Bai *et al. Intell Robot* 2024;4:1-19
DOI: 10.20517/ir.2024.01

Intelligence & Robotics

Research Article

Open Access

Check for updates

Recognition of the behaviors of dairy cows by an improved YOLO

Qiang Bai[1,2,3], Ronghua Gao[1,3], Qifeng Li[1,3], Rong Wang[1,2,3], Hongming Zhang[2]

[1]Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China.
[2]College of Information Engineering, Northwest A&F University, Yangling 712100, Shaanxi, China.
[3]National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China.

Correspondence to: Dr. Ronghua Gao, Beijing Research Center for Information Technology in Agriculture, Postal address: No. 9, Shuguang Garden Middle Road, Haidian District, Beijing 100089, China. E-mail: gaorh@nercita.org.cn

How to cite this article: Bai Q, Gao R, Li Q, Wang R, Zhang H. Recognition of the behaviors of dairy cows by an improved YOLO. *Intell Robot* 2024;4:1-19. https://dx.doi.org/10.20517/ir.2024.01

Received: 16 Oct 2023 **First Decision:** 21 Nov 2023 **Revised:** 4 Dec 2023 **Accepted:** 23 Jan 2024 **Published:** 30 Jan 2024

Academic Editors: Jianjun Ni, Simon X. Yang **Copy Editor:** Yanbing Bai **Production Editor:** Yanbing Bai

Abstract

The physiological well-being of dairy cows is intimately tied to their behavior. Detecting aberrant dairy cows early and reducing financial losses on farms are both possible with real-time and reliable monitoring of their behavior. The behavior data of dairy cows in real environments have dense occlusion and multi-scale issues, which affect the detection results of the model. Therefore, we focus on both data processing and model construction to improve the results of dairy cow behavior detection. We use a mixed data augmentation method to provide the model with rich cow behavior features. Simultaneously refining the model to optimize the detection outcomes of dairy cow behavior amidst challenging conditions, such as dense occlusion and varying scales. First, a Res2 backbone was constructed to incorporate multi-scale receptive fields and improve the YOLOv3's backbone for the multi-scale feature of dairy cow behaviors. In addition, YOLOv3 detectors were optimized to accurately locate individual dairy cows in different dense environments by combining the global location information of images, and the Global Context Predict Head was designed to enhance the performance of recognizing dairy cow behaviors in crowded surroundings. The dairy cow behavior detection model we built has an accuracy of 90.6%, 91.7%, 80.7%, and 98.5% for the four behaviors of dairy cows standing, lying, walking, and mounting, respectively. The average accuracy of dairy cow detection is 90.4%, which is 1.2% and 12.9% higher than the detection results of YOLOV3, YOLO-tiny and other models respectively. In comparison to YOLOv3, the Average Precision evaluation of the model improves by 2.6% and 1.4% for two similar features of walking and standing behaviors, respectively. The recognition results prove that the model generalizes better for recognizing dairy cow behaviors using behavior videos in various scenes with multi-scale and dense environment features.

Keywords: Dairy cow behaviors, dense environment, YOLOv3, multi-scale, attention module

1. INTRODUCTION

Dairy cow behavior is tightly linked to economic rewards and reflects the health, welfare, and growth of the animal. Farms rely on human resources to spot dairy cows behaving abnormally and take appropriate action to prevent breeding losses[1,2]. However, this reliance on human resources for monitoring dairy cow behaviors consumes a lot of time and outstanding expenditure. Consequently, it has become increasingly popular to use information technology to automatically monitor behavior and quantify dairy cow behavior data to assess dairy cows' health[3,4]. This study breaks through the shortcomings of traditional manual observation such as missed detection, time-consuming and labor-intensive and uses visual analysis technology to monitor cow behavior online in real time, providing auxiliary support for individual health analysis.

Researchers use pedometers, neck rings, and other devices to obtain behavior data and realize automatic animal behavior recognition through machine learning algorithms[5-9]. To efficiently extract behavioral information and discern behavioral features within a high-dimensional data space, Ren *et al.* employed a neck ring for gathering data on the neck activity of dairy cows[10]. They perfected the penalty parameters and kernel functions of the support vector machine (SVM) using a firefly search algorithm, with an accuracy of 97.25% for the identification of three behaviors of drinking, eating, and ruminating for the same dairy cow, and 65.12% for experiments with multiple dairy cows. The algorithm is affected by the degree of individual behavior performance of dairy cows, and the universality is low. Wang *et al.* installed a posture sensor on the neck of the sow and obtained three-axis acceleration and angular velocity data to train the dairy cows' posture classification model to recognize three postures: standing, lying, and mounting, and the recognition accuracy was 98.02%[11]. Traditional machine learning algorithms suffer from issues such as poor identification performance, sluggish model inference, contact device research, easy animal stress, and easily broken and damaged devices that result in financial losses. Khin *et al.* combined the feature points positions of cows, bounding box regions, and SVM machine learning methods to classify six behaviors of cows: standing, sitting, eating, drinking, sitting with leg extension, and tail raised[12]. The classification accuracy reached 88.75%. This article provides a new approach to detecting fine-grained behavior in cows.

Deep learning techniques use Convolutional Neural Networks (CNN) to extract visual features[13,14] and solve the problems of Object Detection[15,16], Instance Segmentation[17,18], Object Tracking[19,20], *etc.* The deep learning method has higher accuracy and contactless features and was widely used in the field of animal behavior recognition[21-23]. Yan *et al.* used feature pyramid attention to improve the Tiny-YOLO target detection network to detect individual pigs, and the recognition result was 85.85%[24]. This research compared the model in the circumstances of pig adhesion and obscuration, and the recognition rate was 90% in the case of individual pigs without obscuration and without adhesion and 66.96% in the case of obscuration and adhesion. The aforementioned findings suggest that the data's occlusion situation has a significant impact on the model recognition effect. Li *et al.*[25] labeled 16 key points of bovine body parts and used convolutional pose machine[26], stacked hourglass network[27], and convolutional heat map regression model[28] to predict key points of bovine body parts with the highest score of 0.904, providing a new idea for bovine pose estimation. Wang *et al.*[29] classified three dairy cow behaviors based on AlexNet[30] classification network for lying, walking, and mounting with 100% accuracy, but affected by the change of light at night, the model recall was low at 88.24%. Since there was just one dairy cow in the data used to train the AlexNet model, the impact of background, shading, and surroundings on behavior detection was reduced. Consequently, this model did not properly recognize dairy cow behaviors in communal farming situations. Yin *et al.* used a Bi-directional Feature Pyramid to improve the Bi-directional long short-term memory model to learn the motion features of a single cattle behavior in the video to recognize lying, standing, and

walking for the multi-scale features of cattle targets[31]. The average accuracy of recognition was 95.2%, but the video-based visual behavior recognition model algorithm inference speed is slow and cannot recognize cattle behavior in the video in real time.

In summary, deep learning cow behavior detection research has problems such as model inference delay and simple data samples. In order to solve the problem of difficulty in locating individual cows and multi-scale behavior detection of cows in intensive breeding environment, we introduced Global Context Block (GC Block) and Res2 backbone to improve the YOLOV3 detection model to improve the model's ability to locate cows and detect behaviors. The cow behavior detection model we built has an average accuracy of 90.4% for detecting the four behaviors of cows standing, lying, climbing, and walking, which is greatly improved compared to the detection results of models such as YOLOV3 and YOLO-tiny.

