IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 May 2024, accepted 12 June 2024, date of publication 14 June 2024, date of current version 7 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3414859

RESEARCH ARTICLE

YOLO-DLHS-P: A Lightweight Behavior Recognition Algorithm for Captive Pigs

CHANGHUA ZHONG[1,2], HAO WU[1,2], (Member, IEEE), JUNZHUO JIANG[1,2], CHAOWEN ZHENG[1,2], AND HONG SONG[3]

[1]Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, Sichuan 643000, China
[2]Artificial Intelligence Key Laboratory of Sichuan Province, Zigong, Sichuan 643000, China
[3]School of Automation and Information Engineering, Aba Teachers College, Aba, Sichuan 623002, China

Corresponding author: Hao Wu (11305076@qq.com)

This work was supported in part by the Project of Sichuan Provincial Science and Technology Department under Grant 2022YFS0518 and Grant 2022ZHCG0035; in part by the Artificial Intelligence Key Laboratory of Sichuan Province Foundation under Grant 2023RYY06; in part by the Enterprise Informatization and Internet of things Measurement and Control Technology Key Laboratory Project of Sichuan Provincial University under Grant 2022WYY04; in part by the Talent Introduction Project of Sichuan University of Science and Engineering under Grant 2021RC12; and in part by the Project of Zigong Science and Technology Bureau under Grant 2019YYJC13, Grant 2019YYJC02, and Grant 2020YGJC16.

ABSTRACT To meet the needs of embedded devices for model lightweight and high-precision recognition, this paper proposes a lightweight YOLO-DLHS-P model for pig behavior recognition based on the improved YOLOv8n model. Firstly, the C2f-DRB structure is introduced at the Backbone position, and the sizeable convolutional kernel is used to extend the receptive field to enhance the spatial perception ability of the model, and to enhance the network's ability to capture spatial information while maintaining the number of learnable parameters and computational efficiency; The LSKA attention mechanism is then introduced to be integrated into the SPPF module to construct the SPPF-LSKA structure, which significantly improves the ability of the SPPF module to aggregate features at multiple scales; Then, the downsampling at the Neck position is optimised to the HWD algorithm, which reduces the spatial resolution of the feature map while retaining more useful information and reduces the uncertainty of the information compared with the downsampling method of the baseline model; finally, the Shape-IoU is used to replace the original CIoU, which significantly improves the detection efficiency and accuracy of the model without increasing the extra computational burden. After constructing the improved YOLO-DLHS model, the improved model is then pruned using the LAMP pruning scoring algorithm to obtain a lightweight YOLO-DLHS-P model. The experimental results show that the YOLO-DLHS model improves P, mAP@0.5, and mAP@0.5-0.95 by 4.39%, 1.68%, and 3.97%, respectively, compared to the YOLOv8n model. The YOLO-DLHS-P model improves P, mAP@0.5, and mAP@0.5-0.95 by 3.37%, 1.16%, and 2.11%, and the number of parameters, computation, and model occupancy are substantially reduced by 52.49%, 54.32%, and 49.33%, respectively. Moreover, the FPS of the YOLO-DLHS-P model reaches 79 frames, which has good real-time performance for pig behavior recognition. Therefore, the improved YOLO-DLHS-P in this paper is able to reduce the demand for hardware at the time of deployment under the premise of guaranteed accuracy and provides a lightweight behavioral recognition solution for the intelligent farming of captive pigs.

INDEX TERMS Behavior recognition, YOLOv8, captive hogs, attention mechanisms, lightweight.

I. INTRODUCTION

In modern smart farming, animal behavior recognition technology has become vital for improving farming efficiency and animal welfare [1]. In the beginning, traditional animal behavior monitoring relies on the feeder to determine whether the animal is sick or not through the feeder's experience and observation, which is undoubtedly a time-consuming and labor-intensive method and is not able to achieve 24-hour real-time monitoring. Moreover, in a high-density farming environment, the contact between the feeder and the animal increases the risk of the animal falling ill, posing a potential

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen.

threat to the health of the animal. Further, some farms use Radio Frequency Identification (RFID) technology to wear smart ear tags to the pigs, and when the pigs are in the feeding area, they interact with the signals from the RFID to identify the feeding behavior of the pigs [2]. Some studies have achieved the recognition of four behaviors: walking, feeding, lying, and standing for cows [3] and five behaviors: sitting, standing, walking, grazing, and ruminating for sheep by wearing speed sensors to cows and sheep and collecting acceleration information about the relevant behaviors, and then using algorithms for machine learning [4]. However, the invasive detection methods mentioned above not only increase the cost of farming, but may also cause stress to the animals. Non-invasive identification of animal behavior using computer vision can be a good solution to the above problems [5].

According to relevant studies, behavioral changes in animals are correlated with their health [6]. During the period of illness, the animal will reduce the amount of food intake and exercise [7], and appear to lie down for a long time and so on [8]. When the diseased part of the animal is the internal abdominal organs, most of the animals will relieve the pressure in the abdominal cavity by sitting in a canine position. The climbing behavior of animals may increase skin abrasion, and the pressure of the hindfoot becomes higher to appear lameness hazards. The estrus behavior of animals can be judged by the climbing and straddling behavior that occurs between two animals [9]. Therefore, it is possible to monitor the behavior of animals, which can be used to detect health problems and improve the welfare of animals promptly.

With the development of technology, researchers at home and abroad have used different methods to monitor the behavior of animals. Kashiha et al. [10] used a CCD camera located on the top of the pens together with a water meter to monitor the drinking of pigs. The distance from the pig's head and ears to the waterer and its dwell time at the waterer were accurately measured by the image contour analysis technique to identify the pig's drinking behavior. Yang et al. [11] used a Faster R-CNN network to locate each pig in the pens, correlate the pig's head with its body, and analyze the feeding area with the occupancy of the image pixels to identify the pig's eating behavior, this method resulted in 99.6% accuracy in recognizing pig's eating behavior. Nasirahmadi et al. [12] used RGB cameras to extract pig images through background subtraction and applied a Support Vector Machine (SVM) classifier to identify the side-lying and belly-lying postures of pigs. By calculating the boundaries of each pig, perimeter, and other data, use this as a feature to train SVM and then identify its lying posture. This method achieved an accuracy of 94.4% and 94%, respectively, in the automatic recognition of side-lying and abdominal-lying postures. Subsequently, Nasirahmadi et al. [13] applied the ellipse fitting technique to localize pigs and used the intersection of the long and short axes of the ellipse to define the head, tail, and side positions of the pig. Based on the Euclidean distance between the head, tail, and side of the head and the axis length of the ellipse, the climbing behavior of the pig was successfully identified. Wang et al. [14] proposed a lightweight cow mounting behavior recognition system based on YOLOv5s, which combines the attention mechanism, inverted residual structure, and depth separable convolution of EfficientNetV2, and designed a lightweight backbone network and feature enhancement module, with a MAP of 87.7%. Shang et al. [15] combined the improved SE attention mechanism (Squeeze-and-Exitation Attention Mechanism, SE) and CBAM (Convolutional Block Attention Module, CBAM) attention mechanism to optimize the Mobilenetv3 model, and then combined it with the trajectory recognition algorithm to analyze and judge the cattle's trajectory, and finally identify the cattle's behavior. The algorithm identified the behavior of a variety of livestock, with the highest recognition accuracy reaching 95.17%.

The above research provides ideas for smart animal breeding and has a good recognition rate for a few behaviors of farmed animals. However, in actual animal breeding monitoring, animal behaviors are diverse, and the animal behavior data provided by the models of the above research are slightly single. Lao et al. [16] used a 3D camera combined with a deep image analysis algorithm to identify the lying, sitting, standing, kneeling, eating, and drinking behaviors of sows in the farrowing crates, of which the recognition rate of the kneeling behavior of the sows was 78.1%, and the recognition rate of the rest of the behaviors reached more than 90% recognition rate. Zheng et al. [17] applied the Faster R-CNN algorithm with a deep learning framework to successfully recognize five postures of sows in free-range pens: standing, sitting, chest ambulation, belly ambulation, and side-lying. Li et al. [18] proposed a spatio-temporal convolutional network for multi-behavioral recognition based on the SlowFast network architecture of the spatio-temporal convolutional network (PMB-SCN) for automatic identification and classification of five basic pig behaviors: feeding, lying, moving, scratching and climbing. The highest accuracy achieved by this model was 97.63%. Gu et al. [19] proposed a two-stage recognition method for sheep behavior based on deep learning, combining multi-scale feature aggregation, attention mechanism and deep convolution module, determines whether the sheep behavior is a normal physiological activity or destructive behavior in the detection stage, and uses VGG network to perform a specific classification in the classification stage, which recognizes a total of six behaviors, namely, standing, eating, lying, attacking, biting and climbing. The experimental results show that the method has a mAP of more than 98% in the detection phase and an accuracy of more than 94% in the classification phase, but the memory of the detection model reaches 130 MB.

Although numerous animal behavior recognition algorithms emerge nowadays, problems such as large model parameters, high consumption of computational resources, and excessive main memory occupied by the model arise, which raise the hardware threshold in practical deployment

and are unfriendly to many small and low-capacity devices. Therefore, in this study, by optimizing and adapting the YOLOv8n model architecture and then lightening the model size through model pruning techniques, we not only improve the accuracy of the model in recognizing pig behaviors but also significantly reduce the complexity and running cost of the model, which makes our improved model more suitable to be deployed and used in resource-constrained environments. The contributions of this paper are as follows:

1) In order to make the model more adaptable to pig behavior recognition, this study takes the YOLOv8n model as the baseline model and adjusts and optimizes its structure. Firstly, in the backbone part, this paper introduces the Dilated Reparam Block (DRB) of UniRepLKNet into the C2f structure, aiming to enhance the spatial perception ability of the model by expanding the sensory field, so as to improve the performance of the recognition of pig behaviors. The DRB module combines the parallel large core and the dilated convolutional layer, which is capable of capturing the sparse features, optimizes the computational efficiency by structural reparameterization technique, and finally obtains the optimized C2f-DRB structure.
2) The Large Separable Kernel Attention (LSKA) is integrated into the Spatial Pyramid Pooling Fast (SPPF) structure SPPF to form the SPPF-LSKA structure. The LSKA attention mechanism enhances the network's attention to important features by using large separable convolution kernels and spatially extended convolutions to capture extensive contextual information about the image, generate an attention map, and weigh the original features through the attention map, improve the performance of the pig behavior recognition model.
3) The downsampling part of the YOLOv8 model is improved by introducing the Haar Wavelet Downsampling (HWD) module to replace the stepwise convolution downsampling method of the baseline model. The HWD module utilizes the characteristics of the Haar Wavelet Transform, which effectively retains more feature information of the image in the downsampling process, significantly reduces the loss of spatial information, optimizes the feature extraction effect of the subsequent layers, further improving the overall performance of the pig behavior recognition model.
4) The traditional CIoU loss function is replaced with the Shape-IoU loss function to optimize the performance of the YOLOv8n model in the pig behavior recognition task. The Shape-IoU loss function takes into account the shape and scale differences between the predicted frame and the real frame and defines the loss by accurately calculating the differences in aspect ratio and relative size, which effectively improves the accuracy of bounding box regression. This improvement significantly enhances the sensitivity of the model to changes in the shape of the target, accelerates the convergence of the training process, and reduces the information loss, which significantly improves the detection efficiency and accuracy of the model without adding additional computational burden.
5) The Layer-Adaptive Magnitude-based Pruning (LAMP) scoring method is used to perform the pruning operation on the improved YOLO-DLHS model.LAMP scoring eliminates the need for tedious hyper-parameter tuning and reduces output distortion during re-training, which improves the performance of the pruned model and training efficiency. By model pruning, the parameters, computation, and model size of the improved model are drastically reduced with little change in accuracy, which reduces the hardware requirements of the model for actual deployment and saves costs.

