

Automatic detection of sow estrus using a lightweight real-time detector and thermal images

Haibo Zheng, Hang Zhang, Shuang Song, Yue Wang, Tonghai Liu*

(College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin 300392, China)

Abstract: Determination of ovulation time is one of the most important tasks in sow reproduction management. Temperature variation in the vulva of the sows can be used as a predictor of ovulation time. However, the skin temperatures of sows in existing studies are obtained manually from infrared thermal images, posing an obstacle to the automatic prediction of ovulation time. In this study, an improved YOLO-V5s detector based on feature fusion and dilated convolution (FD-YOLOV5s) was proposed for the automatic extraction of the vulva temperature of sows based on infrared thermal images. For the purpose of reducing the model complexity, the depthwise separable convolution and the modified lightweight ShuffleNet-V2 module were introduced in the backbone. Meanwhile, the feature fusion network structure of the model was simplified for efficiency, and a mixed dilated convolutional module was designed to obtain global features. The experimental results show that FD-YOLOV5s outperformed the other nine methods, with a mean average precision (mAP) of 99.1%, an average frame rate of 156.25 fps, and a model size of only 3.86 MB, indicating that the method effectively simplifies the model while ensuring detection accuracy. Using a linear regression between manual extraction and the results extracted using this method in randomly selected thermal images, the coefficients of determination for maximum and average vulvar temperatures reached 99.5% and 99.3%, respectively. The continuous vulva temperature of sows was obtained by the target detection algorithm, and the sow estrus detection was performed by the temperature trend and compared with the manually detected estrus results. The results showed that the sensitivity, specificity, and error rate of the estrus detection algorithm were 89.3%, 94.5%, and 5.8%, respectively. The method achieves real-time and accurate extraction of sow vulva temperature and can be used for the automatic detection of sow estrus, which could be helpful for the automatic prediction of ovulation time.

Keywords: automatic estrus detection, thermal images, real-time detector, vulva temperature, mixed dilated convolutional

DOI: 10.25165/j.ijabe.20231603.7711

Citation: Zheng H B, Zhang H, Song S, Wang Y, Liu T H. Automatic detection of sow estrus using a lightweight real-time detector and thermal images. *Int J Agric & Biol Eng*, 2023; 16(3): 194–207.

1 Introduction

Accurate estrus detection is a critical determinant of increased profitability and productivity in the swine industry^[1-3]. Also, these profits are closely related to the use of artificial insemination, which contributes to the genetic improvement of the pig herd, but the timing of the insemination is important^[4,5]. Current research has indicated that sows may ovulate 10 to 96 h after estrus and that insemination within 0 to 24 h before ovulation can obtain optimal fertility while reducing the number of inseminations^[6-8]. On the contrary, incorrect insemination time may impair the reproductive performance of the sow^[9]. Alternatively, missing the timing of insemination can reduce the number of piglets produced per sow per year. Either can cause a considerable economic loss^[2,3]. There are various methods of estrus detection, such as the back-pressure test, which investigates the standing response of the sow when the back was pressed^[10], collecting the frequency and duration of sow visits to

a boar^[11,12], monitoring changes in activity^[2], observing the swelling and redness of the sow's vulva^[13] and the riding behavior between pen mates^[14,15]. Unfortunately, these methods to determine the optimal time for insemination still rely on the breeder's relatively subjective evaluation of the sow's behavior and physical signs. Therefore, developing a technique for accurately predicting the timing of ovulation is of great importance since this would allow for more accurate insemination schedules, thus increasing litter rates and reducing the cost of multiple inseminations.

There has been some research on detecting or predicting ovulation in sows, such as ultrasonography for ovulation^[16] and analysis of the relationship between vaginal mucus resistance and estrus^[17]. However, ultrasonography is time-consuming, requires specialized personnel training^[16,18], and does not predict ovulation^[19]. Similarly, the electrical resistance of vaginal mucus varies significantly between individuals and between different measurement locations within the vagina^[17,18,20,21]. More recently, several studies have tested the feasibility of monitoring changes in vulvar skin temperature of sows by infrared thermography as a predictive tool for estrus and ovulation. Sykes et al.^[22] recorded thermal images of vulvar skin temperature in sows and found that both maximum and mean values of vulvar skin temperature were higher for sows in estrus than in diestrus ((36.6±0.2)°C and (33.4±0.3)°C vs. (35.6±0.3)°C and (31.8±0.6)°C, respectively). In a similar method, Luño et al.^[18] combined ultrasonography to observe the changes in the mean temperature of vulvar skin before and after ovulation and showed a significant decrease in the vulvar skin temperature 24-12 h before ovulation and a decrease to the lowest

Received date: 2022-05-31 **Accepted date:** 2022-11-08

Biographies: **Haibo Zheng**, Master, research interest: computer vision, agricultural artificial intelligence, Email: haibo_0219@163.com; **Hang Zhang**, PhD, research interest: intelligent agriculture, Email: zhanghrz@126.com; **Shuang Song**, Master, research interest: artificial intelligence, Email: youzipi_er@163.com; **Yue Wang**, Master, research interest: intelligent agriculture, Email: yuewangyue@163.com.

***Corresponding author:** **Tonghai Liu**, PhD, Professor, research interest: agricultural artificial intelligence, remote sensing. College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin 300392, China. Tel: +86-13920136245, Email: tonghai_1227@126.com.

temperature at 12 h before ovulation. In a different approach, considering the effect of ambient temperature, simultaneously observing the temperature change of vulva and other body parts (gluteal area, udder, and ear base), or the temperature difference change between these body parts and vulvar skin temperature. The results showed that changes in vulvar skin temperature were the most correlated with estrus and that the mean temperature of vulvar skin of gilts and multiparous sows increased at the onset of estrus and decreased before ovulation^[3,23-25]. Moreover, the validity of vulvar skin average temperature variation as a predictor of sow ovulation was evaluated by inseminating sows once when the vulvar skin temperature of diestrous sows was 20% increased^[1] or below 35°C^[26], where the threshold of 35°C resulted from Luño et al.^[18] The results showed comparable reproductive performance with the multiple insemination strategy of the conventional back-pressure test. The above research shows that monitoring the changes in the vulvar skin temperature can help predict the ovulation time of sows and improve the accuracy of insemination time. However, traditional manual temperature extraction is time-consuming, and the operator's skill level can also affect the results of the extraction temperature^[27]. Therefore, the efficiency, accuracy, and stability of vulvar skin temperature extraction still need to be improved.

With the development of deep learning and object detection technology has been rapidly advanced. Currently, object detection networks include two-stage detectors (e.g., Faster RCNN^[28]) and single-stage detectors (e.g., YOLO^[29]). These networks provide new solutions for livestock farming, including animal target detection^[30], animal behavior recognition^[31], animal body size and weight estimation^[32], and animal disease detection^[33]. Zhang et al.^[34] proposed an improved SSD (single-shot multibox detector)^[35] network model to achieve behavioral recognition of grouped pigs in pig houses. Zhang et al.^[33] proposed an EFMYOLOV3 network model for detecting key parts of cows based on YOLO-V3 (You Only Look Once version 3)^[36] and thermal images to achieve localization of key parts of cows' eyes, udder, and head with automatic recognition of mastitis in cows. The average frame rate of the algorithm was 99 frames per second, the average accuracy was 96.8%, and the accuracy of mastitis classification was 83.33%. This also shows that deep learning can be successfully used for object detection in thermal images.

In this study, an automatic temperature extraction method was developed and used for sow estrus detection, which was the first time to describe sow estrus detection using deep learning and thermal images. The method could directly detect and extract the temperature of sows' vulva in thermal images to solve the problem of time-consuming and unstable manual extraction. Firstly, the YOLO-V5s^[37] object detection model was improved and named FD-YOLOV5s to achieve rapid and accurate detection of the vulva and buttocks of sows. The buttocks detection box was used to determine

whether the sow's buttocks posture was tilted to improve the accuracy of vulvar skin temperature extraction. In FD-YOLOV5s, the modified lightweight ShuffleNet-V2^[38] and the depthwise separable convolution^[39] modules were introduced into the backbone network to reduce the number of model parameters. In addition, the number of upsampling and downsampling operations was reduced in the feature fusion process to simplify the network structure, thus increasing the inference speed of the model. Meanwhile, to enhance the model to capture global features, a multi-branch dilated convolution^[40] module was constructed to capture global context information. Then, vulvar temperatures were obtained by mapping the coordinates of the detection box of the temperature matrix of the thermal image and compared with the manually extracted temperatures to verify the effectiveness of the FD-YOLOV5s algorithm for extracting the vulvar temperatures of sows. Finally, the continuous vulvar temperature of the sow was extracted to identify estrus and compared with the manual estrus detection results to verify the accuracy of the method of detecting estrus in sows.

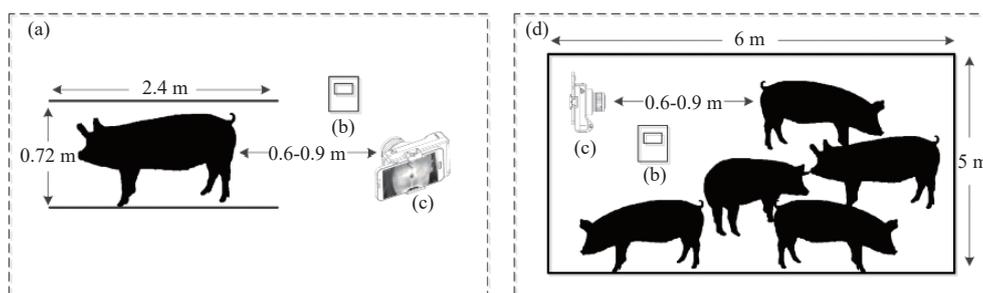
2 Materials and methods

2.1 Dataset

2.1.1 Sows and thermal image data collection

This study was conducted on the sow farm of Anping County Derun Breeding Co., Ltd. in Anping County, Hengshui City, Hebei Province, China. From June to July 2021, 679 multiparous Large White and Landrace sows were selected for thermal image data collection. The sows were situated in the individual housing and the group housing, using the Fotric 225 infrared thermal camera (Fotric, Shanghai, China) with a resolution of 320 pixels (horizontal) × 240 pixels (vertical) and an accuracy of $\pm 2^\circ\text{C}$ to take thermal images from directly behind the sows, and each image was required to contain only one sow. The ambient temperature, relative humidity, and emissivity were adjusted for the thermal imaging camera. We used the VC231 intelligent temperature and humidity recorder (VICTOR, Shenzhen, China) to obtain the ambient temperature and relative humidity in the field with an accuracy of $\pm 0.3^\circ\text{C}$ for ambient temperature and $\pm 2\%$ for relative humidity. The emissivity was set to 0.96 according to Siewert et al.^[41]

As shown in Figure 1, the camera lens was about 0.6-0.9 m away from the sow to ensure that the thermal imager could capture the vulva and buttocks of sows, and 3 to 5 images per sow were taken and stored on a secure digital memory card (32 GB). In addition, Table 1 lists that the error of different shooting distances on the maximum and average vulva temperatures of three sows did not exceed 0.2°C . This indicates that the different shooting distances in this study did not have a significant effect on the temperature measurement results of the infrared thermal camera and were able to obtain the vulva temperature of the sows effectively.



