IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 March 2024, accepted 22 March 2024, date of publication 27 March 2024, date of current version 4 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3382193

RESEARCH ARTICLE

Detection of Leg Diseases in Broiler Chickens Based on Improved YOLOv8 X-Ray Images

XIN ZHANG[1,*], RENWEN ZHU[1,*], WEIGANG ZHENG[1], AND CHANGXI CHEN[1,2]

[1]School of Computer and Information Engineering, Tianjin Agricultural University, Tianjin 300384, China
[2]Key Laboratory of Smart Breeding (Co-Construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Tianjin Agricultural University, Tianjin 300392, China

Corresponding author: Changxi Chen (changxichen@163.com)

This work was supported by the Open Fund of Key Laboratory of Smart Breeding (Co-Construction by Ministry and Province), Ministry of Agriculture and Rural Affairs. The work of Changxi Chen was supported in part by China Agriculture Research System of MOF and MARA under Grant CARS-41, and in part by the National Key Research and Development Program of China under Grant 2022YFD1601900 and Grant 2023YFD2000800.

*Xin Zhang and Renwen Zhu are co-first authors.

ABSTRACT With the excessive selection for body weight (BW) and breast muscle weight (BMW) in broiler chickens, the incidence of leg diseases has gradually increased, which may lead to severe mortality, decreased productivity, and growth restrictions. Traditional methods for detecting leg diseases heavily rely on the interpretation of X-ray images by professionals and scoring methods for chicken gait. However, X-ray images of broiler chicken legs suffer from low background contrast and small, blurry lesion areas, posing significant challenges for traditional target detection methods. This paper proposes an improved algorithm based on the latest YOLOv8 for detecting leg diseases in X-ray images of chicken legs. In the feature extraction phase, Partial Convolution (PConv) is introduced to the C2f module, effectively reducing computational complexity while more accurately extracting spatial features. By incorporating Channel Prior Convolutional Attention (CPCA) into the network backbone, dynamic allocation of attention weights in both channel and spatial dimensions is achieved, preventing the loss of feature details caused by convolution iterations and enhancing the representation capability of small object features. The feature fusion stage introduces a novel Gather-Distribute mechanism (GD), effectively improving the inter-layer information exchange. Additionally, a Partial Convolution-based Shared Weight Detection Head (SharedPConv head) is introduced in the network head, making the model more lightweight and effectively alleviating the overfitting issue. Experimental results demonstrate that the improved method achieves a 7.2% increase in average precision, with a speed of 66.8fps, meeting real-time requirements and performing the detection task more accurately.

INDEX TERMS X-ray, Yolov8, broiler, joint effusion, Tibial Dyschondroplasia.

I. INTRODUCTION

Poultry farming plays a crucial role in meeting global meat demands, with broilers being one of the most commonly raised poultry. However, the prevalence of leg diseases in broilers poses significant challenges to the industry. It is estimated that over 27.6% of chickens exhibit poor mobility, with 3.3% almost unable to walk by the age of 40 days [1]. Internationally, 14% to 30% of broilers are estimated to suffer from lameness, affecting their movement speed and overall health, directly attributed to gait abnormalities [2]. Broiler leg diseases can stem from various factors, including genetic predisposition, rapid growth rates, improper nutrition, and environmental factors. Common leg diseases include physical distortion, Tibial Dyschondroplasia (TD), and Joint Effusion(JE). TD clinically manifests as reduced feeding, gait abnormalities, and movement disorders [3]. These conditions result in significant economic losses and welfare issues in the poultry farming industry, potentially leading to increased mortality in severe cases [4], [5]. Traditionally, the detection of leg diseases in broilers relied on manual clinical examinations and subjective assessments, such as gait scoring methods [6]. These methods are both time-consuming and labor-intensive and are prone to human errors. Therefore,

The associate editor coordinating the review of this manuscript and approving it for publication was Abdel-Hamid Soliman.

there is an urgent need for an automated detection system that provides an objective and reliable assessment of leg health.

In recent years, object detection has demonstrated outstanding performance in various image recognition tasks, bringing accurate and efficient disease detection to the field of medical imaging. Currently, convolutional neural networks (CNNs) have been successfully applied in areas such as brain tumor segmentation [7], lung nodule detection [8], and brain image analysis [9], making significant contributions to the advancement of medical imaging. Deep learning methods play a crucial role in the field of radiology. Despite the remarkable achievements of deep learning in medical image recognition, there are still challenges, particularly in the case of X-ray images of diseases in chicken thighs. Challenges include poor image quality, small and blurry lesion areas, and large aspect ratios. Additionally, in practical applications, handling a large batch of X-ray images simultaneously poses challenges in terms of model complexity and computational load. Moreover, the scarcity and imbalance of target samples lead to issues such as insufficient training data and overfitting during the training process, resulting in poor generalization ability. To tackle these challenges, researchers have proposed several improvement measures. For small object detection, techniques such as feature map scaling and fusion can be employed to prevent the loss of small objects and enhance feature representation [10]. The combination of shallow-level positional information with deep-level semantic information improves feature fusion capabilities [11].

The YOLO series algorithms have been widely applied in object detection due to their accuracy, lightweight design, and scalability [12]. Researchers aim to address the contradiction between semantic feature extraction and target scale and enhance feature extraction and classification capabilities by adopting multi-feature fusion techniques [13]. To further improve the performance of the YOLO algorithm, an anchor-free mechanism can be employed to address the difficulties of positive and negative sample imbalance and hyperparameter tuning [14]. Enhancing the feature learning of the backbone network helps extract more discriminative features [15]. Cross-scale neck fusion attention emphasizes object-specific features [16], while Feature Pyramid Networks improve information flow and sharing across layers [17]. In addition, lightweight convolutional module design, network pruning, and knowledge distillation techniques can accelerate inference speed and improve computational efficiency [18].

The purpose of this study is to prune and optimize the YOLOv8 model to improve its detection capability for small objects while meeting practical requirements. To achieve this goal, we propose the following key modifications.

- The leg diseases to be detected in this study include joint effusion and tibial dyschondroplasia, as illustrated in Figure 1. Specifically, tibial dyschondroplasia is characterized by the overlap of tibial cartilage at the tibiotarsal joint, while joint effusion appears as a semi-transparent grey shadow around the joint. Both diseases exhibit characteristics of small targets in large X-ray images. YOLOv8 is employed as the baseline network, and the Channel Prior Convolutional Attention (CPCA) is introduced into the backbone to dynamically distribute attention weights in both channel and spatial dimensions. This method proves particularly effective in handling low contrast and pronounced organ shape variations in medical images. Additionally, a novel collection-distribution mechanism (GD) called gold-yolo is introduced into the network's neck to enhance partial information fusion capabilities.
- To address the challenges of limited X-ray images and unclear manifestation of the lesion areas for joint effusion and developmental dysplasia of the cartilage, the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm is employed as a data augmentation technique to improve image quality.
- We replace specific C2f layers in the backbone network with C2f_PConv modules. The C2f_PConv module incorporates partial convolution (PConv). Furthermore, to achieve a more lightweight model, we introduce a weight-sharing detection head based on partial convolution (PConv) into the network's head. Through optimization, the proposed lightweight detection head reduces parameters by 30% compared to the original YOLOv8 detection head. The design of these two lightweight convolution modules reduces model complexity and enhances detection efficiency.

In the following sections, we will provide a detailed description of our methodology, including data collection and preprocessing, implementation of the deep learning model, and evaluation of the automated detection system. We will present our experimental results and discuss the impact and potential applications of our proposed approach in the field of broiler health management.

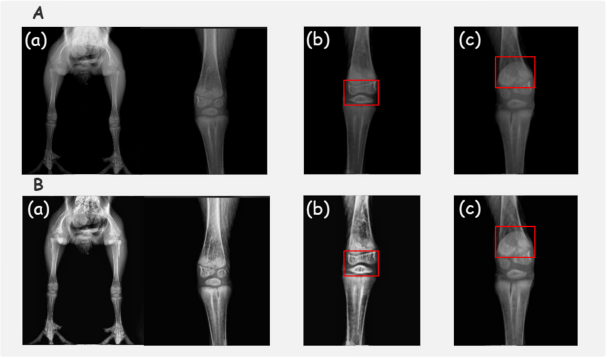


FIGURE 1. Using Clahe data augmentation to compare with the original image, A is the original image, where (a) represents the X-ray image of a broiler chicken with normal leg bones, no joint effusion, and no tibial cartilage dysplasia. (b) Chondrodysplasia. (c) Joint effusion.

II. RELATED WORK

In the field of X-ray object detection, the fundamental aspects of object detection algorithms primarily revolve around three key components: feature extraction, feature fusion, and the integration of various feature processing techniques. These elements collaborate to enhance the capabilities of feature extraction and classification.

A. MULTI-SCALE FEATURES FOR OBJECT DETECTION

Traditionally, features at different levels carry information about the positions of objects of varying sizes. Larger features encompass low-dimensional texture details and the positions of smaller objects, while smaller features contain high-dimensional information and the positions of larger objects. The original concept behind Feature Pyramid Networks (FPN), proposed by Lin et al. [19], is that these distinct information fragments can mutually enhance the performance of the network. FPN, through scale-aware connections and information exchange, offers an efficient architectural design for fusing multi-scale features, thereby improving the detection accuracy of objects of different sizes.

The Feature Pyramid Network (FPN) designed by Wang et al., named Path Aggregation Network (PANet) [20], adopts a bottom-up pathway to facilitate more comprehensive information fusion across different levels. Similarly, EfficientDet [21] introduces a novel repeatable module, BiFPN, to enhance the efficiency of information fusion across different levels. Additionally, Yang et al [22]. extend FPN with an Asymptotic Feature Pyramid Network (AFPN), enabling interactions between non-adjacent layers. Addressing the limitations of FPN in detecting large objects, SFNet [23] aligns features at different levels with the semantic flow to enhance FPN performance in the model.

However, due to the excessive number of paths and indirect interaction methods in the network, the previous FPN-based fusion structures still have drawbacks in low speed, cross-level information exchange, and information loss.