2. RELATED WORK

In this section, we have compiled recent research related to object detection and animal behavior recognition. We conducted an analysis of these articles to emphasize the significance of our work. YOLOV3, as a classic object detection model, has been widely employed by researchers in the agricultural domain for tasks such as animal behavior detection and disease identification. We have gathered information from recent years regarding cow behavior detection and other related studies on animal behavior detection, with specific details outlined in the table below.

Table 1 lists relevant studies on animal behavior recognition, with the "-" symbol indicating that the paper did not provide corresponding data. Some researchers employ hardware-based methods for cow behavior detection. These approaches utilize accelerometers to capture the motion information of cow behavior, extract behavioral features, and differentiate cow behaviors through machine learning methods such as SVM and random forests. For instance, Benaissa *et al.* installed triaxial accelerometers on the legs and neck of cows to obtain behavior information, distinguishing cow behavioral features through SVM[32]. The study achieved an average accuracy of 95.0% in detecting cow lying and standing behaviors. However, accelerometers are susceptible to chemical erosion, leading to damage and decreased accuracy, resulting in a shorter equipment lifespan. Additionally, the high deployment cost arises from the need to install accelerometers for each cow.

Computer vision-based methods can effectively address the aforementioned issues. In the study by Wang *et al.*, monitoring cameras were deployed in a cattle farm to capture surveillance videos[33]. Subsequently, 15,120 images of cows crawling were extracted from these videos to train an improved YOLOV3 model. The accuracy achieved was 99.15%. Although the model, built on a dataset with a limited number of cows and a single scene, demonstrated high accuracy, it may suffer from overfitting due to the constraints of the training dataset, making it unsuitable for detecting cow behavior in diverse scenarios. Zhang *et al.* enhanced the YOLOV3 model, combining it with key point detection for recognizing cow lying and standing behaviors, achieving an accuracy of 99.39%[34]. In this study, the improved YOLOV3 model had a size of 234.96 MB, whereas our model had a smaller size of 123.9 MB.

Ma *et al.* processed surveillance videos to obtain 286 segments of cow behavior videos, including standing, lying down, and walking behaviors[35]. They trained a 3D deep learning model, achieving an accuracy of 95%, with a model size of only 14.3 M. This model is suitable for videos targeting a single cow; however, it cannot be directly applied to surveillance videos to obtain individual cow behavior videos, limiting its deployment feasibility. In Ji's study, monitoring cameras were used to capture video segments of cows' rumination behavior. The behavior was recognized based on FlowNet 2.0, achieving an accuracy of 99.39%.

Table 1. Research related to animal behavior recognition

Publication year	Author	Country	Livestock species and task objectives	Used model and model size	Data collection method	Dataset size	Advantages/disadvantages
2019	Benaissa *et al.*[32]	Belgium	Cows. Cow lying and standing behavior,	Support vector machine model size: -	Leg and neck tri-axial accelerometers	-	Detecting two cow behaviors with an accuracy of 95.0%.
2021	Wang and He[33]	China	Cows. Detecting cow crawling behavior	YOLOV3, model size:-	Surveillance camera	15,120 Images	Ideal dataset, suitable for laboratory environments. Accuracy: 99.15%.
2022	Zhang *et al.*[34]	China	Beef Cattle. Combining object detection and keypoint recognition for recognizing cattle lying and standing behaviors.	YOLOV3, model size: 234.96 MB	Surveillance camera	9,786 Images	Accuracy: 97.18%
2022	Ma *et al.*[35]	China	Dairy cow. Recognizing three behaviors of Dairy cows: lying down, walking, and standing.	Rexnet 3D Model size: 14.3 M	Surveillance Camera	286 segments of cow behavior videos, totaling 181,500 images.	Accuracy is 95%, considering only the behavior of individual cows.
2023	Ji *et al.*[36]	China	Dairy Cows. Detecting rumination behavior in cows.	FlowNet2.0 Model Size: -	Surveillance camera	30 segments of rumination videos for cows, each segment containing 300 images.	Accuracy is 99.39%.
2023	Guo *et al.*[37]	China	Meat pigeons. Detecting five behaviors: self-preening feathers, mutual preening, wing spreading, and others.	YOLOV4 Model Size: 265.3 MB	Surveillance Camera	3,120 images of meat pigeon behavior.	Accuracy: 97.97%. The data is suitable for use in a laboratory environment.
2023	Hu *et al.*[38]	China	Sheep. Detecting three behaviors: feeding, standing, and lying down.	YOLOV5s Model size: 138.3 MB	Camera	1,656 images of sheep behavior.	Accuracy: 91.8%. The number of sheep per image is relatively low.
Ours		China	Dairy cows. Detecting four behaviors: eating, drinking water, lying down, and standing	YOLOV3 Model size: 123.9 MB	Camera	1,033 images of dairy cow behavior.	Accuracy is 91.2% Suitable for detecting cow behavior in a real farming environment.

Guo *et al.* obtained 3,120 images containing five behaviors of meat pigeons, including preening feathers and wing spreading. They achieved an accuracy of 97.97% in detecting meat pigeon behavior based on the YOLOV4 model[37]. Hu *et al.* used handheld cameras to capture 1,656 images containing three behaviors of sheep: standing, feeding, and lying down[38]. The average accuracy of detecting sheep behavior with the YOLOV5s model was 91.8%, but the number of sheep per image was low, making it unsuitable for detecting behavior in a group setting.

We developed a cow behavior detection model based on YOLOV3. Compared with the YOLOV4 and YOLOV5 models in Table 1, the model size has certain advantages. Meanwhile, considering the challenges in the current cow behavior data, especially in the group farming environment, where the occlusion of cow behavior images makes it difficult to accurately locate individual cows, and the difficulty of multi-scale cow behavior detection, we constructed an efficient cow behavior detection model based on YOLOV3.

3. MATERIALS

In September 2021, at Da Di Qun Sheng Cattle Farm in Beijing, we used a 2-megapixel camera to capture videos of cow behavior. The camera was stabilized using a tripod or handheld to obtain a diverse range of cow behavior footage. The video resolution and frame rate were set at 1,920 × 1,080 and 25 frames per second, respectively. We filmed cow behavior at different times, angles, and locations. This pixel configuration enables a clear representation of cow behavior, scene details, and lighting information. The animal use protocol listed below has been reviewed and approved by the Ethics Committee, Northwest A&F University, China.

3.1 Build the dataset

We extract images from the videos, measuring the similarity between two images using Structural Similarity Algorithm (SSIM) and Mean Squared Error (MSE), aiming to obtain a more diverse set of cow behavior images. After manually removing unnecessary images with excessive blurring, we obtained a total of 1,295 cow behavior images. These images were randomly divided into a training set (1,033 images) and a test set (262 images) at an 8:2 ratio. The samples of data are shown in Figure 1.

Samples of dairy cow behaviors in various situations, lighting conditions, and viewpoints are shown in Figure 1A-D. In Figure 1A, a sample of mounting behavior is shown. A sample of standing and mounting around in enough lighting is shown in Figure 1B and C. The dataset's distribution of the number of dairy cows in a single image is examined in Figure 1E, where the ordinate is the number of corresponding photographs, and the abscissa is the number of dairy cows. The majority of the dairy cows in Figure 1E are between 1 and 20, and the images with more than 40 dairy cows are the least. Dairy cow behavior dataset shows signs of a change in degree of intensity. Light, backdrop and multiple scales are among the features of the dairy cow behavior dataset created as detailed above. According to the four criteria of walking, standing, lying, and mounting listed in Table 2, dairy cow behaviors in images were categorized and counted.