The recognition process of the improved algorithm in this paper is illustrated in Figure 1. The YOLO-DLHS-P model is developed by enhancing the YOLOv8n's C2f structure, SPPF structure, downsampling, and loss function, resulting in the C2f-DRB module, SPPF-LSKA module, HWD downsampling module, and Shape-IoU loss function. Subsequently, the model is lightweighted using the LAMP pruning scoring algorithm. Ultimately, the model successfully identifies the behavior of pigs in the pen and outputs the recognition results. The YOLO-DLHS-P pig behavior recognition model presented in this paper is capable of identifying five behaviors: sitting, lying down, standing, crawling, climbing, and eating. This model provides a reference for intelligent pig farming and improving welfare conditions.

II. RESEARCH METHODOLOGY

A. THE YOLOv8n ALGORITHM

YOLOv8 is the latest algorithm of the current YOLO series, which is divided into n, s, m, l, and x, in order from small to large, according to the depth and width of the network, with a total of five models. In this paper, we choose model n. The algorithmic network of the YOLOv8n model consists of four part: Input, Backone, Neck, and Head, respectively. Input part, the base input is 640×640. Backbone This part includes a C2f structure, Conv, and SPPF structure, and it has more hopping layer connections and split operations. The Neck part Adopts the PANet structure, which consists of two networks, the Path Aggregation Network (PAN) and Feature Pyramid Network (FPN), in which the PAN network introduces the path aggregation method, which aggregates the semantic information of the shallow feature map with the deeper feature map, and strengthens the ability to express the multiscale features. The Head part is replaced with a Decoupled-Head structure. In terms of Loss, the strategy of matching positive and negative samples is applied [20], and Distribtion Focal Loss (DFL) [21] is added to the regression loss function. The detailed structure of YOLOv8n is shown in Figure 1.

B. IMPROVED BEHAVIORAL RECOGNITION ALGORITHMS

In this study, the C2f-DRB structure is constructed by introducing the expansion reparameterization block, and the

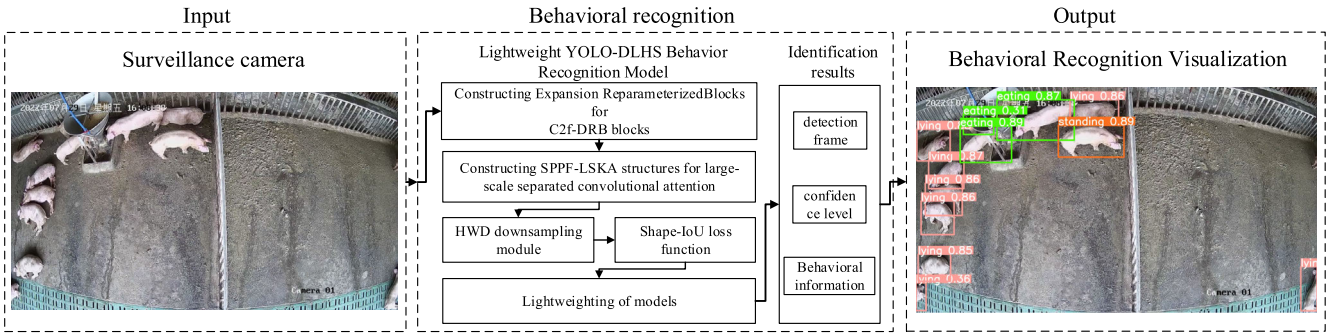


FIGURE 1. Behavior recognition flowchart of the improved algorithm in this paper.

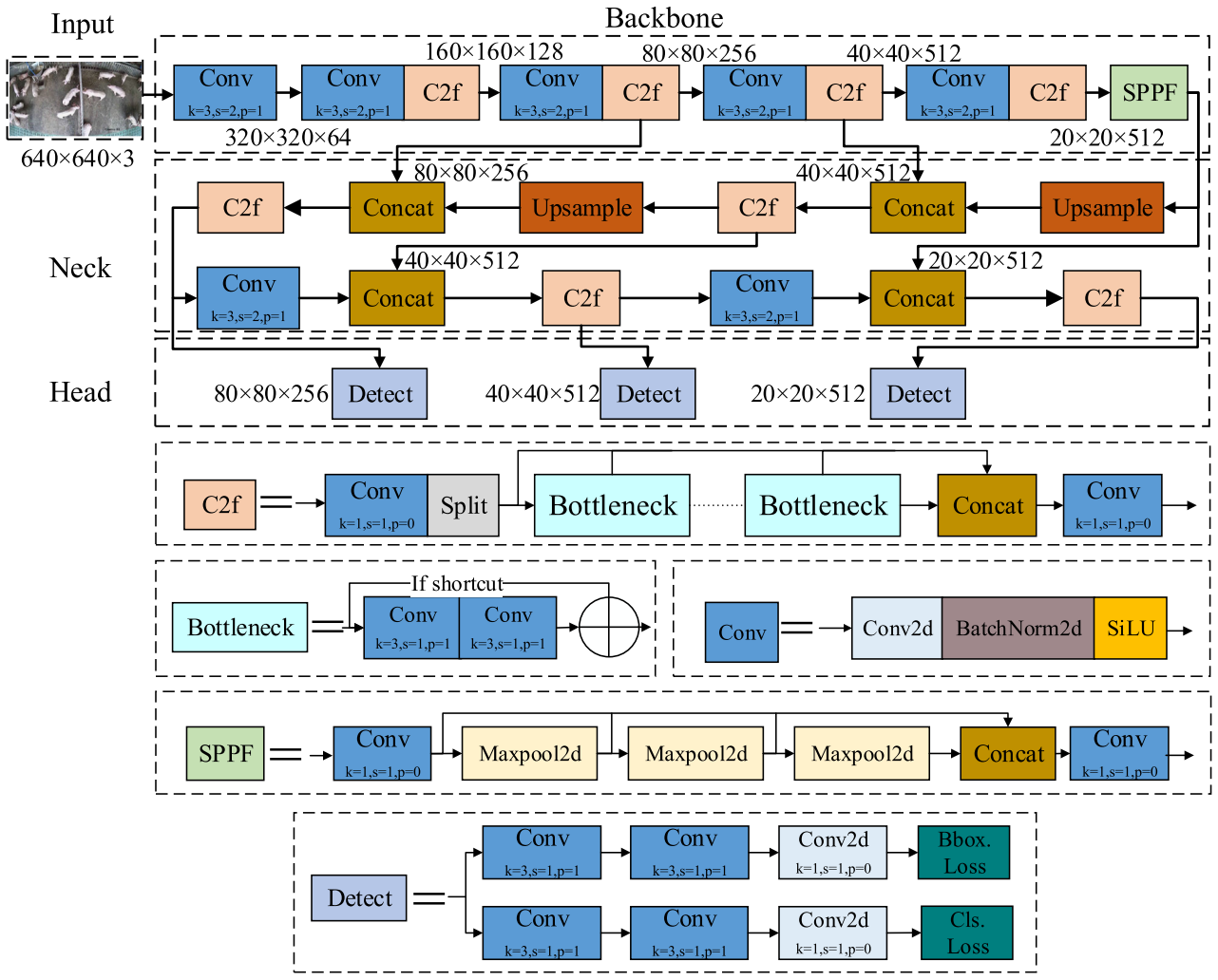


FIGURE 2. YOLOv8n network structure diagram.

structure reparameterization technique is used to improve the computational efficiency and enhance the spatial perception of the pig behavior recognition model; the SPPF-LSKA structure is constructed by using the LSKA attention mechanism, which enhances the details and contextual information related to pig behaviors; the traditional use of the cross-step convolutional downsampling method in the original benchmark model is optimized for HWD downsampling, which preserves the edge and texture information of the image in the downsampling process, reduces the resolution while increasing the number of channels and improves the feature extraction effect; the CIoU in the benchmark model is

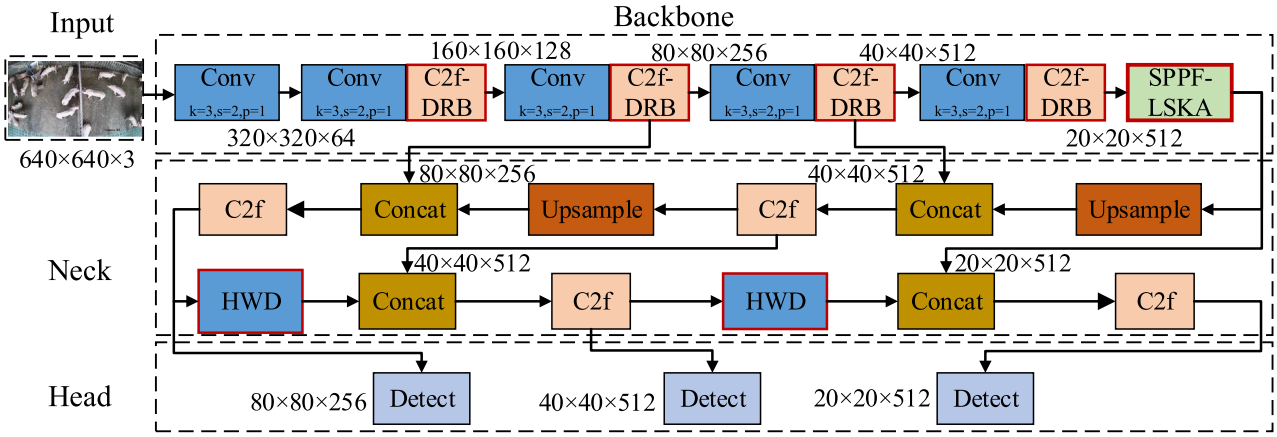


FIGURE 3. Structure of the improved YOLO-DLHS (YOLO-DLHD-P) model.

replaced with Shape-IoU, which reduces the loss of information without increasing the number of parameters and increases the model accuracy. The finally obtained improved YOLO-DLHS model is shown in Figure 3. In order to further lighten the size of the model, in this study, the improved model is pruned to obtain the YOLO-DLHS-P model using LAMP scoring, which is worthwhile in exchange for a significant decrease in the number of parameters, computation, and model occupancy size of the improved model with only a small loss in accuracy. Where improvements have been made in the structural diagram of the YOLO-DLHS model, they have been marked with red line boxes.

III. MODEL IMPROVEMENTS

A. C2f-DRB MODULE

The Bottleneck structure, generally used for deep networks, was first proposed in ResNet [22]. The Bottleneck structure in YOLOv8n first uses a 3×3 Conv for dimensionality reduction to optimize information extraction and reduce redundant data. A 3 × 3 Conv is then used for dimensionality upgrading to extract higher-level features and enhance data representation. The input channel of the first Conv is C1, the output channel is C_, and the input channel of the second Conv is C_, and the output channel is C2. The C_ input channel is 1/2 of C2. If the input channel of C1 is the same as the input channel of C2, then the connection is changed to a residual connection.