Note: a. Individually housed sows; b. VICTOR VC231; c. Fotric 225; d. Group-housed sows.

Figure 1 Schematic diagram of sow thermal infrared collection method

Table 1 Sows's vulva temperature data taken at different distances

Distance/m	Vulvar maximum temperature/°C			Vulvar average temperature/°C		
	Sow 1	Sow 2	Sow 3	Sow 1	Sow 2	Sow 3
0.60	37.7	37.5	36.9	36.0	35.7	35.3
0.75	37.7	37.5	37.0	36.1	35.7	35.3
0.90	37.7	37.5	37.1	36.1	35.8	35.5

In addition, to verify the practicality of the proposed method, 30 multiparous sows were followed for sow estrus detection. From the time the sows entered the pens, the vulva temperature was measured twice a day, approximately 12 h apart, until 2 d after estrus was detected. And the estrus was defined based on the skilled technician's identification of the typical signs of the sow's standing reflex and stinging ear response after a back-pressure test.

2.1.2 Dataset construction

In this study, 3462 images were captured from 649 sows and used to construct the target detection model. Regarding the division of the dataset, instead of using the leave-out method of directly dividing the dataset into two mutually exclusive sets proportionally, this study divided the dataset according to the pig population. Table 2 lists information on the division of the sow dataset, 2758 images from 511 sows served as the training set, 430 images from 93 sows as the validation set, and 274 images from 45 sows served as the test set. Since the images used in the training set, validation set, and test set were from different sows, the data sets were mutually exclusive, and the distribution of samples was almost the same, which met the requirements of data set division and also effectively prevented overfitting. Furthermore, the test set was captured from the group-housed sows and had a more complex background than the validation set, ensuring the proposed model's high generalizability. On the other hand, the overall performance of the proposed model was also evaluated by randomly dividing the dataset into ten equal parts and by ten-fold cross-validation.

Table 2 Division of sow vulva and buttocks data set

Dataset	Number of images	Number of sows
Training set	2758	511
Validation set	430	93
Test set	274	45
Total	3462	649

Infrared thermography provides information about the temperature by measuring the radiation emitted from the target surface. The measurement angle between the thermal imaging camera and the target affects the temperature measurement result^[42]. Consequently, when the sow is standing, the left and right tilt of the buttocks posture (relative to the direction of the infrared thermal camera) will increase or decrease the angle between the sow's vulva and the infrared thermal camera, affecting the measurement of the vulvar skin temperature. Therefore, to ensure the validity of the extracted vulvar skin temperature of the sow, it is necessary to detect the buttocks of the sow to distinguish the buttocks' posture and thus obtain an accurate vulvar skin temperature. Subsequently, the sow vulva and buttocks in the training set, validation set, and test set images were labeled using the Labellmg script open-sourced on GitHub.

2.2 Overall technical route

The overall goal of this study was to accurately and rapidly detect the vulva and buttocks of sows from thermal images and to extract vulvar skin temperature so as to identify sows in estrus. The

technical route of this work is shown in Figure 2, including model construction and evaluation, temperature extraction, and estrus detection. First, the whole dataset was divided and labeled, and the dataset division is presented in Section 2.1.2. Then, to simplify the YOLO-V5s model, the feature extraction module in the backbone network was replaced with an advanced lightweight module, and the structure and modules of the feature fusion network were modified, which further improved the model efficiency under the premise of ensuring the detection accuracy, and the specific implementation of the improvement is discussed in Section 2.3. Next, during temperature extraction, the relative positions of the buttocks detection box and the vulva detection box were used to determine whether the sow's buttocks posture was tilted and then combined with the temperature matrix to obtain the sow vulva temperature, and the details of this process are described in Section 2.4. Finally, a temperature curve was drawn using a fixed length of continuous vulva temperature of the sow, and the trend of this temperature curve was analyzed for estrus detection.

2.3 FD-YOLOV5s

Since the focus of this study is on the extraction of vulva temperature from standing sows, the real-time performance and lightweight performance of the detection model are highly required. Therefore, the most advanced algorithm of the YOLO series, YOLO-V5, was used to detect the vulva and buttocks of the sow. Based on the YOLO-V4^[43] data augmentation algorithm, YOLO-V5 adds adaptive anchor box computation and adaptive image scaling to improve the robustness of the model. Adaptive anchor box calculation can automatically calculate the best anchor box value based on different training datasets. Adaptive image scaling can accelerate object detection by adding minimal black edges when scaling the image. In addition, the YOLO-V5 algorithm includes four network structures, namely YOLO-V5s (small), YOLO-V5m (medium), YOLO-V5l (large), and YOLO-V5x (extra-large), which have the same structure, but with increasing network depth and width in that order. Specifically, The YOLO-V5 network structure consists of three main parts: backbone feature extraction network, neck network, and head. In this study, based on the characteristics of the sow vulva and buttocks dataset, the feature fusion network structure in the neck network of the YOLO-V5s model was modified and a mixed dilated convolutional module was designed to further improve the model detection. Meanwhile, depthwise separable convolution was introduced in the backbone network to reduce the complexity and parameters of the model for future deployment in mobile robots with less computational power. The structure of FD-YOLOV5s is shown in Figure 3.

2.3.1 Lightweight backbone network

To facilitate the deployment of FD-YOLOV5s model on some low-performance devices with small memory, the Conv module and C3 module in the original backbone network is replaced by the depthwise separable convolution (DSC) and the modified Shufflenet-V2 module (S5), respectively, as shown in Figure 4. Among them, the improved backbone network mainly consists of the following modules:

Conv: The Conv module is the smallest component of FD-YOLOV5s and includes the convolution layer, the batch normalization (BN) layer, and the SiLU activation function.

Focus: The focus module is used to reduce the amount of computation and speed up training. It divides the input image into four slices and then splices them in the channel dimension, as shown in Figure 4b, thus converting the width and height information of the image into channel information.

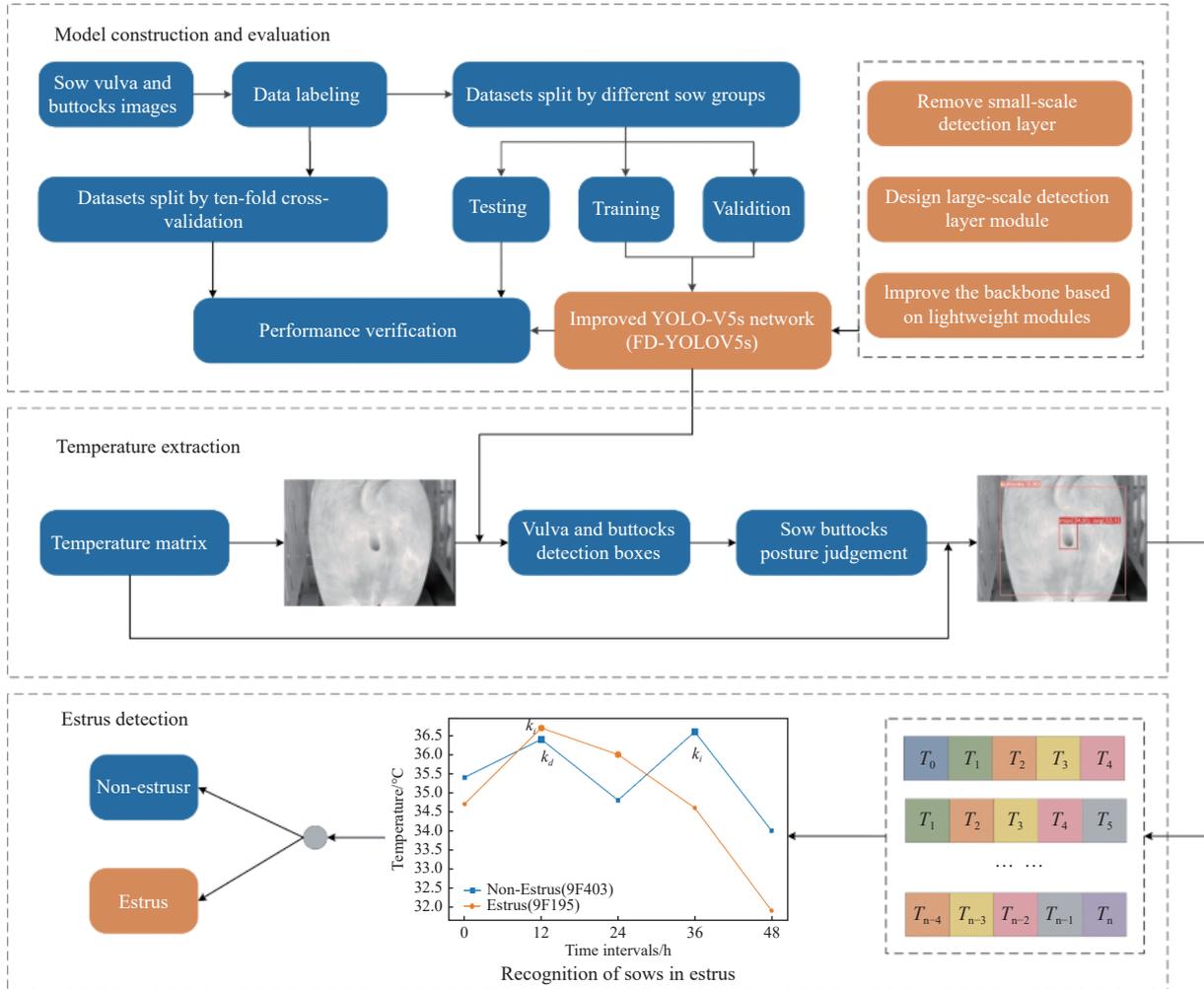
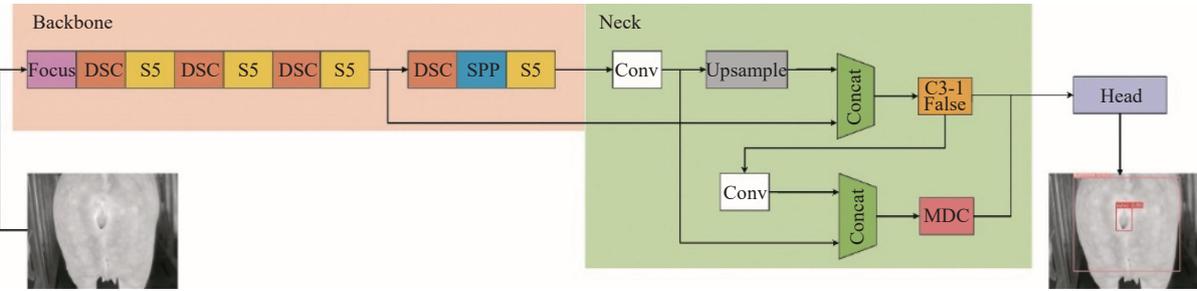