It can be seen from Table 2 that the number of walking was 3,107 and the number of mountings was at least 179. The amount of dairy cow behaviors can be seen in Table 2. The distribution of the number of dairy cow behaviors is the same as the distribution of the number of dairy cow behaviors in a realistic dairy farming environment.

3.2 Hybrid data augmentation method

The expression of dairy cow behavioral information varies greatly depending on the quantity of dairy cow samples and the quantitative differences between dairy cow behaviors, which affect model convergence. To expand the number of samples and give the model rich data on dairy cow behaviors. We employ data augmentation approaches, which boost the model's recognition results. Brightness variation, left-right inversion, and Mosaic[39] were three data augmentation techniques used in this research to improve the original photos. In Figure 2, the enhancement impact is depicted.

The original sample image is displayed in Figure 2A. The result of using left-right flipping, changing the lighting, and enhancing Mosaic data is shown in Figure 2B-D. Due to the high similarity between the left and right flipped images and the original image, it is easy to generate redundant images. To limit the number of comparable photographs, we set $\partial \sim N(0,1)$ (∂ follows a normal distribution with a mean of 0 and a variance of 1) and flip the image left and right when $\partial > 0.5$. To enhance scene and behavior information, the mosaic augmentation approach creates one image from four images in the dataset. The dairy cow behavior image does, however, contain some crowded scenarios. To stop the splicing of numerous dense photos from increasing the number of dairy cows in the image that is not complete, a random number

Table 2. Dairy Cow Behaviors Criteria

Category	Description of behavior	Labels	Number of labels
Walking	Leg cross (got in motion)	Walk	850
Standing	Legs upright to support the body	Stand	7213
Lying	Leg-to-ground contact	Lie	3107
Mounting	Two dairy cows mounting	Mount	179

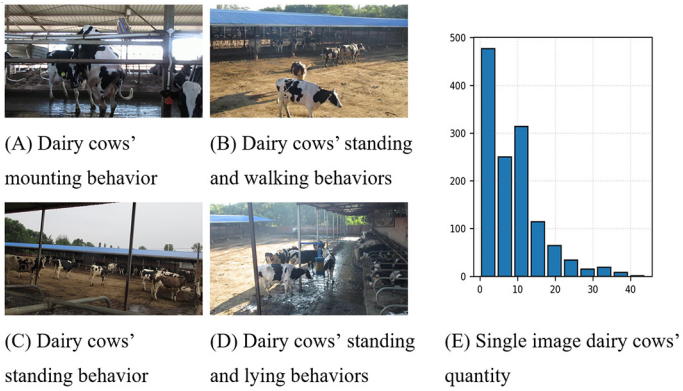


Figure 1. Sample of dairy cows' behaviors dataset.

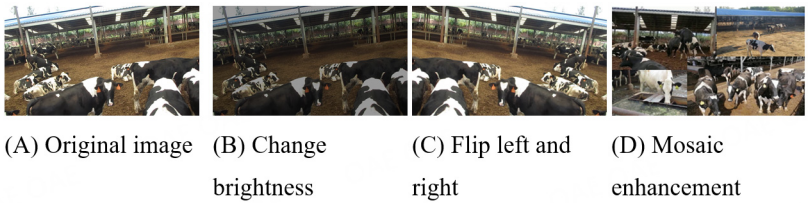


Figure 2. Sample of enhancement dataset.

$\partial \sim N(0,1)$ was generated. Mosaic enhancement at $\partial > 0.5$ to reduce the occurrence of many crippled dairy cow individuals. In order to make the enhanced image show a regular brightness change characteristic, we set the gain parameter $\beta \sim U[-1,1]$ (β obeys the mean distribution of [-1,1]) and change the image brightness on the V-space in HSV for the input image. Expediting model convergence through the application of a hybrid augmentation technique that involves random enhancements to the data pertaining to dairy cow behaviors.

4. METHODS

4.1 Behaviors recognition using YOLOv3

YOLOv3[40] is a One-Stage Object Detection model with high speed to meet the requirements of real-time detection, so it was chosen as the base model to recognize dairy cow behaviors, and its structure is shown in Figure 3.

YOLOv3 extracts dairy cow behavior features from the DrakNet53 and predicts behavior using detectors after feature processing (Neck). For an input dairy cow behaviors image size of 640 × 640, the detector uses convolutional kernels of size 1 × 1 on three feature maps of 20 × 20 pixels, 40 × 40 pixels, and 80 × 80 pixels with abstract semantic information. The parameter $k(k = 3)$ is the number of anchors whose size was

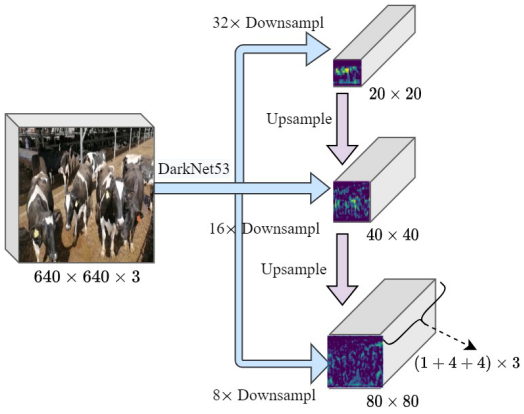


Figure 3. Structural diagram of YOLOv3.

obtained by k-means of the labeled boxes in the dairy cow behavior dataset. The parameter $c(c = 4)$ is the number of behavior categories, ck parameters are the probabilities of the four dairy cow behaviors in the bounding box, $4k$ parameters are the offsets of the bounding box, and k prediction parameters represent the probability that the dairy cow is included in that bounding box. YOLOv3 predicts the bounding box in a regression manner, as shown in Figure 4.

$$b_x = \sigma(t_x) + C_x \tag{1}$$

$$b_y = \sigma(t_y) + C_y \tag{2}$$

$$b_w = P_{us} e^{t_w} \tag{3}$$

$$b_h = P_h e^{t_h} \tag{4}$$

Figure 4 shows a feature map of 20 × 20 size. The anchors are represented by the dashed boxes in Figure 4, and each pixel (Grid Cell) on this feature map corresponds to a 32 × 32 pixel section of the input image. The corresponding variables and their meanings in Figure 4 are shown in Equations 1-4. The detector uses 1 × 1 convolution to generate 3 × 9 different scaled bounding boxes in each Grid Cell, illustrated by a 1:1 aspect ratio bounding box. Each bounding box has four offsets (t_x, t_y, t_w, t_h) relative to the Grid Cell upper-left coordinate (c_x, c_y), which are the center offsets of the network-predicted bounding boxes (t_x, t_y), and a width-height scaling factor (t_w, t_h), which predicts the bounding box size (b_x, b_y, b_w, b_h). The model achieves the categorization and localization of dairy cow behavior recognition by bounding box regression.

4.2 Improve YOLOv3 recognition model

4.2.1 Res2 block feature extraction network

The Bottleneck Block in DarkNet53 employs a convolution of 3 × 3 to extract the feature image of a fixed receptive field, but because it cannot successfully fuse multi-scale data, the model is uneven in how well it can identify the behavior of different-sized dairy cows.