In this paper, we introduce the idea of DRB in UniRepLKNet [23] to enhance the performance of the pig behavior recognition network.DRB This module was originally designed to utilize pa rallel large-core convolutional and dilated convolutional layers to enhance the network's ability to spatially information while maintaining the number of learnable parameters and computational efficiency. capture capability. This module design uses the convolution of the large kernel together with the parallel convolution of the small kernel, and then the corresponding batch normalization (BN) layer after the output is summed up, and by using the structural reparametrization technique after the training is completed, the BN layer can be merged into the convolutional layer, in this design in order to the additional computational overhead not to be increased, an equivalent conversion strategy is proposed, which will be the inflated convolution layer of the small kernel can be equivalent to the large kernel that has a sparse non-inflated (i.e., r = 1) layer of the kernel, i.e., the whole Block can be equivalently converted to a large kernel convolution. And the conversion from an inflated convolutional kernel to an equivalent non-inflated large kernel convolutional kernel is realized by a transpose-convolution operation with the expression shown in Eq. (1) [23]:

$$W' = conv\_transpose2d(W, I, stride = r) \tag{1}$$

where W is the original inflated convolution kernel, I is the unit kernel with scalar 1, r is the expansion rate, and W' is the converted equivalent non-inflated convolution kernel. In this section, a non-expanded small kernel and multiple expanded small kernel layers are utilized to augment a non-expanded large kernel convolutional layer. Its hyperparameters include the size of the large kernel, K, the size of the parallel convolutional layers, k, and the expansion rate, r. As shown in Figure 4, the case with three parallel layers is demonstrated, where $K = 7$, $r = (1,2,3)$, and $k = (5,3,3)$, where the equivalent kernel sizes are (5,5,7), respectively, according to the constraints $(k\text{-}1)r + 1 \leqslant K$. From a parametric point of view, such a dilated layer is equivalent to a non-dilated convolutional layer with a larger sparse kernel, and thus the whole block can be equivalently converted into a single large kernel convolution.

The DRB-Bottleneck structure is obtained by replacing the second convolutional layer in the Bottleneck structure in YOLOv8n with an expansion-weighted parameterized block, as shown in Figure 5.

This improvement strategy will significantly enhance the spatial perception ability of the model, and by expanding the receptive field, it enables the model to capture the details related to pig behavior more effectively, optimizes the inference efficiency of the model, and brings significant

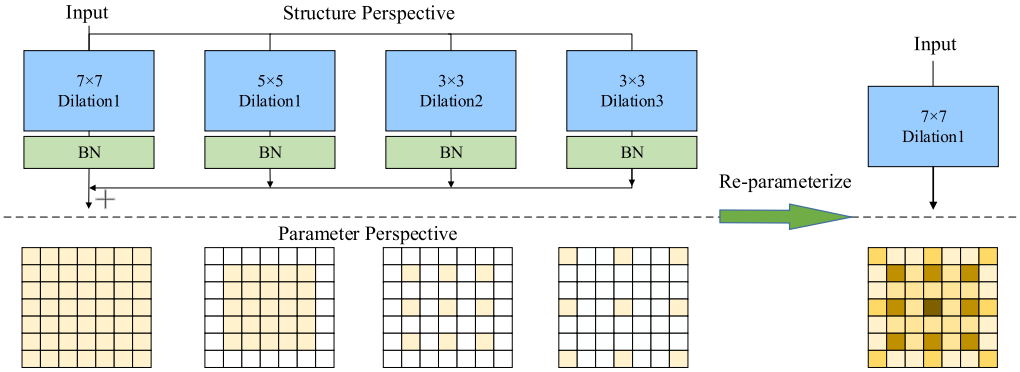


FIGURE 4. Dilated reparam block structure schematic diagram.

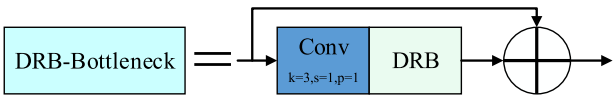


FIGURE 5. DRB-bottleneck structure schematic diagram.

performance improvement for the pig behavior recognition task. Thus the improved C2f-DRB structure diagram is obtained as shown in Figure 6.

B. SPPF-LSKA MODULE

In YOLOv8n, in order to solve the problem that the classical Spatial Pyramid Pooling (SPP) spatial pyramid pooling structure has some redundancy in the computation process, which results in a slower model, the SPPF structure is introduced to better balance the speed and accuracy of detection. In order to make the detection model better serve the pig behavior recognition, the LSKA [24] attention mechanism is integrated into the SPPF structure in YOLOv8n, which makes full use of the high-performance model performance of the LSKA attention mechanism in order to improve the accuracy of the pig behavior detection process, in which the LSKA structure is shown in Figure 7.

LSKA attention mechanism inherits the design of Large Kernel Attention (LKA) attention mechanism, which significantly reduces computational complexity and memory requirements by decomposing the traditional 2D weight kernel into two cascaded 1D separable weight kernels. Compared to LKA, LSKA demonstrates a better speed-accuracy tradeoff at different core sizes. Even when the core size is increased, the inference speed reduction of the LSKA model is significantly lower than that of the LKA model, which suggests that LSKA is able to capture a wider range of image features while maintaining efficient computation and providing similar or better performance. The relevant formulae for the LSKA attention mechanism are shown in Eqs. (2)-(5) [24]:

$$\bar{Z}^C = \sum_{H,W} W^C_{(2d-1)\times 1} * (\sum_{H,W} W^C_{1\times(2d-1)} * F^C) \quad (2)$$

$$Z^C = \sum_{H,W} W^C_{\lfloor \frac{k}{d} \rfloor \times 1} * (\sum_{H,W} W_{1\times \lfloor \frac{k}{d} \rfloor} * \bar{Z}^C) \quad (3)$$

$$A^C = W_{1\times 1} * Z^C \quad (4)$$

$$\bar{F}^C = A^C \otimes F^C \quad (5)$$

where $\bar{z}^C$ represents the output after deep convolution, The H and W below the summation symbols denote the height and width in the feature map, A^C represents the attention map, $\lfloor \cdot \rfloor$ represents the downward rounding operation, d represents the expansion rate, k is the kernel size, $*$ represents the convolution, $\otimes$ represents the Hadamard product, F^C represents the input feature map and C represents the number of input channels.

According to the above formula, it can be seen that the LSKA attention mechanism first uses 1D convolution kernels for spatial separation convolution, then uses the obtained feature maps to generate the attention maps, and finally applies the attention maps to augment the original feature maps, and such a process helps the network to deal with a large range of spatial information more efficiently. The ability of the model to capture key visual features is maintained by cascading 1D convolution kernels instead of traditional 2D convolution kernels. Therefore, when LSKA attention is integrated into the SPPF structure, the model's ability to recognize pig behaviors can be enhanced without adding too much computational burden. The obtained improved SPPF-LSKA structure is shown in Figure 8.

C. HWD DOWNSAMPLING MODULE

In the field of target detection, especially in models involving convolutional neural networks, it is common to employ a stepwise convolutional layer or pooling layer to perform downsampling operations on feature maps. The reason for this is that it is often necessary to create a reduced version of the image, i.e., a thumbnail, in order to ensure that the image fits into a specific display area size. The downsampling operation, which plays a key role in several stages of image processing, not only helps to prevent overfitting of the model but also expands the model's perception of the image. When performing routine detection tasks, this downsampling step

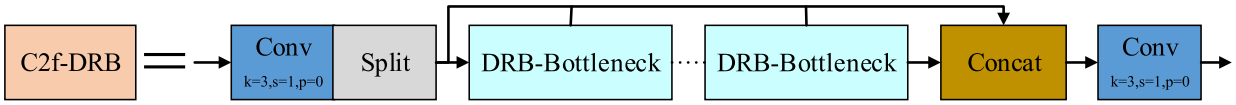


FIGURE 6. C2f-DRB structure schematic diagram.

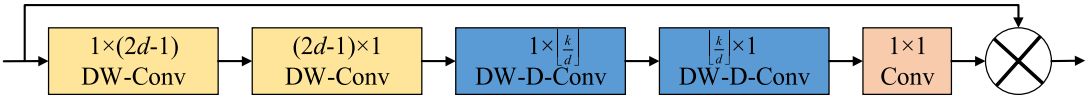


FIGURE 7. LSKA attention mechanism structure schematic.

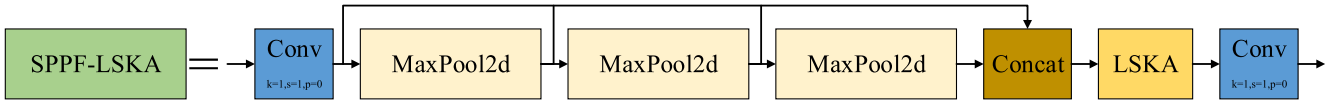


FIGURE 8. SPPF-LSKA structure schematic diagram.

helps to highlight detailed information that is important for the recognition process, while excluding background information that is not relevant to the task.

In this section, improvements are made to downsample the Neck part of the model. While the stepwise convolutional downsampling operation used in the Neck part of the original YOLOv8 model can expand the sensory field and reduce the data dimensions, this method is prone to lead to the loss of important spatial information, especially in fine behavioral recognition, such as the tiny interactive movements among pigs and the subtle changes in their behavioral patterns. In addition, traditional convolutional downsampling tends to cause blurring of boundary information when dealing with image edges or texture details, reducing the accuracy of behavior recognition. To solve these problems, we replace the two downsampled convolutional layers in the Neck part with the HWD [25] module; this module effectively reduces the loss of information during downsampling by retaining more edge and texture information. The structure of the HWD module is shown in Figure 9.

The HWD module consists of two blocks, namely: a lossless feature coding block and a feature representation learning block, the main responsibility of the lossless feature coding block is to transform the features and reduce their spatial resolution, by employing the Haar Wavelet Transform, which is an efficient way to reduce the resolution of the feature mapping while being able to retain all the information. Next, the representation learning block consists of a standard convolutional layer, a batch normalization layer, and a ReLU activation layer, which is used to extract discriminative features and filter redundant information. Where the wavelet basis function and scale function of the first level one-dimensional Haar transform in the Haar wavelet transform can be defined as Eqs. (6)-(9) [25]:

$$\begin{cases} \phi_1(x) = \frac{1}{\sqrt{2}}\phi_{1,0}(x) + \frac{1}{\sqrt{2}}\phi_{1,1}(x) \\ \psi_1(x) = \frac{1}{\sqrt{2}}\phi_{1,0}(x) - \frac{1}{\sqrt{2}}\phi_{1,1}(x) \end{cases} \tag{6}$$

$$\phi_{j,k}(x) = \sqrt{2^j}\phi(2^j x - k), k = 0, 1, \cdots, 2^j - 1 \tag{7}$$

$$\phi_{0,0}(x) = \phi_0(x) = \begin{cases} 0, & x < 0 \\ 1, & 0 \le x < 1 \\ 0, & x \ge 1 \end{cases} \tag{8}$$

$$\begin{cases} \phi_1(x) = \phi_0(2x) + \phi_0(2x-1) \\ \psi_1(x) = \phi_0(2x) - \phi_0(2x-1) \end{cases} \tag{9}$$

where j and k represent the scale (or "order") and ordinate (or orientation when dealing with two-dimensional images) of the Halki function, respectively. Eq. (7) is the definition of $\varphi_{0,0}(x)$. Eq. (9) shows that the level 1 Haar transform can be expressed in terms of the level 0 Haar basis function. And in conjunction with Figure 9, by applying the Haar wavelet transform, the original image is effectively decomposed into four components with halved spatial resolution to capture the low and high frequency (horizontal, vertical, diagonal) information of the image, respectively, in order to retain more image information. This transformation allows the model to maintain information integrity by increasing the number of channels for feature mapping while reducing the resolution, thus providing the model with information-rich and moderate-resolution inputs that optimize feature extraction in subsequent layers. Therefore, the introduction of the HWD module not only improves the accuracy of behavior recognition but also enhances the sensitivity of the model to various behavioral patterns of pigs, thus significantly improving the recognition performance in practical applications.