B. ATTENTION MECHANISM

The attention mechanism plays a crucial role in medical image detection and computer vision. Attention mechanisms typically consist of channel attention and spatial attention. What to focus on is determined by channel attention. Channel attention was pioneered by Hu et al. (2018), who introduced SENet. The core of SENet is the Squeeze-and-Excitation (SE) block, which involves two key steps: global average pooling and channel-wise weighting computation. Through this block, the network can focus on important features. However, SENet only includes channel attention, limiting its ability to selectively focus on important regions. Therefore, Woo et al. proposed the Convolutional Block Attention Module (CBAM) [24], which sequentially combines channel attention and spatial attention. CBAM integrates channel attention and spatial attention in sequence, allowing the network to focus on informative channels and important regions. However, the spatial attention map is computed by compressing channels, resulting in consistent spatial attention weight distribution for each channel when element-wise multiplication is performed with input features. This limitation restricts the adaptive ability of attention since spatial attention weights cannot dynamically adjust based on specific features for each channel.

Addressing this issue, Huang et al. proposed a novel method called Channel Prioritized Convolution Attention (CPCA) [25]. In contrast to CBAM, CPCA employs a depth-wise convolution module as the spatial attention component. The depth-wise convolution module uses bar-shaped convolution kernels of different scales to extract spatial mapping relationships between pixels. The use of multi-scale depth-wise bar-shaped convolution kernels ensures effective information extraction while reducing computational complexity [26]. The channel attention module is initially used to obtain the channel attention map. Subsequently, the depth-wise convolution module sequentially extracts key spatial regions for each channel, obtaining dynamically distributed spatial attention maps for each channel. These dynamically distributed spatial attention maps closely approximate the actual feature distribution for each channel, effectively enhancing network performance.

C. COMPARISON OF METHODS BASED ON REGION AND REGRESSION

Object detection methods are generally divided into two categories: region-based methods and regression-based methods. Region-based methods involve searching for possible lesion areas in images and then running a classifier on these regions to identify targets. For example, Gan et al. [27] used Faster R-CNN and Inception-v4 to detect distal radius fractures, achieving an average Intersection over Union (IOU) value of 0.87, with IOU values not less than 0.5. Jia et al. [28] employed cascade R-CNN as a baseline model for detecting sternal fractures in chest X-ray images. They optimized the detection of small fractures in large X-ray images with significant local variations using attention mechanisms and atrous convolution. Experimental results showed a mean Average Precision (mAP) of 0.71, significantly better than using cascaded R-CNN (mAP = 0.55). Although two-stage object detection algorithms can achieve higher accuracy, they introduce substantial computational overhead and may not meet the real-time requirements of object detection in practice. In contrast, regression-based methods directly predict the position and category of targets. Currently, some researchers have started applying deep learning to object detection in X-ray images. For instance, Han et al. [29] combined a YOLO detector with an InceptionResNet-V2 classifier, achieving good diagnostic performance for breast lesions. The model detected and classified breast lesions in single breast X-ray images in less than 0.025 seconds on average. Wang et al. [30] proposed a YOLO-Xray model for bubble defect detection in chip X-ray images using the YOLOv5 algorithm. By using image segmentation

preprocessing and constructing a defect dataset, the model improved the detection capability of small defects. The YOLO-Xray algorithm achieved an average precision (mAP) of 93.5% on the CXray dataset, a 5.1% improvement over the original YOLOv5, demonstrating state-of-the-art detection accuracy and speed.

The above studies demonstrate the feasibility of deep learning in X-ray image detection. Currently, most deep learning research focuses on the human skeleton and lungs, with limited studies on diseases related to chicken legs. Due to the increasing prevalence of chicken leg diseases, poultry health and welfare are under serious threat. Therefore, there is an urgent need to develop deep learning models specifically for the detection of chicken leg diseases to assist farmers in rapidly and accurately diagnosing and locating diseases in broiler chicken legs.

III. MATERIALS AND METHODS

A. DATASET AND PRE-PROCESSING

1) ESTABLISHMENT OF A DATASET FOR LEG DISEASES IN BROILER CHICKENS

To successfully develop an accurate automatic detection system for chicken leg diseases, establishing a comprehensive and meticulously annotated dataset is crucial as the cornerstone for training and evaluating deep learning models. This study primarily focuses on two specific leg diseases: joint effusion and tibial dyschondroplasia. The construction process of this dataset involves the following steps:

Step 1:Data Collection: Using the HX-100P veterinary diagnostic portable X-ray equipment from Beijing Jiaxinhuixiong Technology Co., Ltd. (Beijing, China), chicken legs were captured. These chickens underwent clinical examinations, and a team of veterinary experts confirmed the presence of each leg disease. The dataset includes only chickens with explicitly diagnosed conditions to ensure the accuracy and reliability of the data.

Step 2:Annotation and Labeling Process: For joint effusion and tibial dyschondroplasia, bounding boxes were placed around the affected joints or regions with abnormal cartilage development. Annotations include disease-specific labels, and As shown in (b) and (c) of A in Figure 1 illustrates the lesion areas indicating the presence of joint effusion or tibial dyschondroplasia.

Step 3:Data Composition: The dataset comprises X-ray images of joint effusion and Tibial Dyschondroplasia. Table 1 summarize the distribution and composition of the dataset, providing detailed information on the number of images for each leg disease category.

By constructing a comprehensive dataset with accurate annotations, we ensure a robust foundation for the proposed automatic detection system. This dataset encompasses various leg diseases, offering effective learning and generalization opportunities for deep learning models. In the following sections, we will discuss data preprocessing techniques used to enhance the dataset's quality and prepare for training the detection model.

2) DATA PREPROCESSING

Data augmentation is a commonly employed technique in the fields of machine learning and computer vision, aimed at expanding the training dataset to improve the model's generalization capability and robustness. In this study, we incorporate the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [21] as a component of data preprocessing to enhance the model's ability to capture subtle details in lesion areas of chicken leg X-ray images.The CLAHE algorithm is based on the concept of histogram equalization, adapting it to local regions of an image to enhance contrast and mitigate uneven illumination. In contrast to traditional global histogram equalization methods, CLAHE demonstrates advantages when dealing with images featuring uneven grayscale distributions, avoiding issues of excessive enhancement while enhancing image details.

During the data preprocessing stage, original X-ray images of chicken legs were input into the CLAHE algorithm, which performed adaptive histogram equalization on local regions of the image. Specifically, we divided the image into multiple overlapping local blocks, applying the CLAHE algorithm calculation to each block to enhance image contrast and details.

By applying the CLAHE algorithm for data augmentation, significant improvements were observed in the task of disease detection in chicken leg X-ray images. As shown in Figure 1, CLAHE-processed images exhibited clearer and brighter features, facilitating the identification and analysis of chicken leg diseases such as joint effusion and tibial dyschondroplasia.

B. METHOD FOR CHICKEN LEG DISEASE DETECTION IN X-RAY IMAGES

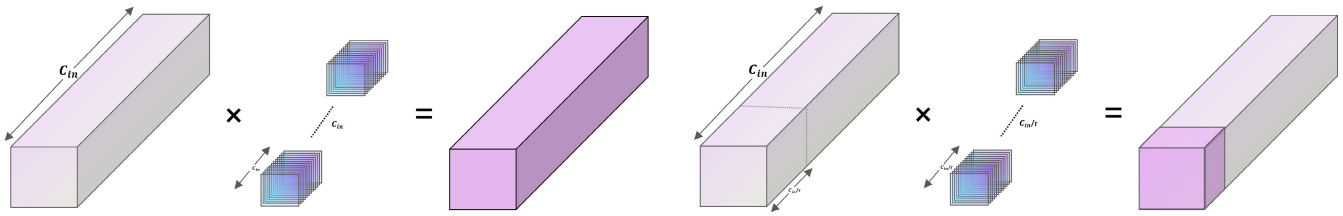
1) ALGORITHM SELECTION

YOLOv8, the latest version of the You Only Look Once (YOLO) series for object detection, has significantly improved its performance in the field through multiple refinements and optimizations. In comparison to earlier YOLO versions, YOLOv8 has achieved notable advancements in both accuracy and speed for object detection. Furthermore, it has demonstrated outstanding performance on the COCO dataset, exhibiting excellent average precision and lower rates of missed detections. Hence, we have chosen YOLOv8 as the baseline model for our research to ensure that our study attains a high level of performance.

The YOLOv8 network architecture is primarily divided into three modules: Backbone, Neck, and Head. In contrast to the previous YOLOv5, the Backbone of the network replaces all C3 modules with C2f modules, introducing more skip connections and additional Split operations to better capture object boundaries and contextual information. The Neck section continues the PAN-FPN structure but removes the CBS 1×1 convolution structure in the upsampling stage of the PAN-FPN from YOLOv5, replacing the C3 modules with C2f modules. The Head section transitions from the original

TABLE 1. Distribution and composition of a dataset on joint effusion and chondrodysplasia in the legs of broiler chickens.

Category	Number of image	After data enhancement	Train	Validation	Test
Tibial cartilage dysplasia	400	800	640	80	80
Joint effusion	400	800	640	80	80
Total	800	1600	1280	160	160



(a): Convolution (b): Partial Convolution

FIGURE 2. For a regular convolution operation (a) and a PConv (partial convolution) operation (b), it can be observed that PConv achieves lower FLOPs compared to the conventional convolution.

coupled head to a decoupled head and shifts from YOLOv5's Anchor-Based form to an Anchor-Free form.

2) THE C2F MODULE INTRODUCES PCONV CONVOLUTION

The original YOLOv8 model is relatively large, featuring a complex backbone network and detection head, which may result in higher computational complexity in resource-constrained environments. Particularly in practical applications, there is a demand for higher efficiency in the X-ray image detection of diseases in broiler chicken legs. The introduction of Partial Convolution (PConv) reduces redundant computations, more effectively extracting spatial features, and significantly decreases the model's parameter count and computational complexity [22]. This adaptation makes the model more suitable for embedded devices and environments with limited resources.

Figure 2(a) illustrates the working principle of conventional convolution and PConv. It can be observed that conventional convolution extracts features by considering all channels, resulting in a FLOPs (floating-point operations per second) calculated by the following formula.

$$H \times W \times K^2 \times C^2 \tag{1}$$

From Figure 2(b) it can be observed that PConv selectively applies conventional convolution to spatially extract features from only a subset of input channels, while leaving the rest unchanged. For consecutive or regular memory access, the first to last consecutive channels (denoted as C_P channels) are calculated as representatives of the entire feature map. Without compromising generality, the FLOPs of PConv are only

$$H \times W \times K^2 \times C_P^2 \tag{2}$$

When a typical fraction $\frac{C_P}{C} = \frac{1}{4}$, Besides, PConv has a smaller amount of memory access, i.e.,

$$H \times W \times 2C_P + K^2 \times C_P^2 \approx H \times W \times 2C_P \tag{3}$$

Which is only $\frac{1}{4}$ of a regular Conv for $r = \frac{1}{4}$ From the above analysis, it can be concluded that PConv significantly improves the model's processing speed and efficiently extracts subtle features from X-ray images.