Res2 Block[41] enhances the DarkNet53 backbone by creating a residual structure with hierarchy within a single residual block. The more robust backbone unifies multi-level sensory fields throughout the convolution process, extracts dairy cow behaviors information, and reduces the impact of multi-scale features on the model's detection ability. The structure of the Res2 Block and Bottleneck Block is shown in Figure 5.

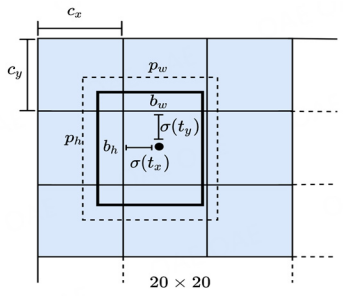


Figure 4. Bounding box regression.

Two 1 × 1 convolutional and one 3 × 3 convolutional processes make up the Bottleneck Block, which extracts feature pictures with fixed receptive fields. To extract picture information, the Res2 Block's structure employs a cascade of three 3 × 3 convolutions. The receptive field of the feature map obtained using two convolutions of 3 × 3 is identical to that of the feature map obtained using a single convolution of 5 × 5. The receptive field of the feature map that was derived using three convolutions of 3 × 3 is identical to that of the feature map that was extracted using a single convolution 7 × 7. As a result, Res2 Block mixes feature maps with varied receptive fields from 1 × 1, 3 × 3, 5 × 5 and 7 × 7. Equation 5 is the definition of the Res2 Block formula.

Suppose x is the input feature image and divide x into four feature maps (x_1, x_2, x_3, x_4) at the channel level, and the number of channels for each feature map is $w \times h \times c/4$. The parameter $i \in \{1,2,3,4\}$; parameter Ki denotes the convolution size of 3 × 3 and the output is yi. The four feature maps y_1, y_2, y_3, y_4 are stitched at the channel level, and the number of channels of the stitched feature maps is adjusted using a convolution kernel size of 1 × 1 and then fused with the input features. As a result, we obtained multi-scale receptive field feature images. In terms of the number of parameters Bottleneck uses 3 × 3 × c parameters and Res2 Block uses 3 × 3 × 3 × $c/4$ parameters, the number of parameters is reduced by 9 × $c/4$. Using a Res2 Block convolution, 1,152 parameters can be decreased for a feature map with 512 channels. In addition to lowering model parameters, creating a backbone based on the Res2 Block enriches multi-scale behavioral features by incorporating added sensory fields, which enhances the model's ability to recognize several scales of activity in dairy cows.

$$\mathbf{y}_i = \begin{cases} \mathbf{x}_i & i = 1 \\ \mathbf{K}_i\left(\mathbf{x}_i\right) & i = 2 \\ \mathbf{K}_i\left(\mathbf{x}_i + \mathbf{y}_{i-1}\right) & 2 < i,, s \end{cases} \tag{5}$$

4.2.2 Global context block

Individual dairy cows are obscured to varying degrees in images of dairy cow behaviors in dense environments, making it more challenging for detectors to find dairy cow targets. The convolution layer extracts local features of the image and ignores the global information of the feature map. The size of the feature image is uniformized using the GC Block[42] via channel compression and dimensional transformation. Then, GC Block creates the long-distance position dependence between the individual dairy cow and the global image by fusing the retrieved global features with the input features.

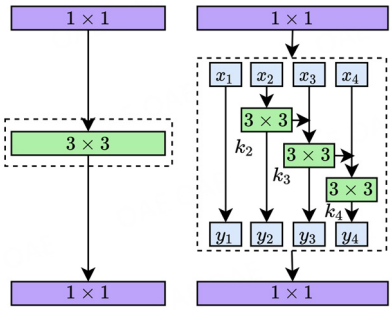


Figure 5. Bottleneck Block (left), Res2 Block (right).

The approach makes use of the information exchange between individual dairy cows and global images, and the model combines dairy cow targets with contextual data to find targets. The localization accuracy of dairy cows in the model can be increased by using GC Block to enhance the YOLO detectors. Figure 6 depicts the GC Block's structural layout.

GC Block consists of three operations: Context modeling, transform, and $\oplus$. Context modeling models the global contextual information of dairy cow behavior images and outputs global feature information. The three operations that make up the GC Block are context modeling, transform, and broadcast element-wise addition designed $\oplus$. Images of dairy cow behaviors are subjected to context modeling, which produces global feature information. Based on the bottleneck structure, the transformer collects various channel correlations and derives channel dependencies. The input features are combined with global contextual features by the $\oplus$ procedure. During the training process, the model creates the individual dairy cows' long-term reliance on the overall image. Equation 6 defines GC Block.

$$\mathbf{z}_i = F\left(\mathbf{x}_i, \delta\left(\sum_{j=1}^{N_p} \alpha_j \mathbf{x}_j\right)\right) \tag{6}$$

Equation 6 uses the sigmoid activation function $\delta(\cdot)$ to quantify the weight information, a_j to signify the global attention weight, N_p to denote the number of global pixel points, X to denote the input feature map, and $F(\cdot)$ to denote the $\oplus$ operation. Let the input feature map dimension be $C \times H \times W$; C is the number of channels, H is the feature map height, W is the feature map width, and the $C \times H \times W$ feature map is mapped to $C \times 1 \times 1$ dimensional global features by Context modeling. The Transformer structure uses a convolution of size 1×1 to compress the global features and obtain the dependencies between channels, where the compression multiplier r is 16 to enhance the GC Block generalization. The Layer Normalization regularization operation is used to output the global context information with dimension $C \times 1 \times 1$. The $F(\cdot)$ operation fuses the global location features with the input features in the form of broadcasts, fusing the information obtained in GC Block to improve the accuracy of the model's localization of dairy cows.

4.2.3 GC_Res2 YOLOv3 model

The shade between individual dairy cows and the multi-scale feature will have an impact on the model recognition outcomes for the intensive breeding scenario. This is done by enhancing the backbone and optimizing the detectors to create the dairy cow behavior recognition model, which is represented by the GC_Res2 YOLOv3 structure in Figure 7.

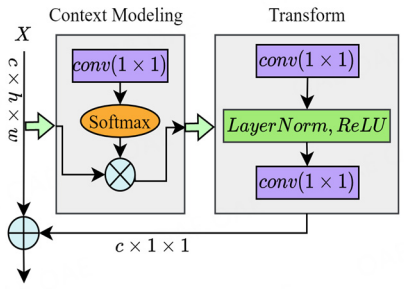


Figure 6. GC Block module structure.

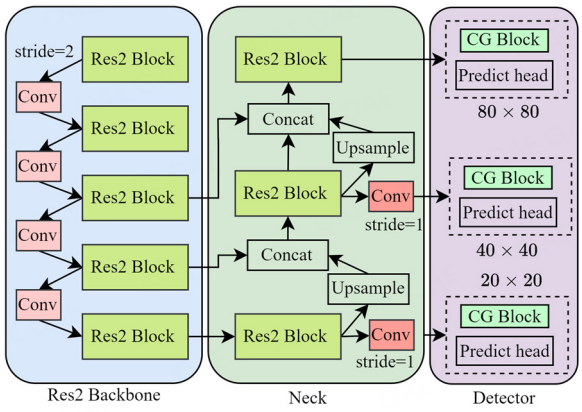


Figure 7. Schematic diagram of GC_Res2 YOLOv3 structure.

