PIONEER
ACADEMIC PUBLISHING LIMITED

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 2 April 2025

A Study on Multi-Target Dairy Cow Feeding Behavior Recognition Based on Improved YOLOv7

Ruilong Kui[2], Weiping Luo[1,2] & Yapeng Zhang[1,2]

[1] Hubei Key Laboratory of Digital Textile Equipment, Wuhan Textile University, Wuhan 430200, China
[2] School of Mechanical Engineering and Automation, Wuhan Textile University, Wuhan 430200, China
Correspondence: Weiping Luo, Hubei Key Laboratory of Digital Textile Equipment, Wuhan Textile University, Wuhan 430200, China; School of Mechanical Engineering and Automation, Wuhan Textile University, Wuhan 430200, China.

doi:10.56397/JPEPS.2025.04.05

Abstract

To make the research on multi-target dairy cow feeding behavior recognition in pastures more lightweight and improve the detection accuracy and inference speed of the model, this paper proposes a lightweight and improved algorithm YOLOv7-CDD based on the YOLOv7 object detection model. Firstly, the algorithm adds the CA attention mechanism module to the last layer of all backbone extraction networks to replace the original output layer, resulting in better detection performance and higher accuracy without the need for manual threshold adjustment. Secondly, DSConv is introduced to replace some conventional convolutions (3×3 convolutions) in the back-bone network and in the multi-branch stacking module (Multi_Concat_Block), further reducing the number of model parameters without compromising detection accuracy. Finally, the dynamic detection head Dynamic Head is added, enhancing the expression capability of the target detection head and further improving detection accuracy without increasing computational complexity. Experimental results show that the YOLOv7-CDD model achieves an accuracy of 98.4%, a recall rate of 98.3%, and an mAP@0.5 of 99.3%, representing improvements of 2.8%, 2.6%, and 3.1%, respectively, compared to the YOLOv7 model, while significantly reducing model parameters and GFLOPs, demonstrating that YOLOv7-CDD meets the application requirements in pastures.

Keywords: multi-target, YOLOv7, lightweight, attention mechanism, cow feeding behavior

1. Introduction

The eating behavior of dairy cows is of great research significance in the process of dairy farming. By analyzing the eating behavior of dairy cows, we can understand the preference of dairy cows for different feeds, feed consumption and other information, so as to better meet the nutritional needs of dairy cows and improve production efficiency (Eleonora F, Alberto R, Mirco C. et al., 2023; Zou J. & Arshad RM., 2024). Abnormal changes in the eating behavior of dairy cows may be a precursor to certain health problems, such as anorexia and digestive system

diseases (Li Z, Zhu Y, Sui S. et al., 2024). By monitoring the eating behavior of dairy cows, these problems can be discovered and treated in time, reducing the incidence and mortality of diseases (Xing Yongxin, Sun Youdong & Wang Tianyi, 2022). By observing the eating behavior of dairy cows, we can understand the feed intake, feeding speed, preferences and other information of dairy cows, so as to optimize feeding management and feed rationing and improve the production performance and health of dairy cows (Song Huaibo, Li Rong, Wang Yunfei et al., 2023; Bai Qiang, Gao Ronghua, Zhao Chunjiang et al., 2022; Wang Zheng, Xu Xingshi, Hua Zhixin et al., 2022). By monitoring and analyzing the eating behavior of dairy cows, a large amount of production data can be obtained, and prediction models and decision support systems can be established based on these data, providing scientific basis for farm managers to help make more accurate decisions, such as feed rationing, disease prevention and control, and improve the efficiency and sustainability of agricultural production (Qin Lifeng, Zhang Xiaoqian, Dong Mingxing et al., 2021). Therefore, it is very necessary to develop a method for monitoring the eating behavior of multi-target cows based on machine vision. Liu Yuefeng et al. proposed a better sparse subnetwork screening method based on the YOLOv3 amplitude iteration pruning algorithm to realize the detection of cow eating behavior. This method illustrates the feasibility of reducing the cost of cow behavior monitoring tasks through amplitude iteration pruning technology, verifies the effectiveness of screening better sparse subnetworks from the cow eating behavior recognition model based on the lottery hypothesis, and provides a reference for reducing the cost of animal behavior monitoring tasks. However, this model focuses more on the lightweight of the model and does not pay much attention to improving the accuracy of the model. Song Lvming et al. proposed a new method for detecting small samples of glass surface defects by adding a convolutional attention mechanism module and a pre-detection head to YOLOv7, and used image enhancement methods such as random Gaussian noise, Mix-up, random filling images and random splicing to expand the samples and balance the samples. The improved model improves the efficiency of detecting glass surface defects in engineering to a certain extent, but this model does not pay attention to the problem of model lightweight. Zhang Zhen et al. proposed a lightweight apple detection model based on YOLOv7, adding partial convolution (PConv) and efficient channel attention (ECA) modules to the model, and using the Sparrow search algorithm (SSA) during model training to further improve the detection accuracy of the model. This laid the foundation for unmanned intelligent apple picking. Deng Changzheng et al. proposed an infrared image recognition algorithm for substation equipment based on YOLOv7-Tiny, introducing a lightweight bottleneck structure GhostNetV2BottleNeck to replace part of the CBS module, and embedding the CA attention mechanism in the feature extraction stage, replacing the network coordinate loss function with SIoU Loss. This laid the foundation for subsequent substation fault diagnosis. The above algorithm considers changing the backbone convolution layer, backbone extraction network and model pruning, so some operations are complex and some workloads are large. It is possible to consider using a more simplified and efficient optimization technology to lightweight the model.

