EasyChair Preprint

№ 15928

Aggressive Behavior Recognition for Group-Housed Pigs

Chia Duo Wang, Yea Shuan Huang and Chang Wu Yu

March 20, 2025

Aggressive Behavior Recognition for Group-housed Pigs

Chia Duo Wang , Yea Shuan Huang, and Chang Wu Yu

CSIE, Chung-Hua University, Hsinchu, Taiwan 30012, R.O.C.
jaredcdwang@gmail.com, yeashuan@chu.edu.tw, cwyu@chu.edu.tw

Abstract. This paper proposes a method for automatically identifying aggressive behavior in pigs using fixed-position monitoring, featuring a hybrid architecture that integrates object detection, tracking, and behavioral analysis. In the proposed architecture, a stationary camera captures images, with YOLO detecting pig locations and DeepSORT tracking their movements to identify individuals potentially exhibiting aggression. This process generates five-second video clips of individual pigs, which are then processed by a behavioral analysis module based on Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network. Experimental results demonstrate that the proposed method achieves approximately 90% accuracy in recognizing aggressive behavior in pigs on the test dataset.

Keywords: Penned pigs, Behavior recognition, Deep learning, Yolo, DeepSORT, LSTM.

1 Introduction

Aggressive behavior is a common issue in captive pig herds, often causing significant harm such as injuries that can lead to wound infections or, in severe cases, death. This behavior is a major source of economic loss for farms. Moreover, the stress linked to aggression has been shown to affect the fertility of breeding pigs negatively (Kongsted, 2004). As such, aggression among pigs is a critical concern for animal health, welfare, and farm profitability (Faucitano, 2001).

Several researchers have employed computer vision techniques to monitor aggressive behavior in pigs. For example, Viazzi et al. (2014) used segmentation regions from pig motion history images to extract two key features for each region: the average exercise intensity, represented by the mean pixel difference, and the total exercise volume, measured by the total number of pixel differences. They then applied linear discriminant analysis to classify individual aggressive interactions based on these features. Hakansson et al. (2023) developed a computer vision-based method for detecting pig tail-biting, leveraging convolutional neural networks (CNNs) to extract spatial information and integrating two additional networks—long short-term memory (LSTM) and CNNs—to enhance detection. Similarly, Gao et al. (2023) proposed a hybrid model combining CNNs with gated recurrent units (GRUs), incorporating a spatiotemporal attention mechanism to identify aggressive pig behavior more effectively.

In this study, we propose an integrated architecture that combines object detection and tracking technologies, leveraging a deep learning framework based on convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to identify aggressive behaviors. The system provides two key outputs: (1) the presence or absence of attacking events in the monitored footage and (2) the identities of the aggressive pigs.

2 Proposed Method

In this study, a single frame is an $M \times N$ image (M is the number of horizontal pixels and N is the number of vertical pixels in each image) captured by a fixed camera on a pig farm and presented in Fig. 1.

Fig. 1. Pig farm environment.

Initially, we employed the original $M \times N$ video images as the training and testing data for the pig aggressive behavior discrimination network like the Aggression identification of individual pig module in the system architecture (presented in Fig. 2). However, the accuracy of discrimination was not good. To address this, we enhanced our system architecture by incorporating a pig detection and tracking module, along with a pig movement trajectory filtering and image generation module (as illustrated in Fig. 2). This refinement allowed the training and testing of the pig aggression discrimination network to focus on individual pigs rather than the broader area of the husbandry compartment. Consequently, the system more effectively identifies instances of aggression within the video footage.

2.1 Pig Aggressive Behavior Discrimination System Architecture

Given that aggression is closely correlated to motion patterns, which are crucial for identifying target behaviors, we exploit differences between adjacent image frames as input to the discrimination network. This approach facilitates the extraction of features necessary for accurately detecting aggressive behavior. The system architecture is depicted in Fig. 2.