Three detectors, Neck and the Res2 backbone make up GC_Res2 YOLOv3. Res2 Block and convolutional block make up the Res2 backbone. To enhance the outcomes of multi-scale dairy cow behavior recognition, the backbone uses five Res2 Blocks to extract features with numerous scales on feature maps at various sizes. The GC Block is used in the detector section of the model to optimize the detectors and perform multi-scale behavior recognition of dairy cows in dense environments. This increases the model's localization accuracy for individual dairy cows. To significantly increase the bounding box model's forecast precision. A prediction box that is more similar to the actual bounding box was created using the loss function L_{CIoU}, which has the similarity ratio of the bounding box, as outlined in Equations 7-10.

$$L_{CIOU} = 1 - R_{IOU} + R_{CIOU} \tag{7}$$

$$R_{CIOU} = \frac{\rho^2\left(b, b_{gt}\right)}{l^2} + \alpha v \tag{8}$$

$$\alpha = \frac{v}{1 - R_{IOU} + v} \tag{9}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w_{Rt}}{h_{gt}} - \arctan\frac{w}{h}\right)^2 \tag{10}$$

The variable b in Equation 8 is the predicted bounding box, the variable b_{gt} is the ground truth, and the variable l is the minimum closed diagonal length of b and b_{gt}. The variable a in Equation 11 is the weighting

coefficient, and the variable v is similar in length and width. The variable $\rho^2(b, b_{gt})$ in Equation 9 is the Euclidean distance between b and b_{gt}. The variables w, h in Equation 10 are the width and height of b, and the variables w_{gt} and h_{gt} are the width and height of b_{gt}. CIOU adds the similarity ratio between the length and width of the predicted frame and the real frame to the penalty term of loss. This parameter can enhance the model's responsiveness to individual dairy cow position information in intensive training conditions.

5. RESULTS AND DISCUSSION

5.1 Implementation details

Three 16 GB Tesla P100 GPUs were used in experiments on a deep learning framework with Python 3.8 and Pytorch 1.7.1. Using the same dairy cow behavior dataset, YOLOv3, YOLOv3-Tiny, Res2 YOLOv3 and GC_Res2 YOLOv3 models were trained. The Res2 YOLOv3 model is a YOLOv3 model perfected using the Res2 backbone, and the GC_Res2 YOLOv3 model is obtained by adding the GC Block module to the Res2 YOLOv3. Set the input image size to 640 × 640 pixels, learning rate to 0.01, batch-size to 16, and epoch to 200 to train YOLOv3, YOLOv3-Tiny, Res2 YOLOv3, and GC_Res2 YOLOv3.

5.2 Comparison of model experiments

In the experiment, five indicators of Model size, Precision, Recall, and mAP@.5 (Mean Average Precision), and mAP@.5:.95 were used to evaluate the performance of different models. The information provided by mAP, which integrates P and R indicators, was the most informative of these. When IoU is 0.5, the model marks the detection results as positive samples, as shown by the mAP@.5 symbol. The average AP (Average Precision) value achieved when IoU is 0.5, 0.55, 0.65, and 0.95 IoU, respectively, is known as the mAP@.5:.95. It is possible to use it to evaluate how closely the model detection target corresponds to the actual. Table 3 shows the outcomes of the experiment.

The GC_Res2 YOLOv3 and Res2 YOLOv3 models significantly outperform YOLOv3. The YOLOv3-Tiny model performs poorly, and the model size is the smallest, as seen in Table 3. In comparison to YOLOv3, the mAP@.5 indices of GC_Res2 YOLOv3 and Res2 YOLOv3 both increased by 1.2 and 0.9%, respectively. GC_Res2 YOLOv3 is 1.2% superior to YOLOv3 in terms of mAP.5:.95 value, which is 0.685. According to past studies, the GC_Res2 YOLOv3 model provides better results for identifying dairy cow behaviors. Regarding P and R metrics, both the GC_Res2 YOLOv3 and Res2 YOLOv3 models showed a drop in P but a noticeable increase in R. The Res2 YOLOv3 showed a 1.8% increase in the R, showing that the model is more concerned with the R and detects a greater number of correct dairy cow behaviors. Res2 Block's usage of multiple convolutions of size 3 × 3 to minimize model size is supported by the fact that the Res2 YOLOv3 model size is reduced by 12.26 MB when compared to the YOLOv3 model weights. The effectiveness of the modified technique is demonstrated by the data in Table 3.

The recognition results of a dairy cow's single behavior can demonstrate the model's capacity to identify and categorize dairy cow behavior. The APs of all models for identifying the four dairy cow behaviors are listed in Table 4. The results of the standing, walking, and mounting behavior recognition have been significantly enhanced by the GC_Res2 YOLOv3 and Res2 YOLOv3. The AP of the GC_Res2 YOLOv3 model is increased by 1.4%, 2.6%, and 1% for standing, walking, and mounting detection compared to the YOLOv3, demonstrating that the model is better able to identify the multi-scale behavior of dairy cows.

The mounting behavior's AP with the highest score is 99.3% for YOLOv3-Tiny. The YOLOv3-Tiny shallow network model only uses six convolutional kernels in its feature extraction network to extract dairy cow behaviors features. These behavior features are more detailed but lack abstract semantic features. Because the mounting target scale is larger than the other behaviors and the model gathers more data about the

Table 3. Comparison of experimental results of different models

Model	Precision	Recall	mAP@.5	mAP@.5:.95	Model size
YOLOv3	**0.913**	0.851	0.892	0.673	117.83 MB
YOLOv3-Tiny	0.799	0.669	0.775	0.421	17.40 MB
Res2 YOLOv3	0.909	**0.869**	0.901	0.677	**105.57 MB**
GC_Res2 YOLOv3	0.912	0.857	**0.904**	**0.685**	123.90 MB

Table 4. Comparison results on the behaviors of individual types of dairy cows

AP	YOLOv3	YOLOv3-Tiny	Res2 YOLOv3	GC_Res2 YOLOv3
Stand	0.892	0.857	0.904	**0.906**
Lie	**0.919**	0.866	0.917	0.917
Walk	0.781	0.754	0.803	**0.807**
Mount	0.975	**0.993**	0.979	0.985

shape and texture of the mounting behavior, *etc*. So, the AP of YOLOv3-Tiny for dairy cow mounting behavior recognition is significantly improved when compared to the other.

Other networks, such as YOLOv3, harvest features from a deeper backbone, resulting in the acquisition of semantic data but a dearth of detailed features. The model is, therefore, unable to incorporate both detailed and semantic characteristics. As a result, YOLOv3-Tiny was notably different from other models in its ability to recognize mounting behavior. The difference in AP for mounting behavior recognition between GC_Res2 YOLOv3 and YOLOv3-Tiny, however, is 0.8%, and the difference between Res2 YOLOv3 and YOLOv3-Tiny is smaller than YOLOv3.

The four models discussed above perform less well in detecting dairy cows' walking behavior, as seen in Table 4. Therefore, we select GC_Res2 YOLOv3 to predict dairy cow behaviors on the test dataset with the parameters set to an IoU threshold of 0.45 and a confidence level of 0.25 to evaluate why the four models in Table 4 have low AP for recognizing walking behavior. Based on the model recognition outcomes, the confusion matrix was created to examine the GC_Res2 YOLOv3 model for walking behavior, as shown in Figure 8.