As shown in Figure 3(c) the C2f module consists of two convolutional layers and n bottlenecks, including numerous skip connections and additional split operations. These operations facilitate the fusion of features extracted from different depth layers, enabling the model to simultaneously process low-level and high-level features. However, this characteristic also adds to the computational complexity of the network. During both training and inference, handling additional feature map merging and channel adjustments may result in increased computational and memory requirements.

In this paper, we replaced the first regular convolution in the bottleneck of the C2f module with PConv, and used two 1×1 convolutions to replace the second regular convolution, forming the C2f_PConv module. However, it is important to note that replacing all C2f modules with C2f_PConv modules in the YOLOv8 network structure may potentially lead to a decrease in network accuracy. To address these challenges, we conducted experiments to validate detection accuracy and inference speed. Through this process, we found that replacing the second and third C2f modules and the fourth C2f module in the backbone network with C2f_PConv modules achieved optimal optimization results. This modification preserves task-relevant information for extracting critical features related to broiler chicken joint effusion and tibial dyschondroplasia, while disregarding less crucial details. Figure 3 illustrates the improved overall network architecture of YOLOv8 designed in this study,

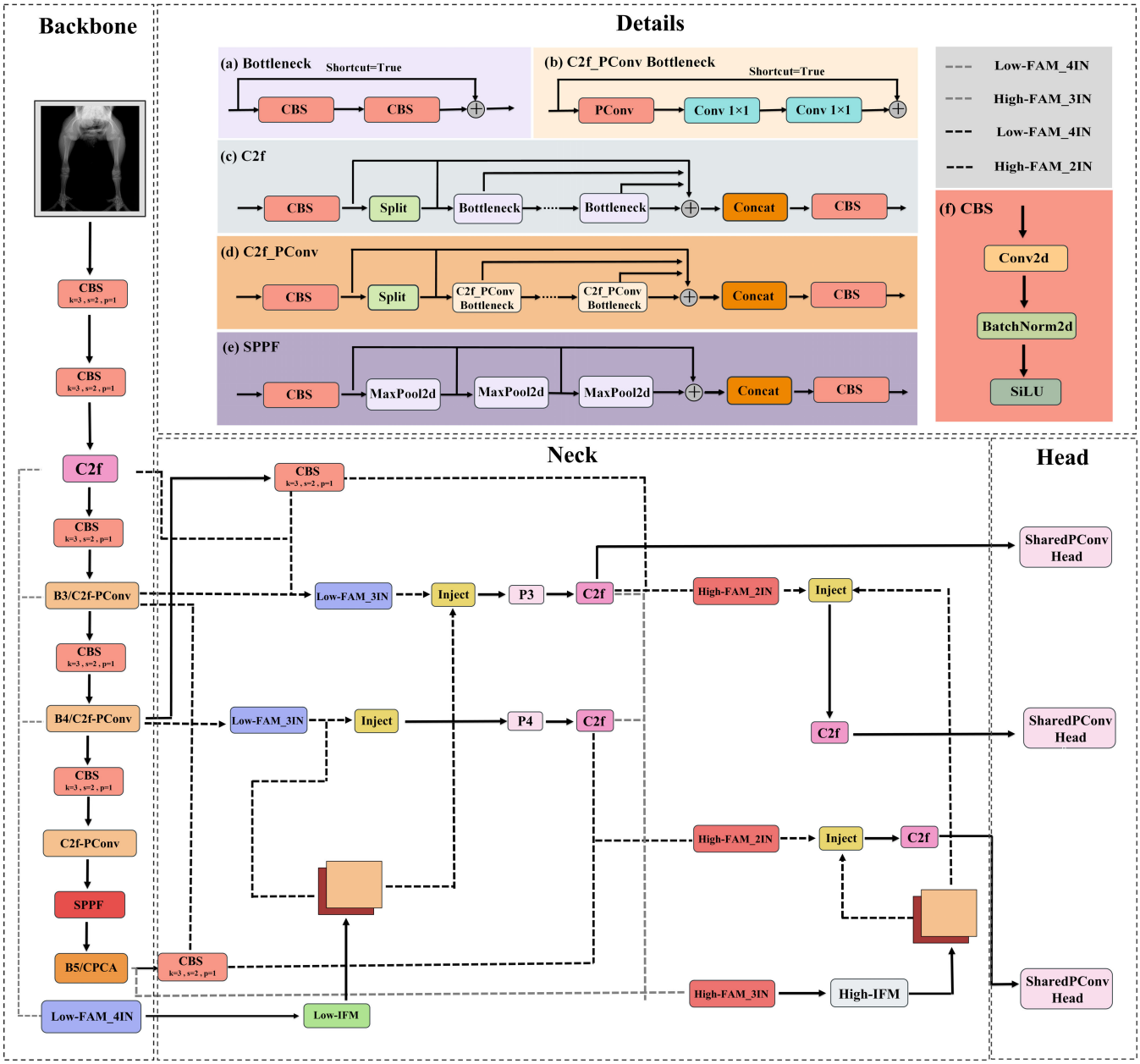


FIGURE 3. Improved YOLO v8s structure schematic.

providing detailed information on both the C2f module and C2f_PConv module.

3) PARTIAL CONVOLUTION-BASED SHARED WEIGHT DETECTION HEAD (SHAREDPCONV HEAD)

The detection head of YOLOv8 inherits the decoupling design from YOLOX, where the parameters of the classification branch and regression branch are independently trained, endowing the model with stronger feature learning capabilities. However, this decoupling design results in a substantial parameter count and computational load; it is estimated that the YOLOv8 detection head occupies approximately 30% of the entire model's parameters and computational resources. When addressing the challenge of detecting leg diseases in X-ray images, there is an urgent need to enhance the efficiency and accuracy of the model.

To address this issue, we drew inspiration from the weight-sharing concept in RetinaNet and adopted a coupled design to replace the decoupling design of the YOLOv8 detection head. Simultaneously, we utilized Partial Convolution (PConv) to replace the original 3×3 regular convolution in the YOLOv8 detection head, creating a novel Partial Convolution-based Shared Weight Detection Head (SharedPConv head). Figure 4 provides detailed information

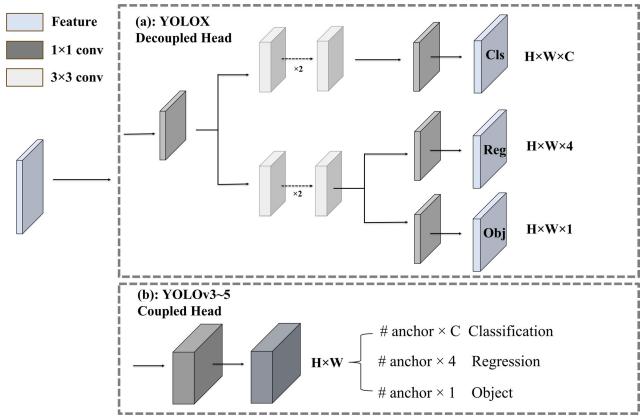


FIGURE 4. Decoupling head and coupling head design principles.

on decoupling and coupled designs, while Figure 5(a) illustrates the specific network structure of the SharedPConv head. By adopting the coupled design, both branches share a set of parameters until the final layer, significantly reducing the model's complexity and parameter count. This enhances the model's inference speed and more effectively improves the detection performance of broiler chicken leg diseases.

4) CHANNEL PRIOR CONVOLUTIONAL ATTENTION (CPCA)

The attention mechanism is a technology that enhances the extraction of key information by directing focus toward important regions of input objects. Currently, several mainstream attention mechanisms such as SE-Net, SimAM, and CPCA have been validated, achieving significant advancements [33], [34]. Particularly noteworthy is the CPCA attention mechanism, which dynamically distributes attention weights in both channel and spatial dimensions. It accomplishes this by extracting spatial relationships through deep convolutional modules without disrupting channel priors. This approach demonstrates remarkable efficacy in handling features like low contrast and pronounced organ shape variations in medical images. The channel prior convolutional attention module performs channel attention and spatial attention sequentially, as shown in Figure 6. Given an intermediate feature map $F \in \mathbb{R}^{C\times H\times W}$ as input, the channel attention module (CA) first infers a 1D channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$. M_C is then element-wise multiplied with the input feature F, and the channel attention values are broadcast along the spatial dimension to obtain the refined feature $F_c \in \mathbb{R}^{C\times H\times W}$ with channel attention. The spatial attention module (SA) processes to generate a 3D spatial attention map $M_s \in \mathbb{R}^{C\times H\times W}$. The final output feature $\hat{F} \in \mathbb{R}^{C\times H\times W}$ is obtained by element-wise multiplying M_s with F_c. The overall attention process can be summarized as:

$$\begin{aligned} F_c &= \mathrm{CA}(F) \otimes F, \\ \hat{F} &= \mathrm{SA}\,(F_c) \otimes F_c \end{aligned} \tag{4}$$

where $\otimes$ represents element-wise multiplication.

- Channel Attention: The module employs the approach proposed by CBAM [24] to implement channel attention. Spatial information is extracted from the feature map through average pooling and max pooling operations. This process generates two independent channel attention vectors. Subsequently, these vectors are input into a shared multi-layer perceptron (MLP). The outputs of the shared MLP are combined through element-wise summation to obtain the channel attention map. The computation of the channel attention can be summarized as:

$$\mathrm{CA}(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \tag{5}$$

where σ denotes the sigmoid function

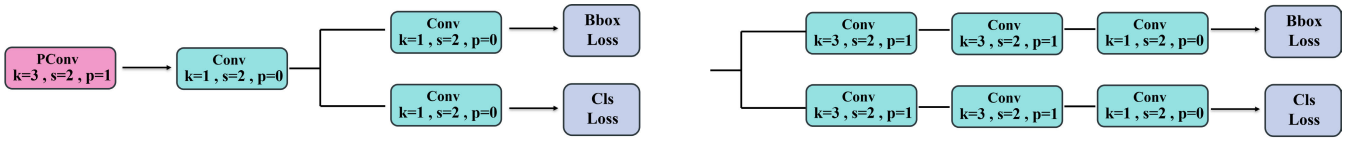
- Spatial Attention: To avoid enforcing consistency in the spatial attention maps for each channel, CPCA utilizes deep convolution to capture spatial relationships between features. This ensures the preservation of inter-channel relationships while reducing computational complexity. A multi-scale structure is employed to enhance the convolutional operation's ability to capture spatial relationships. Channel fusion is performed at the end of the spatial attention module using a 1×1 convolution, resulting in a more refined attention map. The computation method for spatial attention is as follows:

$$\mathrm{SA}(F) = \mathrm{Conv}_{1\times1}\left(\sum_{i=0}^{3} \mathrm{Branch}_i(\mathrm{DwConv}(F))\right) \tag{6}$$

where, DwConv represents depth-wise convolution, and Branch_i, $i \in \{0, 1, 2, 3\}$ represents the i-th branch.