In order to solve the problem of complex operation and large processing volume of the above model, this paper applies neural network to the recognition of dairy cow eating behavior, and uses image recognition technology based on YOLOv7 model to integrate CA attention mechanism, DSConv convolution and Dynamic Head dynamic detection head to propose a more lightweight dairy cow eating behavior recognition detection model YOLOv7-CDD. This model uses the pasture dairy cow eating data set to train and test the model, and compares it with other commonly used detection models to achieve the compression of the dairy cow eating behavior recognition model, while ensuring that the model performance is not affected or even better. Finally, the model is trained and tested, and the detection effect is compared with other detection models, hoping to provide a reliable new idea for the research on pasture dairy cow eating behavior recognition.

2. Materials and Methods

2.1 Construction of YOLOv7-CDD Model

2.1.1 CA Attention Mechanism Module

When existing attention mechanisms (such as CBAM, SE, etc.) obtain channel attention, they generally use global maximum pooling or

average pooling to process channels. Although this can maintain the most important features of the input feature map and reduce the risk of overfitting, it will also lose the spatial position information of the object. The CA attention mechanism (Li Yuwei, Fu Rui & Liu Fan, 2024) embeds the position information into the channel information. The implementation of the CA attention mechanism is mainly divided into two parallel stages: global average pooling of the input information with a width of w and a number of channels of c and a height of h and a number of channels of c respectively to obtain two feature layers, namely, feature maps in the wide dimension and feature maps in the high dimension, as shown in equations (1) and (2).

$$z_c^h(h) = \frac{1}{W}\sum_{0\le i<W} x_c(h,i) \quad (1)$$

$$z_c^w(w) = \frac{1}{H}\sum_{0\le j<H} x_c(j,w) \quad (2)$$

Then the two parallel stages are merged to transpose the width and height to the same dimension and stacked, the features of width and height are merged together, and the convolution normalization activation function is used for feature extraction, see formula (3).

$$\mathrm{f} = \delta\left(\mathrm{F}_1\left(\left[\mathrm{z}^{\mathrm{h}}, \mathrm{z}^{\mathrm{w}}\right]\right)\right) \quad (3)$$

Where f is the intermediate feature map, which is used to store spatial information in the horizontal and vertical directions, $\mathrm{f}\in\mathrm{R}^{\mathrm{C/r}\times(\mathrm{H}+\mathrm{W})\times 1}$ and, δis a nonlinear activation function. Along the spatial dimension, f is cut into height and width, $f^h \in R^{C/r\times H\times 1}$ and then the number of channels is adjusted $f^w \in R^{C/r\times 1\times W}$ to be consistent with the number of channels in the input feature map using 1×1 convolution, and the sigmoid function is used to obtain the final attention weights g^h and g^w, see equations (4) and (5).

$$g^h = \sigma\left(F_h(f^h)\right) \quad (4)$$

$$g^w = \sigma\left(F_w(f^w)\right) \quad (5)$$

Where F_h and F_w are 1×1 convolutions, is the sigmoid activation function, g^h and g^w are the adjusted attention weights. Finally, multiplying the weight by the input feature map can obtain the re-weighted feature map. The output formula of Coordinate Attention is shown in formula (6).

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

Where $y_c(i,j)$ is the output feature map, $x_c(i,j)$ is the input feature map, $g_c^w(j)$ and $g_c^h(i)$ is the attention weights in the horizontal and vertical directions. The CA attention mechanism usually does not need to perform global calculations on all positions, but dynamically adjusts the attention weights based on the relevance of the input data. Therefore, the introduction of the CA attention mechanism can reduce the amount of calculation and improve the efficiency of model detection. The structural flow of the CA attention mechanism is shown in Figure 1.

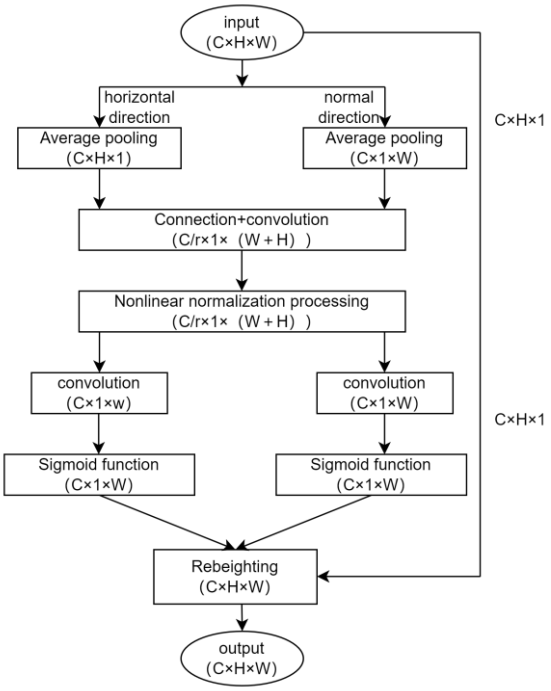


Figure 1. CA attention mechanism structure flow chart

Note: C, H, W are the number of channels, width, and height of the input feature map, and r is the reduction factor.