The horizontal axis represents labeled dairy cow behavior, while the vertical axis represents model-predicted dairy cow behavior in Figure 8. Due to the similarities between walking and standing behaviors and the model's obfuscation of these differences, as seen in Figure 8, GC_Res2 YOLOv3 identifies some walking as standing behavior. The model, therefore, interprets a portion of the dairy cow's walking activity as standing behavior. GC_Res2 YOLOV3 outperformed the three network models YOLOv3, YOLOv3-Tiny, and Res2 YOLOV3 for walking recognition, with scores ranging from 0.781 to 0.807. According to the experimental findings, this model is better to detect dairy cow behaviors in crowded, multiscale situations.

5.3 Hybrid data augmentation comparison experiment

The enhanced and unenhanced dairy cow behavior dataset was examined using YOLOv3 and YOLOv3-Tiny, and the experimental findings are provided in Table 5 to show the impact of mixed data augmentation methods on the results of the experiment.

Table 5. Comparison of data augmentation experiments

Status	Model	Precision	Recall	mAP@.5	mAP@.5:.95
Unenhanced	YOLOv3	0.819	0.823	0.865	0.593
	YOLOv3-Tiny	0.67	0.714	0.73	0.355
Hybrid Data Augmentation	YOLOv3	**0.913**	**0.851**	**0.892**	**0.673**
	YOLOv3-Tiny	0.868	0.819	0.868	0.569

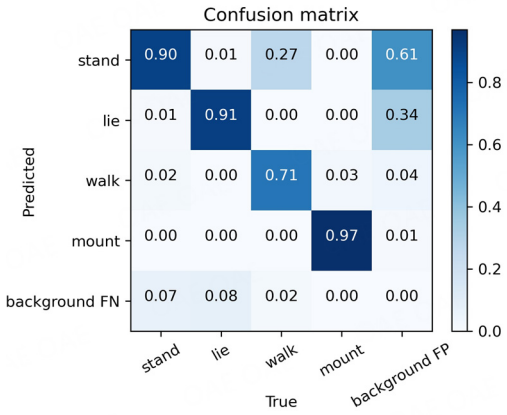


Figure 8. Confusion Matrix.

Table 5 shows an improvement in the mAP@.5:.95 measures of YOLOv3 and YOLOv3-Tiny of 8% and 21.4%, respectively, demonstrating the efficacy of the hybrid data augmentation technique.

5.4 Res2 block performance verification

To examine the effect of the Res2 Block on the model in the stages of feature extraction and feature processing, the Bottleneck Block of the backbone and Neck in Darknet53 were swapped out for the relevant Res2 Blocks. The experimental findings are displayed in Table 6.

The data comparison in Table 6 indicates that the mAP@.5 for Models 1 and 2 has increased by 1% and 0.7%, respectively, compared to YOLOv3. This result suggests that Res2 Block contributes to the feature extraction and feature processing stages. The Res2 Block, however, has a bigger disparity between the prediction frame produced by the model and the true frame at IoU > 0.5, as seen by a lower mAP@.5:.95 for Model 2 compared to YOLOv3, which is 1.7% lower. The Bottleneck Block was swapped out for the Res2 Block in the backbone and neck to create the Res2 YOLOv3 model. The mAP@.5 and mAP@.5:.95 of this model both increased, proving a greater contribution of the Res2 Block to the feature extraction and processing steps.

5.5 Loss for bounding box regression

The GC_Res2 YOLOv3 model was trained, and the localization loss values during training were recorded using a combination of GC Block and three loss functions, DIoU, CIoU, and GIoU, respectively, as shown in Figure 9. This was done to confirm the localization performance of GC Block and CIOU for individual dairy cows in dense situations.

Table 6. Res2 Block performance verification

Model	Backbone	Neck	mAP@.5	mAP@.5:.95
YOLOv3			0.892	0.673
Model 1	√		0.902	0.674
Model 2		√	0.899	0.656
Res2 YOLOv3	√	√	0.901	0.677

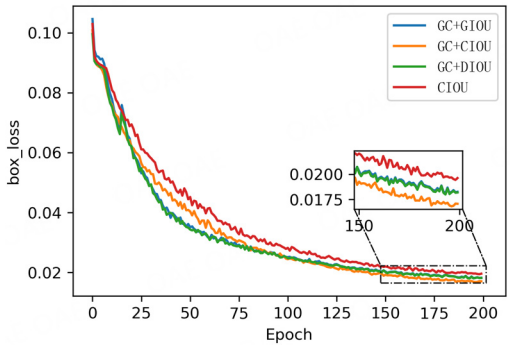


Figure 9. Loss for Bounding Box Regression under different IoU calculation methods.

In Figure 9, the horizontal point represents the number of times the model has been trained to 200, and the vertical coordinate represents the model's localization loss value. The Global Context (GC) module may significantly increase the localization accuracy of individual dairy cows, as can be seen from the two curves of CIoU and GC+CIoU. The model with the GC module has a smaller value of localization loss. The localization loss value decreases more quickly for the two localization loss curves, GC+GIoU and GC+DIoU, during the 13-85 stage. After 85 epochs, the localization loss for GC+CIoU continues to decline, and the localization loss reaches its lowest value after the model training is finished. Figure 9 illustrates that the CIoU and GC modules are superior at identifying specific dairy cows inside the model.

5.6 Discussion

To increase the dairy cow behavior recognition accuracy based on the dairy cow behavior dataset, this work created and improved GC_Res2 YOLOV3, although there are still certain drawbacks, which are shown in Figure 10.

The results of GC_Res2 YOLOV3 recognizing dairy cow behaviors on different scales are shown in Figure 10. The dairy cow's mounting behavior has been recognized by GC_Res2 YOLOV3 in Figure 10A and B. Dairy cow mounting is the cooperative behavior of two dairy cows. The model will recognize the behavior of the two dairy cows independently when the mounting features are not immediately apparent. As shown in Figure 10A, GC_Res2 YOLOV3 recognized dairy cow mounting behavior as standing and mounting behavior. For the behavior recognition result of a single dairy cow, the recognition result is correct, but we prefer the model to recognize it as a mounting behavior. The recognition of the little target dairy cows' mounting behavior is shown in Figure 10B. Only the dairy cow mounting label is visible in Figure 10B since the dairy cow's target is small and other behavior labels will cover it. The distant mounting behavior can be recognized by GC_Res2 YOLOV3, but the outcome is unstable because certain detection frames have lower scores. Low-scoring detection boxes would be filtered out when presented; therefore, this does not affect the display. When the features of the mounting behavior are not instantly obvious, GC_Res2 YOLOV3 will recognize the behavior of two dairy cows alone rather than recognizing it as a mounting

(A) GC_Res2 YOLOV3 recognizes the mounting behavior of big target dairy cows

(B) GC_Res2 YOLOV3 recognizes the mounting behavior of small target dairy cows

(C) GC_Res2 YOLOV3 detects dairy cow standing and walking behavior

(D) GC_Res2 YOLOV3 detects multi-scale dairy cow behaviors in multiple scenes

Figure 10. GC_Res2 YOLOV3 results for dairy cow behavior recognition.

behavior, according to the analysis of the behavior of the dairy cows in Figure 10A and B. Future efforts should be made to improve this path.