In this paper, we investigate two diseases–joint effusion and tibial dyschondroplasia–that typically manifest as relatively small targets in large X-ray images. This characteristic leads to the continuous accumulation of complex background information during convolutional operations, resulting in a significant amount of redundant information. This situation may cause the model to overlook sufficient attention to the lesion areas, leading to the issue of false negatives. Additionally, as mentioned earlier, while PConv convolution has made progress in significantly reducing model complexity and computational burden, it has a limitation in that it only convolves over a subset of channels, thereby diminishing information exchange between channels.

To address these challenges, we introduced the Channel Prior Convolutional Attention (CPCA) into the backbone of the YOLOv8 object detection model. Specifically, we embedded the CPCA attention mechanism after the Spatial Pyramid Pooling Fusion (SPPF) module, thereby enhancing the bidirectional flow of information between the backbone network and the feature fusion layers. Additionally, this module employs multi-scale depth convolution to effectively extract spatial relationships, enabling better capture of subtle features in the regions of joint effusion and tibial dyschondroplasia.



(a): SharedPConv Head (b): The head of YOLOv8

FIGURE 5. Detailed network structure diagram of SharedPConv head.

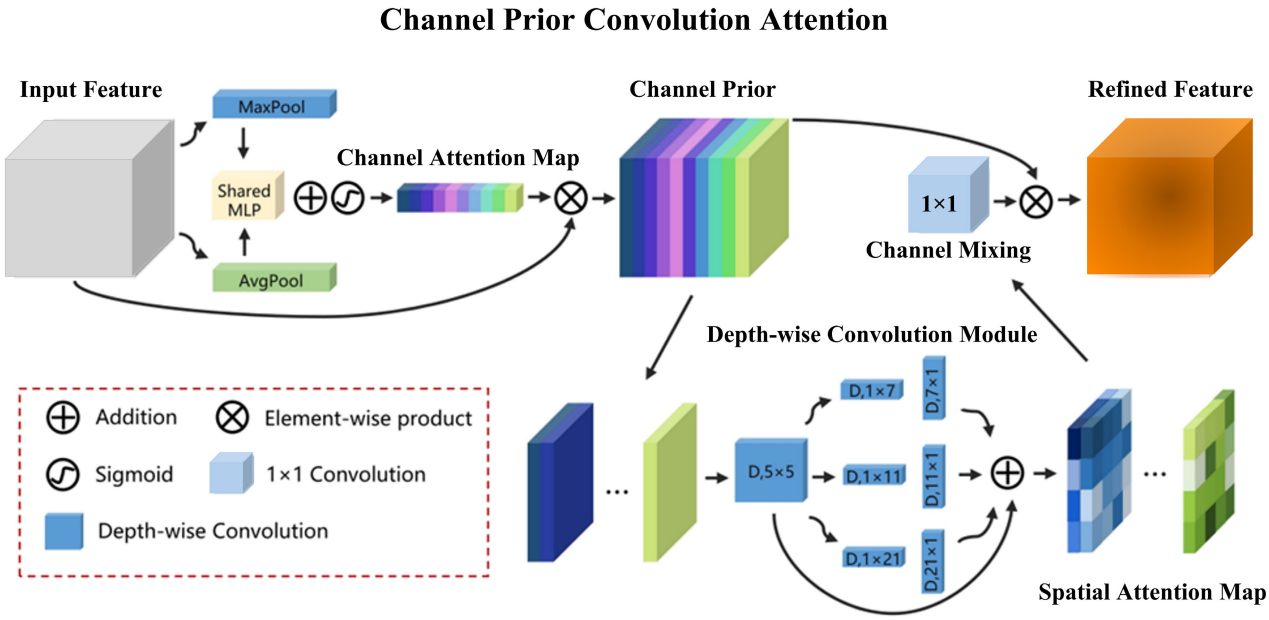


FIGURE 6. Channel Prior Convolutional Attention (CPCA) features an overall structure comprising sequential placement of channel attention and spatial attention. Spatial information of the feature maps is aggregated by the channel attention through operations such as average pooling and max pooling. The spatial information is subsequently processed through a shared MLP (Multi-Layer Perceptron) and added to produce the channel attention map. The channel prior is obtained by element-wise multiplication of the input feature and the channel attention map. Subsequently, the channel prior is inputted into the depth-wise convolution module to generate the spatial attention map. The convolutional module receives the spatial attention map for channel mixing. Ultimately, the refined features are obtained as the output by element-wise multiplication of the channel mixing result and the channel prior. The channel mixing process contributes to enhancing the representation of features.

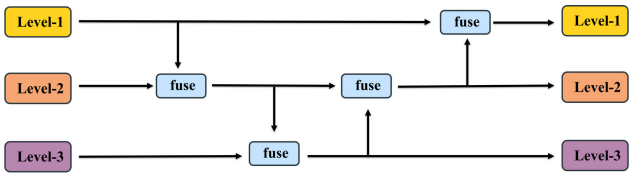


FIGURE 7. Traditional neck structure.

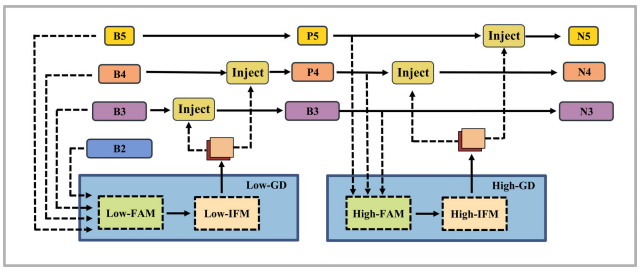


FIGURE 8. The architecture of the proposed gold-YOLO.

5) CROSS-LEVEL FEATURE FUSION

The disease characterized by joint effusion in the leg of broiler chickens typically manifests as mild effusion, presenting small and indistinct features on X-ray images. The YOLO series neck structure, as depicted in Figure 7, employs a traditional FPN structure, which comprises multiple branches for multi-scale feature fusion. However, it only fully fuse features from neighboring levels, for other layers information it can only be obtained indirectly 'recursively'. This results in a significant loss of small-scale information, such as lesions in the area affected by broiler chicken joint effusion disease, during the computation process, leading to a considerable number of missed detections. To avoid information loss in the transmission process of traditional FPN structures, We have introduced a novel GatherDistribute (GD) mechanism inspired by Gold-YOLO [35]. Specifically, the process gathers and distributes correspond to three modules: Feature

Alignment Module (FAM), Information Fusion Module (IFM), and Information Injection Module(Inject).

- The gathering process involves two steps. Firstly, the FAM collects and aligns features from various levels. Secondly, IFM fuses the aligned features to generate global information.
- Upon obtaining the fused global information from the gather process, the inject module distributes this information across each level and injects it using simple attention operations, subsequently enhancing the branch's detection capability.

To enhance the model's ability to detect objects of varying sizes, The method developed two branches: low-stage gather-and-distribute branch (Low-GD) and high-stage gather-and-distribute branch (High-GD). As shown in Figure 8, the neck's input comprises the feature maps B2, B3, B4, and B5 extracted by the backbone, where $B_i \in \mathbb{R}^{N\times C_{Bi}\times R_{Bi}}$.The batch size is denoted by N, the channels by C, and the dimensions by R = H T× W. Moreover, the dimensions of R_{B2}, R_{B3}, R_{B4}, and R_{B5} are R, $\frac{1}{2}R$, $\frac{1}{4}R$, and $\frac{1}{8}R$, respectively.

Low-stage gather-and-distribute branch: In this branch, the output B2, B3, B4, and B5 features from the backbone are selected for fusion to obtain high resolution features that retain small target information. The structure is shown in Figure 9(a).

a: LOW-STAGE FEATURE ALIGNMENT MODULE

In the Low-Level Feature Alignment Module (Low-FAM), the input features undergo downsampling through the use of average pooling (AvgPool) operations, achieving a uniform size. By resizing the features to the smallest feature size of the group$\left(R_{B4} = \frac{1}{4}R\right)$, we obtain F_{align} .

b: LOW-STAGE INFORMATION FUSION MODULE

The low-stage information fusion module (Low-IFM) design comprises multi-layer reparameterized convolutional blocks (RepBlock) and a split operation. Specifically, RepBlock takes F_{align} (channel = sum $(C_{B2}, C_{B3}, C_{B4}, C_{B5})$) as input and produces F_{fuse} (channel = $C_{B4} + C_{B5}$). The middle channel is an adjustable value (e.g., 256) to accommodate varying model sizes. The features generated by the RepBlock are subsequently split in the channel dimension into F_{inj_P3} and F_{inj_P4}, which are then fused with the different level's feature.

The formula is as follows:

$$F_{\text{align}} = Low\_FAM([B2, B\,3, B4, B5]) \quad (7)$$

$$F_{\text{fuse}} = \text{RepBlock}\left(F_{\text{align}}\right) \quad (8)$$

$$F_{\text{inj\_P3}}\,, F_{\text{inj\_P4}} = \text{Split}\left(F_{\text{fuse}}\right) \quad (9)$$

c: INFORMATION INJECTION MODULE

In order to inject global information more efficiently into the different levels, This module draws inspiration from the segmentation experience and employs attention operations to fuse the information, as illustrated in Figure 9(c).

