IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 April 2023, accepted 2 May 2023, date of publication 11 May 2023, date of current version 30 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275432

RESEARCH ARTICLE

Joint Feature Target Detection Algorithm of Beak State Based on YOLOv5

LINAN ZU, XIAOYU CHU, QIAOMEI WANG, YUNPENG JU, AND MINGYUE ZHANG
College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China
Corresponding author: Mingyue Zhang (zyy_2011@163.com)

This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2021QF031.

ABSTRACT Accurate grasp of chicken body temperature can effectively improve the success rate of caged chicken breeding, by monitoring the number of open-mouthed chickens as a percentage of the total number of chickens and can directly determine whether the chicken body temperature is appropriate. There is no relevant solution to this requirement at present, so this paper proposes a joint feature target detection algorithm based on YOLOv5 to detect the opening and closing state of the chicken mouth. The algorithm improves the YOLOv5 network in the following ways: 1) The improved ResC module is used to reconstruct the backbone network of YOLOv5, which diversifies feature scales and enhances the ability of target feature extraction; 2) Integrate the Transformer module with the four-layer feature pyramid to expand the range of feature fusion and improve the accuracy of feature extraction; 3) The joint feature verification(JFV) module is designed to improve the detection accuracy of small targets by adopting the idea of joint verification of small targets and large targets. Finally, the improved network is used to detect the opening and closing state of the chicken beak on the test set, which is derived from the actual cage chicken breeding environment. The results show that the average accuracy (mAP) of the improved RJ-YOLOv5 algorithm is 85.6%, and the detection accuracy is 7.1% higher than the YOLOv5 algorithm; The video detection frame rate reaches 69 FPS, which can meet the requirements of real-time monitoring of chicken farms.

INDEX TERMS YOLO, computer vision, deep learning, poultry, caged chicken.

I. INTRODUCTION

Since the 21st century, the level of large-scale breeding of broilers has been continuously improved, which has greatly improved the production of broilers [1]. However, High-density large-scale broiler breeding will lead to trampling, disease, and other problems, and if the sick and dead chickens can not be cleaned up in time, infectious diseases will bring serious consequences, and even threaten the production efficiency of the entire farm. In the process of chicken breeding, due to the short growth cycle of caged broilers, the comfort level of the chicken body directly affects its growth quality and mortality. Therefore, the environment of the chicken house should be maintained within the range of the appropriate body temperature of the chicken as far as possible to reduce its morbidity and mortality [2]. However, the ambient temperature and humidity value cannot be directly equivalent to the chicken body temperature [3], and the proportion of open-mouthed chickens in the total number of chickens in the chicken house can determine whether the current ambient temperature and humidity are appropriate, When the breeding temperature is suitable, evacuation is uniform and some chickens open their mouths to breathe. When the chicken group senses a higher temperature, the heat stress performance of the flock is very obvious, and a large number of chickens open their mouths to breathe [4], [5]. that is, when the proportion of open-mouthed chickens in the total number of chickens in the chicken house is kept within a certain threshold, it can effectively reduce the mortality of chickens caused by the temperature and humidity discomfort, thus improving the quality of broilers, and increasing production efficiency.

It is unrealistic to monitor the opening and closing state of the chicken's mouth manually in real-time, while the camera can monitor multiple animals at the same time, the working time is unlimited and there is no visual fatigue. Therefore,

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

machine vision is increasingly widely used in poultry detection systems [6]. The existing research on poultry detection based on machine vision mostly focuses on the detection of poultry behavior [7], For example, Leroy et al. quantified the posture of hens and generated continuous data by identifying the six behaviors of hens (standing, sitting, scratching, pecking, sleeping, combing) [8]. Marco et al. utilized image processing to identify comfortable behaviors and unpopular behaviors to measure animal welfare [9]. Christian et al. proposed a method using deep learning combined with machine vision to evaluate the feather status of laying hens [6].