Figure 2 Schematic of the sow estrus identification process



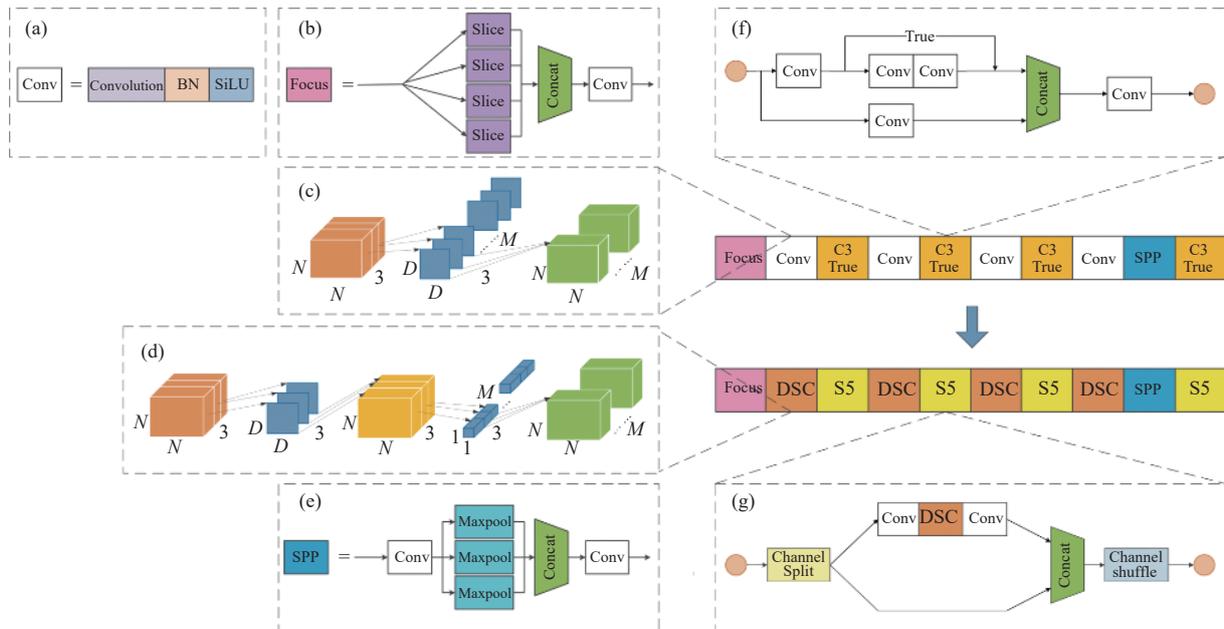
Note: DSC: Depthwise separable convolution; S5: Modified ShuffleNet-V2; SPP: Spatial Pyramid Pooling; Conv: Convolution; MDC: Mixed Dilated Convolution.

Figure 3 FD-YOLO-V5s network structure

DSC: The DSC is a strategy for decomposing convolution, which splits the standard convolution into two parts: deep convolution and point-by-point convolution. Figure 4c and Figure 4d show the process of standard convolution and depthwise separable convolution, respectively, where N and D denote the length and width of the feature map and convolution kernel, respectively, and M denotes both the number of convolution kernels and the number of channels of the output feature map, that is, the input feature map dimension is $N \times N \times 3$, the output feature map dimension is $N \times N \times M$ and the convolution kernel size is $D \times D$. The number of parameters for the standard convolution is $D \times D \times 3 \times M$, and the number of parameters for the DSC is $D \times D \times 1 \times 3 + 1 \times 1 \times 3 \times M$. It can be seen that the number of parameters of DSC is about $1/2D + 1/M$ of the standard convolution, which can simplify the size of the model.

SPP: The SPP module consisting of 5×5 , 9×9 , and 13×13

maximum pooling layers converts the input feature maps into fixed-size feature vectors to enhance the multi-scale fusion of the network. S5: The S5 module is obtained by improving the ShuffleNet-V2 module by replacing 3×3 DSC with 5×5 DSC to increase the perceptual field of the network. As shown in Figure 4g, first, it uses the “Channel Split” operation to divide the channel dimension of the input feature map equally into two branches, one branch contains two 1×1 convolutions and one 5×5 DSC, and the other branch remains unchanged. Then, the outputs of the two branches are combined by the “Concat” operation. Finally, a “Channel Shuffle” operation is used to disrupt the order of the output channels to exchange the feature information of the two branches. This structure can enhance the feature information exchange between different channels and reduce the computational complexity and memory usage of the model.



Note: a. Conv module; b. Focus module; c. Standard convolution module; d. Depthwise separable convolution module; e. Spatial pyramid pooling module; f. C3 module; g. Modified Shufflenet-V2 module; BN: Batch Normalization; SiLU: Sigmoid Linear Unit; SPP: Spatial Pyramid Pooling; C3: Improved bottleneck layer containing 3 standard convolutional layers; DSC: Depthwise separable convolution; S5: Modified Shufflenet-V2.

Figure 4 Improved FD-YOLOV5s backbone network

In addition, Table 3 lists the number of parameters before and after the improved backbone network for each layer. With the same number of input and output channels, the number of parameters of the improved backbone network is reduced by about 77%.

Table 3 Comparison of the number of parameters before and after the backbone improvement

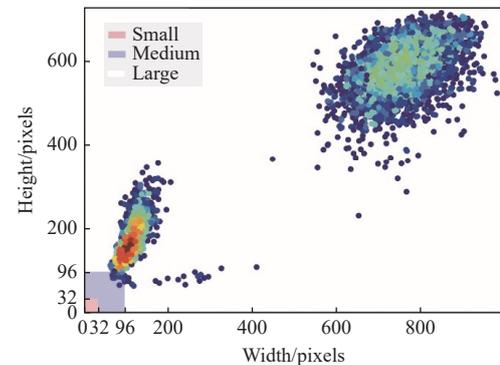
No.	Input channels	Output channels	Original module	Replaced module	Parameters	
					YOLO-V5s	FD-YOLOV5s
0	3	32	Focus	--	3520	3520
1	32	64	Conv	DSC	18 560	704
2	64	64	C3	S5	18 816	3040
3	64	128	Conv	DSC	73 984	1408
4	128	128	C3	S5	156 928	32 528
5	128	256	Conv	DSC	295 424	2816
6	256	256	C3	S5	625 152	110 208
7	256	512	Conv	DSC	1 180 672	5632
8	512	512	SPP	--	656 596	656 596
9	512	512	C3	S5	1 182 720	139 008
Total					4 212 372	955 460

2.3.2 Improvement of the neck network

Feature Pyramid Networks (FPN)^[44] and Path Aggregation Networks (PAN)^[45] are used as the neck network of FD-YOLOV5s for enhanced feature fusion. FPN transfers and merges high-level semantic feature information from top to bottom through up-sampling. PAN transfers location feature information from bottom to top. Both are used simultaneously to fully integrate the low-level and high-level features, which effectively enhances the model's ability to detect objects of different scales.

To improve the real-time performance and detection capability of FD-YOLOV5s, the neck network was improved by changing the FPN+PAN network structure by considering the characteristics of the dataset. In the field of object detection, the most common definition of target size currently comes from the common dataset MS COCO dataset, which classifies objects into three categories based on area size, including small objects ($area \leq 32^2$ pixels),

medium objects ($32^2 < area \leq 96^2$ pixels) and large objects ($area > 96^2$ pixels). Figure 5 shows the object size distribution in the training set of this study. As can be seen from Figure 5, there were no small objects in this dataset, so the feature maps responsible for the detection of small targets can be considered for deletion. On the contrary, there were more large objects in this dataset, and the size distribution was not dense enough, hence the detection of large objects needs more attention.



Note: The horizontal and vertical coordinates indicate the length and width of the object detection box, respectively; The three color areas indicate the range of small, medium, and large objects respectively.

Figure 5 Distribution of the size of objects in the dataset

Therefore, The C3 module in front of the 20×20 feature map was replaced with our designed mixed dilated convolution (MDC) module to improve the detection capability of the feature map for large targets. Moreover, under the premise of ensuring the detection task, the up-sampling and down-sampling operations were reduced, and the 80×80 feature map was deleted, which further reduces the computational cost of the model. The structure of the neck network before and after the improvement is shown in Figure 6.