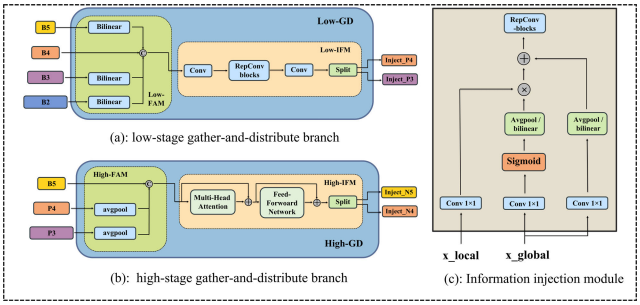


FIGURE 9. Gather-and-distribute structure. In (a), the Low-FAM and Low-IFM is low-stage feature alignment module and low-stage information fusion module in low-stage branch, respectively. In (b), the High-FAM and High-IFM is high-stage feature alignment module and high-stage information fusion module, respectively.

Specifically, this module inputs both local information (which refers to the feature of the current level) and global inject information (generated by IFM), denoted as F_{local} and F_{inj}, respectively. This module uses two different Convs with F_{inj} for calculation, resulting in $F_{\text{global\_embed}}$ and F_{act}. While $F_{\text{global\_embed}}$ is calculated with F_{local} using Conv. The fused feature F_{out} is then computed through attention. Due to the size disparity between F_{local} and F_{global}, this module employs either average pooling or bilinear interpolation to scale $F_{\text{global\_embed}}$ and F_{act} based on the size of F_{inj}, ensuring proper alignment. At the conclusion of each attention fusion, a RepBlock is added to further extract and integrate information.

In low stage, F_{local} is equal to Bi, so the formula is as follows:

$$F_{global\_act\_Pi} = \text{resize}\left(\text{Sigmoid}\left(\text{Conv}_{\text{act}}\left(F_{inj\_Pi}\right)\right)\right) \quad (10)$$

$$F_{global\_embed\_Pi} = \text{resize}\left(\text{Conv}\,v_{global\_embed\_Pi}\left(F_{inj\_Pi}\right)\right) \quad (11)$$

$$F_{att\_fuse\_Pi} = \text{Conv}_{local\_embed\_Pi}(Bi) * F_{global\_act\_Pi} + F_{global\_embed\_Pi} \quad (12)$$

$$Pi = \text{Rep}\,Block\left(F_{att\_fuse\_Pi}\right) \quad (13)$$

High-stage gather-and-distribute branch: The High-GD fuses the features P3, P4, P5 that are generated by the Low-GD, as shown in Figure 9(b).

d: HIGH-STAGE FEATURE ALIGNMENT MODULE

The high-stage feature alignment module (High-FAM) consists of avgpool, which is utilized to reduce the dimension of input features to a uniform size. Specifically, when the size of the input feature is $\{R_{P3}, R_{P4}, R_{P5}\}$, avgpool reduces the feature size to the smallest size within the group of features$\left(R_{P5} = \frac{1}{8}R\right)$. Since the transformer module extracts high-level information, the pooling operation facilitates information aggregation while decreasing the computational requirements for the subsequent step in the Transformer module.

e: HIGH-STAGE INFORMATION FUSION MODULE

The high-stage information fusion module (High-IFM) comprises the transformer block (explained in greater detail below) and a splitting operation, which involves a three-step process: (1) the F_{align} , derived from the High-FAM, are combined using the transformer block to obtain the F_{fuse} . (2) The F_{fuse} channel is reduced to sum (C_{P4}, C_{P5}) via a Conv 1×1 operation. (3) The F_{fuse} is partitioned into F_{inj_N4} and F_{inj_N5} along the channel dimension through a splitting operation, which is subsequently employed for fusion with the current level feature.

The formula is as follows:

$$F_{\text{align}} = \text{High\_FAM}([P3, P4, P5]) \quad (14)$$

$$F_{\text{fuse}} = \text{Transformer}\left(F_{\text{align}}\right) \quad (15)$$

$$F_{inj\_N4}, F_{inj\_N5} = \text{Split}\left(\text{Conv } 1 \times 1\left(F_{\text{fuse}}\right)\right) \quad (16)$$

The transformer fusion module in Eq. 15 comprises several stacked transformers, with the number of transformer blocks denoted by L. Each transformer block includes a multi-head attention block, a Feed-Forward Network (FFN), and residual connections.

f: INFORMATION INJECTION MODULE

The information injection module in High-GD is exactly the same as in Low-GD. In high stage, F_{local} is equal to Pi, so the formula is as follows:

$$F_{global\_act\_Ni} = \text{resize}\left(\text{Sigmoid}\left(\text{Conv}_{\text{act}}\left(F_{inj\_Ni}\right)\right)\right) \quad (17)$$

$$F_{global\_embed_N i} = \text{resize}\left(\text{Conv}_{global\_embed\_Ni}\left(F_{inj\_Ni}\right)\right) \quad (18)$$

$$F_{att\_fuse\_Ni} = \text{Conv}_{local\_embed\_Ni}(Pi) * F_{global\_act\_Ni} + F_{global\_embed\_Ni} \quad (19)$$

$$Ni = \text{Rep}\,Block\left(F_{att\_fuse\_Ni}\right). \quad (20)$$

In this paper, as shown in Figure 1, we replaced the original FPN structure in the yolov8 neck with the Gather-and-Distribute mechanism (GD). To further enhance the interconnectivity of cross-level information, We added two Feature Alignment Modules (FAM) with three inputs each in the Low-Level Gathering and Distributing Branch. The formulation can be expressed as:

$$F_i = \text{Low\_}FAM([B(i-1), B\,i, \text{Conv}(B(i+1))]) \quad (21)$$

where Fi serves as Bi in Equation 12, specifically, as illustrated in Figure 8(c), Fi is utilized as the x_local input in the Information Injection Module. Therefore, in this paper, Equation 9 should be modified to:

$$F_{att\_fuse\_Pi} = Conv_{local\_embed\_Pi}(Fi) * F_{global\_act\_Pi} + F_{global\_embed\_Pi}. \quad (22)$$

Similarly, in the High-stage gather-and-distribute branch, two additional Feature Alignment Modules were added, and a C2f structure was introduced after all Information Injection Modules. The formula can be expressed as:

$$F_i = \text{High\_}FAM([C2f(pi), \text{Conv}(B(i+1))]) \quad (23)$$

where Fi serves as Pi in Equation 9, Therefore, in this paper, Equation 9 should be modified to:

$$F_{\text{att\_fuse\_Ni}} = Conv_{\text{local\_embed\_Ni}}(Fi) * F_{\text{global\_act\_Ni}} + F_{\text{global\_embed\_Ni}} \quad (24)$$

Through this improvement, the effectiveness of information fusion and transmission has been strategically increased, thereby better addressing the issue of indistinct joint effusion lesion areas.

TABLE 2. Comparison of experimental test results.

Hyperparameter Options	Setting
Input Resolution	640×640
Initial Learning Rate(lr0)	0.01
Learning Rate Float(lrf)	0.01
Momentum	0.937
Weight_decay	0.0005
Batch_size	64
Activation Function	SiLU
Epoch	500

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL CONDITIONS AND PARAMETER SETTINGS

The experimental setup employed the Linux operating system, equipped with an NVIDIA GeForce RTX 3090 graphics processing unit (GPU) with 24GB of memory. Computation was conducted using PyTorch version 1.13.1+cu116, and the Python language environment was configured to version 3.8.0.

To enhance the model's performance during training, we introduced learning rate decay methods into the training process. This approach involved adjusting the update speed of the model's parameters using an initial learning rate (lr0). Furthermore, a learning rate coefficient (lrf) was applied to regulate the decay of the learning rate throughout the training. The final learning rate was determined by multiplying the initial learning rate by the coefficient. To ensure an adequate number of training steps, the iteration was set to 500. During training, the learning rate gradually decreased, fostering model stability, facilitating smooth convergence, and minimizing fluctuations to attain the optimal solution. Table2 provides an overview of the hyperparameter configurations utilized during model training.

B. EVALUATION INDICATORS

To assess the performance of the improved detection model, it is crucial to choose appropriate metrics. The evaluation metrics employed in this paper include recall (R), average precision (AP), and mean average precision (mAP). These

TABLE 3. Comparison of experimental test results.

Model Type	Depth	Width	Training time(hour)	mAP(0.5)
YOLOv8s	0.33	0.50	**0.710**	**0.865**
YOLOv8m	0.67	0.75	1.941	0.870
YOLOv8l	1.00	1.00	2.146	0.857
YOLOv8x	1.33	1.25	3.465	0.859

evaluation metrics are calculated as follows:

$$\begin{aligned} AP &= \int_0^1 P \, \mathrm{d}R \\ \mathrm{mAP} &= \frac{\sum_{i=1}^{N} \mathrm{AP}_i}{N} \\ \mathrm{Recall} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \\ \mathrm{Precision} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \end{aligned} \tag{25}$$

TP (true positive) represents the number of detected diseases in X-ray images with diseases, FP represents the number of detected diseases in X-ray images without diseases, FN represents the number of undetected diseases in X-ray images with diseases, and TN represents the number of undetected diseases in X-ray images without diseases. The AP value corresponds to the area under the precision-recall (P-R) curve. mAP is the average precision obtained by averaging the AP values for detecting two types of chicken leg diseases: joint effusion and tibial dyschondroplasia. The number of defect categories, N, is 2. A higher mAP indicates better defect detection performance and higher recognition accuracy.

In addition to performance metrics, model complexity needs to be considered, including the number of parameters (Mb) and the number of Floating Point Operations (FLOPs). For a standard convolution, the calculations for parameters and FLOPs are as follows:

$$\begin{aligned} \text{Parameters} &= (K_h * K_w * C_{\text{in}}) * C_{\text{out}} + C_{\text{out}} \\ \text{FLOPs} &= [(K_h * K_W * C_{\text{in}}) * C_{\text{out}} + C_{\text{out}}] * (H * W) \end{aligned} \tag{26}$$

where, K_h and K_w denote the height and width of the convolution kernel, C_in, C_out represents the number of input channels, represents the number of output channels, and H and W are the height and width of the output feature vector. Typically, K_w and K_h have the same size.

C. EXPERIMENTAL RESULTS AND ANALYSIS

1) ANALYSIS OF ORIGINAL MODEL PERFORMANCE

The YOLOv8 model has been developed in four variants: YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. YOLOv8s is the model with the minimum depth and width, while the other three models are products that deepen and extend YOLO v8s. Smaller network models demand lower performance requirements for mobile terminals and are easier to deploy.