In the area of machine vision-based target detection algorithms, many improvements have been proposed over the years to improve detection performance. Zhao et al. improved the path aggregation network (PANet) based feature pyramid network by combining the RFB feature enhancement module and ASFF feature fusion strategy to obtain richer feature information for the adaptive fusion of multi-scale features (RA-PANet) [10]. Wang et al. constructed a new feature enhancement module (FEM) and used the attention mechanism to help the detection network focus more on useful features [11]. Dai et al. improved the problem of feature similarity loss during fusion by replacing Conv with CrossConv in the C3 module and enhanced the feature representation [12]. Liu et al. removed the feature layers and prediction heads with poor feature extraction ability, and a new feature extractor with strong feature extraction ability was integrated into the network [13]. Zhao and Zhang et al. used confidence weights to fuse detection frames in multi-resolution images to improve detection accuracy under occlusion conditions [14].Wang et al. introduced convolutional block attention module to the neck, weakening the feature extraction of complex backgrounds by assigning weights and enhancing the representation ability of target features [15]. Li et al. introduced an improved attention module into the residual block, achieving focus on objects and suppression on backgrounds [16].

Among the above detection algorithms, traditional image processing algorithms have low performance and robustness and rely on human experience, while deep learning-based detection algorithms, although some improvement in performance and robustness is observed, for the detection of chicken mouth closure, there are still significant limitations. The reasons are as follows. Reason 1: There is no obvious difference between the opening and closing of the chicken's beak. The difference between the two may only be whether there is a gap between the upper and lower beaks. Reason 2: Free movement of chickens in the cage can cause problems such as mutual occlusion of targets and large-scale changes; Reason 3: When the chicken's mouth opens, the complicated background features between the upper and lower beaks will blend in, which will greatly interfere with the detection of the state of the chicken's mouth; Reason 4: Existing improvements are mostly limited to the structure of the network itself, and the enhancement effect is not obvious.

The main purpose of this study was to solve the detection of small differences and complex background targets and to provide a method for monitoring the living status of caged chickens in commercial cage chicken farming. The method is designed as follows:

(1) Reconstruct the YOLOv5 backbone network and add the improved ResC module to make the extended receptive field reach the feature scale diversity when extracting features, and enhance the feature extraction ability of small targets;
(2) Integrate the Transformer module with the four-layer feature pyramid, expand the range of feature fusion, and make the network better learn the dependency between features, to enrich the context information and improve the accuracy of feature extraction;
(3) The joint feature verification module JFV is designed, which adopts the idea of joint feature verification of associated small targets and large targets, and integrates this module with the network to form an end-to-end network, reducing the false detection rate and improving the detection accuracy of small targets.

At present, each platform does not provide standard data of caged chicken images. The nearly 2000 images used in this experiment are from the actual caged chicken breeding environment, which can be used as a reference data set for subsequent research in the field of intelligent breeding. At the same time, the research results can be combined with the environmental control system of the breeding farm to assist the environmental controller to monitor the physiological signs of the chicken growth process, which has practical research significance and is also a new trend in the development of the breeding industry.

II. RELATED WORK

The target detection model based on deep learning is divided into one-stage detection and two-stage detection algorithms. Among them, the two-stage detection algorithm is mainly RCNN series, and the first-stage detection algorithm is mainly YOLO series. YOLOv5 network has faster detection speed and higher accuracy compared with the earlier versions such as YOLOv4 and YOLOv3. Therefore, this paper selects YOLOv5 as the basic network for improvement.

The YOLOv5 network structure consists of four parts: Input, Backbone, Neck, and Output. The input side mainly includes Mosaic data enhancement, image size processing, and adaptive frame calculation. Backbone mainly contains BottleneckCSP and Focus modules for extracting image feature information. The neck is to augment the features extracted by the backbone network, including the FPN structure and PAN structure, to achieve top-down and bottom-up approaches for a fusion of feature information. The output is used to output the location and category information of the detected targets.