The MDC module enhances the detection of large targets by introducing dilated convolution. Dilated convolution expands the convolution kernel's receptive field by changing the convolution

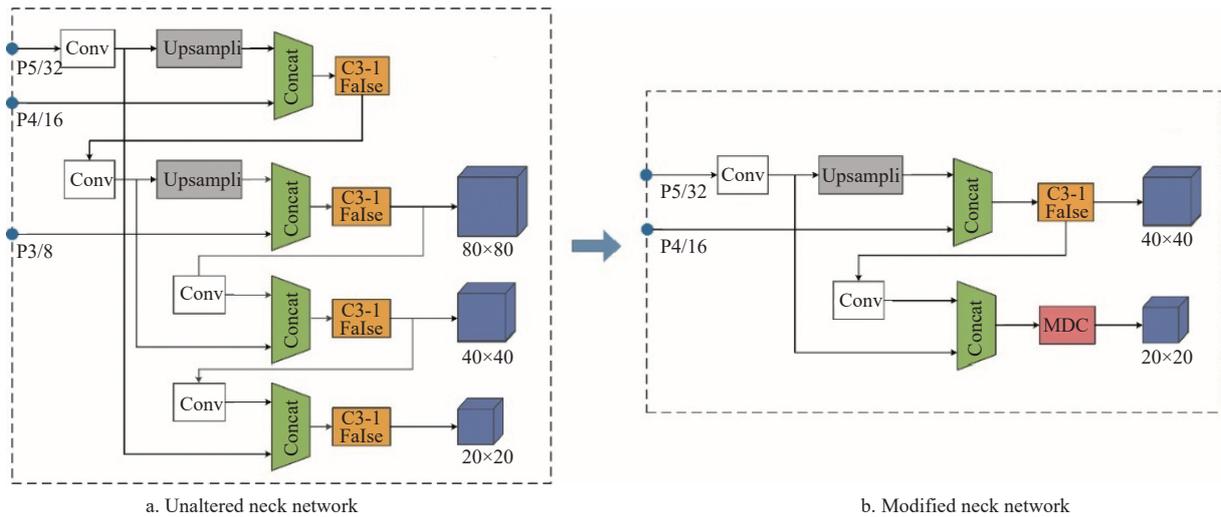


Figure 6 Improved FD-YOLOV5s neck network

kernel’s internal space. Compared with conventional convolution, the number of parameters of the dilated convolution does not increase with the expansion of the receptive field^[46], and the dilated convolution has a hyperparameter called the dilation rate, which needs to be set manually. Figure 7a shows the receptive field of a 3×3 convolution kernel at different scale dilation rates. There is no difference between the dilated convolution and the conventional

convolution when the dilation rate is equal to 1, but when the dilation rate is 2, the receptive field is 5×5. Similarly, when the dilation rate is 3, the receptive field becomes 7×7. Therefore, the receptive field of the dilated convolution increases with the dilation rate, and different scales of the dilation rate can extract feature information at different scales.

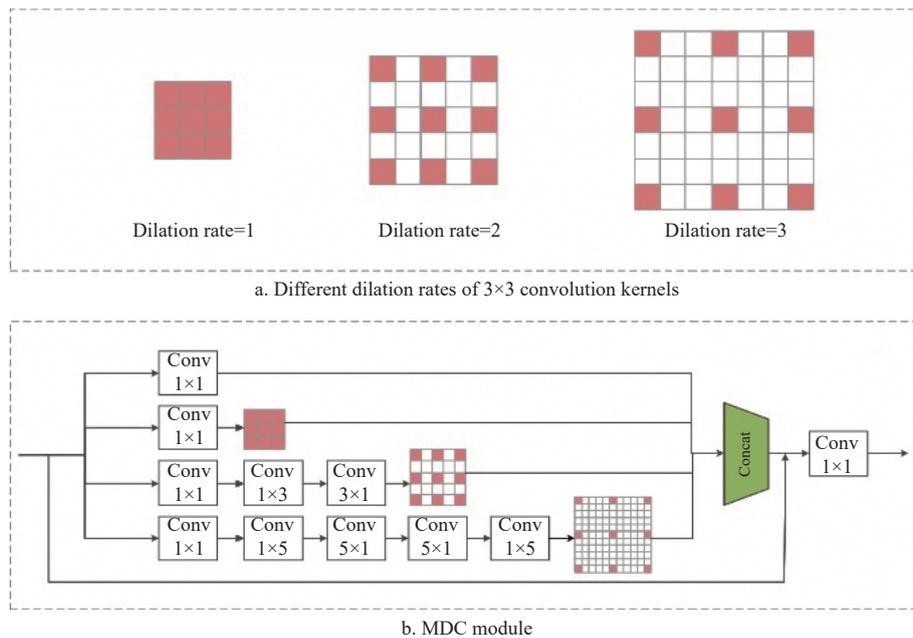


Figure 7 Different dilation rates of dilation convolution and the structure of the MDC module

The MDC module structure was inspired by Inception-V3^[47] network and can be seen in Figure 7b. First, the channel dimension of the input feature map was equally divided into four branches by 1×1 convolution, and different convolution depths and asymmetric convolution of each branch increased the diversity of semantic information. In addition, dilated convolution was introduced at the end of the three branches to increase the receptive field. The dilated convolution rate varied between branches, enabling the construction of diverse receptive fields so that the feature maps contained multi-scale information. Then, more global feature information was extracted by fusing the feature maps of different receptive fields. Finally, the residual structure was used to enhance feature delivery

and reduce the risk of gradient disappearance. Moreover, there were three dilated convolutions in the MDC module, and the dilation rates were defined as 1, 2, and 5, respectively, and the setting was experimentally verified to be optimal in Section 3.1.3.

2.3.3 Detection head

The head inherits the head structure from YOLO-V3, and the loss function consists of three parts: bounding box loss, classification loss, and confidence loss. Specifically, the complete intersection over union (CIoU) loss^[48] was used as the loss function of the bounding box, which better describes the regression of the rectangular box. Binary cross-entropy was used as a loss function for categorical loss and confidence loss to calculate category

probabilities and target confidence scores.

2.4 Temperature extraction

In this study, the infrared thermal image acquired by the infrared thermal camera has a resolution of 990×720 instead of 320×240 , as shown in Figure 8a. The reason for this is that in order to give the user a clear and intuitive view of the temperature distribution in the thermal image, the manufacturer processed the image through operations such as interpolation algorithms, adding temperature bars and pseudo color after the infrared thermal camera captured the 320×240 raw temperature data to increase the resolution of the image to 990×720 . To detect the temperature more accurately and quickly, the obtained temperature matrix was converted into a grayscale image using the calculation Equation (1).

$$\text{gray_img}[i, j] = \frac{M[i, j] - \min(M)}{\max(M) - \min(M)} \times 255, \quad i \in [1, 240], j \in [1, 320] \quad (1)$$

where, $\text{gray_img}[i, j]$ is the gray scale of row i and column j in the gray scale image, $M[i, j]$ is the temperature value of row i and column j in the temperature matrix, $\max(M)$, and $\min(M)$ are the highest and lowest temperature in the temperature matrix, respectively. The grayscale image obtained after applying this formula to the temperature matrix is shown in Figure 8b.

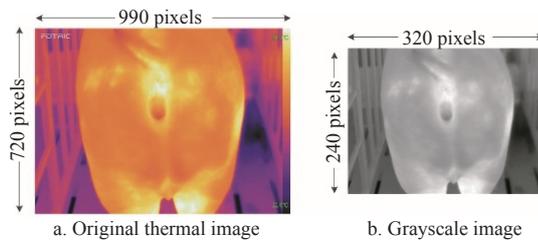


Figure 8 Image pre-processing

The extraction of sow vulvar skin temperature consists of three steps. Firstly, the grayscale image is input into the trained model to obtain the coordinates of the buttocks and vulva detection boxes in the image. Secondly, the relative positions of the buttocks detection box and vulva detection box are used to determine whether the buttocks are tilted or not, and this process is described in Algorithm 1. Finally, the maximum and average temperatures of the vulva area are calculated by mapping the coordinates of the vulva detection box to the temperature matrix. The steps of Algorithm 1 are as follows:

Input: The result of FD-YOLOV5s, i.e., the bounding box of the vulva and buttocks of the sow;

Output: Sow's buttocks posture;

Step 1 Define the x-axis coordinates of the upper left and lower right corners of the vulva bounding box and the buttocks bounding box as v_{left} , v_{right} , and b_{left} , b_{right} , respectively;

Step 2 Judge the sow's buttocks posture. When $v_{\text{left}} < (b_{\text{left}} + (b_{\text{right}} - b_{\text{left}})/3)$, the sow's buttocks inclined to the left, when $b_{\text{right}} > (b_{\text{right}} - (b_{\text{right}} - b_{\text{left}})/3)$, the sow's buttocks inclined to the right. If none of the above judgment conditions were satisfied, it means that the sow's buttocks showed no inclination;

Step 3 Stop the procedure if the sow buttocks inclined. On the contrary, continue to extract vulva temperature.

2.5 Estrus detection

The variation in sow vulva temperature can help analyze sow estrus and ovulation. There was a circadian rhythm in the vulva temperature of sows, with low values in the morning and high values in the evening^[18]. However, the average duration of estrus in sows was 52.6 h (range: 30-72 h)^[49], and vulvar temperature

increased with the onset of estrus and decreased significantly 36-12 h before ovulation^[3,23]. Therefore, when using vulva temperature to monitor sow estrus, the vulva temperature in the morning and in the afternoon was extracted first when the sow was standing to feed or drink once each by FD-YOLOV5s algorithm and stored in the database. Subsequently, a temperature curve was plotted using the last five consecutive vulval temperatures of the sow (each approximately 12 h apart), and the inflection point of this curve was calculated using the kneed algorithm^[50]. Finally, the trend of the temperature curve was obtained by the inflection point and the temperature value to judge whether the sow was in estrus or not, and the process is described in Algorithm 2. The above method was applied to monitor the 30 multiparous empty-pregnant sows mentioned in Section 2.1.1, and the schematic diagram of the estrous and non-estrous temperature profiles is shown in Figure 9. The steps of Algorithm 2 are as follows:

Input: The values and inflection points of the temperature curve;

Output: Sow status (estrus or non-estrus);

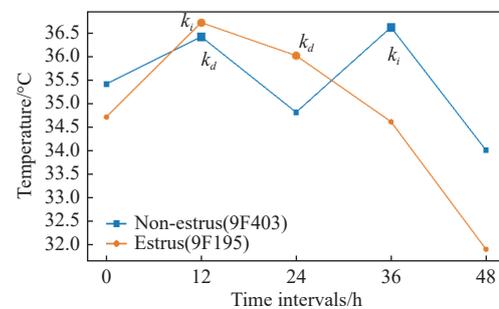
Step 1 Define the temperature value array, the position of the increasing inflection point and the position of the decreasing inflection point of the temperature curve as A_i , K_i and K_d respectively;

Step 2 Judge the shape of the temperature curve. When $A_i[0] \leq A_i[1]$ and $A_i[1] \leq A_i[2]$, there is a bump in the curve;

Step 3 Judge the trend of the temperature curve. When $K_i < K_d$, the curve is rising first and then falling;

Step 4 When K_i is not the first and penultimate point of the temperature curve, it means the sow is in estrus;

Step 5 If the above conditions are not met at the same time, the sow is in non-estrus.