This section investigates the impact of model depth and width on the detection of diseases in broiler chicken legs in X-ray images. In deep learning models, a more complex structure and deeper depth typically lead to better detection performance. However, in the most complex models, small sample data may not demonstrate optimal detection. To design a model with the best cost-effectiveness, we trained four different models, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, using a dataset processed through Clahe data augmentation. The model metrics are presented in Table 3.

The experimental results indicate that the average precision (mAP) of YOLOv8l and YOLOv8x is lower than that of YOLOv8s. This suggests that, for the detection of diseases in broiler chicken legs under X-ray images, a deeper neural network structure may lead to overfitting issues, reducing the model's generalization. Although the mAP of YOLOv8m is only 0.5% higher than YOLOv8s, considering the training time cost and hardware requirements, we decided to use YOLOv8s as the base detection model.

2) LIGHTWEIGHT STRUCTURE COMPARATIVE ANALYSIS

In this section, to assess the effectiveness and feasibility of the C2f_PConv module and the Partial Convolution-based Shared Weight Detection Head (SharedPConv head), we conducted ablative experiments. To ensure the precision of the ablative experiments, the model's operational environment and hyperparameters remained consistent, and specific experimental results are detailed in Table 4. In the table, C2f_PConv represents the enhancement of introducing Partial Convolution (PConv) into the C2f module of the original YOLOv8 model, while SharedPConv denotes our designed Shared Weight Coupling Head based on Partial Convolution. Params indicate the number of model parameters, GFLOPs represent the model's billion floating-point operations per second, and mAP is the average detection precision across all target classes.

Analysis of the performance of the C2f_PConv module reveals that the Partial Convolution (PConv) in this module operates only on specific channels of the feature map. Consequently, compared to the baseline model, the improved model showed a reduction of 12.6% in parameters and a decrease of 13.4% in GFLOPs. Further analysis of the experimental results indicates that the improved model achieved a 4.0% increase in mAP and a 1.3% improvement in precision. This validates that the enhancement of this module does not compromise model precision despite the reduction in parameter count.

The analysis of the performance of the Partial Convolution-based Shared Weight Detection Head (SharedPConv head) shows that, as presented in Table 4, the improved model exhibited a reduction of 13.7% in parameters and a decrease of 24.3% in GFLOPs. Experimental results demonstrate that by replacing the decoupled design of the YOLOv8 detection head with a coupled design, parameter sharing was achieved, significantly reducing model complexity. Further analysis of experimental results shows slight improvements in mAP,

TABLE 4. Comparison of experimental test results.

Algorithms	Parameters	GFLOPs	Precision	Recall	mAP(0.5)
yolov8s	11126358	28.4	0.884	0.889	0.865
+C2f_PConv	9721760	24.6	0.897	0.888	**0.905**
+SharedPConv Head	9608534	21.5	**0.907**	0.895	0.88
ours	**8163734**	17.3	0.898	0.898	0.899

TABLE 5. Embedding CPCA attention mechanism into detection regions in different regions of the network.

Models	Parameters	Precision	Recall	mAP(0.5)
yolov8s	11126358	0.884	0.889	0.865
yolov8s-backbone	11577046	0.917	0.883	**0.902**
yolov8s-neck	11743030	0.911	0.885	0.885

TABLE 6. Comparison of detection performance of different feature fusion algorithms in this dataset.

Models	Methods	Precision	Recall	mAP(0.5)
yolov8s	FPN	0.884	0.889	0.865
yolov8s+AFPN	AFPN	0.887	0.898	0.880
yolov8s+GFPN	GFPN	0.917	0.872	0.867
yolov8s+BiFPN	BiFPN	0.884	0.916	0.887
yolov8s+GD	GD	0.917	0.908	0.884

TABLE 7. Statistical results of ablation experiments.

Data Enhancement	C2f_PConv block	GD	SharedPConv	CPCA	mAP(0.5)	mAP Increase
✓					0.865	
✓	✓				0.905	4.0%
✓		✓			0.884	1.9%
✓			✓		0.880	1.5%
✓				✓	0.902	3.7%
✓	✓	✓	✓	✓	0.937	7.2%

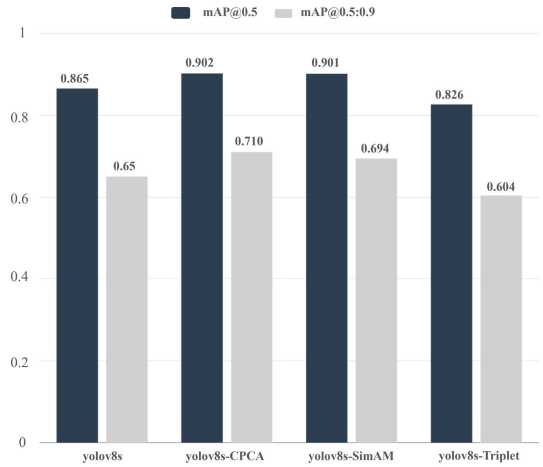


FIGURE 10. Comparison of various attention mechanisms under YOLOv8s.

precision, and recall, supporting the effectiveness of the model enhancement.

Performance analysis of the combined modules reveals that the integration of the C2f_PConv module and the Partial Convolution-based Shared Weight Detection Head (SharedPConv head), compared to the baseline model, resulted in a 26.6% reduction in parameters, with corresponding increases of 3.4% and 2.9% in mAP and precision, respectively. Furthermore, comparing the combined model with models incorporating only the C2f_PConv module or the SharedPConv head individually, it is evident that the combined model did not suffer significant precision loss while achieving reductions of 16% and 15% in parameter count, respectively. This indicates that, compared to individual component modules, the model is more lightweight, successfully validating the superiority of the combined module over single-component modules, rather than a simple accumulation of accuracy.

3) COMPARATIVE ANALYSIS OF ATTENTION MECHANISMS

Figure 10 presents a comparison of detection results for the YOLOv8s detection model with different attention mechanisms embedded in the backbone. The experiment utilized X-ray images from a dataset of chicken leg diseases for training.

As shown in the table, introducing the YOLOv8 model, the Channel Prior Convolutional Attention (CPCA), and the SimAM attention mechanisms can improve the network's detection accuracy. However, with the introduction of the Triplet attention mechanism, there is a slight decrease in detection accuracy compared to the original algorithm. The CPCA attention mechanism achieves the optimal performance, increasing mAP by 3.7% compared to the original algorithm. In summary, the CPCA attention mechanism, employing multiscale depth convolutional modules, effectively extracts features from medical images with low contrast and significant organ shape variations, thereby enhancing feature learning capabilities.

Furthermore, to thoroughly investigate whether embedding CPCA in the model backbone is the optimal choice, this experiment conducted a detailed comparison by embedding CPCA separately in the backbone and feature fusion layer of the YOLOv8s object detection model. The experimental results are presented in Table 5: embedding CPCA in the feature fusion layer improves detection accuracy by 3.3%, and mAP (0.5) increases by 2%. Embedding CPCA in the backbone achieves an mAP value of 90.2%, with only a slight increase in overall model parameters. The improvement is particularly significant in the detection model. Through comparative analysis of the experimental results, we conclude that embedding CPCA in the YOLOv8s model effectively enhances the correlation between low-dimensional and high-dimensional features, thereby achieving more accurate target localization and recognition.

When the CPCA attention mechanism is embedded in the feature fusion layer, the detection accuracy improves by 3.3%, and mAP(0.5) increases by 2%. Embedding the CPCA attention mechanism in the network's backbone results in a higher mAP(0.5) of 90.2%, with only a slight increase in the overall model parameters. Through comparative experimental results, it is evident that embedding the CPCA attention mechanism into the backbone of yolov8s allows for more accurate localization and recognition of targets.

4) COMPARATIVE ANALYSIS OF FEATURE FUSION LAYERS

In this section, we aim to explore the impact of different feature fusion methods on the dataset by embedding AFPN, GFPN, BiFPN, and the GatherDistribute (GD) mechanism into the neck of the YOLO v8s detection model. The experimental results are presented in Table 6.

From Table 6, it is evident that when adopting AFPN, GFPN [36], BiFPN, and GD, the model's detection accuracy is superior to the original model with FPN structure. However, through a comprehensive analysis of Precision, Recall, and mAP(0.5), we found that the GatherDistribute mechanism (GD) achieved the most balanced detection results on this dataset. Compared to the original model, the precision improved by 3.3%, and there was a slight increase in recall. By comparing the experimental results, we conclude that the GatherDistribute mechanism (GD), by employing a unified module to gather and fuse information from different levels and then distributing it to different levels, successfully avoids the information loss problem caused by the lack of inter-layer communication in traditional FPN structures and their variants (such as AFPN, GFPN, BiFPN). Therefore, embedding the GatherDistribute mechanism (GD) into the neck of YOLOv8s can more accurately locate and identify features of the affected areas.

5) ABLATION EXPERIMENT

To visually observe the impact of different improvement modules on the model's performance, this section conducted ablation experiments for validation. Specifically, various improvements were introduced into the original yolov8s model, including Clahe data augmentation, CPCA attention mechanism, C2f_PConv module, cross-level feature fusion through the Gather-and-Distribute mechanism (GD), and a detection head based on shared PConv weights. The detection performance was compared while keeping the dataset and training parameters constant.

The experimental results are presented in the Table 6. By introducing Clahe data augmentation, CPCA attention mechanism, C2f_PConv module, Gather-and-Distribute mechanism (GD), and a detection head based on shared PConv weights, each improvement module contributed to an increase in the model's average precision. When yolov8s is integrated with these five improvement modules to form the final chicken leg disease detection model, under conditions nearly identical to yolov8s parameters, the overall mAP increased by 10.6%.

Table 7 demonstrates that each improvement module contributes positively to the model's performance, and the integration of these modules results in a notable enhancement in overall mAP for the chicken leg disease detection model.

6) ANALYSIS OF DETECTION RESULTS

In this section, we focus on detecting chicken joint effusion and tibial dyschondroplasia diseases by randomly selecting different types of X-ray images. The detection results of the original yolov8s model and the improved yolov8s model are illustrated in Figure 11.