2.1.2 DSConv

Distribution Shifting Convolution (Jia Xueying, Zhao Chunjiang, Zhou Juan et al., 2023) (DSConv) is a variant of depthwise separable convolution and has been widely used in the field of computer vision. Its working principle is shown in Figure 2. Depthwise separable convolution can be divided into two steps. The first step is channel-by-channel convolution and the second step is point-by-point convolution. Ordinary convolution requires convolution on each channel, while depthwise convolution only performs convolution on a single channel and applies an

independent convolution kernel to each channel. Point-by-point convolution is a 1×1 convolution. Like regular convolution operations, it applies a convolution kernel on all channels to fuse the results of depthwise convolution. The advantage of DSConv over traditional depthwise separable convolution is that it uses learnable convolution kernels to improve model performance (Xu Hongwei, Li Ran & Zhang Jiaxu, 2024). DSConv decomposes the traditional convolution kernel into two components: variable quantization kernel (VQK) and distribution shift. It achieves lower memory usage and higher speed by storing only the integer value in VQK, while maintaining the same output as the original convolution by applying kernel-based and channel-based distribution shifts (Niu Weihua & Wei Yali, 2024). Therefore, DSConv is introduced into this model to make the model faster and occupy less memory, thus realizing a lightweight structure of the model.

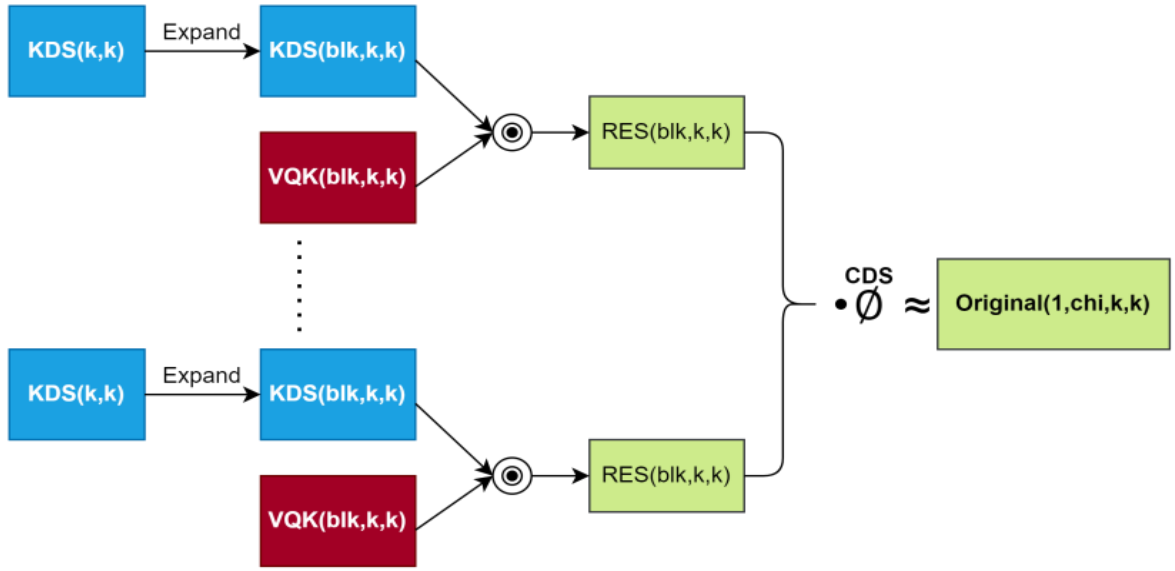


Figure 2. DSConv working principle diagram

Note: In the figure, ◎ represents the Hadamard operator (or element operator).

In the figure above, the size of the original convolution tensor is recorded as (c h_o, c h_i, k, k), where c is the number of channels in the lower layer, c is the number of channels in the current layer, and k is the height and width of the kernel.

In this model, DSConv is introduced to replace the 3×3 conventional convolution in the multi-branch stacking module Multi_Concat_Block in Backbone and Neck for lightweight improvement, so as to design a lighter and faster network model. The replaced D-Multi_Concat_Block module is shown in Figure 3.

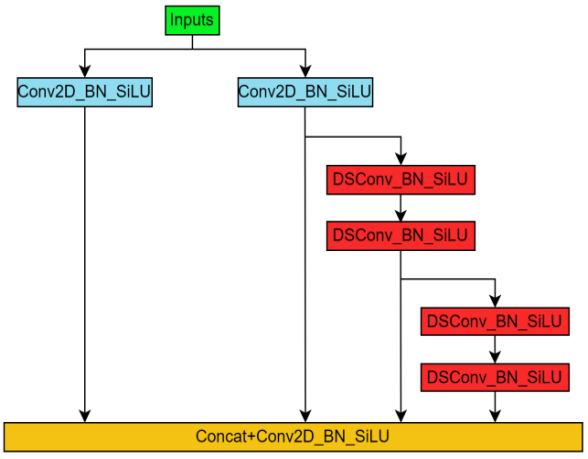


Figure 3. Improved D-Multi_Concat_Block

2.1.3 Dynamic Head

The traditional detection head attention can only solve one of the problems in scale perception, space perception and task perception. For a given feature tensor, its generalized attention can be expressed as formula (7):

$$W(F) = \pi(F) \cdot F \quad (7)$$

Where $\pi(\cdot)$ is the attention function, which is implemented by the fully connected layer, but this method has a large amount of calculation and is time-consuming and laborious in practical application. The solution of the dynamic detection head Dynamic Head is to convert the above attention into three sequences, each of which focuses on only one dimension. The calculation formula can be shown as formula (8):

$$W(F) = \pi_C(\pi_S(\pi_L(F) \cdot F) \cdot F) \cdot F \quad (8)$$

Among them $\pi_C(\cdot)$, $\pi_S(\cdot)$, $\pi_L(\cdot)$ represent the attention in the three dimensions of C, S, and L respectively.