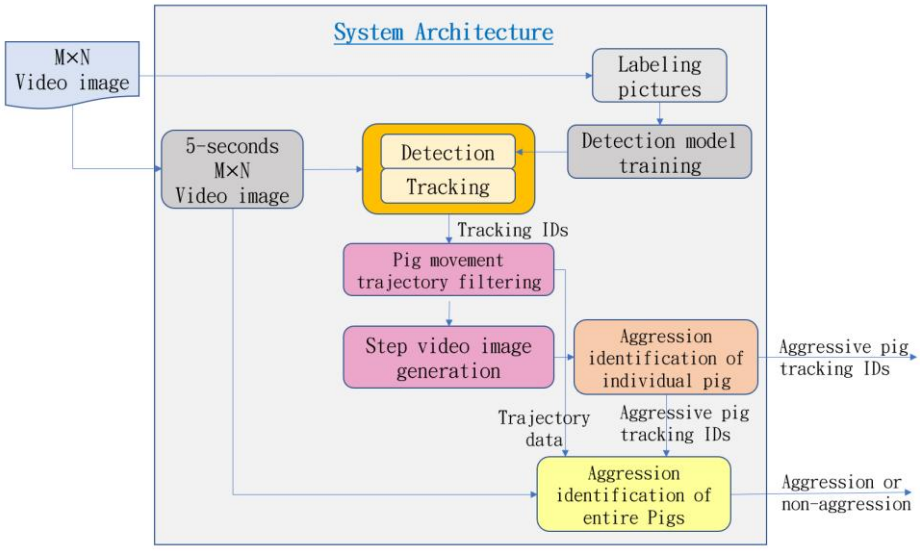

Fig. 2. The system architecture of the proposed method comprises several key components: pig detection and tracking, pig movement trajectory filtering, step video image generation, aggression identification of individual pig, and aggression identification of entire Pigs.

2.2 Pig Detection and Tracking

In our system, pig detection is implemented using YOLOv5. We prepared labeled images to train the YOLOv5 model for pig detection, achieving a detection accuracy of over 98%. Following the training of the YOLOv5 pig detection model, the DeepSORT module is employed in conjunction with YOLOv5 to track pig movement trajectories. We input $M \times N$ video images with a duration of 5 seconds into the DeepSORT module, which then outputs the movement track information for each pig in the video. The detected pig movement trajectories are visualized on the image, as illustrated in Fig. 3, with different track IDs represented by distinct colors.

In Fig. 3, the yellow box highlights the trajectory of an aggressive pig. Observations reveal that aggressive behavior often exhibits drastic changes in the pig's moving direction and circular movement patterns. The trajectory in the red box displays a linear jump, indicating ID switching from DeepSORT. The green box represents the movement of a pig but is depicted into three distinct track IDs in different time segments.

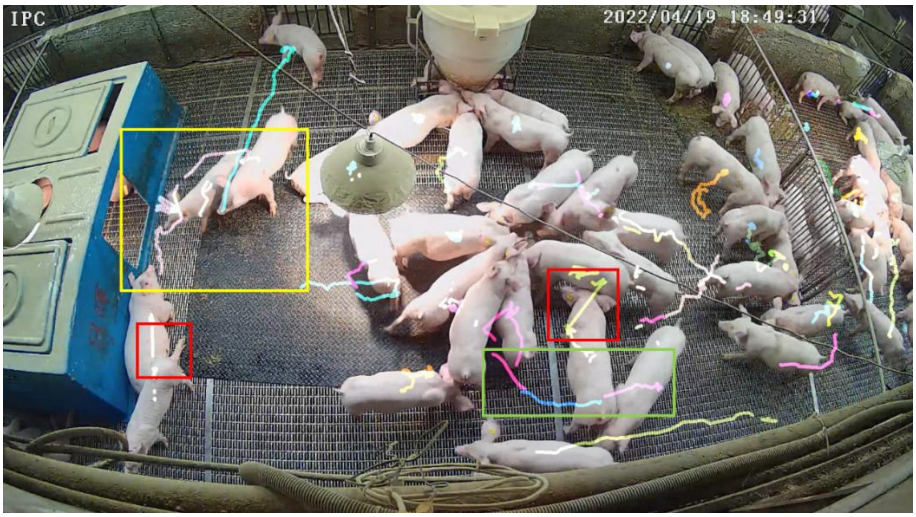

Fig. 3. Pig movement trajectory. The variation in color intensity is influenced by the background color.