D. LOSS FUNCTION

In target detection, Intersection over Union (IoU) is one of the important metrics to measure the performance of target detection algorithms. IoU is used to evaluate the degree of overlap between the predicted bounding box and the real target bounding box, and thus to determine whether the predicted box has correctly captured the location and size of the target. IoU is calculated by calculating the area of the intersection of the predicted box and the real box divided by

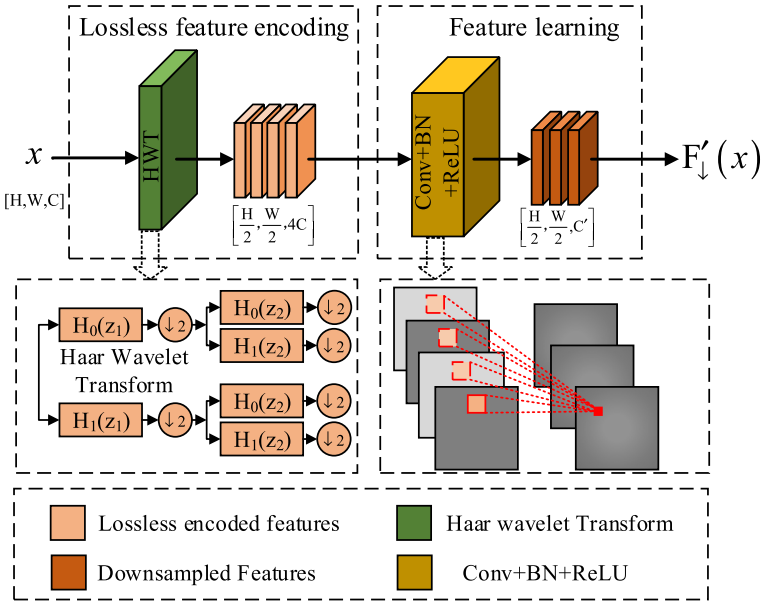


FIGURE 9. HWD module structure schematic.

their concatenated area, as shown in Eq. 10 [26]:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \tag{10}$$

where B denotes the predicted bounding box, B^{gt} denotes the true bounding box, and the numerator denotes the area of the intersection region of the predicted bounding box and the true bounding box. The denominator represents the area of the concatenation region of the predicted and real bounding boxes.

In the YOLOv8n model, CIoU [26], as one of the regression loss functions for the detection frames, is deficient in dealing with the mismatched orientations of the predicted and real frames in the actual treatment of group-housed hogs' behavioral recognition task. This leads to the prediction frames shifting in incorrect directions during the training process, which leads to slow convergence and inefficiency, and ultimately affects the model performance, and there are some ambiguities in the CIoU when describing the vertical and horizontal intersection and merger ratios of the bounding box, and it also does not take into account how to balance the processing of difficult and easy samples.

To solve the above problems, this paper introduces Shape-IoU [27] to replace CIoU in the YOLOv8n model, as opposed to the traditional loss function which mainly considers the geometric relationship between the predicted frame and the real frame and calculates the loss by taking into account the relative positions and shapes of the bounding boxes, but ignores the influence of the inherent attributes of the bounding boxes' shapes and scales on the regression results, while the Shape-IoU can calculate the loss by focusing on the shape of the bounding box itself and its scale, thus making the regression of the bounding box more accurate. Shape-IoU defines equations as in Eqs. (11)-(15) [27]:

$$ww = \frac{2 \times (w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \tag{11}$$

$$hh = \frac{2 \times (h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \tag{12}$$

$$distance^{shape} = hh \times (x_c - x_c^{gt})^2/c^2 + ww \times (y_c - y_c^{gt})^2/c^2 \tag{13}$$

$$\Omega^{shape} = \sum_{t=w,h} \left(1 - e^{-\omega_t}\right)^{\theta}, \theta = 4 \tag{14}$$

$$\begin{cases} \omega_w = hh \times \dfrac{|w - w^{gt}|}{max(w, w^{gt})} \\ \omega_h = ww \times \dfrac{|h - h^{gt}|}{max(h, h^{gt})} \end{cases} \tag{15}$$

where *scale* is the scale factor, which is related to the proportion of the target in the dataset, and *ww*and *hh* are the weighting coefficients in the horizontal and vertical directions, respectively, whose values are related to the shape of the GT frame. The final obtained loss function is defined as Eq. (16) [27].

$$L_{Shape-IoU} = 1 - IoU + distance^{shape} + 0.5 \times \Omega^{shape} \tag{16}$$

E. IMPROVED MODEL PRUNING

Pruning the model has the advantage of reducing the number of model parameters, computation, and memory occupied by the model, and favorably reduces the requirement of hardware parameters when deployed in practice. In this study, model pruning is used to compress the improved YOLO-DLHS model.

TABLE 1. Experimental environment parameters.

Sports event	Parameters
Operating system	Windows 10
Programming language	Python3.7
CPU	AMD EPYC 7302
GPU	Nvidia Ampere A100
CUDA	11.3
Deep Learning Framework	Pytorch-1.10.0

LAMP [28] is a new pruning scoring criterion that does not require additional hyper-parameter tuning and can be applied to a wide range of network architectures and datasets. LAMP scoring is based on the magnitude of the weights, but is tuned by a specific ''model-level'' distortion metric, which allows the pruning process to be more focused on reducing the output distortion, thus improving the retraining performance of the model after pruning. model retraining performance after pruning. The LAMP score definition formula is shown in Eq. (17) [28]:

$$score(u; W) = \frac{(W[u])^2}{\sum_{v \geq u} (W[v])^2} \tag{17}$$

where $W[u]$ denotes the u-th weight and the denominator denotes the sum of the squares of all the weights from index u to the end of the list of weights in that layer. The LAMP pruning decision is formulated as in Eq. (18) [28]:

$$(W[u])^2 > (W[v])^2 \Rightarrow score(u; W) > score(v; W) \tag{18}$$

From the above equation, it can be seen that the square of W[u] is greater than the square of $W[v]$, and it can be obtained that the LAMP score of $W[u]$ will also be greater than the score of $W[v]$. Therefore, it can be seen that larger weights correspond to higher LMAP scores, while LAMP scores that are relatively lower will be pruned away. The process of calculating the LAMP score and its application to global pruning is shown in Figure 10.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT

The parameters for this experiment are set to 300 rounds of model training iterations using the SGD optimizer with a batch size of 32, the initial learning rate is set to 0.01, and the size of the input image is 640 × 640 (pixels), and all of the experiments in this paper are done under the same computer. Other parameters of the experimental environment are shown in Table 1.

B. COMPOSITION OF THE DATASET

The dataset in this study was collected from a large hog farm in Zigong, Sichuan Province, China. Videos of different numbers of hogs in different pens at different shooting angles were collected between 10:00 and 17:00 in the morning. Most of the videos in the dataset were captured by a 2-megapixel dome camera with a frame rate of 15 frames per second and a resolution of 1920 × 1080 pixels. A small portion of the video was captured by a handheld mobile video capture device. Using ffmpeg software, the captured video is segmented into images, images with similar content are removed by taking frames at intervals, and manual screening is used to remove blurred and distorted images, and finally the dataset for this study is constructed. In this paper, LabelImg software was used to label the pig pictures, and the saved labeling format was YOLO format. The labeled infoboxes included the coordinate positions and behavioral categories of the hogs in the pictures. Eventually, a total of 2,222 pictures were manually labeled. The detailed division of the dataset is shown in Table 2.

TABLE 2. Data set segmentation.

Data type	Quantities
Train	1777
Validation	222
Test	223
Total	2222

There are 6 behaviors in the dataset, namely: sitting, lying down, standing, climbing, crawling across and eating.The 6 behaviors are labeled with a total of 15,983 number of behaviors in 2,222 images, and the distribution of the number of each behavior is shown in Figure 11.

Figure 12 illustrates examples of the six behaviors in the dataset, sitting, lying down, standing, climbing, crawling across, and eating, at some of the angles.

C. EVALUATION INDICATORS

In order to evaluate the detection effect of the improved YOLOv8n model algorithm, this paper uses the detection accuracy rate (Precision, P), the detection rate (Recall, R), the average precision (AP), the mean average precision (mAP), the number of parameters (Parameters), Giga Floating-point Operation Per Second (GFLOPS), and ModelSize are the evaluation metrics. Among them, the time of model execution is measured using the unit of GFLOPS, Frames Per Second (FPS) and the larger the amount of computation, the more computational resources the model needs to use. The formulas for P, R, AP, and mAP are shown below:

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$AP = \int_0^1 P(R)dR \tag{21}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{22}$$

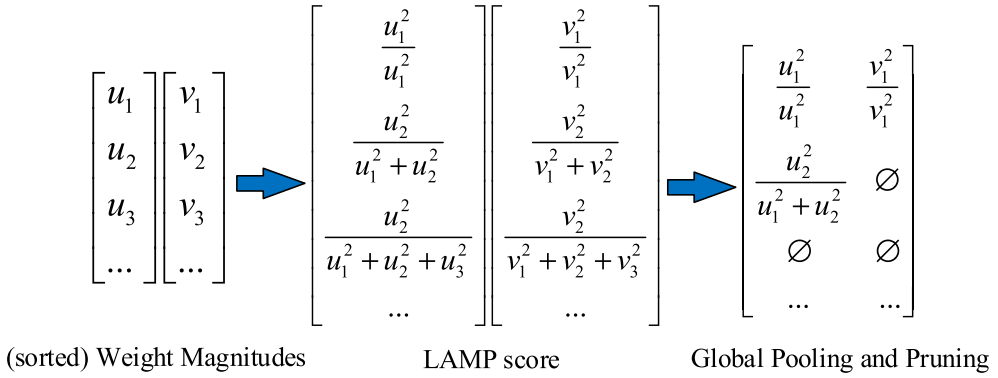


FIGURE 10. Schematic diagram of the LAMP score calculation process and its application to global pruning.

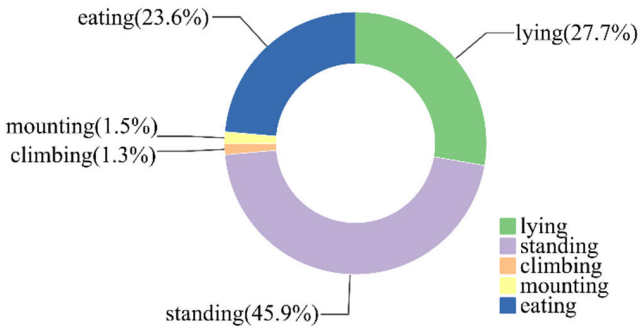


FIGURE 11. Distribution of the number of pig behaviors.