The dynamic detection head combines scale awareness, spatial awareness and task awareness by coherently combining multiple self-attention mechanisms between feature levels of scale awareness, spatial locations of spatial awareness, and within the output channels of task awareness, significantly This significantly improves the representation ability of target detection heads. In the Dynamic Head framework, the output of Backbone is regarded as a three-dimensional tensor: level × space × channel, where level is the feature level, space is the width and height product of the feature layer, and channel is the number of channels. Dynamic Head deploys attention mechanisms separately in specific dimensions. That is, the scale-aware attention module scale-aware attention (level-wise) is deployed on the feature level. Feature maps at different levels correspond to different target scales. Adding attention at the feature level can enhance the scale-aware ability of target detection (Xu Ming, Qu Taipeng & Jiang Yanji, 2024), its calculation formula is as formula (9);

$$\pi_L(F) \cdot F = \sigma(f(\frac{1}{SC}\sum_{S,C} F)) \cdot F \quad (9)$$

Where is a linear function, which is approximated by 1 f(·)× 1 convolution $\sigma(x) = \max(0, \min(1, \frac{x+1}{2}))$ in Dynamic Head and is a hard-sigmoid function.

Deploy the spatial-aware attention module spatial-aware attention (spatial-wise) on the spatial dimension space. Different spatial positions correspond to the geometric transformation of the target. Increasing attention on the spatial dimension can enhance the spatial position perception ability of the target detector (Qu Chenyang & Cheng Yanyun, 2024). Its calculation formula is as shown in formula (10);

$$\pi_S(F) \cdot F = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K} w_{l,k} \cdot F(l; p_k + \Delta p_k; C) \cdot \Delta m_k \quad (10)$$

Where k is the number of sparsely sampled locations, is $p_k + \Delta p_k$ the position moved by the Δm_k self-learned spatial offset to focus on a discriminative region, Δp_k and is the importance measure of the self-learned location p_k, which are all learned from the input features at the median level of F.

The task-aware attention module (channel-wise) is deployed on the channel. Different channels correspond to different tasks. Adding attention to the channel can enhance the object detection's perception of different tasks (Cui Liqun & Cao Huawei, 2024). Its calculation formula is shown in formula (11).

$$\pi_C(F) \cdot F = max(\alpha^1(F) \cdot F_c + \beta^1(F), \alpha^2(F) \cdot F_c + \beta^2(F)) \quad (11)$$

Where F_c is the feature slice of the C channel, $[\alpha^1, \beta^1, \alpha^2, \beta^2]^T = \theta(\cdot)$ is the hyperfunction for learning to control the activation threshold, and $\theta(\cdot)$ is used similarly to Dynamic relu. First, global pooling is performed on the L×S dimension, then two fully connected layers and a normalization layer are used, and finally the output is normalized to using the shifted sigmoid function $[-1,1]$.

Since the above three attention mechanisms are random, we can use formula (8) to serialize and stack them multiple times. The network details of its working principle are shown in Figure 4.