2.3 Pig movement trajectory filtering

Given our video images are captured from a fixed camera monitoring an area, generating individual $n \times n$ (where n is the number of horizontal or vertical pixels) video images for these pigs would impose a substantial burden on the system. Therefore, we focus on eliminating abnormal trajectories potentially affected by ID switching and small-movement trajectories indicating resting pigs. Consequently, we established two filtering rules: (1) a trajectory with a movement distance over *DX* between adjacent frames is likely to be indicative of ID switching; and (2) the accumulated distance of pig movement multiplied by the direction change angle between adjacent frames, when exceeding *DXA*, suggests a possible aggressive behavior. Suppose *T* is the total number of pig track IDs after detection and tracking, *t* is the *t*-th pig track ID, C_j^t is the center position of the *j*-th tracked frame of *t*, and N_t is the total number of tracked frames of *t*, D_j^t and θ_j^t are the displacement distance and displacement angle from C_{j-1}^t to C_j^t.

$$D_j^t = \left| C_j^t - C_{j-1}^t \right|$$
$$\theta_j^t = \angle \overrightarrow{C_{j-1}^t C_j^t} \ (\angle \text{ is a vector angle}) \quad (1)$$

According to the above definition, if the following two conditions are met, the *t*-th pig track ID will be an aggressive candidate.

$$1.\ \forall_{j=1}^{N_t} D_j^t < DX$$
$$2.\ \sum_{j=1}^{N_t} (\theta_j^t - \theta_{j-1}^t) \times D_j^t > DXA \quad (2)$$

where *DX* and *DXA* are two thresholds for filtering out aggressive pigs.

2.4 Step video image generation

Suppose the *t*-th trajectory is filtered to be a possible aggressive pig tracked ID, an $n \times n$ video image will be generated from the video image generator, and entered into the pig attack behavior discrimination module to further identify whether it is an aggressive behavior or not. If there is no pig trajectory with possible aggressive behavior, then no $n \times n$ video image will be generated, and the $M \times N$ video image can be judged to be non-aggressive. An $n \times n$ video image sequence is as shown in Fig. 4.

Fig. 4. An $n \times n$ video image sequence. An $n \times n$ video image is extracted from its original $M \times N$ video image, and its center is aligned to the mean position of $\{C_1^t, C_2^t, \cdots, C_{N_t}^t\}$.

Let CX_j^t and CY_j^t represent the X and Y coordinates of C_j^t, and TCX^t and TCY^t represent the center position of the $n \times n$ video image of the *t*-th tracked pig ID in the $M \times N$ video image, and the TCX^t needs to be greater than $n/2$ and less than $M - n/2$ and the TCY^t needs to be greater than $n/2$ and less than $N - n/2$ to avoid capturing the picture beyond a $M \times N$ image.

$$TCX_{temp}^t = \frac{\max_{j=1\ldots jcount}(CX_j^t) - \min_{j=1\ldots jcount}(CY_j^t)}{2}$$

$$TCX^t = \begin{cases} \frac{n}{2} & \text{, if } TCX_{temp}^t < \frac{n}{2} \\ TCX_{temp}^t & \text{, if } \frac{n}{2} \le TCX_{temp}^t \le M - \frac{n}{2} \\ M - \frac{n}{2} & \text{, if } TCX_{temp}^t > M - \frac{n}{2} \end{cases}$$

$$TCY_{temp}^t = \frac{\max_{j=1\ldots jcount}(CY_j^t) - \min_{j=1\ldots jcount}(CY_j^t)}{2}$$

$$TCY^t = \begin{cases} \frac{n}{2} & \text{, if } TCY_{temp}^t < \frac{n}{2} \\ TCY_{temp}^t & \text{, if } \frac{n}{2} \le TCY_{temp}^t \le N - \frac{n}{2} \\ N - \frac{n}{2} & \text{, if } TCY_{temp}^t > N - \frac{n}{2} \end{cases} \quad (3)$$

The $n \times n$ video image of the t-th tracked pig ID is extracted from the $M \times N$ video image. To alleviate the image background interference in judging the pig attack behavior, only $n/2 \times n/2$ image centered at (CX_j^t, CY_j^t) is extracted from the $M \times N$ video image and the rest of the $n \times n$ video image is set to be 0. Suppose $f_{M\times N}(x,y)$ is the image value of the $M \times N$ video image at position (x,y), $f_{n\times n}^t(x,y)$ is the image value of the $n \times n$ video image of the t-th tracked pig ID at position (x,y), and $\mathbb{Z}(x,y)$ is a Boolean value to indicate whether $f_{n\times n}^t(x,y)$ is 0 or not.

$$\mathbb{Z}(x,y) = \begin{cases} 1 & \text{, if } \begin{pmatrix} x > CX_j^t - TCX^t + \frac{n}{4} \\ x \le CX_j^t - TCX^t + \frac{3n}{4} \\ y > CY_j^t - TCY^t + \frac{n}{4} \\ y \le CY_j^t - TCY^t + \frac{3n}{4} \end{pmatrix}, \\ 0 & \text{, else.} \end{cases} \tag{4}$$

and

$$f_{n\times n}^t(x,y) = \begin{cases} f_{M\times N}\left(TCX^t - \frac{n}{2} + x, TCY^t - \frac{n}{2} + y\right) & \text{, if } \mathbb{Z}(x,y) = 1; \\ 0 & \text{, else.} \end{cases} \tag{5}$$