From Figure 11(c)-(d), it can be observed that under the complex background interference of low-contrast X-ray images, YOLOv8s exhibits instances of missed detection with low confidence in detecting lesions related to chicken joint effusion disease. Visual results for the detection of chicken cartilage malformation disease show cases of false positives. In contrast, in Figure 11(a)-(b), the improved YOLOv8s model in this study is not affected by non-uniform complex backgrounds. It successfully detects 100% of the lesion areas of joint effusion in the images, providing more precise bounding boxes with higher confidence and without any instances of false positives. The comparison indicates that the improved YOLOv8s model delivers more accurate detection results, capturing crucial information in the lesion areas of chickens and demonstrating excellent generalization performance.

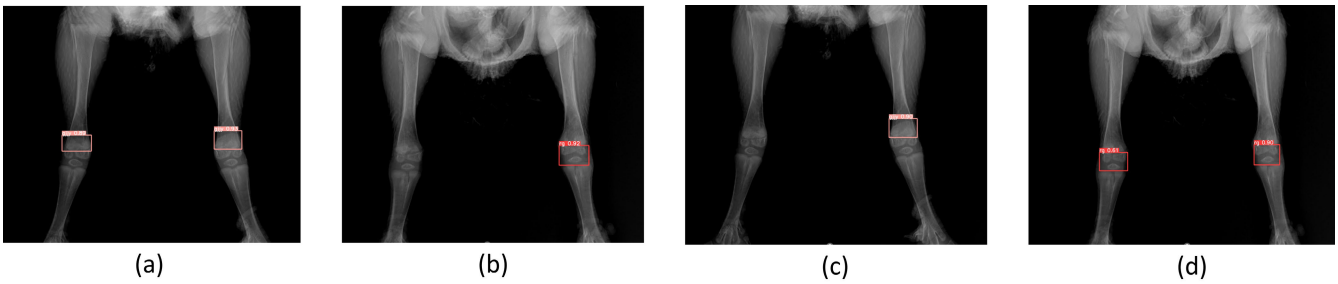


FIGURE 11. Comparison of the improved YOLOv8s with the original YOLOv8s detection results is shown in Figure, where (a) provides examples of detecting chicken joint effusion disease on the baseline network YOLOv8s, (b) presents examples of detecting chicken tibial dyschondroplasia disease on the baseline network YOLOv8s, (c) illustrates examples of detecting chicken joint effusion disease on the improved YOLOv8s proposed in this paper, and (d) demonstrates examples of detecting chicken tibial dyschondroplasia disease on the improved YOLOv8s.

TABLE 8. Performance comparison results between the improved algorithm and other algorithms.

Models	FPS	Precision	recall	mAP(0.5)
YOLOv5	77.5	0.889	0.873	0.815
YOLOv7 [37]	**127.6**	0.864	0.897	0.818
YOLOv8	80.6	0.884	0.889	0.865
Faster-RCNN	36.3	0.087	0.026	9.25
Cascade-RCNN [38]	23.2	0.850	0.821	0.831
Ours	66.8	**0.928**	**0.902**	**0.937**

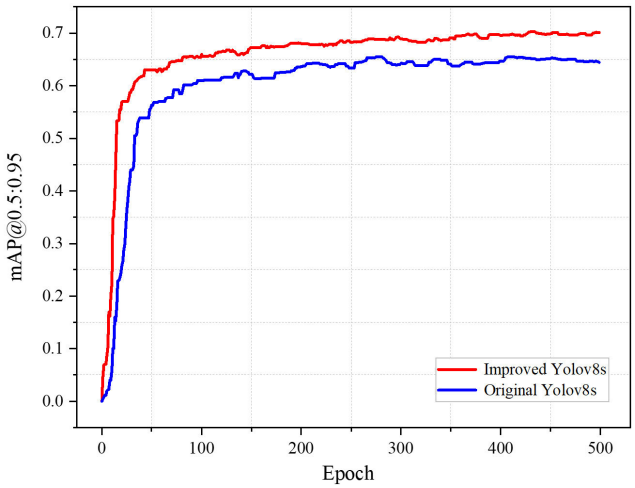


FIGURE 12. Comparison of two YOLO V8S detection models mAP@0.5:0.95.

Figures 12,13 present the loss functions and mAP@0.5:0.95 curves of the yolov8s detection model before and after improvement during the training process on the training set and validation set. From Figures 13, it can be observed that the improved yolov8s network consistently exhibits lower loss values in classification, bounding box prediction, and DFL loss, presenting an earlier inflection point. On the validation set, the improved yolov8s model demonstrates stability without oscillations during iterations, indicating better overall model performance.

Figure 12 illustrates that the curve of the improved yolov8s is consistently above the curve of the original yolov8s, indicating an overall higher detection accuracy for the improved yolov8s network. Notably, the learning curve of the improved network is smoother, indicating better stability.

In summary, the improved yolov8s detection model in this study has significantly enhanced accuracy, optimizing the detection performance of the network and achieving accurate detection of chicken joint effusion and tibial dyschondroplasia diseases.

V. DISCUSSION

In this section, we validated the performance of the proposed method by comparing it with several mainstream object detection algorithms under the same experimental environment. Table 8 presents the quantitative comparison results, where mAP denotes the mean average precision across all target classes in the model. FPS represents the detection speed of the model. The improved YOLOv8s model proposed in this study, with the introduction of the CPCA attention mechanism in the backbone of the network and the replacement of the FPN structure with the GD mechanism in the neck of the network, has a slightly lower parameter count and average inference time than some mainstream one-stage object detection algorithms (YOLOv5s, YOLOv8s). However, the model proposed in this research achieves higher mAP, accuracy, and recall, reaching 93.7%, 92.8%, and 90.2%, respectively, outperforming other models. It's worth noting that in the experiments, we used the Faster R-CNN and Cascade-RCNN models for object detection. However, we observed that these models did not achieve satisfactory performance on our X-ray dataset of broiler chicken leg diseases. In light of this phenomenon, we conducted the following analysis: In our study, after CLAHE data augmentation, the number of images per class

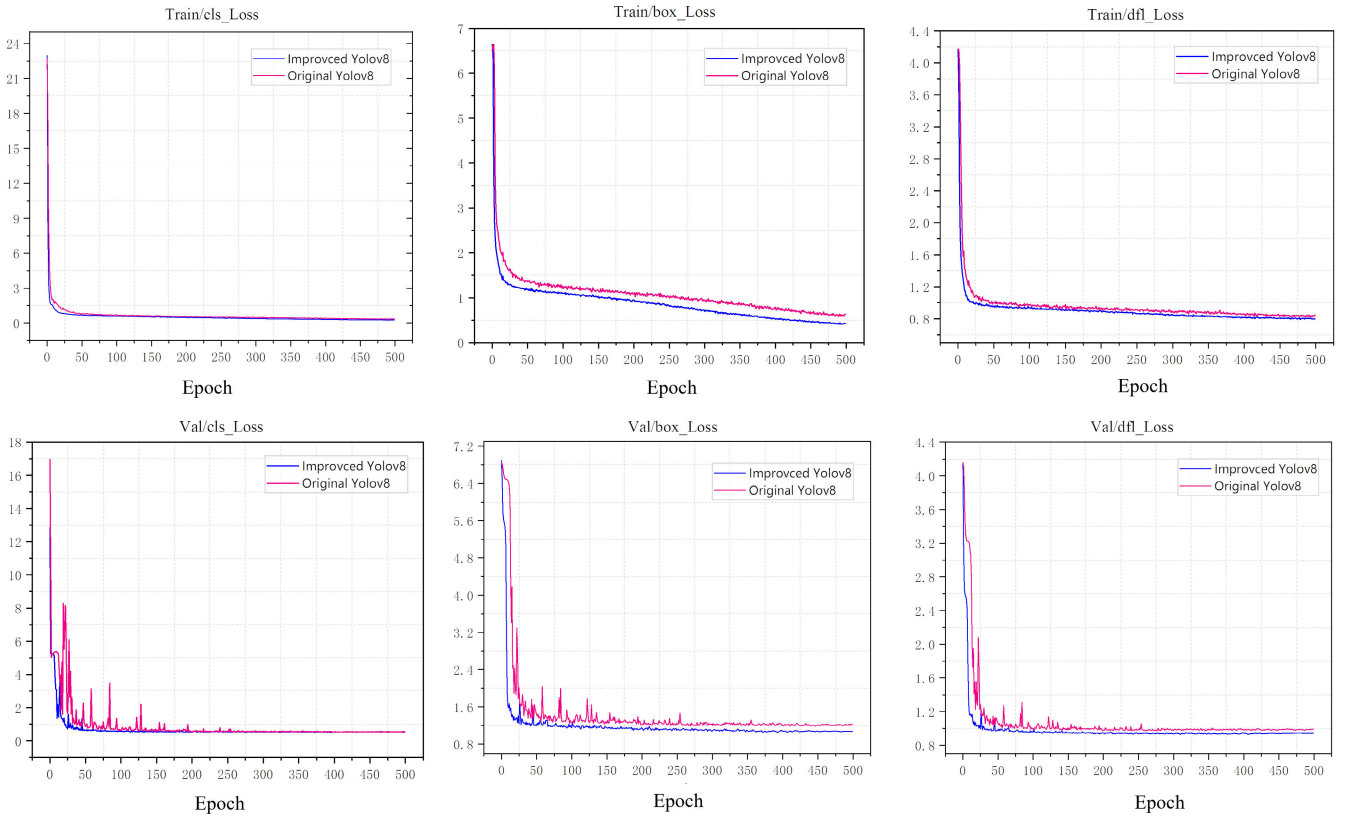


FIGURE 13. Effect diagram of improved yolov8s and original yolov8s loss function.

was only 800. Since Faster R-CNN and Cascade-RCNN models belong to two-stage object detection algorithms and involve multiple convolutional layers in the network structure, when the training data is limited, the model may overfit to the noise in the training data and fail to generalize to new data, resulting in poor performance on this dataset.

While our improved model has achieved an overall increase in detection accuracy, there are still some limitations. As shown in Table 8, the parameter count and image processing time of our model are not optimal. When detecting a large number of X-ray images of broiler chicken leg diseases simultaneously, it still requires a certain amount of time for processing. Therefore, our proposed method needs further optimization. In the future, we will consider using model compression and pruning to alleviate the YOLO series networks, reducing memory overhead and model file size to balance performance and inference speed.