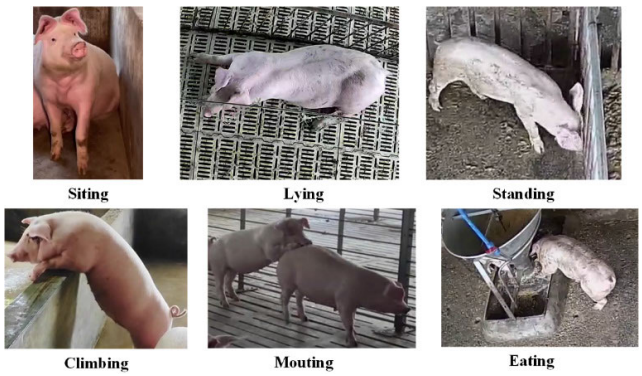


FIGURE 12. Example graph of hog behavior at some angles.

where FN means that the model detects positive samples as negative samples, FP means that the model predicts negative samples as positive samples, and TP means that the model successfully detects positive samples that are really present. Where mAP@0.5 denotes the mAP value at the intersection and concurrency ratio threshold of 0.5, which is used to assess the accuracy of the model in localizing the bounding box. mAP@0.5-0.95 then calculates the average mAP value from the intersection and concurrency ratio threshold of 0.5 to 0.95 (in steps of 0.05), which provides an assessment of the comprehensive performance of the model at different levels of localization accuracy.

D. ABLATION EXPERIMENT

Ablation experiments can prove the effectiveness of the model improvement, and in this section, we conduct ablation experiments for multiple groups using YOLOv8n as the original baseline model, and P, mAP@0.5, mAP@0.5-0.95, number of parameters, computation and model size as the evaluation indexes, respectively. As shown in Table 3. In the column of model YOLO-D stands for the YOLO-D model constructed on the basis of the benchmark model by improving the C2f and introducing the C2f-DRB, and the naming of the latter is the same. According to Table 3, it can be seen that improving C2f to C2f-DRB structure in the backbone network, P, mAP@0.5, and mAP@0.5-0.95 were improved by 0.77%, 0.47%, and 2.56%, respectively, compared to the baseline model, which was attributed to the inclusion of inflated heavy parameterized block, which enlarged the sensory field and captured the spatial information more efficiently, and the C2f-DRB enhanced the model's behavioral sensitivity and facilitated the accurate classification of behavioral actions. The SPPF structure was then replaced with the SPPF-LSKA structure integrating the LSKA attention mechanism, and the P, mAP@0.5,and mAP@0.5-0.95 were improved by 2.59%, 1.26%, and 2.17%, respectively, compared with the baseline model, indicating that the large separable convolutional kernel enhances the model's feature extraction ability while focusing on the behavioral key features more effectively. Next, the downsampling convolution in the benchmark model is replaced with the HWD module, and P, mAP@0.5, and mAP@0.5-0.95 are improved by 2.28%, 1.60%, and 2.76%, respectively, compared to the benchmark model, because HWD retains more useful feature information in the process of downsampling, which effectively reduces the loss of information in the downsampling process. The improved YOLO-DLHS model is obtained by replacing the loss function of the original baseline model with Shape-IoU. The addition of Shape-IoU enhances the sensitivity of the model to changes in the shape of the

TABLE 3. Ablation experiment.

Model	C2f-DRB	SPPF-LSKA	HWD	Shape-IoU	Prune	P /%	mAP@0.5 /%	mAP@0.5-0.95 /%	Param /M	Computation /GFLOPS	Model size /MB
YOLOv8n	—	—	—	—	—	90.74	93.08	71.55	3.01	8.10	5.98
YOLO-D	√	—	—	—	—	91.51	93.55	74.11	2.78	7.50	5.63
YOLO-DL	√	√	—	—	—	93.33	94.34	73.72	3.05	7.70	6.15
YOLO-DLH	√	√	√	—	—	93.02	94.68	74.31	3.00	7.60	5.96
YOLO-DLHS	√	√	√	√	—	**95.13**	**94.76**	**75.52**	3.00	7.60	5.96
YOLO-DLHS-P	√	√	√	√	√	94.11	94.24	73.66	**1.43**	**3.70**	**3.03**

target, accelerates the convergence of the training process, and improves the model accuracy. Up to this point, the YOLO-DLHS model optimized for the benchmark model after the four improvement points are added at the same time, P, mAP@0.5, and mAP@0.5-0.95 are improved by 4.39%, 1.68%, and 3.97% respectively compared to the benchmark model.

In order to further lighten the model, based on the improved model YOLO-DLHS, the LAMP pruning scoring algorithm is utilized for pruning the improved model, and finally, the lightened YOLO-DLHS-P model of YOLO-DLHS is obtained. The YOLO-DLHS-P model is relative to the YOLO-DLHS model, the P, mAP@0.5, and mAP@0.5-0.95 only decreased by 1.02%, 0.52%, and 1.86%, respectively, while the number of parameters, computation, and model size decreased dramatically by 52.33%, 51.32%, and 49.16% respectively. The YOLO-DLHS-P model compared to the YOLOv8n model, P, mAP@0.5 and mAP@0.5-0.95 improved by 3.37%, 1.16%, and 2.11%, and the number of parameters, computation, and model size are reduced by 52.49%, 54.32%, and 49.33%, respectively. The YOLO-DLHS-P model obtained by the YOLO-DLHS model with only a little sacrifice of accuracy still has higher recognition accuracy than the YOLOv8n model, and the number of parameters, computation, and model size has decreased a lot, so the present ablation experiments can be obtained that the YOLO-DLHS model is optimized through the four improvement points, and better recognition accuracy is obtained compared to the YOLOv8n model. YOLOv8n better pig identification model. The YOLO-DLHS-P model is better for practical deployment, with substantially lower hardware requirements for deployment and higher recognition accuracy than the YOLOv8n model. The YOLO-DLHS model and the YOLO-DLHS-P model demonstrate the effectiveness of the improvements.

To further validate the efficiency of the YOLO-DLHS and YOLO-DLHS-P models in this study relative to the YOLOv8n improvement. This section also compares the heat map results of the images from the YOLOv8n model, the YOLO-DLHS model, and the YOLO-DLHS-P model in different scenarios, as shown in Figure 13. The heat map not only enhances the transparency and interpretability of the model decisions by highlighting the image regions that the algorithm regards as important but also helps to optimize the algorithm performance by revealing possible regions that the algorithm may have overlooked or mislabeled. The red color in the heat map indicates regions that the model pays more attention to, while the blue color indicates relatively less attention.

In Figure 13, (a) the scene is an overhead shot of a large pig breeding pen. The environment is characterized by a large number of pigs, and involves two situations of dense and dispersed pigs; (b) the scene is an overhead shot of a small pig breeding pen. The characteristic of this environment is the high-density breeding of pigs in a small environment; the scenes (c) and (d) are pictures trimmed from the scene (a) at different times, divided into pig conditions and dense conditions. Comprehensively analyzing the heat maps in the four scenes (a), (b), (c), and (d) in Figure 13, it can be clearly seen that the YOLO-DLHS model and the YOLO-DLHS-P model become more attentive to the pigs, and the contour information of the pigs becomes more obvious compared to the YOLOv8n model, especially the YOLO-DLHS model, which is the model with the highest attention to the pigs. The red area of the heat map is also more concentrated. This shows that the YOLO-DLHS model has the best ability to identify behaviors through the four optimizations. As for the YOLO-DLHS-P model, the red area of the heat map is slightly less concentrated after pruning compared to the YOLO-DLHS model, but compared to the YOLOv8n model, the YOLO-DLHS-P model pays better attention to the pigs.

E. COMPARISON EXPERIMENT

1) IMPROVED POSITION COMPARISON EXPERIMENT

In the process of pig identification model improvement, the improvement module may have different effects at different locations in the network. In the structure of YOLOv8n, C2f, and downsampling convolution are present in both the Backbone position and Neck position, and the selected place to optimize can be derived from the improvement position ablation experiment. Where A, B, and AB represent the improved C2f-DRB structure replacing all C2f structures in the Backbone position, replacing all C2f structures in the Neck position, and replacing all C2f structures in both the Backbone position and the Neck position, respectively. A', B,' and A'B' represent the improved downsampled HWD structure replacing the Backbone position, respectively all convolutions except the first convolution, replacing all

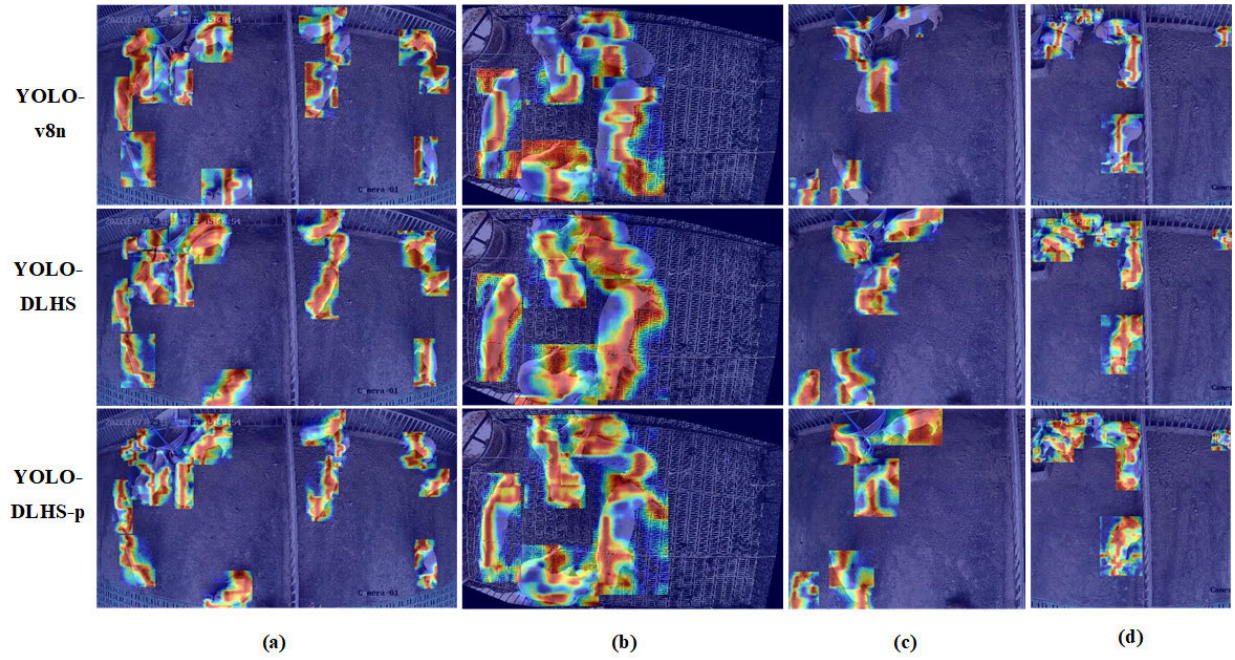


FIGURE 13. Comparison of heat maps without scenarios.

TABLE 4. Improvement location comparison.

Improved location (C2f-DRB/HWD)	P /%	R /%	mAP@0.5 /%	mAP@0.5-0.95 /%
A	91.51	**92.30**	**93.55**	**74.11**
B	89.95	87.75	92.25	70.71
AB	94.70	89.46	93.54	71.21
A'	91.88	91.66	93.86	72.45
B'	**93.02**	**93.41**	**94.68**	**74.31**
A'B'	92.35	89.00	93.67	71.86

convolutions in the Neck position, and replacing all convolutions in the Backbone position except the first convolution and in the Neck position. The experimental results are shown in Table 4.