The constructed $n \times n$ video image not only can have the important image information of the individually tracked pig, but also can obtain its movement information. During the $M \times N$ video image, if a certain frame i loses the tracking of the t-th tracked pig ID, the center of the corresponding $n/2 \times n/2$ image of this frame is the last valid tracked (CX_j^t, CY_j^t) of this pig ID, where $j < i$. In a generated $n \times n$ video image, each $n/2 \times n/2$ video image may be shown in different positions according to the movement of the corresponding tracked pig. Therefore, the generated $n \times n$ video image is called "step video image", which can provide not only spatial information (the image of an individual pig and interaction with other nearby pigs) but also spatiotemporal information (the movement information of this pig) for the detection of pig aggression. An example of an $n \times n$ step video image sequence is shown in Fig. 5.

Fig. 5. an $n \times n$ step video image sequence. Each $n/2 \times n/2$ video image may be shown in different positions according to the movement of the corresponding tracked pig.

2.5 Identification of pig aggression

After generating the $n \times n$ step video image of a potentially aggressive pig, it is input into the pig aggressive behavior discrimination network module to distinguish whether there is an aggressive behavior or not. The adopted pig aggressive behavior discrimination network is referenced to the discriminant network designed by Z. Islam et al (2021) (Fig. 6. Pig aggressive behavior discrimination network architecture).

We use adjacent frame difference (Fig. 7) to be the input of SepConvLSTM (Separable Convolutional Long Short-Term Memory), which SepConvLSTM to extract the spatiotemporal features of the temporal changes between adjacent frames. The difference of adjacent frames mainly encodes the temporal information of motion by highlighting the image change of a tracked pig, which not only extracts spatial features from each time step image, but also learns the time change, captures the difference between consecutive frames, and generates a powerful spatiotemporal feature map to distinguish whether there is aggression in the video.

Suppose S_j is the $n \times n$ image the j-th frame of a step video image，A_j is the area of the $n/2 \times n/2$ images of S_j and B_j denotes the rest area of S_j discussed in Sec. 2.4. Then, the adjacent frame difference D_j between S_j and S_{j-1} is designed as

$$D_j(x,y) = \begin{cases} \left|S_j(x,y) - S_{j-1}(x,y)\right| & \text{, if } (x,y) \in A_j \text{ and } (x,y) \in A_{j-1}; \\ 0 & \text{, otherwise.} \end{cases} \tag{6}$$

The input of the discriminant network is an $n \times n$ step video image, which contains 31 differential frames with a resolution of 224×224×3, and local spatiotemporal features are generated from the feature map output by MobileNet CNN, and the spatial features of the shape of 7×7×56 are extracted, so we obtain spatial feature streams of 31×7×7×56. Then, we use SepConvLSTM with 64 filters, and it will output a 7×7×64 feature map containing spatial and temporal information. Finally, the feature map is passed to a fully connected layer for classification. When the system completes the discrimination of individual $n \times n$ step video images, and outputs the track ID of the aggressive pig in the original $M \times N$ video image, we use the entire pig aggressive discrimination module to aggregate the $n \times n$ step video image discrimination results.

FigSuppose that in a 5-second $M \times N$ video image, the detection and tracking module generates *tcount* pig track IDs after filtering, so the image generator generates *tcount* $n \times n$ step video images of tracked pigs. Suppose t represents the *t-th* individually tracked pig, and P_t represents the aggression prediction of the *t-th* individually tracked pig by the discriminant network, then the entire pig discrimination result is as follows:

$$P_t = \begin{cases} 1 & \text{, if the } t-\text{th individually tracked pig is aggressive;} \\ 0 & \text{, else} \end{cases}.$$

$$\text{Entire pigs aggression result} = \begin{cases} \text{aggression} & \text{, if } \sum_{t=1}^{tcount} P_t > 0; \\ \text{non} - \text{aggression} & \text{, if } \sum_{t=1}^{tcount} P_t = 0. \end{cases} \tag{7}$$