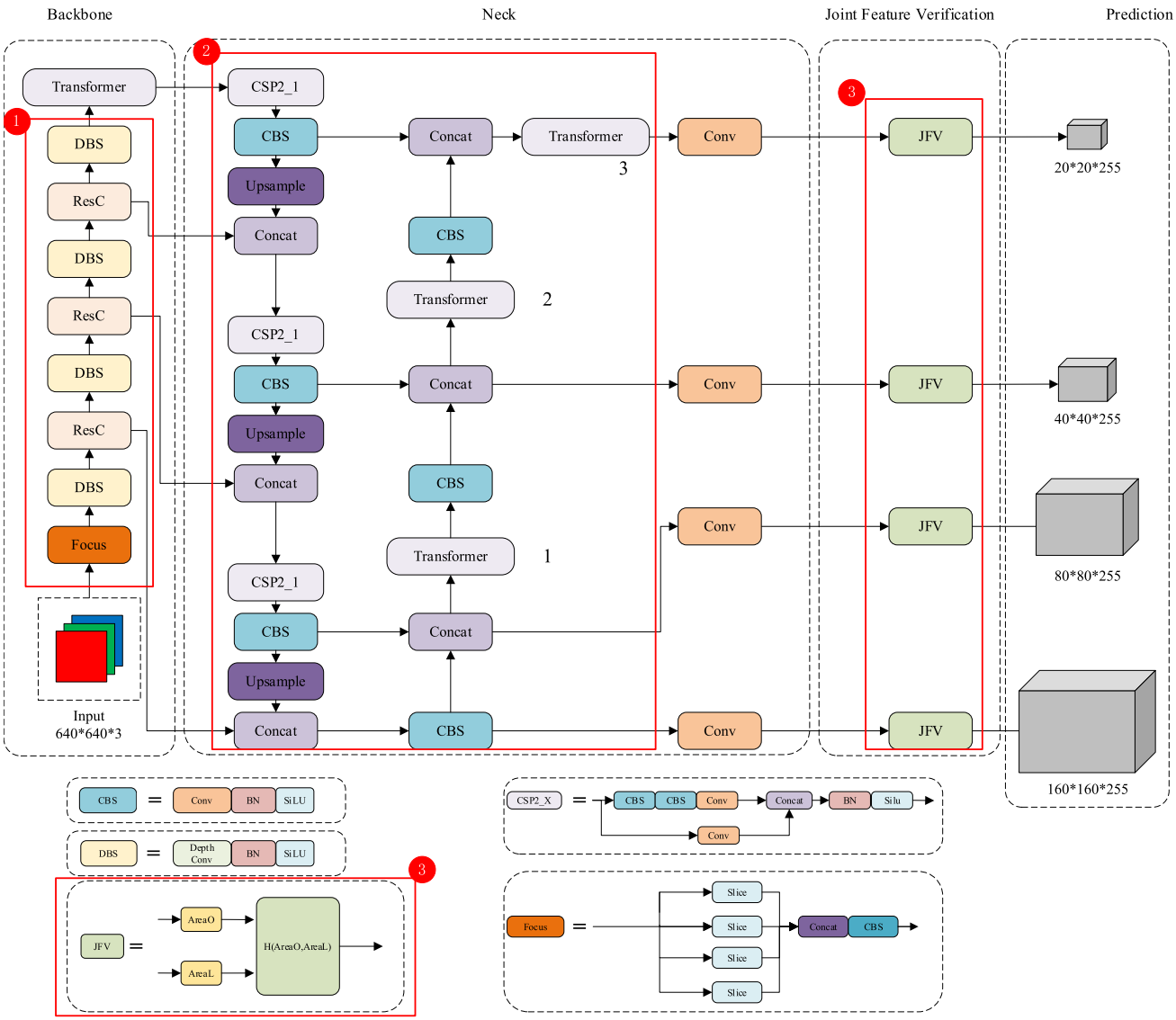

FIGURE 1. Improved RJ-YOLOv5 network structure.

III. METHOD

In this paper, relevant improvements are made based on the YOLOv5 model. The backbone network of YOLOv5 is proposed to be reconstructed using the ResC module, which fuses the C3 module with the multi-layer residual structure of Res2Net [17] to improve the extraction capability of the multi-scale features of the network; to make the fused information of convolution more adequate, the Transformer is fused at the end of the backbone network and in the improved four-layer feature pyramid, respectively, to enhance the global information extraction capability, reduces the interference of background on target features, and improves the model detection performance; the joint feature verification model (JFV) is proposed and integrated at the output of the network for secondary verification of the detection results to improve the model accuracy. The structure of the improved RJ-YOLOv5 network is shown in Fig.1. The above improvements will be presented separately in the following paragraphs.