Note: k_i is the inflection point of temperature increase; k_d is the inflection point of temperature decrease; 9F403 and 9F195 are the ear tag numbers of the sows, the estrus curve is plotted from the 48-h temperature values when estrus was detected in 9F195, and the non-estrus curve is plotted from the random 48-h temperature values when estrus was not detected in 9F403.

Figure 9 Temperature change curve of estrus and non-estrus

2.6 Evaluation metrics

2.6.1 Evaluation of the model performance

In order to verify the effectiveness of the improved algorithm, the evaluation of this study was carried out using five metrics: precision (P), recall (R), mean average precision (mAP), F1-score, and frames per second (FPS). The P and R are expressed in Equations (2)-(4). Before that, the number of true positives (TP), false positives (FP), and false negatives (FN) were need to be determined based on the intersection over union (IoU) (which was set to 0.5) between the detected box and ground-truth box and the confidence level (which was set to 0.2) threshold. The F1-score considers both P and R to calculate the score given in Equation (5).

$$\text{IoU}(R, R') = \frac{|R \cap R'|}{|R \cup R'|} \quad (2)$$

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (3)$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (4)$$

$$\text{F1-score} = \frac{2 \times P \times R}{P + R} \times 100\% \quad (5)$$

where, R is the area of the detected bounding box; R' is the area of the ground-truth bounding box; TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives, respectively; TP and FP denote the number of bounding boxes with IoU values greater than 0.5 and IoU values less than 0.5 between the detected and ground-truth boxes, respectively, and FN denotes IoU values equal to 0 for detected and ground-truth boxes, i.e., the number of undetected ground-truth boxes. F1-score is the harmonic mean of precision and recall.

The AP_i denotes the average precision of the i -th category with a value equal to the region under the P - R curve, calculated as in Equation (6). The definition of mAP is given in Equation (7) and indicates the average precision of the C categories. The k in mAP_k stands for the IoU threshold; mAP_{50} is the average of the precision of all categories when the IoU threshold is 0.5; $mAP_{50:95}$ is the average of 10 values of mAP_{50} , mAP_{55} , mAP_{60} , ..., mAP_{90} , mAP_{95} , as shown in Equation (8). In addition, the mAP in the following refers to mAP_{50} .

$$AP_i = \int_0^1 P_i(R_i) dR_i \quad (6)$$

$$\text{mAP} = \frac{1}{C} \sum_{i=1}^C AP_i \quad (7)$$

$$\text{mAP}_{50:95} = \frac{1}{10} (\text{mAP}_{50} + \text{mAP}_{55} + \dots + \text{mAP}_{90} + \text{mAP}_{95}) \quad (8)$$

2.6.2 Evaluation of temperature extraction and estrus detection

To assess the accuracy of the temperature extracted by FD-YOLOV5s, the manually extracted temperature m_i was compared with the temperature a_i automatically extracted by the algorithm. The following evaluation indicators were used: coefficient of determination (R^2), mean absolute error (MAE) and mean absolute percentage error (MAPE). R^2 is the squared Pearson correlation coefficient between manual extraction and automatic detection to measure the linear association between the two values; MAE is the mean of the absolute difference between manual extraction and automatic detection; MAPE is the mean of the absolute percentage error between manual extraction and automatic detection. The MAE and MAPE are expressed in Equations (9) and (10). To evaluate the effectiveness of estrus recognition in sows, the error rate, specificity, and sensitivity of estrus detection were calculated based on the results of the proposed algorithmic detection and manual detection, as expressed in Equations (11)-(13).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |m_i - a_i| \quad (9)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{m_i - a_i}{m_i} \right| \times 100\% \quad (10)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100\% \quad (11)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{FN} + \text{TP}} \times 100\% \quad (12)$$

$$\text{Errorrate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100\% \quad (13)$$

where, TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives, respectively; Specificity indicates the proportion of negative samples with correct predictions to total negative samples; Sensitivity indicates the proportion of positive samples with correct predictions to the total positive samples; Error rate represents the proportion of samples with incorrect predictions to the total sample. In the calculation, estrus and non-estrus were set as positive and negative, respectively.

3 Results and discussion

3.1 Vulva and buttocks detection

3.1.1 Training result analysis

In this study, Python, PyTorch, an Nvidia GeForce RTX 1080TI GPU, and an Intel(R) Xeon(R) Gold 6132 CPU @ 2.60GHz have been utilized to implement the proposed FD-YOLOV5s. The experimental environment is listed in Table 4. In model training, the training hyperparameters of the model were set with batch sizes of 16 and 300 iterations. The momentum, initial learning rate, and weight decay used the original parameters from the YOLO-V5 network. Figure 10 shows the changes in loss and mAP of FD-YOLOV5s during the training process. From Figure 10, the loss and mAP values decreased and increased rapidly before 50 iterations, respectively, and stabilized as the number of iterations increased. The final loss value stabilized at around 0.06, and mAP stabilized at around 99%. As the loss value and mAP gradually stabilized, the model gradually converged, and the training achieved the expected results.

Table 4 Experimental environment

Configuration	Parameter
Programming language	Python 3.8
Library and wrapper	PyTorch 1.9
CPU	Intel(R) Xeon(R) Gold 6132 CPU@2.60 GHz
GPU	Nvidia GeForce RTX 1080TI
Operating system	Ubuntu 18.06

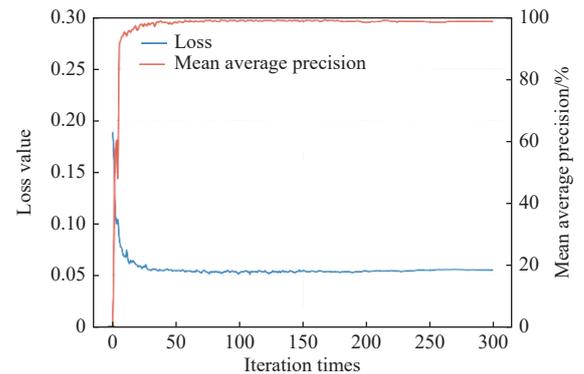


Figure 10 Variation of loss value and mAP value of the FD-YOLOV5s model during training

Figure 11 and Table 5 statistically analyze the model parameters of YOLO-V5s and FD-YOLOV5s and their performance on the validation set. As shown in Figure 10, FD-YOLOV5s is significantly better than YOLO-V5s in terms of model lightweight. Specifically, there are 76.9%, 71.8%, 81.1%, and 12.3% reductions in the number of parameters, model size, floating-point operations (FLOPs), and inference time, respectively, which facilitate the deployment of the model in embedded devices.

Moreover, as shown in Table 5, the proposed FD-YOLOV5s method outperformed YOLO-V5s for all the quantitative indexes of the two detected classes. Compared with YOLO-V5s, FD-YOLOV5s improved the mean F1-score and mAP by 0.3% and 0.7% to 97.6% and 99.2%, respectively.

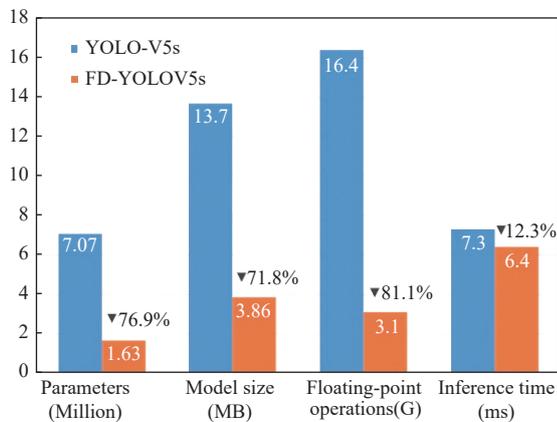


Figure 11 Comparison of model parameters of YOLO-V5s and FD-YOLOV5s

Table 5 Comparison of performance (F1-score and AP) in detecting vulva and buttocks on the validation dataset

Model	F1-score/%		MeanF1-score/%	AP/%		mAP/%
	Vulva	Buttocks		Vulva	Buttocks	
YOLO-V5s	98.7	95.7	97.3	99.6	97.5	98.5
FD-YOLOV5s	99.0	96.2	97.6	99.8	98.5	99.2

3.1.2 Effects of S5 and DSC modules

In Table 6, the impact of using the S5 module and DSC in the proposed FD-YOLOV5s detector backbone network has been mentioned. According to Table 6, the mAP of FD-YOLOV5s on the validation set increased from 98.5% to 99.1% when the original ShuffleNet-V2 (S3) module was used instead of the C3 module. Furthermore, the mAP was further improved when the modified S5 module was applied. In addition, when the algorithm used DSC instead of the Conv module for down sampling, mAP decreased slightly with the model size reducing by 2.92M. Thus, these methods increase efficiency.

Table 6 Effects of S5 and DSC modules in the proposed FD-YOLOV5s detector

Method	mAP/%	Size/MB
C3, Conv	98.5	13.7
S3, Conv	98.8	10.4
S5, Conv	99.1	10.4
S5, DSC	98.9	7.48

3.1.3 Effects of MDC module and feature map discarding

In order to detect large targets in the image, this research improved the neck network of YOLO-V5s, designed the MDC module to replace the C3 module before the large target detection feature map, and deleted the feature map responsible for small target detection to reduce the number of model parameters. The results are listed in Table 7. After deleting the feature map and re-clustering the initial anchor box, the mAP of the validation set rose from 98.5% to 98.7%, indicating that two feature maps were also applicable to this task. On the other hand, when replacing the C3 module with the MDC module, mAP increased to 99.4%. The

results show that the improvement of the large target feature map effectively improves the detection accuracy.

Table 7 Effects of MDC module and feature map discarding in the proposed FD-YOLOV5s detector

Method	mAP/%	Size/MB
F3, C3	98.5	13.7
DR, C3	98.7	9.4
DR, MDC	99.4	9.5

Note: F3: Original three feature maps; DR: deleting the feature map responsible for small target detection and re-clustering the initial anchor box.

In addition, the detection performance of MDC was also tried to evaluate when choosing different dilation rates. As listed in Table 8, the best detection performance was achieved when the dilation rates were 1, 2, and 5. It not only obtained global information to improve the detection accuracy of the buttocks but also took into account the detection of the vulva. However, as the dilation rate continues to increase, such as when the dilation rates were 1, 3, and 5, the average precision of the vulva decreased from 99.7% to 99.2%. Therefore, the dilation rates in the MDC module were selected as 1, 2, and 5.