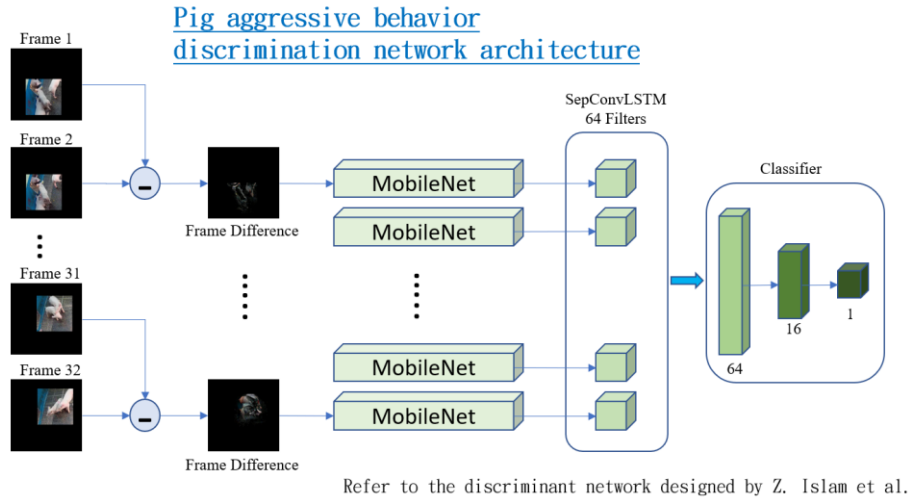

Fig. 6. Pig aggressive behavior discrimination network architecture uses MobileNet for spatial feature extraction, and SepConvLSTM, in which the SepConvLSTM layer learns to encode spatial information in series to generate spatiotemporal feature maps and pass them to the classification layer.

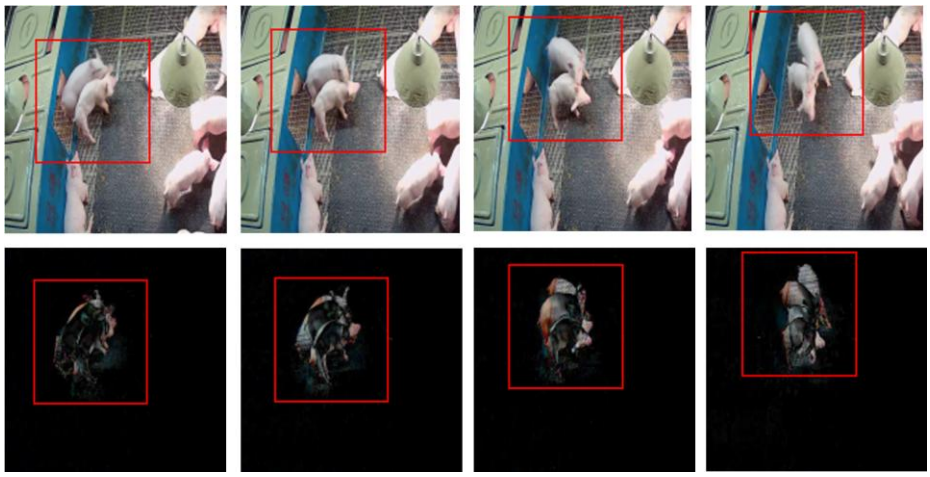

Fig. 7. The adjacent image frame difference streaming.

3 Experiment

We obtained $M \times N$ video images taken by the manufacturer itself, and M=1920, N=1080 and n=640 in our experiments. As shown in Fig. 2, we use the pig detection model trained by Yolo v5 for DeepSORT, and execute the pig detection and tracking on the prepared 234 1920×1080 5 seconds video images with aggressive behavior and

234 1920×1080 5 seconds video images without aggressive behavior, generate the movement trajectory data of individual pigs. Then, these 1920×1080 video images were processed with trajectory data after the pig movement trajectory filtering, to generate 640×640 individual pig step video images.

In general, if a pig is identified as aggressive, individual video images will be generated for both the aggressive pig and its opponents involved in the fight. But, in some poor situations such as incorrect tracking trajectory, or difficult to distinguish aggressive behavior, some step video images should be filtered out and not used for training and testing the discrimination model. In total, 324 aggressive individual pig step video images and 1129 non-aggressive step video images were generated. Among them, 205 aggressive step video images and 717 non-aggressive step video images are used for training; 51 aggressive step video images and 179 non-aggressive step video images are used for validation; 68 aggressive step video images and 233 non-aggressive step video images are used for testing.