VI. CONCLUSION

This article introduces an enhanced YOLOv8 object detection model tailored to the distinctive features of joint effusion and chondrodysplasia in broiler chickens as observed in X-ray images. We employed the Clahe data augmentation method to highlight the features of lesion areas in X-ray images. Additionally, to further optimize model performance, we introduced the CPCA attention mechanism and C2f_PConv module into the network's backbone. In the feature fusion layer, we utilized the Gather-and-Distribute mechanism (GD). Finally, we introduced a detection head based on shared PConv weights, significantly enhancing the model's ability to extract features from lesion areas.

Our proposed automatic detection method for chicken leg diseases has gained initial recognition in the poultry industry, assisting farmers in assessing whether chickens have joint effusion and tibial dyschondroplasia. However, there are still some limitations that need further improvement. Firstly, we plan to enhance the generalization capability of the deep learning model by adding more data to the dataset. Although data augmentation methods were applied to the current dataset, the limited size of the original data still imposes constraints on the model's performance. Secondly, we will focus on model pruning and distillation processes to reduce model complexity and achieve further lightweighting.

REFERENCES

[1] T. G. Knowles, S. C. Kestin, S. M. Haslam, S. N. Brown, L. E. Green, A. Butterworth, S. J. Pope, D. Pfeiffer, and C. J. Nicol, "Leg disorders in broiler chickens: Prevalence, risk factors and prevention," *PLoS ONE*, vol. 3, no. 2, Feb. 2008, Art. no. e1545, doi: 10.1371/journal.pone.0001545.

[2] A. W. Bassler, C. Arnould, A. Butterworth, L. Colin, I. C. De Jong, V. Ferrante, P. Ferrari, S. Haslam, F. Wemelsfelder, and H. J. Blokhuis, "Potential risk factors associated with contact dermatitis, lameness, negative emotional state, and fear of humans in broiler chicken flocks," *Poultry Sci.*, vol. 92, no. 11, pp. 2811–2826, Nov. 2013, doi: 10.3382/ps.2013-03208.

[3] K. Mehmood, H. Zhang, K. Li, L. Wang, M. U. Rehman, F. Nabi, M. K. Iqbal, H. Luo, M. Shahzad, and J. Li, "Effect of tetramethylpyrazine on tibial dyschondroplasia incidence, tibial angiogenesis, performance and characteristics via HIF-1α/VEGF signaling pathway in chickens," *Sci. Rep.*, vol. 8, no. 1, pp. 24–95, Feb. 2018, doi: 10.1038/s41598-018-20562-3.

[4] R. F. Wideman, K. R. Hamal, J. M. Stark, J. Blankenship, H. Lester, K. N. Mitchell, G. Lorenzoni, and I. Pevzner, "A wire-flooring model for inducing lameness in broilers: Evaluation of probiotics as a prophylactic treatment," *Poultry Sci.*, vol. 91, no. 4, pp. 870–883, Apr. 2012, doi: 10.3382/ps.2011-01907.

[5] T. Xu, K. Yue, C. Zhang, X. Tong, L. Lin, Q. Cao, and S. Huang, "Probiotics treatment of leg diseases in broiler chickens: A review," *Probiotics Antimicrobial Proteins*, vol. 14, no. 3, pp. 415–425, Nov. 2021, doi: 10.1007/s12602-021-09869-2.

[6] K. E. Kittelsen, B. David, R. O. Moe, H. D. Poulsen, J. F. Young, and E. G. Granquist, "Associations among gait score, production data, abattoir registrations, and postmortem tibia measurements in broiler chickens," *Poultry Sci.*, vol. 96, no. 5, pp. 1033–1040, May 2017, doi: 10.3382/ps/pew433.

[7] R. Mehta and T. Arbel, "3D U-Net for brain tumour segmentation," in *Proc. Int. MICCAI. Brainlesion. Workshop*, vol. 11384, Jan. 2019, pp. 254–266.

[8] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognit.*, vol. 85, pp. 109–119, Jan. 2019, doi: 10.1016/j.patcog.2018.07.031.

[9] J. Bernal, K. Kushibar, D. S. Asfaw, S. Valverde, A. Oliver, R. Martí, and X. Lladó, "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review," *Artif. Intell. Med.*, vol. 95, pp. 64–81, Apr. 2019.

[10] H.-B. Chen, S. Jiang, G. He, B. Zhang, and H. Yu, "TEANS: A target enhancement and attenuated nonmaximum suppression object detector for remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 632–636, Apr. 2021, doi: 10.1109/LGRS.2020.2983070.

[11] L. Hou, K. Lu, J. Xue, and L. Hao, "Cascade detector with feature fusion for arbitrary-oriented objects in remote sensing images," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, London, U.K., Jul. 2020, pp. 1–6, doi: 10.1109/ICME46284.2020.9102807.

[12] A. Krizhevsky, I. Sutskever, and G.-E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[13] L. Shen, B. Lang, and Z. Song, "CA-YOLO: Model optimization for remote sensing image object detection," *IEEE Access*, vol. 11, pp. 64769–64781, 2023, doi: 10.1109/ACCESS.2023.3290480.

[14] P. Das, A. Chakraborty, R. Sankar, O. K. Singh, H. Ray, and A. Ghosh, "Deep learning-based object detection algorithms on image and video," in *Proc. 3rd Int. Conf. Intell. Technol. (CONIT)*, Hubli, India, Jun. 2023, pp. 1–6, doi: 10.1109/conit59222.2023.10205601.

[15] D. Xu and Y. Wu, "FE-YOLO: A feature enhancement network for remote sensing target detection," *Remote Sens.*, vol. 13, no. 7, p. 1311, Mar. 2021, doi: 10.3390/rs13071311.

[16] L. Chen, W. Shi, and D. Deng, "Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images," *Remote Sens.*, vol. 13, no. 4, p. 660, Feb. 2021, doi: 10.3390/rs13040660.

[17] C. Li, B. Luo, H. Hong, X. Su, Y. Wang, J. Liu, C. Wang, J. Zhang, and L. Wei, "Object detection based on global-local saliency constraint in aerial images," *Remote Sens.*, vol. 12, no. 9, p. 1435, May 2020, doi: 10.3390/rs12091435.

[18] S. Chen, R. Zhan, W. Wang, and J. Zhang, "Learning slimming SAR ship object detector through network pruning and knowledge distillation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1267–1282, 2021, doi: 10.1109/JSTARS.2020.3041783.

[19] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[20] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "PANet: Few-shot image semantic segmentation with prototype alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9197–9206.

[21] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.

[22] G. Yang, J. Lei, Z. Zhu, S. Cheng, Z. Feng, and R. Liang, "AFPN: Asymptotic feature pyramid network for object detection," 2023, *arXiv:2306.15988*.

[23] X. Li, "Semantic flow for fast and accurate scene parsing," in *Proc. Eur. Conf. Comput. Vis.*, Aug. 2020, pp. 775–793.

[24] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, p. 319.

[25] H. Huang, Z. Chen, Y. Zou, M. Lu, and C. Chen, "Channel prior convolutional attention for medical image segmentation," 2023, *arXiv:2306.05196*.

[26] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 603–612.

[27] K. Gan, D. Xu, Y. Lin, Y. Shen, T. Zhang, K. Hu, K. Zhou, M. Bi, L. Pan, W. Wu, and Y. Liu, "Artificial intelligence detection of distal radius fractures: A comparison between the convolutional neural network and professional assessments," *Acta Orthopaedica*, vol. 90, no. 4, pp. 394–400, Apr. 2019, doi: 10.1080/17453674.2019.1600125.

[28] Y. Jia, H. Wang, W. Chen, Y. Wang, and B. Yang, "An attention-based cascade R-CNN model for sternum fracture detection in X-ray images," *CAAI Trans. Intell. Technol.*, vol. 7, no. 4, pp. 658–670, Jan. 2022, doi: 10.1049/cit2.12072.

[29] M. A. Al-Antari, S.-M. Han, and T.-S. Kim, "Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105584, doi: 10.1016/j.cmpb.2020.105584.

[30] J. Wang, B. Lin, G. Li, Y. Zhou, L. Zhong, X. Li, and X. Zhang, "YOLO-xray: A bubble defect detection algorithm for chip X-ray images based on improved YOLOv5," *Electronics*, vol. 12, no. 14, p. 3060, Jul. 2023, doi: 10.3390/electronics12143060.

[31] S. F. Tan and N. A. M. Isa, "Exposure based multi-histogram equalization contrast enhancement for non-uniform illumination images," *IEEE Access*, vol. 7, pp. 70842–70861, 2019, doi: 10.1109/ACCESS.2019.2918557.

[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] 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.

[34] 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.

[35] C. Wang, W. He, Y. Nie, J. Guo, C. Liu, K. Han, and Y. Wang, "Gold-YOLO: Efficient object detector via Gather-and-Distribute mechanism," 2023, *arXiv:2309.11331*.

[36] X. Xu, Y. Jiang, W. Chen, Y. Huang, Y. Zhang, and X. Sun, "DAMO-YOLO : A report on real-time object detection design," 2022, *arXiv:2211.15444*.

[37] 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," 2022, *arXiv:2207.02696*.

[38] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.

XIN ZHANG received the bachelor's degree from the Department of Computer Science and Technology, Tianjin Agricultural University, in June 2022. He is currently pursuing the degree with the School of Computer and Information Engineering, Tianjin Agricultural University. He has a certain level of innovation ability and has applied for a national invention patent. His research interests include image processing based on deep learning and object detection in computer vision.

RENWEN ZHU graduated from the Department of Computer Science and Technology, Tianjin Agricultural University, in June 2022, where he is a graduate student with the College of Computer and Information Engineering. He is able to think and solve problems independently, with a certain level of innovative practical ability. His main research areas include blockchain application development and template-based system research.

WEIGANG ZHENG received the bachelor's degree from the Department of Internet of Things Engineering, Tianjin Agricultural University, in June 2022, where he is currently a graduate student with the School of Computer and Information Engineering. He possesses certain innovative and hands-on practical abilities. In 2021, he was authorized as the first inventor to publish a utility model patent, and he applied for one national invention patent and one utility model patent in 2023. His research interests include image processing based on deep learning and object detection based on multimodal data fusion.

CHANGXI CHEN received the Ph.D. degree from the School of Mechanical Engineering, Beijing Institute of Animal Science and Veterinary Medicine, Chinese Academy of Agricultural Sciences, China, in 2011. He is currently a Professor with the School of Computer and Information Engineering, Tianjin Agricultural University. His research interests include machine learning, big data analysis, and agricultural information technology.