The outcomes of GC_Res2 YOLOV3 recognized dairy cows' walking and standing behaviors are displayed in Figure 10C. The walking behavior was recognized as standing behavior in the first image of Figure 10C. The standing behavior was recognized as walking in the second image of Figure 10C. GC_Res2 YOLOV3 successfully recognized the standing and walking behavior of dairy cows in the last image of Figure 10C. GC_Res2 YOLOV3 is easy to confuse dairy cows' walking and standing behavior. The result was the same as that shown in Figure 8. The walking dairy cow in Figure 10C is a small target, and the behavioral features that GC_Res2 YOLOV3 has acquired are biased more toward standing behavior, which causes mistakes in behavior recognition. If you look closely at Figure 10C, you will notice that the standing dairy cow's legs are slightly crossed. Based on the features of the leg crossing, GC_Res2 YOLOV3 recognized this as walking behavior. GC_Res2 YOLOV3 successfully recognized dairy cows' standing and walking behaviors in the last image of Figure 10C. The outcomes of recognizing the multi-scale behavior of dairy cows in various scenarios are shown in Figure 10D. GC_Res2 YOLOV3 could correctly recognize dairy cow behaviors.

In conclusion, GC_Res2 YOLOV3 can recognize multi-scale dairy cow behaviors in a variety of situations. Standing and walking of dairy cows have visually similar features, which will increase the difficulty of recognizing these two behaviors in multi-scale situations. Although GC_Res2 YOLOV3 makes a few mistakes when recognizing dairy cows' standing and walking behavior, it performs better than the unimproved model in recognizing these behaviors. Future studies will use the 3D model to recognize two behaviors and extract the spatial and temporal dimension features of dairy cow behaviors. The 3D model can better distinguish between dairy cows' standing and walking behaviors because, in the time dimension, dairy cows are in motion and have rich dynamic features, but the dynamic aspects of dairy cows' standing are less visible.

5.7 Video analysis of dairy cow behaviors based on GC_Res2 YOLOv3

To confirm GC_Res2 YOLOv3's generalization for dairy cow behavior identification, 37 segments of approximately 115 minutes of dairy cow behavior videos were acquired in various settings. The collected video also shows multi-scale, lighting, density fluctuations, and other properties. Figure 11 shows the GC_Res2 YOLOv3 model's recognition outcomes.

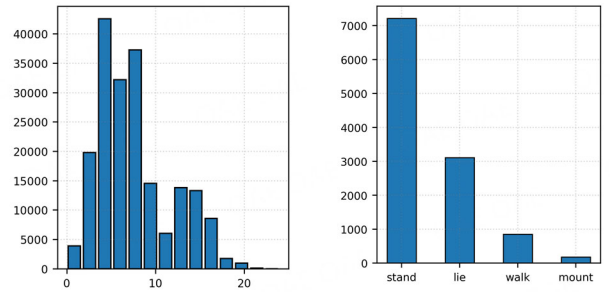
The scenes in Figure 11A-H are similar to those in the training set, and Figure 11I-X are the dairy cow behavior images in the new scene. It can be seen from Figure 10 that the recognition effect of Figure 11A-H is the best. It is possible to discriminate between the dairy cow behaviors and find the dairy cows with obvious targets in Figure 11I-X. The preceding recognition results show that GC_Res2 YOLOv3 can more precisely recognize the behaviors of dairy cows in many settings, proving the model's strong generalizability. Figure 12 depicts the statistical outcomes of the GC_Res2 YOLOv3 model used to analyze the behaviors of dairy cows in the preceding video.

Figure 12A depicts the distribution of the number of dairy cows in all video frames; the ordinate is the total number of video frames, and the abscissa is the number of dairy cows. Figure 12A demonstrates that there are typically five to ten dairy cows in each image, with more than ten being a rarity. The statistics of the GC_Res2 YOLOv3 model's behavior prediction findings for the dairy cows in the video are displayed in Figure 12B. The ordinate is the total number of dairy cows in the video, and the ordinate is the behavior category of the dairy cows. From Figure 12B, the number of standing, lying, walking, and mounting shows a downward trend. Dairy cow behaviors are strongly correlated with a particular period. Dairy cows exhibit more lying and standing but less walking movement during the day. It is another piece of evidence that the GC_Res2 YOLOv3 model can successfully predict the behaviors of dairy cows in the group breeding setting because the statistical results in Figure 12B are congruent with the scenario in the dairy cow farm.

6. CONCLUSIONS

This YOLO-based research addresses challenges in densely populated dairy farming, overcoming multi-scale detection and individual dairy cow localization difficulties. Through refinements in YOLOv3, the introduction of the Res2 backbone enhances the model's capability to capture nuanced behavioral features, while the integration of the GC Block strengthens localization. Our dairy cow behavior detection model achieves high accuracy, with rates of 90.6%, 91.7%, 80.7%, and 98.5% for standing, lying, walking, and mounting, respectively. The remarkable average accuracy of 90.4% outperforms YOLOv3, YOLO-tiny, and other models by 1.2% and 12.9%, highlighting the superior performance of our model. The recognition performance for the similar actions of walking and standing increased by 2.6% and 1.4%, respectively. GC_Res2 YOLOv3 establishes a feature extraction backbone with Res2 Block, integrating multiple receptive fields. This effectively addresses the challenge of low model recognition results caused by the multi-scale behaviors of dairy cows. Additionally, the model incorporates an improved detector to enhance its ability to identify specific dairy cows in diverse dense scenes. The identification and visualization results in a new scenario further demonstrate the model's strong generalizability.

Figure 11. Detection video samples.



(A) Distribution of the number of dairy cows in the video

(B) Distribution of dairy cow behaviors

Figure 12. Statistics of dairy cow behaviors in the videos.

DECLARATIONS

Authors' contributions

Made substantial contributions to the research methodology, conceptual, and simulation and wrote and edited the original draft: Bai Q

Performed critical review, comments, and revisions and provided administrative, technical, and material support: Gao R, Li Q, Wang R, Zhang H

Availability of data and materials

The data supporting the findings of this study are accessible from the corresponding author, but access is subject to restrictions due to the data being utilized under license for the current study. Consequently, these data are not publicly available. For inquiries regarding the datasets used and analyzed during the current study, please contact the corresponding author, Ronghua Gao, who will consider reasonable requests.

Financial support and sponsorship

This research is financially supported by Technological Innovation Capacity Construction of Beijing Academy of Agricultural and Forestry Sciences (KJCX20230204).

Conflict of Interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

The animal use protocol listed below has been reviewed and approved by the Ethics Committee, Northwest A&F University, China (IFA-2022001).

Consent for publication

Not applicable.