As can be seen from Table 4, when C2f-DRB replaces all C2f structures in the Backbone location, the data of P, R, and mAP@0.5 and mAP@0.5-0.95 are more balanced with respect to the other locations, and the values of R, and mAP@0.5 and mAP@0.5-0.95 are higher with respect to the locations B and AB. Similarly, when the downsampled HWD structure position is at B', it has higher detection accuracy.

2) COMPARATIVE PRUNING EXPERIMENTS

Different pruning strategies for the same model will give different results for the accuracy of the model after pruning. Speed-up represents the compression rate of the computation under the LAMP pruning scoring algorithm. For example, LAMP-1.5 represents that the amount of computation before pruning is 1.5 times the amount of computation after pruning under the LAMP pruning scoring algorithm, and the same for the later ones. For model pruning, it is important to minimize the number of parameters, the amount of computation, and the size of the model while maintaining accuracy as much as possible. Therefore, this section compares the different compression rates of computation under the same pruning algorithm by comparing several groups, and the detailed data of each item is shown in Table 5.

Through Table 5, it can be seen that when the compression ratio of the calculation amount is between 1.5 and 3.5, mAP@0.5 remains at a high level, but except for LAMP-1.5 and LAMP-2.0, the P and mAP@0.5-0.95 of other compression ratios have been greatly reduced. If the accuracy is neglected in order to maximize the compression ratio of the calculation amount, it is meaningless in practical applications. Therefore, in this study, LAMP-2.0 was selected as the lightweight model after improving the model under the condition of weighing the compression ratio of the calculation amount and the loss of precision.

3) C2f IMPROVED STRUCTURE COMPARISON

In order to verify the effectiveness of C2f-DRB improvement at the Backbone location, this section improves the structure of several C2f for comparison, C2f-DWR is constructed by introducing a Dilation-wise Residual structure [29], C2f-DBB is constructed by introducing DiverseBranchBlock [30] structure, C2f-iRMB is constructed by introducing the Inverted Residual Mobile Block [31] structure, C2f-Faster is constructed by introducing the FasterBlock [32] structure, C2f-OREPA is constructed by introducing the Online Convolutional Re-parameterization [33] structure is constructed,

TABLE 5. Comparison of compression rates for different compute volumes.

Speed-up	P /%	mAP@0.5 /%	mAP@0.5-0.95 /%	Param /M	Computation /GFLOPS	Model size /MB
LAMP-1.5	94.42	93.80	74.23	1.90	4.80	3.95
LAMP-**2.0**	94.11	**94.24**	73.66	1.43	3.70	3.03
LAMP-2.5	92.36	94.20	72.64	1.28	2.90	2.75
LAMP-3.0	90.61	93.03	71.97	0.96	2.50	2.15
LAMP-3.5	91.40	94.02	72.28	0.81	2.10	1.86
LAMP-4.0	91.53	91.71	68.28	0.72	1.80	1.66
LAMP-4.5	88.25	91.45	67.89	0.63	1.60	1.50
LAMP-5.0	90.04	89.62	67.01	0.55	1.40	1.35

TABLE 6. Different approaches to C2f improvement.

Framework	R /%	mAP@0.5 /%	mAP@0.5-0.95 /%
C2f-DWR	90.50	93.08	74.19
C2f-DBB	89.59	93.51	73.67
C2f-iRMB	88.90	92.95	72.02
C2f-Faster	88.51	92.28	71.21
C2f-OREPA	85.52	93.04	72.66
C2f-AKConv	89.82	92.89	71.59
C2f-DRB	**92.30**	**93.55**	74.11

and C2f-AKConv is constructed by introducing Variable Kernel Convolution [34]. The data of the specific comparison is shown in Table 6.

Through comparison, it can be seen that the mAP@0.5-0.95 index of the C2f-DRB structure is almost the same as the first in the table, and its R and mAP@0.5 are both the first. All the improved structures of C2f that choose the C2f-DRB structure as the Backbone position are better.

4) COMPARISON OF SPPF IMPROVED STRUCTURES

Integrating different attentional mechanisms in SPPF can make the accuracy of pig behavior recognition different. This section compares the effect of different attention mechanisms integrated into the SPPF structure, various attention mechanisms include MEA [35], DAT [36], SNA [37], TA [38], LA [39], SimAM [40], SE [41], and the detailed comparison is shown in Table 7.

According to the table, the SPPF-LSKA structure composed of the LSKA attention mechanism introduced in this study has better performance on the two indicators of mAP@0.5 and mAP@0.5-0.95. According to the comprehensive judgment, the SPPF-LSKA structure is more suitable for constructing the behavior recognition model of pigs.

TABLE 7. SPPF integrates different attention mechanisms.

Framework	P /%	mAP@0.5 /%	mAP@0.5-0.95 /%
SPPF-EMA	93.36	93.76	70.95
SPPF-DAT	91.18	91.74	71.35
SPPF-SNA	94.49	93.53	71.57
SPPF-TA	92.97	93.90	73.05
SPPF-LA	92.48	93.78	72.80
SPPF-SimAM	91.60	94.19	73.66
SPPF-SE	91.59	93.59	73.40
SPPF-LSKA	93.33	**94.34**	**73.72**

TABLE 8. Different downsampling algorithms.

Arithmetic	P /%	R /%	mAP@0.5 /%	mAP@0.5-0.95 /%
ContextGuidedDown	90.02	88.32	93.04	72.77
SPDConv	92.06	89.55	93.41	71.64
v7DS	91.40	84.44	90.85	68.32
Adown	90.54	88.78	92.30	69.37
Conv	93.33	89.81	94.34	73.72
HWD	93.02	**93.41**	**94.68**	**74.31**

5) OMPARISON OF DOWNSAMPLING ALGORITHMS

In this section, a comparison table of different downsampling algorithms in Table 8 is constructed by substituting downsampling algorithms at the same location. In the table, ContextGuidedDown algorithm is the downsampling using Light-weight Context Guided DownSample in CGNet [42], SPDConv [43] consists of a Space to Depth (SPD) layer and a non-spanning convolutional (Conv) layer. v7DS is the downsampling algorithm of YOLOv7 [44] and Addown is the downsampling algorithm of YOLOv9 [45] conv is the downsampling convolution of YOLOv8n.

In the table, the accuracy of P and YOLOv8 n of the down-sampling HWD structure is not much different, while

TABLE 9. Loss function comparison.

IoU	P /%	mAP@0.5 /%	mAP@0.5-0.95 /%
DIoU	93.06	93.74	72.17
GIoU	90.64	93.47	72.84
EIoU	94.41	93.03	73.28
CIoU	93.02	94.35	73.31
SIoU	91.29	93.95	74.36
PIoU-v1	90.72	94.42	73.60
PIoU-v2	92.15	93.82	72.61
WIoU-v1	92.54	93.54	72.26
WIoU-v2	91.32	93.42	73.00
WIoU-v3	93.72	93.97	73.28
Focal-CIoU	88.94	92.75	72.03
Focal-SIoU	90.10	93.85	73.08
Focal-EIoU	93.02	92.68	71.65
Focal-GIoU	86.70	92.82	70.32
Focal-Shape-IoU	88.58	93.42	72.66
Shape-IoU	**95.13**	**94.76**	**75.52**

R, mAP@0.5, and mAP@0.5-0.95 are better than other down-sampling algorithms in the table, so the HWD structure is more suitable for the down-sampling of this study.

6) LOSS FUNCTION COMPARISON

To verify the efficiency of the loss function of the improved pig behavior recognition model, the v1 and v2 versions of DIoU [26], GIoU [46], EIoU [47], CIoU, SIoU [48], PIoU [49], the v1, v2 and v3 versions of WIoU [50], and various combinations of IoU with Focal [51]deformation are compared through experiments. The detailed values of different loss functions are shown in Table 9.

The experimental results show that the three indexes of P, mAP@0.5, and mAP@0.5-0.95 of Shape-IoU are better than other loss functions. Therefore, the introduction of Shape-IoU in the model can improve the accuracy of pig behavior recognition.

7) COMPARISON OF PRUNING ALGORITHMS

This study also compares several model pruning methods, namely, Slim [52], Group_slim [53], Group_sl [54], L1 [55], Group_taylor [56], and Group_norm [54]. The comparison details are shown in Table 10. Through the comparison of the pruning algorithms, we can clearly see that the LAMP pruning scoring algorithm has higher values of all the indicators compared to the other algorithms, and therefore it can be concluded that the LAMP pruning scoring algorithm is more suitable for this study.

F. COMPARISON OF DIFFERENT RECOGNITION ALGORITHMS

1) QUANTITATIVE COMPARISON

In order to compare the level of behavioral recognition of the improved models in this paper, this section selects the

TABLE 10. Comparison of pruning algorithms.

Pruning algorithm	P /%	mAP@0.5 /%	mAP@0.5-0.95 /%
Slim	93.22	93.51	73.31
Group_slim	90.67	93.91	72.98
Group_sl	93.27	92.82	69.81
L1	86.48	92.60	73.27
Group_taylor	93.61	93.63	73.26
Group_norm	93.87	94.02	73.30
LAMP	**94.11**	**94.24**	**73.66**

current mainstream detection models to be compared under the same conditions: the Faster R-CNN [57], SSD [58], CenterNet [59], and YOLOv8n benchmark models. The detailed comparison is shown in Table 11.

As can be seen from Table 11, both models YOLO-DLHS and YOLO-DLHS-P proposed on the self-constructed dataset of this study have higher mAP@50 metrics compared to the mainstream algorithms Faster R-CNN, SSD, CenterNet, and YOLOv8n, with the YOLO-DLHS model having the highest mAP@50 metrics and the highest AP metrics for Siting, Lying, and Eating also have the highest AP metrics. The AP metrics of Standing and Mounting are only a little bit lower than the highest metrics, which is acceptable in the experiment. Although the introduced DRB and LSKA structures of the YOLO-DLHS model extend the sensory field of the model, HWD improves the model's sensitivity to details, and ShapeIOU optimizes the shape and scale matching of the bounding box, it is possible that these optimizations are insufficient in the complex situations of specific behaviors, and this is the reason why the AP metrics of Climbing are lower relative to the highest metrics. The YOLO-DLHS-P model is a lightweight model obtained by model pruning of the YOLO-DLHS model, which has only a little lower AP metrics and mAP@50 metrics for each behavior recognition, in exchange for a significant decrease in the number of parameters, computation, and model size, which are 98.96%, 99.00%, compared to the Faster-RCNN algorithm, 97.20%, 94.56%, 94.11%, and 94.75% compared to the SSD algorithm, and 95.62%, 94.71%, and 97.56% compared to the CenterNet algorithm, respectively, but the YOLO-DLHS-P model still has a higher mAP@50 metric than the above-compared algorithms. Moreover, the YOLO-DLHS-P model also improves the detection speed by the lightweight method, and the FPS reaches 79 frames, which is better compared with all the algorithms in Table 11, and has a better real-time performance for the behavioral recognition of pigs.

