, G.I., Yoo, H., Chung, K., 2023. SlowFast based real-time human motion recognition with action localization[J]. Comput. Syst. Sci. Eng. 47 (2), 2135–2152.
Lao, F., Brown-Brandl, T., Stinn, J.P., et al., 2016. Automatic recognition of lactating sow behaviors through depth image processing[J]. Comput. Electron. Agric. 125, 56–62.
Leonard, S.M., Xin, H., Brown-Brandl, T.M., et al., 2019. Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls[J]. Comput. Electron. Agric. 163, 104866.
Li, Y., Qi, X., Saudagar, A.K.J., et al., 2023. Student behavior recognition for interaction detection in the classroom environment[J]. Image Vis. Comput. 136, 104726.
Li, B., Xu, W., Chen, T., et al., 2023. Recognition of fine-grained sow nursing behavior based on the SlowFast and hidden Markov models[J]. Comput. Electron. Agric. 210, 107938.
Lin T Y, Maire M, Belongie S, et al., (2014). Microsoft coco, Common objects in context [C]//Computer Vision–ECCV 2014, 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing. 740-755.
Liu, J.J., Hou, Q., Cheng, M.M., et al., 2020. Improving convolutional networks with self-calibrated convolutions[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 10096–10105.
Liu, Y., Xiao, D., Zhou, J., et al., 2023. AFF-YOLOX: An improved lightweight YOLOX network to detect early hatching information of duck eggs[J]. Comput. Electron. Agric. 210, 107893.
Luo, Y., Xia, J., Lu, H., et al., 2024. Automatic recognition and quantification feeding behaviors of nursery pigs using improved YOLOV5 and feeding functional area proposals[J]. Animals 14 (4), 569.
Matthews, S.G., Miller, A.L., PlÖtz, T., et al., 2017. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring[J]. Sci. Rep. 7 (1), 17582.
Redmon J, Farhadi A., (2018). Yolov3, An incremental improvement[J]. arXiv preprint arXiv,1804.02767.
Subedi, S., Bist, R., Yang, X., et al., 2023. Tracking pecking behaviors and damages of cage-free laying hens with machine vision technologies[J]. Comput. Electron. Agric. 204, 107545.
Sun C, Shrivastava A, Vondrick C, et al., (2018). Actor-centric relation network[C]// Proceedings of the European Conference on Computer Vision (ECCV). 318-334.
Sun, G., Liu, T., Zhang, H., et al., 2023. Basic behavior recognition of yaks based on improved SlowFast network[J]. Eco. Inform. 78, 102313.
Tong, Z., Song, Y., Wang, J., et al., 2022. Videomae, Masked autoencoders are data-efficient learners for self-supervised video pre-training[J]. Adv. Neural Inf. Proces. Syst. 35, 10078–10093.
Tran, D., Bourdev, L., Fergus, R., et al., 2015. Learning spatiotemporal features with 3d convolutional networks[C]. Proceedings of the IEEE international conference on computer vision 4489–4497.
Tu, S., Cai, Y., Liang, Y., et al., 2024. Tracking and monitoring of individual pig behavior based on YOLOv5-Byte[J]. Comput. Electron. Agric. 221, 108997.
Viazzi, S., Ismayilova, G., Oczak, M., et al., 2014. Image feature extraction for classification of aggressive interactions among pigs[J]. Comput. Electron. Agric. 104 (57–62), 503.
Wang Q, Wu B, Zhu P, et al., (2020). ECA-Net, Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 11534-11542.
Wang, R., Gao, R., Li, Q., et al., 2023. A lightweight cow mounting behavior recognition system based on improved YOLOv5s[J]. Sci. Rep. 13 (1), 17418.

Wang, M., Larsen, M., Bayer, F., et al., 2021. A PCA-based frame selection method for applying CNN and LSTM to classify postural behaviour in sows[J]. Comput. Electron. Agric. 189, 106351.
Wang, X., Yang, K., Ding, Q., et al., 2024. TQRFormer, tubelet query recollection transformer for action detection[J]. Image Vis. Comput. 147, 105059.
Woo S, Park J, Lee J Y, et al., (2018). Cbam, Convolutional block attention module[C]// Proceedings of the European conference on computer vision (ECCV). 3-19.
Wu, W., Liu, H., Li, L., et al., 2021. Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image[J]. PLoS One 16 (10), e0259283.
Zeng, F., Li, B., Wang, H., et al., 2023. Detection of calf abnormal respiratory behavior based on frame difference and improved YOLOv5 method[J]. Comput. Electron. Agric. 211, 107987.
Zhang, Y., Cai, J., Xiao, D., et al., 2019. Real-time sow behavior detection based on deep learning[J]. Comput. Electron. Agric. 163, 104884.
Zwicker, B., Weber, R., Wechsler, B., et al., 2015. Degree of synchrony based on individual observations underlines the importance of concurrent access to enrichment materials in finishing pigs[J]. Appl. Anim. Behav. Sci. 172, 26–32.