REFERENCES

1. Hu H, Dai B, Shen W, et al. Cow identification based on fusion of deep parts features. *Biosystems Engineering* 2020;192:245-56. DOI
2. Kumar S, Pandey A, Sai Ram Satwik K, et al. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* 2018;116:1-17. DOI
3. Uenishi S, Oishi K, Kojima T, et al. A novel accelerometry approach combining information on classified behaviors and quantified physical activity for assessing health status of cattle: a preliminary study. *Applied Animal Behaviour Science* 2021;235:105220. DOI
4. Niloofar P, Francis DP, Lazarova-molnar S, et al. Data-driven decision support in livestock farming for improved animal health, welfare and greenhouse gas emissions: overview and challenges. *Computers and Electronics in Agriculture* 2021;190:106406. DOI
5. Shen W, Cheng F, Zhang Y, Wei X, Fu Q, Zhang Y. Automatic recognition of ingestive-related behaviors of dairy cows based on triaxial acceleration. *Information Processing in Agriculture* 2020;7:427-43. DOI
6. Benaissa S, Tuyttens FA, Plets D, et al. Classification of ingestive-related cow behaviors using RumiWatch halter and neck-mounted accelerometers. *Applied Animal Behaviour Science* 2019;211:9-16. DOI
7. Barker ZE, Vázquez Diosdado JA, Codling EA, et al. Use of novel sensors combining local positioning and acceleration to measure feeding behavior differences associated with lameness in dairy cattle. *J Dairy Sci* 2018;101:6310-21. DOI
8. Arablouei R, Currie L, Kusy B, Ingham A, Greenwood PL, Bishop-hurley G. In-situ classification of cattle behavior using accelerometry data. *Computers and Electronics in Agriculture* 2021;183:106045. DOI
9. Benaissa S, Tuyttens F, Plets D, et al. Calving and estrus detection in dairy cattle using a combination of indoor localization and accelerometer sensors. *Computers and Electronics in Agriculture* 2020;168:105153. DOI
10. Ren XH, Liu G, Zhang M, Si YS, Zhang XY, MA L. Dairy cattle's behaviour recognition method based on support vector machine classification model. *Transactions of the Chinese Society for Agricultural* 2019;50:290-6. DOI
11. Wang K, Liu CH, Duan QL. Identification of sow oestrus behaviour based on MFO-LSTM. *Transactions of the Chinese Society of Agricultural Engineering* 2020;36:211-9. DOI
12. Khin MP, Zin TT, Mar CC, Tin P, Horii Y. Cattle pose classification system using deeplabcut and svm model. 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE); 2022 Oct 494-5; Osaka, Japan.
13. Minar MR, Naher J. Recent advances in deep learning: an overview. Available from: https://arxiv.org/abs/1807.08169 [Last accessed on 25 Jan 2024].
14. Jiang B, Chen S, Wang B, Luo B. MGLNN: semi-supervised learning via multiple graph cooperative learning neural networks. *Neural Netw* 2022;153:204-14. DOI
15. Zhang HM, Fu ZY, Han WT, Yang G, Niu DD, Zhou XY. Detection method of maize seedlings number based on improved YOLO. *Transactions of the Chinese Society for Agricultural* 2021;52:221-9. DOI
16. Roy AM, Bhaduri J. DenseSPH-YOLOv5: An automated damage detection model based on densenet and swin-transformer prediction

head-enabled YOLOv5 with attention mechanism. *Advanced Engineering Informatics* 2023;56:102007. DOI
17. Hu Z, Yang H, Lou T. Dual attention-guided feature pyramid network for instance segmentation of group pigs. *Computers and Electronics in Agriculture* 2021;186:106140. DOI
18. Xiao J, Liu G, Wang K, Si Y. Cow identification in free-stall barns based on an improved mask R-CNN and an SVM. *Computers and Electronics in Agriculture* 2022;194:106738. DOI
19. Bonneau M, Vayssade J, Troupe W, Arquet R. Outdoor animal tracking combining neural network and time-lapse cameras. *Computers and Electronics in Agriculture* 2020;168:105150. DOI
20. Su Q, Tang J, Zhai M, He D. An intelligent method for dairy goat tracking based on Siamese network. *Computers and Electronics in Agriculture* 2022;193:106636. DOI
21. Chen C, Zhu W, Steibel J, et al. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. *Computers and Electronics in Agriculture* 2020;169:105166. DOI
22. Chen C, Zhu W, Steibel J, Siegford J, Han J, Norton T. Classification of drinking and drinker-playing in pigs by a video-based deep learning method. *Biosystems Engineering* 2020;196:1-14. DOI
23. Chen C, Zhu W, Steibel J, Siegford J, Han J, Norton T. Recognition of feeding behaviour of pigs and determination of feeding time of each pig by a video-based deep learning method. *Computers and Electronics in Agriculture* 2020;176:105642. DOI
24. Yan HW, Liu ZY, Fui QL, Hu ZW. Multi-target detection based on feature pyramid attention and deep convolution network for pigs. *Transactions of the Chinese Society of Agricultural Engineering* 2020;36:193-202. DOI
25. Li X, Cai C, Zhang R, Ju L, He J. Deep cascaded convolutional models for cattle pose estimation. *Computers and Electronics in Agriculture* 2019;164:104885. DOI
26. Wei S, Ramakrishna V, Kanade T, Sheikh Y. Convolutional pose machines. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 4724-32.
27. Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer Vision – ECCV 2016. Cham: Springer International Publishing; 2016. pp. 483-99.
28. Bulat A, Tzimiropoulos G. Human pose estimation via convolutional part heatmap regression. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer Vision – ECCV 2016. Cham: Springer International Publishing; 2016. pp. 717-32.
29. Wang SH, He DJ, Liu D. Automatic recognition method of dairy cow estrus behaviour based on machine vision. *Transactions of the Chinese Society for Agricultural* 2020;51:241-9. DOI
30. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90. DOI
31. Yin X, Wu D, Shang Y, Jiang B, Song H. Using an efficientnet-LSTM for the recognition of single cow's motion behaviours in a complicated environment. *Computers and Electronics in Agriculture* 2020;177:105707. DOI
32. Benaissa S, Tuyttens FAM, Plets D, et al. On the use of on-cow accelerometers for the classification of behaviours in dairy barns. *Res Vet Sci* 2019;125:425-33. DOI
33. Wang SH, He DJ. Estrus behavior recognition of dairy cows based on improved YOLO v3 model. *Transactions of the Chinese Society for Agricultural Machinery* 2021;52:141-50. DOI
34. Zhang HM, Li YH, Zhou LX, Wang R, Li SQ, Wang HY. Multi-target skeleton extraction method of beef cattle based on improved YOLO v3. *Transactions of the Chinese Society for Agricultural Machinery* 2022;53:285-93. DOI
35. Ma S, Zhang Q, Li T, Song H. Basic motion behavior recognition of single dairy cow based on improved Rexnet 3D network. *Computers and Electronics in Agriculture* 2022;194:106772. DOI
36. Ji JT, Liu QH, Gao RH, Li QF, Zhao KX, Bai Q. Ruminant behavior analysis method of dairy cows with improved flownet 2.0 optical flow algorithm. *Transactions of the Chinese Society for Agricultural Machinery* 2023;54:235-42. DOI
37. Guo JJ, He GH, Xu LQ, Liu TL, Feng DC, Liu SY. Pigeon behavior detection model based on improved YOLO v4. *Transactions of the Chinese Society for Agricultural Machinery* 2023;54:347-55. DOI
38. Hu T, Yan R, Jiang C, et al. Grazing sheep behaviour recognition based on improved YOLOV5. *Sensors* 2023;23:4752. DOI PubMed PMC
39. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: optimal speed and accuracy of object detection. *ArXiv* 2020; preprint arXiv:10934. DOI
40. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.p.779-788.
41. Gao SH, Cheng MM, Zhao K, Zhang XY, et al. Res2Net: a new multi-scale backbone architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence; 2021.p.652-662.
42. Cao Y, Xu J, Lin S, et al. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. Proceedings of the IEEE/CVF international conference on computer vision workshops; 2019.