The test dataset was tested with the model of the best accuracy of the validation dataset, and the accuracy rate of individual pig aggressive behavior was 92.691%, the precision rate was 85.938%, and the recall rate was 80.882%. The Confusion Matrix of the individual pig test dataset is shown in Fig. 8.

The individual pig test results cannot fully represent the original 1920×1080 video images. Therefore, we use the entire pig aggression discrimination module to aggregate the 640 × 640 step video image discrimination results generated by the original 1920×1080 video image of the test dataset. The 52 original 1920×1080 video images in the test dataset were detected by the system for whether there was an aggression, and the detection results were summarized by the entire pig aggression discrimination module to obtain a confusion matrix, which is presented in Fig. 8. The confusion matrix of entire pigs test result showed that the accuracy rate was 88.46%, the precision rate was 91.67%, and the recall rate was 84.62%.

Individual pig		Prediction	
		nonFight	Fight
Truth	nonFight	224	9
	Fight	13	55

Entire Pigs		Prediction	
		nonFight	Fight
Truth	nonFight	24	2
	Fight	4	22

Fig. 8. The confusion matrices of the individual pig and the entire pigs' test dataset

3.1 Comparison and discussion

A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the performance of a binary classifier model (can be used for multi-class classification as well) at varying threshold values. The ROC curve is the plot of the true positive rate (TPR) against the false positive rate (FPR) at each threshold setting. (Wikipedia).

In this study, the discriminant network tried 1920×1080 video images, 640×640 video images, and 640×640 step video images, and the ROC results are as shown in Fig. 9. ROC comparison with three different video images.

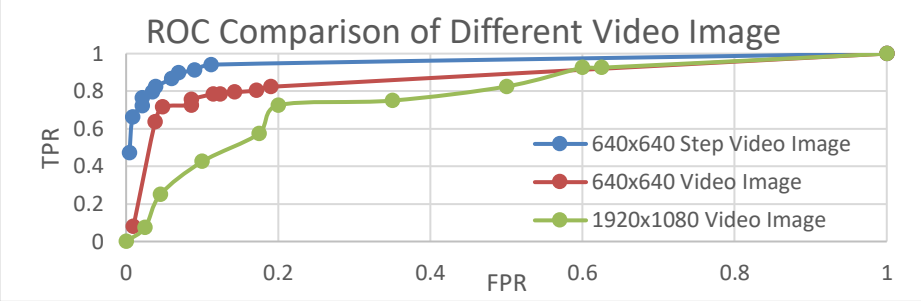

Fig. 9. ROC comparison with 640×640 Step Video Image, 640×640 Video Image and 1920×1080 Video Image.

4 Conclusions

In this paper, we propose a novel and effective method to detect aggressive behavior in the surveillance video of pig farms, and the proposed pig detection, tracking, filtering and individual pig video generation programs can help the discrimination network to more effectively distinguish the temporal and spatial characteristics, and achieve high recognition accuracy.

The current system mainly uses the difference of adjacent frames to capture the fast-moving characteristics of pig aggressive behavior, so the accuracy of system discrimination has its limit and cannot be universally applied to a wide range of behavior discrimination, but its advantage is that the calculation speed is fast enough and does not require high-end equipment. In the future, it is worth changing the aggressive behavior discrimination system from adjacent frames difference to pig posture as the core of discrimination, which should improve the discrimination accuracy again, and could try to discriminate other behaviors. It is an interesting and worthy research topic.

References

1. Gao, Y.; Yan, K.; Dai, B.; Sun, H.; Yin, Y.; Liu, R.; Shen, W. Recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model with spatio-temporal attention mechanism. Comput. Electron. Agric. 2023, 205, 107606.
2. Hakansson, F.; Jensen, D.B. Automatic monitoring and detection of tail-biting behavior in groups of pigs using video-based deep learning methods. Front. Vet. Sci. 2023, 9, 1099347.
3. Kongsted, A.G. Stress and fear as possible mediators of reproduction problems in group housed sows: a review. Acta Agric. Scand., Sect. A– Anim. Sci. 2004, 54 (2), 58–66.
4. Viazzi S, Ismayilova G, Oczak M, et al. Image feature extraction for classification of aggressive interactions among pigs[J]. Computers and Electronics in Agriculture, 2014, 104: 57－62. H
5. Z. Islam, M. Rukonuzzaman, R. Ahmed, M.. Kabir and M. Farazi, "Efficient Two-Stream Network for Violence Detection Using Separable Convolutional LSTM," *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, 2021, pp. 1-8, doi: 10.1109/IJCNN52387.2021.9534280.