Table 8 Results for the MDC module with different dilation rates

Dilation rates	AP/%		mAP/%
	vulva	buttocks	
1, 2, 3	99.5	98.4	98.9
1, 2, 4	99.6	98.2	98.9
1, 2, 5	99.7	99.1	99.4
1, 3, 4	99.3	98.9	99.1
1, 3, 5	99.2	98.8	99.0

3.1.4 Comparison of different models of YOLO-V5 detector

To evaluate the effectiveness and generalization of the proposed method, the performance of the YOLO-V5s applying different improvement strategies and YOLO-V5 on the test set were compared, and the results are listed in Table 9. Among the YOLO-V5 algorithms, YOLO-V5x had the best detection effect. However, after 1) replacing the C3 and Conv modules in the backbone network with the S5 and DSC modules, respectively, and 2) removing the small target detection layer and adding the MDC module to the large target detection layer were performed on YOLO-V5s, the highest mAP, R and F1-score were obtained. More specifically, R and mAP reached 99.2% and 99.1%, respectively, and F1-score was the same as YOLO-V5x. The results show that the model detection accuracy is effectively improved through a series of improvement strategies, and the model size was reduced.

Table 9 Comparison of detection parameters between YOLO-V5s with different improvement strategies and YOLO-V5

Model	P/%	R/%	F1-score/%	mAP/%	Size/MB
YOLO-V5x	97.6	98.5	98.1	98.6	166
YOLO-V5l	97.5	97.3	97.3	97.5	89.3
YOLO-V5m	96.9	97.9	97.3	98.2	40.4
YOLO-V5s	96.3	97.7	97.0	97.8	13.7
YOLO-V5s+①	95.7	98.6	97.1	98.1	7.48
YOLO-V5s+①+②(FD-YOLOV5s)	97.0	99.2	98.1	99.1	3.86

Note: 1) The C3 and Conv modules in the backbone were replaced with S5 and DSC modules, respectively; 2) The small target detection layer was removed and the MDC module was added to the large target detection layer.

In addition, to objectively evaluate the model's performance, ten-fold cross-validation was performed on the YOLO-V5s before and after the improvements. First, the dataset was randomly and equally divided into ten parts, nine of which were used for training the model and the rest for validation, and repeated ten times to ensure that each part constituted a validation set. Then, the FD-YOLOV5s and YOLO-V5s were trained on the same training set and tested on the same test set. The comparison of the performance of FD-YOLOV5s and YOLO-V5s in ten-fold cross-validation was shown in Table 10. The results showed that the proposed FD-YOLOV5s algorithm achieved higher accuracy compared to YOLO-V5s. The average values of F1-Score and mAP were 98.4% and 99.4%, respectively, which were both higher than YOLO-V5s. Moreover, the standard deviation (SD) was also smaller than YOLO-V5s.

Table 10 Comparison of the performance of YOLO-V5s and the proposed FD-YOLOV5s detector on tenfold cross-validation

No.	No. of Training/ Validation Images	YOLO-V5s		FD-YOLOV5s	
		mAP/%	F1-score/%	mAP/%	F1-score/%
1	3115/347	99.1	98.2	99.1	98.2
2	3115/347	99.3	98.3	99.4	98.3
3	3116/346	98.8	97.8	99.4	98.5
4	3116/346	99.0	97.8	99.2	98.4
5	3116/346	99.0	97.7	99.6	98.3
6	3116/346	99.1	98.3	99.5	98.8
7	3116/346	98.5	97.3	99.6	98.3
8	3116/346	99.4	98.2	99.1	98.5
9	3116/346	99.1	98.0	99.1	97.9
10	3116/346	99.2	98.3	99.8	98.6
Mean±SD	-	99.1±0.24	98.0±0.32	99.4±0.23	98.4±0.23

3.1.5 Comparison of detection methods

To verify the reliability and robustness of the FD-YOLOV5s detector, we compared it with nine mainstream target detection models: Faster RCNN (Vgg16)^[51], Faster RCNN (MobileNetv2)^[52], SSD 300 (ResNet50)^[53], ReanaNet^[54], YOLO-V3, Tiny YOLO-V3, YOLO-V4, Tiny YOLO-V4, and YOLOX-s^[55]. Each of these nine models was trained using the training set, and the performance of the different detection algorithms was evaluated on the test set.

The results of each detection model are listed in Table 11. Analysis of the data shows that FD-YOLOV5s detection accuracy was better than all the other nine algorithms and had the lowest number of parameters and smallest model size. Although Tiny YOLO-V3 and Tiny YOLO-V4 obtained better FPS, the mAP was still much lower than ours. The mAP for FD-YOLOV5s was 11.7%, 5.1% higher than Tiny YOLO-V3 and Tiny YOLO-V4, respectively. In addition, the mAP of ReanaNet was comparable to

Table 11 Comparison of detection parameters of different methods

Model	mAP ₅₀ /%	mAP ₅₀₋₉₅ /%	Size/MB	FPS	
				GPU	CPU
Faster RCNN (Vgg16)	96.40	69.00	334.00	18.87	1.16
Faster RCNN (MobileNetv2)	97.90	71.30	628.00	27.93	1.46
SSD 300	97.90	72.20	102.00	99.01	5.87
ReanaNet	99.00	73.20	245.10	18.76	1.03
YOLO-V3	93.30	67.10	235.10	84.75	10.13
Tiny YOLO-V3	87.40	57.20	33.10	400.00	58.82
YOLO-V4	96.20	70.30	244.00	50.51	4.43
Tiny YOLO-V4	94.00	66.10	22.50	312.50	48.54
YOLOX-s	94.20	71.60	68.50	69.93	8.28
FD-YOLOV5s	99.10	74.70	3.86	156.25	18.93

the proposed FD-YOLOV5s, but its model size was about sixty-three times larger than that of FD-YOLOV5s, and the inference speed was also the slowest among the ten methods. Hence, judging the detection effect, computational speed, and model size together, FD-YOLOV5s is the most reliable and suitable algorithm for embedded systems and portable devices for detecting the vulva and buttocks of sows.

3.2 Performance of temperature extraction

To verify the accuracy of the proposed algorithm for temperature extraction, the differences between manually extracted temperatures and automatically detected temperatures were compared. Specifically, 100 randomly selected thermal infrared images from the test set were used, and the temperatures acquired by using the AnalystIR 4.13 software (Fotric, Shanghai, China) were compared to the results detected by the proposed FD-YOLOV5s algorithm. In manual extraction, the thermal images were loaded into the AnalystIR 4.13 software, and the vulva region is selected by the mouse. Selected the maximum and average temperatures of the selected region provided by this software as a criterion to evaluate the proposed FD-YOLOV5s algorithm. Furthermore, to reduce subjective errors when manually selecting regions, the vulva temperatures were extracted by three people, and their mean values were taken as the gold standard.

As illustrated in Figure 12, for the maximum and average temperatures of the sow vulva, the FD-YOLOV5s model detection, and manual extraction showed essentially the same trend. Moreover, the MAE of the maximum and average temperature of the vulva was 0.03 and 0.14, and the MAPE was 0.1% and 0.42%, respectively, indicating that the proposed FD-YOLOV5s algorithm extracted the maximum and average vulva temperatures with a high accuracy rate.

Linear regression was also performed between manual extraction and automatic detection, as shown in Figure 13. Using the proposed FD-YOLOV5s detection model, the R^2 found by Pearson correlation was 0.995 and 0.993 for vulvar maximum and average temperatures, respectively, indicating a close correlation between the two measures.

Furthermore, 94% and 99% of the test files had errors within 0.1°C and 0.3°C for vulvar maximum and average temperatures, respectively, as shown in Figure 14. Therefore, the proposed FD-YOLOV5s algorithm can replace the manual extraction of vulvar maximum and average temperatures in the infrared thermal images of sows, showing the feasibility of applying the algorithm to automatically extract vulva temperature based on the infrared thermal images of sows.

The detection and temperature extraction results of the FD-YOLOV5s detector have been visualized in Figure 15. According to Figures 15a-15c, it can be seen that the FD-YOLOV5s detector can detect the vulva in various states very well. In addition, as shown in Figure 15d, the detector can determine whether the sow's buttocks are tilted to filter inaccurate vulvar skin temperatures and improve the reliability of vulvar skin temperature extraction.

3.3 Evaluation of sow estrus detection results

To further verify the practicality of the proposed algorithm, the results of manual estrus detection were used as the validation set. Among the 30 sows monitored, estrus was detected in 28 sows. The results of the estrus recognition algorithm are listed in Table 12. The algorithm used maximum and average temperature fragments for estrus detection with sensitivities of 89.3% and 85.7%, specificities of 94.5% and 95.9%, and error rates of 5.8% and 4.7%, respectively. The results showed that the sensitivity of the

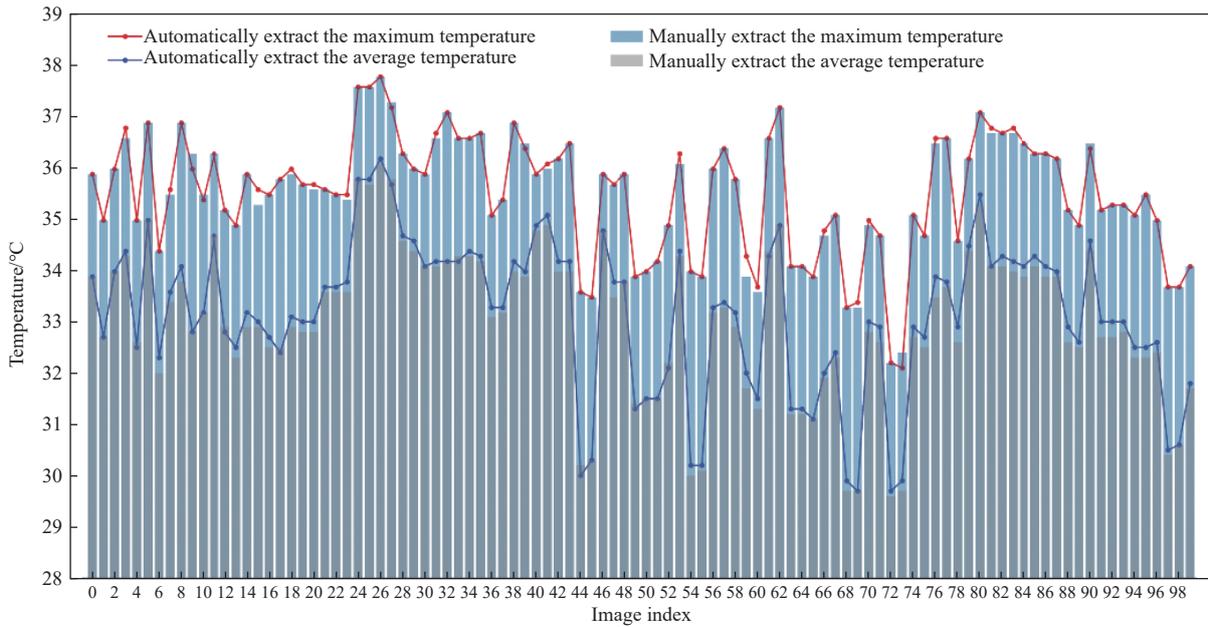


Figure 12 Comparison of the temperature extraction results between manually extracted values and automatic detection methods