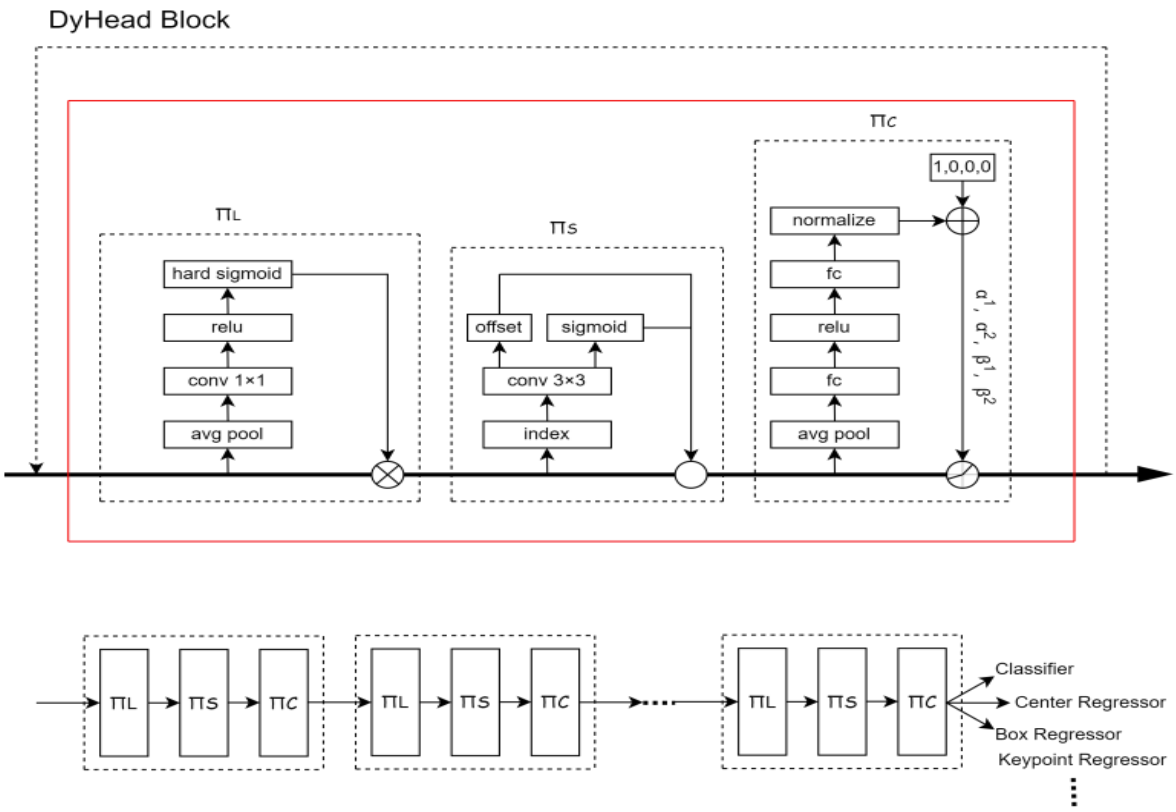


Figure 4. Dynamic Head detailed design network details

2.1.4 YOLOv7-CDD Model

This model adds the CA attention mechanism module to the last layer of all backbone extraction networks in Backbone to replace the original output layer. The detection effect after replacement is better than directly replacing the C3 module of all backbone extraction networks in Backbone with the CA module. It is more accurate and simplified, and does not require manual adjustment of thresholds. And DSConv was introduced to replace some of the conventional convolutions (3×3 convolution) in the Backbone Network and some of the conventional convolutions (3×3 convolution) in the multi-branch stacking module Multi_Concat_Block, without reducing the model detection accuracy. Under the premise, the number of parameters of the model can be further reduced. Finally, adding the dynamic detection head Dynamic Head can significantly improve the expression ability of the target detection head without increasing the amount of calculation. The structure of the improved algorithm YOLOv7-CDD network model is shown in Figure 5.

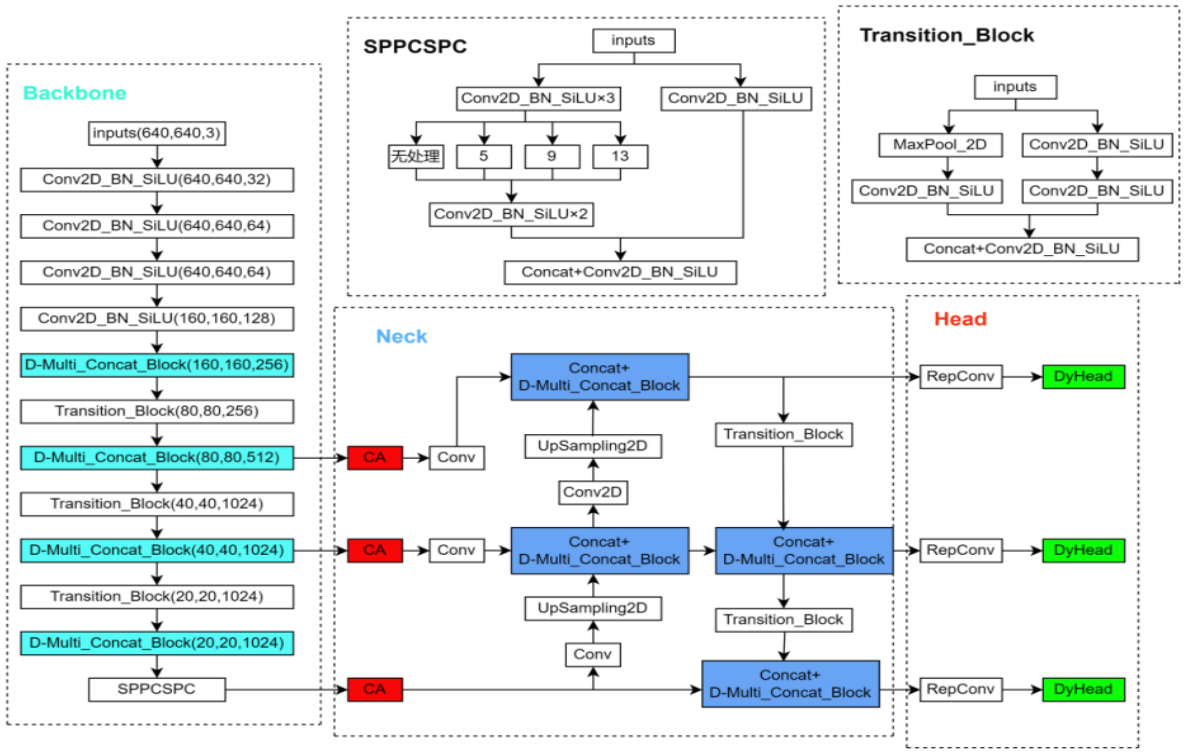


Figure 5. YOLOv7-CDD network model structure

2.2 Experimental Data and Methods

2.2.1 Experimental Environment and Parameter Settings

Experimental environment: The system used in this experiment is Windows 10, Intel(R) Core (TM) i7-9700k-3.6GHz (CPU), NVIDIA-GeForce-RTX-2080Ti-11G*2 (GPU), and 48GB (RAM). The GPU is used to accelerate model training. The software used is PyCharm 2022, CUDA 11.6, Python version 3.8, and the framework is PyTorch version 1.12.0.