Therefore, the YOLO-DLHS model and the lightweight YOLO-DLHS-P model obtained in this study have the best-combined performance behavior of the above algorithms, and under the same conditions, the improved

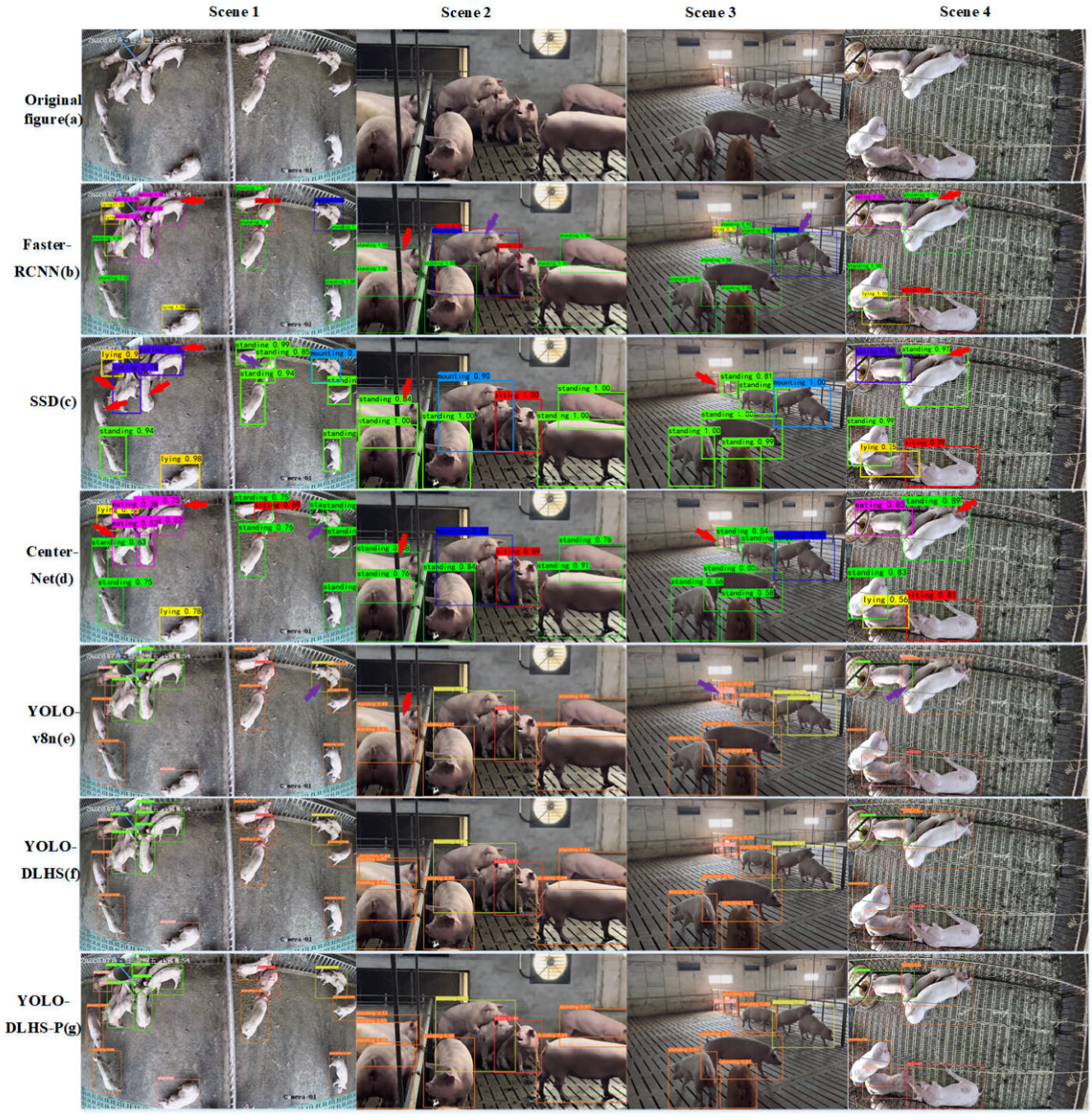


FIGURE 14. Algorithm visualization comparison chart.

algorithm in this study can achieve a higher recognition rate of behavioral recognition of pigs, and the lightweight model of the improved algorithm in this study substantially reduces the hardware requirements for the deployment of the model, under the conditions of maintaining a high level of accuracy. Requirements for model deployment while maintaining high accuracy. In particular, the YOLO-DLHS-P model also demonstrates better real-time performance on pig behavior, which facilitates behavioral recognition results after deployment.

2) QUALITATIVE COMPARISON

A visual comparative analysis of the models can visualize the accuracy of the algorithm recognition. In this section, a comparison is made with an actual scenario of captive hogs. As shown in Figure 14. Scene 1 is the overhead view of the large circle of pig farming, Scene 2 is the horizontal shooting view of the large circle of pig farming, Scene 3 is the written side shooting view of the large circle of pig farming, and Scene 4 is the overhead view of the small circle of pig farming. Among them, the red arrow in the figure

TABLE 11. Performance comparison of different recognition models.

Mould	*AP*/%						mAP@50	Param	Computation	Model size	FPS
	Siting	Lying	Standing	Climbing	Mounting	Eating	/%	/M	/GFLOPS	/MB	/frame.s^{-1}
Faster-RCNN	88.37	96.37	97.54	60.00	94.12	98.35	89.12	137.10	370.41	108.00	20.41
SSD	78.15	89.26	92.23	98.89	81.46	94.93	89.15	26.29	62.80	93.10	67.78
CenterNet	89.19	94.22	94.86	99.12	81.79	96.18	92.56	32.67	69.98	124.00	39.65
YOLOv8n	87.18	95.98	96.63	88.83	91.51	98.38	93.08	3.01	8.10	5.98	70.93
YOLO-DLHS	**91.36**	**96.64**	97.34	90.75	93.79	**98.67**	**94.76**	3.00	7.60	5.96	65.88
YOLO-DLHS-P	90.19	96.35	97.42	90.10	92.79	98.58	94.24	**1.43**	**3.70**	**3.03**	**79.15**

represents missed detection, and the purple arrow represents misdetection.

The comparison shows that in Scene 1, Faster R-CNN, SSD, and CenterNet all have different degrees of missed detection, and both CenterNet and YOLOv8n have misdetections when it comes to the crawling spanning behavior of the pigs. CenterNet identifies the two crawling spanning pigs as standing, and YOLOv8n has one more detection of the standing behavior of the frame. In Scene 2, Faster R-CNN, SSD, CenterNet, and YOLOv8n all miss-detect the pig on the left side of the scene. In Scene 3, Faster R-CNN misdetected the piglet with crawling straddling behavior and adds the behavioral frame of standing to the hind pig with crawling straddling behavior, SSD and CenterNet miss the detection of the piglet with lying behavior occurring in the distance, and YOLOv8n duplicates the behavioral information frames of the in piglet with standing and lying behaviors occurring in the distance, although it is identified. In Scene 4, when there is a small pig and a large pig close to each other, Faster R-CNN, SSD, and CenterNet, all missed detection, and the YOLOv8n algorithm misdetected and recognized the small pig as lying down. In the four scenarios mentioned above, neither the YOLO-DLHS model nor the YOLO-DLHS-P model showed leakage or misdetection, which further verified the effectiveness of the research-improved model and the improved lightweight model in pig behavior recognition.

V. CONCLUSION

In this study, by improving the C2f at the Backbone location and introducing the inflated reparameterization block, the BN layer can be merged into the convolutional layer by using the convolution of the large kernel together with the parallel convolution of the small kernel and then summing up the corresponding BN layers after the output, and then using the reparameterization technique. The inflated convolutional layer with a small kernel can be equated to the non-inflated convolutional layer with a sparse large kernel, reducing the computational overhead. The C2f-DRB structure obtained after the improvement is able to capture the multi-scale feature information of pig behavior, which improves the recognition accuracy. The emergence of the SPPF-LSKA structure maintains the model's ability to capture key visual features of the pig by replacing the traditional 2D convolutional kernel by cascading 1D convolutional kernels in the LSKA attention level mechanism. The accuracy of the model is further improved without increasing the arithmetic power too much. The traditional down-sampling convolution of the benchmark model is replaced by the HWD structure. By combining the Haar wavelet transform, the original image is effectively decomposed into four components with spatial resolution halved to capture the information of the image's low-frequency and high-frequency multiangles, respectively, and the completeness of the information is maintained by increasing the number of channels of the feature mapping, which provides the model with information-rich inputs with moderate resolution. The subsequent feature extraction is enhanced by down-sampling the HWD structure. In order to avoid the deficiency of the benchmark model in dealing with the direction of mismatch between the predicted and real frames when actually dealing with the task of behavior recognition of group-housed hogs. This leads to the prediction frames shifting in incorrect directions during the training process, which leads to slow convergence and inefficiency and ultimately affects the model performance problem. Shape-iou is introduced in this study, which is enough to compute the loss by focusing on the shapes of the frames themselves with their own scales, which makes the regression of the frames more accurate. Without adding extra computational burden, the detection efficiency and accuracy of the model are significantly improved. The improved YOLO-DLHS model is finally constructed. Compared with the baseline model, the improved YOLO-DLHS model has P, mAP@0.5 and mAP@0.5-0.95 increased by 4.39%, 1.68% and 3.97% respectively.

In the context of actual deployment and the demand for lightweight models in embedded devices. This study further proposes a lightweight YOLO-DLHS-P model based on the improved YOLO-DLHS model to address the common problems of large number of parameters, high computational cost and high memory usage in pig behavior recognition models. Using the LAMP pruning scoring algorithm, the improved YOLO-DLHS model was pruned to delete unimportant channels in the model, thereby significantly reducing the number of model parameters, calculations, and memory size. Through fine-tuning and retraining, we obtained The YOLO-DLHS-P model was developed. Compared with the baseline model, the YOLO-DLHS-P model has P, mAP@0.5 and mAP@0.5-0.95 increased by 3.37%, 1.16% and 2.11% respectively. The number of parameters, calculation amount and model occupancy have been significantly reduced, respectively, by 52.49%, 54.32%, 49.67%. And the FPS index of the YOLO-DLHS-P model reaches 79 frames,

which is higher than the YOLO-DLHS model, Faster R-CNN, SSD, CenterNet and YOLOv8n benchmark models, while the YOLO-DLHS-P model only has YOLO- The DLHS model has low accuracy with subtle differences. Therefore, compared with the YOLOv8n model, the YOLO-DLHS-P model has higher accuracy, better real-time performance, and is more conducive to actual deployment.

The improved model in this study achieves improvement in all performance indexes while significantly reducing the number of parameters, model occupancy, and computational requirements, reflecting the superiority of the improved algorithm in this study. It shows the prospect of application on embedded devices in the behavioral recognition scenarios of captive hogs, which is conducive to the development of smart farming. However, this study also has certain limitations. First of all, although the improved model performs well, its robustness in different actual environments still needs further verification. In addition, although the improvement and lightweight processing of the model are effective, there may still be room for optimization in processing complex scenes and long-term continuous monitoring. In the future, we will explore model performance in more complex environments, improve its robustness and prediction, and combine it with other deep learning technologies and optimization methods to further reduce the total cost of model calculations and improve real-time processing capabilities.

REFERENCES

[1] W. Hao, W. Han, M. Han, and F. Li, "A novel improved YOLOv3-SC model for individual pig detection," *Sensors*, vol. 22, no. 22, p. 8792, Nov. 2022.