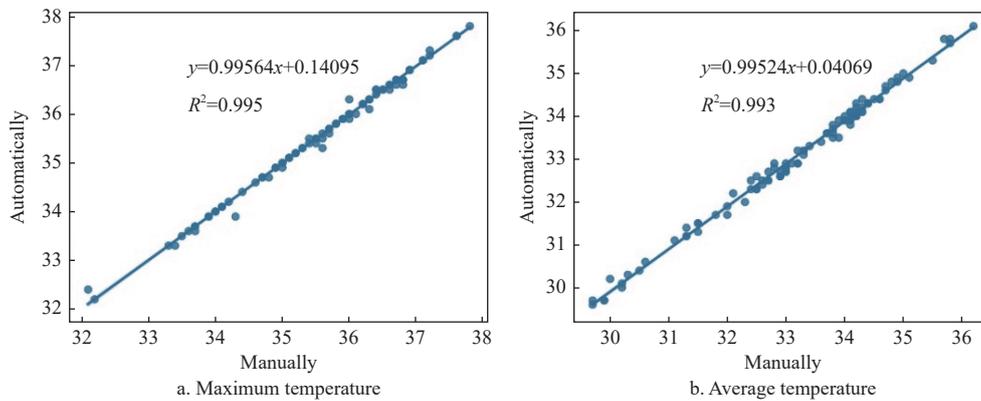


Figure 13 Linear regression between manually and automatically extracted temperatures

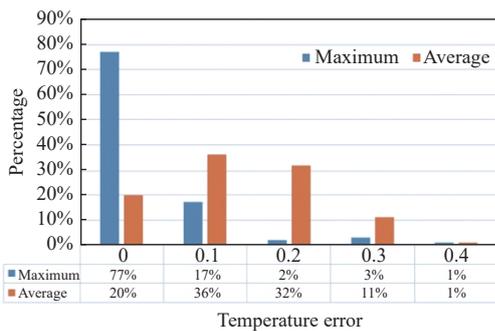


Figure 14 Error between the vulvar maximum and average temperatures extracted manually and those extracted by FD-YOLOV5s

maximum vulva temperature was higher compared with the average vulva temperature, and the maximum vulva temperature was less influenced by environmental factors in actual production, so the maximum vulva temperature was used for sow estrus detection with better results. Moreover, applying the maximum temperature of the vulva for estrus detection, the number of manual estrus detections could be reduced by more than 90%, accepting that 1.1% of sows were not detected during estrus (1.1% false negatives). Therefore, monitoring the maximum vulvar temperature may be more suitable for sow estrus detection.

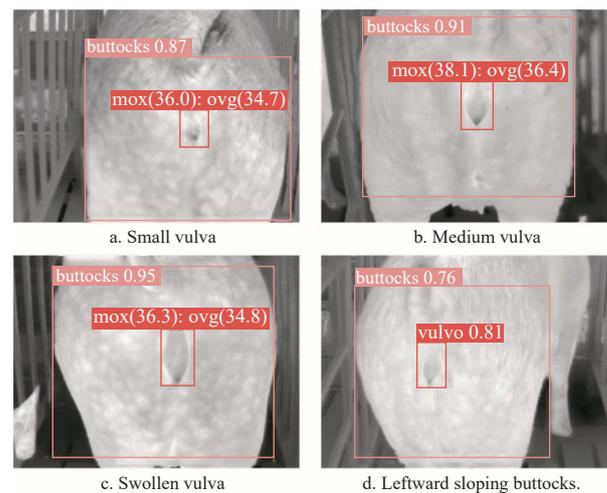


Figure 15 Visual detection and temperature extraction results of the proposed FD-YOLOV5s detector

Table 12 Results of sow estrus detection

Temperature	Total TP	Total FP	Total TN	Total FN	Error rate/%	Specificity/ %	Sensitivity/ %
Maximum	25	24	416	3	5.8	94.5	89.3
Average	24	18	420	4	4.7	95.9	85.7

3.4 Discussion

To date, many studies have been published on methods for detecting estrus in sows based on infrared thermography, thus providing a good research foundation. However, most studies on the extraction of sow skin temperature and the analysis of sow estrus are still based on manual manipulation. In this study, we described a new method for sow estrus detection. The key novelty of this method is the combination of deep learning and infrared thermography, which solves the problems of time consuming and unstable when extracting temperature manually, and achieves automatic detection of sow estrus by judging the trend of vulva temperature change in sows within 48 h.

From the example results of the estrus test (see Figure 9), it can be seen that the sow's vulvar skin temperature increases and then decreases during estrus, which is consistent with the conclusions drawn from previous studies^[3,23]. The difference is that we collected thermal infrared images of sows in their natural state, without additional cleaning and wiping of stains on their vulva before the shooting, which may also be one of the reasons why the sensitivity of the average vulva temperature in the estrus detection results (see Table 12) is lower than the maximum vulva temperature, but this approach is closer to the actual production scenario. In addition, when performing estrus detection on sows, estrus was judged by analyzing the trend of sow vulva temperature segments over a fixed period of time, instead of using only temperature increments^[1] or temperature thresholds^[26], which can not only solve the problem of significant differences between the vulva temperatures of sows and gilts^[23] but also greatly avoid the interference brought by environmental factors such as geography and season, so the method has a certain degree of universality.

The proposed FD-YOLOV5s network reached 99.1% detection mAP for positioning the key parts of the sow and the detection speed was 156.25 FPS, which was better than YOLO-V5 (see Table 9) and other mainstream object detection models (see Table 11). In addition, the algorithm test results in Section 3.2 showed that the error of the automatically extracted vulva temperature based on FD-YOLOV5s was within 0.4°C from the manual extraction, especially the highest temperature error of 94% of the test files was only 0.1°C, as shown in Figure 14. This means that the detection algorithm used in this study can extract the maximum vulvar temperature of the sow quickly and accurately. More importantly, the sensitivity of using the maximum temperature to detect estrus in sows reached 89.3%. Moreover, the model size of 3.86 MB is more convenient and can be applied to low-cost devices, reflecting a better application value. Therefore, the detection method proposed in this paper can successfully apply deep learning and infrared thermography to automate sow estrus detection. However, there are still some limitations to this study that need to be further addressed through follow-up work. First, the animal experiments in this study were completed within a month of the sow farm when only multiparous empty-pregnant sows were about to enter estrus, which limited the richness of the data set to the extent that was unable to verify the performance of the proposed estrus detection model on weaned sows and gilts. This urgently needs to be perfected in the future. Secondly, from the results of sow estrus detection, the error rate was high when judging sows in estrus based on the trend of the temperature curve obtained by the simple inflection point detection algorithm. Therefore, using machine learning methods to extract finer-grained change features from temperature segments to further improve the accuracy of estrus detection is the focus of our later work.

4 Conclusions

In this study, a real-time and lightweight object detection algorithm, FD-YOLOV5s, was proposed for the first time for the automatic extraction of sow vulva temperature from thermal images to achieve automatic detection of sow estrus. Experimental results showed that FD-YOLOV5s achieved 99.1% mAP on the test set, with detection speeds of 156.25 FPS on GPU and 18.93 FPS on CPU, respectively, and model size of only 3.86 MB. In addition, compared to commonly-used deep learning-based target detection network models, the FD-YOLOV5s detection algorithm was much faster while maintaining high accuracy. On the other hand, the maximum error in vulvar maximum and average temperatures was 0.1°C and 0.3°C for 94% and 99% of the 100 test images, respectively. The coefficients of determination for FD-YOLOV5s automatic detection and manual extraction of vulvar maximum and average temperatures were 99.5% and 99.3%, MAE were 0.03 and 0.14, and MAPE were 0.1% and 0.42%, respectively. Thus, the FD-YOLOV5s algorithm is a practical method for sow vulva and buttocks detection, providing accurate, real-time information on the vulva temperature of the sow and filtering inaccurate vulva temperature for monitoring the temperature changes of the sow during estrus. The estrus detection algorithm had a sensitivity, specificity, and error rate of 89.3%, 94.5%, and 5.8%, respectively, which can greatly reduce the number of manual estrus detection. In all, the method can quickly and effectively extract the vulva temperature and automatically detect estrus in sows. However, the detection accuracy of the sow estrus detection algorithm based on simple temperature trend changes is not ideal. In future work, we plan to monitor more sows, increase the sample size of estrus detection, and explore machine learning-based estrus recognition algorithms to extract more fine-grained temperature change features to improve the accuracy of sow estrus detection and make it more practical.

Acknowledgements

This work was financially supported by the Tianjin Key Research and Development Program Science and Technology Support Key Project (Grant No. 20YFZCSN00220), the Central Government Leading Local Science and Technology Development Special Project (Grant No. 21ZYCGSN00590), and the Inner Mongolia Autonomous Region Science and Technology Department Project (Grant No. 2020GG0068).

[References]

- [1] Lee J H, Lee D H, Yun W, Oh H J, An J S, Kim Y G, et al. Quantifiable and feasible estrus detection using the ultrasonic sensor array and digital infrared thermography. *Journal of Animal Science and Technology*, 2019; 61(3): 163–169.
- [2] Johnson J S, Shade K A. Characterizing body temperature and activity changes at the onset of estrus in replacement gilts. *Livestock Science*, 2017; 199: 22–24.
- [3] Weng R C. Variations in the body surface temperature of sows during the post weaning period and its relation to subsequent reproductive performance. *Asian-Australas Journal of Animimal Sciences (AJAS)*, 2020; 33(7): 1138–1147.
- [4] Knox R V. Artificial insemination in pigs today. *Theriogenology*, 2016; 85(1): 83–93.
- [5] Sandu M, Mantea Ş, Ipate I, Kruzslieka M, Chirişescu V. Study upon the moment of ovulation in sows to establish the optimum moment for semen inoculation. *Anim Sci Biotechnol*, 2012; 45: 346–348.
- [6] Cassar G, Kirkwood R N, Poljak Z, Bennett-Steward K, Friendship R M. Effect of single or double insemination on fertility of sows bred at an induced estrus and ovulation. *J Swine Health Prod*, 2005; 13(5): 254–258.