Parameter setting: the number of model training iterations is 150, and the batch size is 32.

2.2.2 Dataset

The dataset used in this study is a field scene shot of a ranch in Inner Mongolia, which contains a total of 7166 1280×720 pixel pictures of dairy cows eating and not eating. In the experiment, the ratio of training set, test set and validation set are set to 6:3:1, that is, there are 4300 training set pictures, 2150 test set pictures, and 716 validation set pictures. The data sample is shown in Figure 6.

(a) (b) (c) (d)

Figure 6. Dairy Cow Feeding Data Set

Note: (a) Eating during the day; (b) not eating during the day; (c) eating at night; (d) not eating at night.

The data annotation software Labelimg was used to annotate the cows in the data, and in order to identify the cows' eating behavior, the cows were divided into cow eating (eating cows) and cows (not eating cows). When cows eat, they stick their heads out of the fence to eat. In Figure 5(b), there are two states of cows sticking their heads out of the fence. The left one is eating, and the right one has the desire to eat. For the convenience of the experiment, the cows sticking their heads out of the fence are marked as eating (cow eating). In Figure 5(a), there are three different states of cows. For the convenience of the experiment, they are all marked as not eating cows.

2.2.3 Evaluation Metrics

This experiment uses precision P, recall R, and mean average precision mAP to evaluate the detection efficiency and performance of the YOLOv7-CDD model. The calculation formula is as follows:

$$P = \frac{TP}{TP+FP} \tag{12}$$

$$R = \frac{TP}{TP+FN} \tag{13}$$

$$AP = \int_0^1 P(R)dR \tag{14}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{15}$$

Where TP is the number of cows that are correctly predicted to be eating; FP is the number of cows that are incorrectly predicted to be eating; FN is the number of cows that are not predicted to be eating; AP is the average precision of the experiment; N is the number of sample categories.

3. Experimental Results Analysis

3.1 Comparative Experiment on Performance of Different Attention Mechanisms

In order to verify the effectiveness of introducing CA attention mechanism in YOLOv7 in testing the feeding behavior dataset of dairy cows in the pasture, SE attention mechanism, EAM attention mechanism and ECA attention mechanism were used to improve the YOLOv7 model, and the same dataset was tested and compared under the same experimental environment. The comparison results are shown in Table 1.

Table 1. Comparative test of attention mechanisms

Model	P%	R%	mAP%	Parameter quantity	FLOPS/G
YOLOv7	95.6	95.7	96.2	3.721×10^7	1.051×10^{11}
YOLOv7-SE	96.3	96.6	97.1	3.774×10^7	1.023×10^{11}
YOLOv7-EAM	96.9	97.5	97.8	3.842×10^7	1.102×10^{11}
YOLOv7-ECA	96.8	96.7	97.6	3.741×10^7	1.044×10^{11}
YOLOv7-CA	98.3	98.0	98.9	3.715×10^7	1.055×10^{11}

In this m Analysis and comparison results show that after embedding CA attention, the accuracy of the model is 98.3%, which is 2.7%, 2.0%, 1.4%, and 1.5% higher than other models respectively; the recall rate is 98.0%, which is higher than other models. The model improved by 2.3%, 1.4%, 0.5%, and 1.3% respectively; the average accuracy-to-mean ratio was 98.9%, which was improved by 2.7%, 1.8%, 1.1%, and 1.3% respectively compared with embedding other models; and embedding CA attention. After force, the GFLOPs are larger and the detection speed is

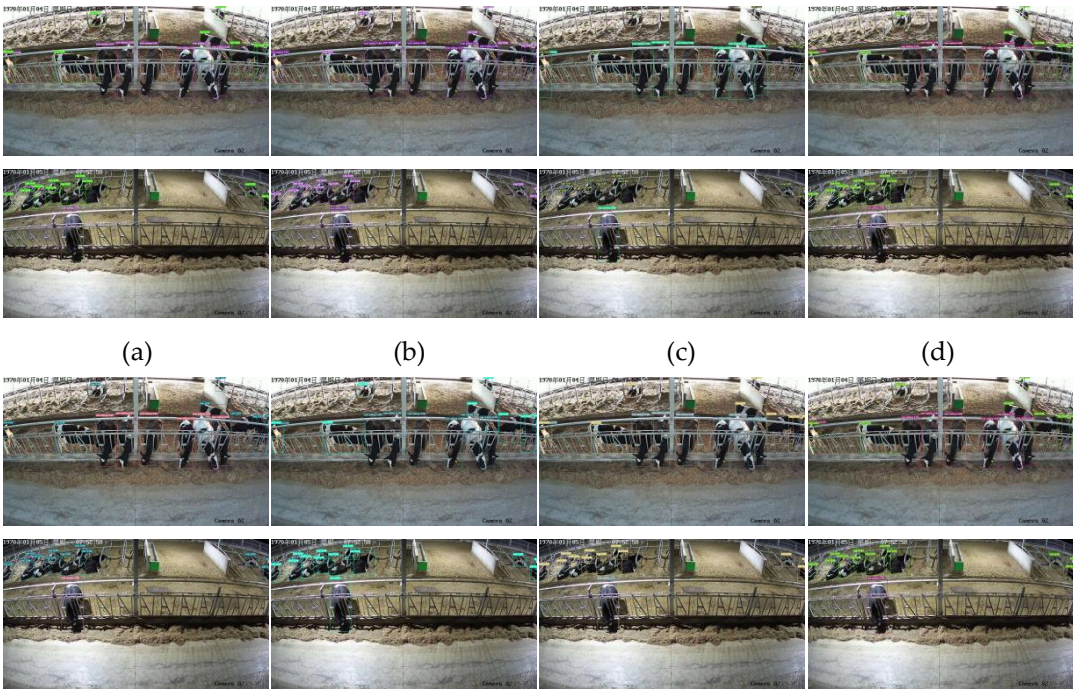
faster. Therefore, the effect of YOLOv7-CA on the model is more significant.