[2] M. Marcon, L. Brossard, and N. Quiniou, "Precision feeding based on individual daily body weight of group-housed pigs with an automatic feeder developed to allow for restricting feed allowance," *Precis. Livestock Farming*, vol. 15, no. 592, p. e601, 2015.

[3] D.-N. Tran, T. N. Nguyen, P. C. P. Khanh, and D.-T. Tran, "An IoT-based design using accelerometers in animal behavior recognition systems," *IEEE Sensors J.*, vol. 22, no. 18, pp. 17515–17528, Sep. 2022.

[4] S. Hou, T. Wang, D. Qiao, D. J. Xu, Y. Wang, X. Feng, W. A. Khan, and J. Ruan, "Temporal–spatial fuzzy deep neural network for the grazing behavior recognition of herded sheep in triaxial accelerometer cyber-physical systems," *IEEE Trans. Fuzzy Syst.*, early access, May 8, 2024, doi: 10.1109/TFUZZ.2024.3398075.

[5] A. Nasirahmadi, S. A. Edwards, and B. Sturm, "Implementation of machine vision for detecting behaviour of cattle and pigs," *Livestock Sci.*, vol. 202, pp. 25–38, Aug. 2017.

[6] D. Mellor, "Updating animal welfare thinking: Moving beyond the 'five freedoms' towards 'a life worth Living,'" *Animals*, vol. 6, no. 3, p. 21, Mar. 2016.

[7] C. Munsterhjelm, M. Heinonen, and A. Valros, "Effects of clinical lameness and tail biting lesions on voluntary feed intake in growing pigs," *Livestock Sci.*, vol. 181, pp. 210–219, Nov. 2015.

[8] B. Krsnik, R. Yammine, Ž. Pavič ić, T. Balenović, B. Njari, I. Vrbanac, and I. Valpotić, "Experimental model of enterotoxigenic *Escherichia coli* infection in pigs: Potential for an early recognition of colibacillosis by monitoring of behavior," *Comparative Immunol., Microbiol. Infectious Diseases*, vol. 22, no. 4, pp. 261–273, Oct. 1999.

[9] L. Rydhmer, G. Zamaratskaia, H. K. Andersson, B. Algers, R. Guillemet, and K. Lundström, "Aggressive and sexual behaviour of growing and finishing pigs reared in groups, without castration," *Acta Agriculturae Scandinavica, A-Animal Sci.*, vol. 56, no. 2, pp. 109–119, Jun. 2006.

[10] M. Kashiha, C. Bahr, S. A. Haredasht, S. Ott, C. P. H. Moons, T. A. Niewold, F. O. Ödberg, and D. Berckmans, "The automatic monitoring of pigs water use by cameras," *Comput. Electron. Agricult.*, vol. 90, pp. 164–169, Jan. 2013.

[11] Q. Yang, D. Xiao, and S. Lin, "Feeding behavior recognition for group-housed pigs with the faster R-CNN," *Comput. Electron. Agricult.*, vol. 155, pp. 453–460, Dec. 2018.

[12] A. Nasirahmadi, B. Sturm, A.-C. Olsson, K.-H. Jeppsson, S. Müller, S. Edwards, and O. Hensel, "Automatic scoring of lateral and sternal lying posture in grouped pigs using image processing and support vector machine," *Comput. Electron. Agricult.*, vol. 156, pp. 475–481, Jan. 2019.

[13] A. Nasirahmadi, O. Hensel, S. A. Edwards, and B. Sturm, "Automatic detection of mounting behaviours among pigs using image analysis," *Comput. Electron. Agricult.*, vol. 124, pp. 295–302, Jun. 2016.

[14] R. Wang, R. Gao, Q. Li, C. Zhao, W. Ma, L. Yu, and L. Ding, "A lightweight cow mounting behavior recognition system based on improved YOLOv5s," *Sci. Rep.*, vol. 13, no. 1, p. 17418, Oct. 2023.

[15] C. Shang, F. Wu, M. Wang, and Q. Gao, "Cattle behavior recognition based on feature fusion under a dual attention mechanism," *J. Vis. Commun. Image Represent.*, vol. 85, May 2022, Art. no. 103524.

[16] F. Lao, T. Brown-Brandl, J. P. Stinn, K. Liu, G. Teng, and H. Xin, "Automatic recognition of lactating sow behaviors through depth image processing," *Comput. Electron. Agricult.*, vol. 125, pp. 56–62, Jul. 2016.

[17] C. Zheng, X. Zhu, X. Yang, L. Wang, S. Tu, and Y. Xue, "Automatic recognition of lactating sow postures from depth images by deep learning detector," *Comput. Electron. Agricult.*, vol. 147, pp. 51–63, Apr. 2018.

[18] D. Li, K. Zhang, Z. Li, and Y. Chen, "A spatiotemporal convolutional network for multi-behavior recognition of pigs," *Sensors*, vol. 20, no. 8, p. 2381, Apr. 2020.

[19] Z. Gu, H. Zhang, Z. He, and K. Niu, "A two-stage recognition method based on deep learning for sheep behavior," *Comput. Electron. Agricult.*, vol. 212, Sep. 2023, Art. no. 108143.

[20] C. Feng, Y. Zhong, Y. Gao, M. R. Scott, and W. Huang, "TOOD: Task-aligned one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3490–3499.

[21] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21002–21012.

[22] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," 2016, *arXiv:1603.08029*.

[23] X. Ding, Y. Zhang, Y. Ge, S. Zhao, L. Song, X. Yue, and Y. Shan, "UniRepLKNet: A universal perception large-kernel ConvNet for audio, video, point cloud, time-series and image recognition," 2023, *arXiv:2311.15599*.

[24] K. W. Lau, L.-M. Po, and Y. A. U. Rehman, "Large separable kernel attention: Rethinking the large kernel attention design in CNN," *Expert Syst. Appl.*, vol. 236, Feb. 2024, Art. no. 121352.

[25] G. Xu, W. Liao, X. Zhang, C. Li, X. He, and X. Wu, "Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation," *Pattern Recognit.*, vol. 143, Nov. 2023, Art. no. 109819.

[26] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12993–13000.

[27] H. Zhang and S. Zhang, "Shape-IoU: More accurate metric considering bounding box shape and scale," 2023, *arXiv:2312.17663*.

[28] J. Lee, S. Park, S. Mo, S. Ahn, and J. Shin, "Layer-adaptive sparsity for the magnitude-based pruning," 2020, *arXiv:2010.07611*.

[29] H. Wei, X. Liu, S. Xu, Z. Dai, Y. Dai, and X. Xu, "DWRSeg: Rethinking efficient acquisition of multi-scale contextual information for real-time semantic segmentation," 2022, *arXiv:2212.01173*.

[30] X. Ding, X. Zhang, J. Han, and G. Ding, "Diverse branch block: Building a convolution as an inception-like unit," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10881–10890.

[31] J. Zhang, X. Li, J. Li, L. Liu, Z. Xue, B. Zhang, Z. Jiang, T. Huang, Y. Wang, and C. Wang, "Rethinking mobile block for efficient attention-based models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 1389–1400.

[32] J. Chen, S.-H. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H.-G. Chan, "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 12021–12031.

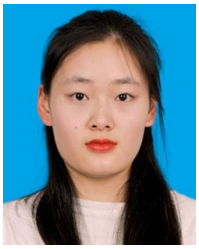
[33] M. Hu, J. Feng, J. Hua, B. Lai, J. Huang, X. Gong, and X. Hua, "Online convolutional reparameterization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 558–567.

[34] X. Zhang, Y. Song, T. Song, D. Yang, Y. Ye, J. Zhou, and L. Zhang, "AKConv: Convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters," 2023, *arXiv:2311.11587*.
[35] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
[36] Z. Xia, X. Pan, S. Song, L. E. Li, and G. Huang, "Vision transformer with deformable attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4784–4793.
[37] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "SegNext: Rethinking convolutional attention design for semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 1140–1156.
[38] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. Hou, "Rotate to attend: Convolutional triplet attention module," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3138–3147.
[39] Y. Bondarenko, M. Nagel, and T. Blankevoort, "Quantizable transformers: Removing outliers by helping attention heads do nothing," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 75067–75096.
[40] L. Yang, R. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. 38th Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 11863–11874.
[41] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[42] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "CGNet: A light-weight context guided network for semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 1169–1179, 2021.
[43] R. Sunkara and T. Luo, "No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, Sep. 2022, pp. 443–459.
[44] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.
[45] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.
[46] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.
[47] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022.
[48] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," 2022, *arXiv:2205.12740*.
[49] Z. Chen, K. Chen, W. Lin, J. See, H. Yu, and Y. Ke, "PIoU loss: Towards accurate oriented object detection in complex environments," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K. Cham, Switzerland: Springer, 2020, pp. 195–211.
[50] Z. Tong, Y. Chen, Z. Xu, and R. Yu, "Wise-IoU: Bounding box regression loss with dynamic focusing mechanism," 2023, *arXiv:2301.10051*.
[51] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
[52] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, "Learning efficient convolutional networks through network slimming," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2755–2763.
[53] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group lasso and a sparse group lasso," 2010, *arXiv:1001.0736*.
[54] G. Fang, X. Ma, M. Song, M. Bi Mi, and X. Wang, "DepGraph: Towards any structural pruning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16091–16101.
[55] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. Peter Graf, "Pruning filters for efficient ConvNets," 2016, *arXiv:1608.08710*.
[56] P. Molchanov, A. Mallya, S. Tyree, I. Frosio, and J. Kautz, "Importance estimation for neural network pruning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11256–11264.
[57] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2016.
[58] W. Liu, D. Anguelov, D. Erhan, and C. Szegedy, "SSD: Single shot MultiBox detector," in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Oct. 2016, pp. 21–37.
[59] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.

CHANGHUA ZHONG received the B.S. degree from Binjiang College, Nanjing University of Information Science and Technology, Nanjing, China, in 2021. He is currently pursuing the M.S. degree with Sichuan University of Science and Engineering, Yibin, China. His research interest includes image processing.

HAO WU (Member, IEEE) received the B.S. degree in electrical engineering and automation from Southwest Jiaotong University, Chengdu, China, in 2003, the M.S. degree in pattern recognition and intelligent systems from Sichuan University of Science and Engineering, and the Ph.D. degree in power systems and automation from Southwest Jiaotong University. His research interests include big data, artificial intelligence, deep learning, image processing, modern signal processing and artificial intelligence technology in power systems, transmission (distribution) grid fault diagnosis and fault positioning technology, intelligent distribution grid protection, control and microgrid technology, and conditional intelligent monitoring technology for electrical equipment.

JUNZHUO JIANG received the B.S. degree from Chengdu Technological University, Chengdu, China, in 2022. He is currently pursuing the M.S. degree with Sichuan University of Science and Engineering, Yibin, China. His research interests include signal processing and power quality disturbances.



CHAOWEN ZHENG received the B.S. degree from Sichuan University of Science and Engineering, Yibin, China, in 2022, where she is currently pursuing the M.S. degree. Her research interests include fault diagnosis and flexible dc distribution networks.

HONG SONG is currently a Professor and a Reserve Candidate for the Academic Leader in Sichuan. His research interests include big data, artificial intelligence and deep learning, image processing, modern signal processing and artificial intelligence technology in power systems, transmission (distribution) grid fault diagnosis and fault positioning technology, intelligent distribution grid protection, control and microgrid technology, and conditional intelligent monitoring technology for electrical equipment.