- [7] Kemp B, Soede N M. Consequences of variation in interval from insemination to ovulation on fertilization in pigs. *Journal of Reproduction and Fertility Supplement*, 1997; 52(Supp.): 79–89.
- [8] Knox R V, Esparza-Harris K C, Johnston M E, Weibel S K. Effect of numbers of sperm and timing of a single, post-cervical insemination on the fertility of weaned sows treated with OvuGel®. *Theriogenology*, 2017; 92: 197–203.
- [9] Terqui M, Guillouet P, Maurel M C, Martinat-Botté F. Relationship between peri-oestrus progesterone levels and time of ovulation by echography in pigs and influence of the interval between ovulation and artificial insemination (AI) on litter size. *Reproduction Nutrition Development*, 2000; 40(4): 393–404.
- [10] Soede N M, Langendijk P, Kemp B. Reproductive cycles in pigs. *Animal Reproduction Science*, 2011; 124(3-4): 251–258.
- [11] Ostersen T, Cornou C, Kristensen A R. Detecting oestrus by monitoring sows' visits to a boar. *Computers and Electronics in Agriculture*, 2010; 74(1): 51–58.
- [12] Lei K D, Zong C, Du X D, Teng G H, Feng F Q. Oestrus analysis of sows based on bionic boars and machine vision technology. *Animals*, 2021; 11(6): 1485.
- [13] Langendijk P, Soede N M, Bouwman E G, Kemp B. Responsiveness to boar stimuli and change in vulvar reddening in relation to ovulation in weaned sows. *Journal of Animal Science*, 2000; 78(12): 3019–3026.
- [14] Pedersen L J. Sexual behaviour in female pigs. *Hormones and Behavior*, 2007; 52(1): 64–69.
- [15] Pedersen L J, Rojkittikhun T, Einarsson S, Edqvist L E. Postweaning grouped sows: Effects of aggression on hormonal patterns and oestrous behaviour. *Applied Animal Behaviour Science*, 1993; 38(1): 25–39.
- [16] Kauffold J, Rautenberg T, Richter A, Waehner M, Sobiraj A. Ultrasonographic characterization of the ovaries and the uterus in prepubertal and pubertal gilts. *Theriogenology*, 2004; 61(9): 1635–1648.
- [17] Rezáč P, Vasičková D, Pöschl M. Changes of electrical impedance in vaginal vestibule in cyclic sows. *Animal Reproduction Science*, 2003; 79(1-2): 111–119.
- [18] Luño V, Gil L, Jerez R A, Malo C, González N, Grandía J, et al. Determination of ovulation time in sows based on skin temperature and genital electrical resistance changes. *Veterinary Record*, 2013; 172(22): 579.
- [19] Soede N M, Hazeleger W, Kemp B. Follicle size and the process of ovulation in sows as studied with ultrasound. *Reproduction in Domestic Animals*, 1998; 33(3-4): 239–244.
- [20] Cornou C. Automated oestrus detection methods in group housed sows: Review of the current methods and perspectives for development. *Livestock Science*, 2006; 105(1-3): 1–11.
- [21] Hidalgo D M, Cassar G, Manjarin R, Dominguez J C, Friendship R M, Kirkwood R N. Relationship between vaginal mucus conductivity and time of ovulation in weaned sows. *Canadian Journal of Veterinary Research*, 2015; 79(2): 151–154.
- [22] Sykes D J, Couvillion J S, Cromiak A, Bowers S, Schenck E, Crenshaw M, et al. The use of digital infrared thermal imaging to detect estrus in gilts. *Theriogenology*, 2012; 78(1): 147–152.
- [23] Scolari S C, Clark S G, Knox R V, Tamassia M A. Vulvar skin temperature changes significantly during estrus in swine as determined by digital infrared thermography. *Journal of Swine Health and Production*, 2011; 19(3): 151–155.
- [24] Simões V G, Lyazrhi F, Picard-Hagen N, Gayrard V, Martineau G P, Waret-Szkuta A. Variations in the vulvar temperature of sows during proestrus and estrus as determined by infrared thermography and its relation to ovulation. *Theriogenology*, 2014; 82(8): 1080–1085.
- [25] Weng R C, Ndwanwe S B. Application of modern estrus detection protocols in small scale Hybrid Black pig production systems. *Journal of Agricultural and Crop Research*, 2020; 8(6): 120–131.
- [26] Luño V, Gil L, Olaciregui M, Grandía J, Ansó T, De Blas I. Fertilisation rate obtained with frozen-thawed boar semen supplemented with rosmarinic acid using a single insemination timed according to vulvar skin temperature changes. *Acta Veterinaria Hungarica*, 2015; 63(1): 100–109.
- [27] Lu M, He J, Chen C, Okinda, C, Shen M, Liu L, et al. An automatic ear base temperature extraction method for top view piglet thermal image. *Comput Electron Agric*, 2018; 155: 339–347.
- [28] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017; 39(6): 1137–1149.
- [29] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, 2016; pp.779–788. doi: 10.1109/CVPR.2016.91.
- [30] Zhou C, Lin K, Xu D M, Liu J T, Zhang S, Sun C H, et al. Method for segmentation of overlapping fish images in aquaculture. *Int J Agric & Biol Eng*, 2019; 12(6): 135–142.
- [31] Yang Q M, Xiao D Q, Cai J H. Pig mounting behaviour recognition based on video spatial-temporal features. *Biosystems Engineering*, 2021; 206: 55–66.
- [32] Li G X, Liu X L, Ma Y F, Wang B B, Zheng L H, Wang M J. Body size measurement and live body weight estimation for pigs based on back surface point clouds. *Biosystems Engineering*, 2022; 218: 10–22.
- [33] Zhang X D, Kang X, Feng N N, Liu G. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. *Computers and Electronics in Agriculture*, 2020; 178: 105754.
- [34] Zhang Y, Cai J, Xiao D, Li Z, Xiong B. Real-time sow behavior detection based on deep learning. *Computers and Electronics in Agriculture*, 2019; 163: 104884.
- [35] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. SSD: Single Shot Multibox Detector. In: Proceedings of the European Conference on Computer Vision (2016ECCV), Amsterdam: Springer, Cham, 2016; pp.21–37. doi: 10.1007/978-3-319-46448-0_2.
- [36] Redmon J, Farhadi A. Yolov3: An incremental improvement. *arXiv preprint*, 2018. arXiv: 1804.02767. doi: 10.48550/arXiv.1804.02767.
- [37] Jocher G, Stoken A, Borovec J, NanoCode012, Chaurasia A, TaoXie, et al. Ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervisely and YouTube integrations (v5.0). Zenodo, 2021. Available: <https://zenodo.org/record/4679653>. Accessed on [2021-6-18].
- [38] Ma N, Zhang X, Zheng H T, Sun J. Shufflenet v2: Practical guidelines for efficient CNN architecture design. In: Proceedings of the European Conference on Computer Vision (2018ECCV), Munich, Germany: Springer, 2018; pp.122–138. doi: 10.1007/978-3-030-01264-9_8.
- [39] Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*, 2017. arXiv: 1704.04861. doi: 10.48550/arXiv.1704.04861.
- [40] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint*, 2015. arXiv: 1511.07122. doi: 10.48550/arXiv.1511.07122.
- [41] Siewert C, Dänicke S, Kersten S, Brosig B, Rohweder D, Beyerbach M, et al. Difference method for analysing infrared images in pigs with elevated body temperatures. *Zeitschrift für Medizinische Physik*, 2014; 24(1): 6–15.
- [42] Liu H, Shen H, Ci W B, Qi Z, Zhou C, Yao J X, et al. Effects of the distance and test angle on the precision of infrared temperature measurement. In: IOP Conference Series: Earth and Environmental Science, Bristol: IOP Publishing, 2022; 983(1): 012025.
- [43] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection. *arXiv*: 2004.10934, 2020. doi: 10.48550/arXiv.2004.10934.
- [44] Lin T-Y, Dollár P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA: IEEE, 2017; pp.936–944. doi: 10.1109/CVPR.2017.106.
- [45] Liu S, Qi L, Qin H F, Shi J P, Jia J Y. Path aggregation network for instance segmentation. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA: IEEE, 2018; pp.8759–8768. doi: 10.1109/CVPR.2018.00913.
- [46] Lu Z, Bai Y, Chen Y, Su C, Lu C, Zhan T, et al. The classification of gliomas based on a pyramid dilated convolution resnet model. *Pattern Recogn Letters*, 2020; 133: 173–179.
- [47] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, 2016; pp.2818–2826. doi: 10.1109/CVPR.2016.308.
- [48] Zheng Z H, Wang P, Liu W, Li J Z, Ye R G, Ren D W. Distance-IoU loss: Faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, California, USA: AAAI Press, 2020; 34(7): 12993–13000.
- [49] Almeida F R, Novak S, Foxcroft G R. The time of ovulation in relation to estrus duration in gilts. *Theriogenology*, 2000; 53(7): 1389–1396.
- [50] Satopaa V, Albrecht J, Irwin D, Raghavan B. Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In: 2011 31st

- international conference on distributed computing systems workshops, Minneapolis, MN, USA: IEEE, 2011; pp.166–171. doi: 10.1109/ICDCSW.2011.20.
- [51] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint, 2014. arXiv: 1409.1556. doi: 10.48550/arXiv.1409.1556.
- [52] Sandler M, Howard A, Zhu M L, Zhmoginov A, Chen L C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA: IEEE, 2018; pp.4510–4520. doi: 10.1109/CVPR.2018.00474.
- [53] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, 2016; pp.770–778. doi: 10.1109/CVPR.2016.90.
- [54] Lin T Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020; 42(2): 318–327.
- [55] Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: Exceeding YOLO series in 2021. arXiv preprint, 2021. arXiv: 2107.08430. doi: 10.48550/arXiv.2107.08430.