3.2 Ablation Experiment

In order to more comprehensively verify the effectiveness of the three improved methods proposed in this paper, a series of ablation experiments are carried out in the same experimental environment and experimental equipment. The experimental data are shown in Table 2, and the experimental results are shown in Figure 7.

Table 2. Ablation experiments

Model	P(%)	R(%)	MAP@0.5(%)	Parameter quantity/M	FLOPS/G
YOLOv7	95.6	95.7	96.2	3.721×10^7	1.051×10^{11}
YOLOv7-CA	98.3	98.1	98.9	3.715×10^7	1.055×10^{11}
YOLOv7-DSConv	98.2	98.1	98.8	3.259×10^7	6.59×10^{10}
YOLOv7-DyHead	98.2	97.7	98.8	3.642×10^7	1.024×10^{11}
YOLOv7-CA+DSConv	98.1	97.9	98.7	3.274×10^7	6.62×10^{10}
YOLOv7-CA+DyHead	98.1	98.2	98.8	3.657×10^7	1.027×10^{11}
YOLOv7-DSConv+DyHead	97.8	98.1	98.9	3.642×10^7	6.84×10^{10}
YOLOv7-CDD	98.4	98.3	99.3	3.657×10^7	6.88×10^{10}



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 7. Comparison of detection effects of different algorithms in the same scene

Note: (a) YOLOv7; (b) YOLOv7-CDD; (c) YOLOv7-CA; (d) YOLOv7-DSConv; (e) YOLOv7-DyHead; (f) YOLOv7-CA+DSConv; (g) YOLOv7-CA+DyHead; (h) YOLOv7-DSConv+DyHead.

It can be seen from the above table that the accuracy rate of the original YOLOv7 model is lower than that of other models. After adding the CA attention mechanism based on the YOLOv7 model, Map@0.5 increased by 2.7 percentage points, which improved the accuracy of the model and slightly reduced the calculation amount, making the model partially lightweight. After adding DSConv distributed offset convolution, mAP@0.5 increased by 2.6 percentage points and the number of parameters dropped a lot. GFLOPs were significantly

reduced and the calculation amount was slightly reduced. This shows that adding the module can reduce the calculation amount of the model. After adding the Dynamic Head dynamic detection head, the attention mode is systematically considered in the head design to obtain better performance, and mAP@0.5 is increased by 2.6 percentage points. The improved YOLOv7-CDD algorithm increased mAP@0.5 by 3.1 percentage points, and the number of parameters and the number of floating-point operations were significantly reduced, indicating that this algorithm model takes up less memory resources and has better detection performance.

3.3 Comparative Experiment

In order to further verify the objectivity and effectiveness of the improved YOLOv7-CDD network model, different models were compared on the same data set under the same experimental conditions. Under the same configuration environment and training parameters, the comparative experimental results of the improved YOLOv7-CDD network model in this paper and other network models are shown in Table 3.

Table 3. Comparative experiments

Model	P(%)	R(%)	MAP@0.5(%)	Parameter quantity/M	FLOPS/G
YOLOv4	94.3	94.8	95.1	6.313×10^7	6.94×10^{10}
YOLOv5	94.7	94.8	95.4	7.174×10^6	5.99×10^{10}
YOLOv7	95.6	95.7	96.2	3.721×10^7	1.051×10^{11}
YOLOv8	97.9	98.3	99.1	3.006×10^7	8.9×10^{10}
YOLOv7-CDD	98.4	98.3	99.3	3.657×10^7	6.88×10^{10}

From the analysis of Table 3, we can see that the YOLOv7-CDD algorithm proposed in this paper has improved all indicators compared with other algorithms, and the number of parameters and floating-point operations have decreased significantly, and the amount of operations is less than that of other algorithms. The visualization results show that the algorithm proposed in this paper makes the model lighter and the model detection accuracy is improved, and the comprehensive detection effect is better than other algorithms.

4. Conclusion

This paper proposes the YOLOv7-CDD model based on the YOLOv7 target detection model. The model adds the CA attention mechanism and DSConv distribution offset convolution to make the model lighter, and adds Dynamic Head to the head, which makes the model more accurate in recognizing the behavior of eating cows. The accuracy of the improved model is 98.4%, the recall rate is 98.3%, and the mAP@0.5 is 99.3%, which are all improved compared with the original YOLOv7 model. However, although this improved model is more lightweight, the corresponding FPS will be reduced. In order to better optimize the model, in future research, the number of FPS frames and the diversification of data sets will be further considered to achieve more innovative research in the recognition of cow eating behavior.

Author Contributions

Conceptualization, W.L. and R.K.; methodology, W.L. and R.K. and Y.Z.; software, R.K; validation, R.K, W.L. and P.Z.; formal analysis, W.L. and R.K.; investigation, R.K; resources, R.K; data curation, R.K; writing — original draft preparation, R.K; writing — review and editing, W.L.; visualization, R.K; supervision, W.L.; project administration, Y.Z.; funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by National Natural Science Foundation of China, NSFC Fund No.62103309.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The dataset cannot be publicly disclosed due to privacy concerns. However, we can provide models and codes. The models and codes that support the findings of this study can be accessed by contacting the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Featured Application

This paper can help pasture enterprises or herders to save some costs and make it more convenient for them.

References

Bai Qiang, Gao Ronghua and Zhao Chunjiang et al. (2022). Multi-scale behavior recognition method of dairy cows based on improved YOLOV5s network. *Transactions of the Chinese Society of Agricultural Engineering*, *38*(12), 163-172.

Cui Liqun, Cao Huawei. (2024, May). Improved YOLOv7 for aerial image target detection. *Computer Engineering and Applications*, 1-11.

Deng Changzheng, Liu Mingze and Fu Tian et al. (2024, March). Infrared image recognition of substation equipment based on improved YOLOv7-Tiny. I*nfrared Technology*, 1-8.

Eleonora F, Alberto R and Mirco C et al. (2023). Eating time of dairy cows: a study focusing on commercial farms. *Italian Journal of Animal Science*, *22*(1), 1023-1032.

Jia Xueying, Zhao Chunjiang and Zhou Juan et al. (2023). Online detection of citrus surface defects based on improved YOLOv7 model. *Transactions of the Chinese Society of Agricultural Engineering*, *39*(23), 142-151.

Li Yuwei, Fu Rui and Liu Fan. (2024). Improved YOLOv7 lightweight traffic sign detection algorithm. *Journal of Taiyuan University of Technology*, *55*(01), 195-203. DOI: 10.16355/j.tyut.1007-9432.2023BD009.

Li Z, Zhu Y and Sui S et al. (2024). Real-time detection and counting of wheat ears based on improved YOLOv7. *Computers and Electronics in Agriculture*, *218*, 108670.

Liu Yuefeng, Bian Haodong and He Yingjie et al. (2022). Multi-target dairy cow eating behavior recognition method based on amplitude iterative pruning. *Transactions of the Chinese Society of Agricultural Machinery*, *53*(02), 274-281.

Niu Weihua, Wei Yali. (2024). Aerial small target detection algorithm based on improved YOLOv 7. *Electro-Optics & Control*, *31*(01), 117-122.

Qin Lifeng, Zhang Xiaoqian and Dong Mingxing et al. (2021). Moving cow target extraction based on multi-feature fusion correlation filtering. *Transactions of the Chinese Society of Agricultural Machinery*, *52*(11), 244-252.

Qu Chenyang, Cheng Yanyun. (2024, January). Traffic sign detection algorithm based on improved YOLOv7. *Microelectronics and Computers*, 1-11.

Song Huaibo, Li Rong and Wang Yunfei et al. (2023). Severely occluded beef cattle target recognition method based on ECA-YOLO v5s network. *Transactions of the Chinese Society of Agricultural Machinery*, *54*(03), 274-281.

Song Lvming, Liu Mingqin and Li Xiangbin et al. (2024, January) Research on glass surface defect detection method based on improved YOLOv7. *Mechanical and Electrical Engineering Technology*, 1-10.

Wang Zheng, Xu Xingshi and Hua Zhixin et al. (2022). Lightweight dairy cow estrus behavior recognition based on YOLOv5n and channel pruning algorithm. *Transactions of the Chinese Society of Agricultural Engineering*, *38*(23), 130-140.

Xing Yongxin, Sun Youdong and Wang Tianyi. (2022). Individual recognition of dairy cows based on improved SSD algorithm. Computer *Engineering and Applications*, *58*(02), 208-214.

Xu Hongwei, Li Ran and Zhang Jiaxu. (2024). Lake floating object detection algorithm based on improved YOLOv7. *Modern Electronic Technology*, *47*(01), 105-110. DOI: 10.16652/j.issn.1004-373x.2024.01.019.

Xu Ming, Qu Taipeng and Jiang Yanji. (2024, May). Improved YOLOv7 traffic sign detection algorithm in complex scenes. *Computer Engineering*, 1-11.

Zhang Zhen, Zhou Jun and Jiang Zizhen et al. (2024) Apple recognition method in natural orchard environment based on improved YOLO v7 lightweight model. *Transactions of the Chinese Society of Agricultural Machinery*, *55*(03), 231-242+262.

Zou J, Arshad RM. (2024). Detection of whole body bone fractures based on improved YOLOv7. *Biomedical Signal Processing and Control*, *91*, 105995.