A. RECONSTRUCTION OF BACKBONE

The images acquired in the chicken coop due to the shooting environment usually have the problems of mutual occlusion between chickens, large changes in the target scale, and the complex background of the chicken coop environment, which lead to the low detection accuracy of the existing models. The multi-scale feature representation of convolutional neural networks is important in object detection tasks and can significantly improve model performance. The C3 module of the original YOLOv5 network maps the input feature map in the channel dimension into two parts, where the BottleNeck structure can increase the difference in the combination of the

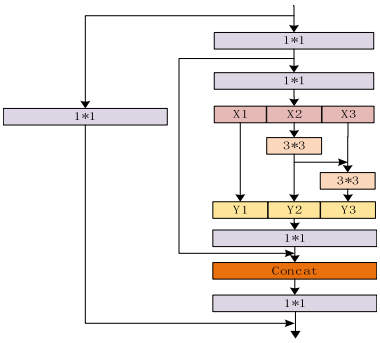

FIGURE 2. ResC module incorporating C3 module and Res2Net module.

propagated gradient information. However, this combination approach still has limitations in the fine-grained level feature representation and the extraction of multi-scale features. Therefore, to optimize the multi-scale representation capability of the model at the granularity level, increase the receptive field, and improve the learning capability of the network, the C3 module of the original YOLOv5 backbone network is fused with the multi-level residual structure of Res2Net [17] to obtain the ResC module in this paper.

As shown in Fig.2, the ResC module first divides the input into two branches, one branch passing through only a normal convolution module, and the other branch entering the multi-level residual structure after passing through a standard convolution layer, and finally concatting the two branches [18]. The multilevel residual structure performs 1∗1 convolution on the input, divides the generated feature map into 3 feature subsets equally, and the feature map size of each feature subset remains the same, but the number of channels is 1/3. The first feature subset does not perform 3∗3 convolution, and the direct output is noted as Y1, the second and third feature subsets are summed with the output of the previous subset and perform 3∗3 convolution operation, noted as Y2 and Y3 respectively, and finally, all outputs are stitched together.

As shown in Fig.1, the ResC module replaces the C3 module in the original network, while the Transformer Encoder module [19] is added at the end of the backbone network to realize the reconfiguration of the backbone part, the new backbone network can change the perceptual field at a finer granularity level, improving the extraction capability of features at different scales and better access to details and global features.

B. FUSION TRANSFORMER AND FOUR-LAYER FEATURE PYRAMID

The feature pyramid of the original YOLOv5 network eventually results in three layers of feature maps with different scale sizes, where the larger-scale feature maps reflect the smaller-scale targets in the image [20]. Although YOLOv5 detection on a multi-scale feature map improves the recognition accuracy of the target, the complex environment of the chicken coop and the distance of the recognition target from the camera makes the target image too small in the image, and

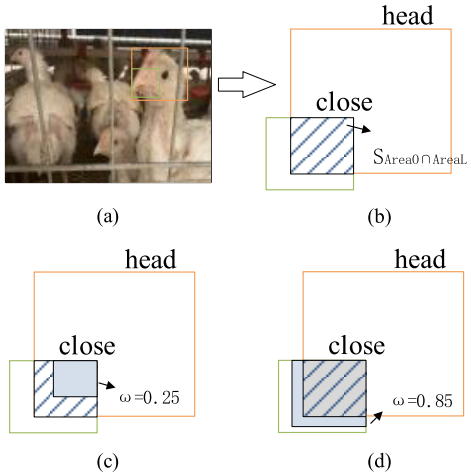

FIGURE 3. Improved RJ-YOLOv5 network structure.

problems such as missed detection and false detection occur when using YOLOv5 for small-scale target detection [21]. Therefore, in order to further improve the performance of the model in detecting targets, this paper proposes a four-scale feature pyramid model incorporating the Transformer self-attention mechanism.

As shown in Fig.1, this paper adds a layer of large-scale feature map 160∗160 on top of the original three-layer pyramid and adds Transformer Encoder modules, i.e., modules 1, 2, and 3, after the feature map stitching. Among them, the multi-headed self-attention mechanism of the Transformer can input the feature information of the feature graph in parallel to achieve the modeling of the dependencies between features and obtain rich contextual information [22]. The four-scale feature pyramid can improve the scale detection range of the network model. Therefore, the four-scale feature pyramid model incorporating Transformer Encoder does feature fusion to improve the low-level semantic information of features while increasing the global information of features and focusing attention on important features, thus reducing the interference of background features on target features and improving the global perception ability of the model.

C. JOINT FEATURE VERIFICATION MODEL

The improvement of the above network is to improve the extraction capability of target features by resetting the network structure while the target remains unchanged. To further optimize the detection accuracy, this paper designs a joint feature verification model from the perspective of extending the target features according to the target characteristics. The model is designed to go beyond the idea of improving the network structure itself to further improve the performance of the network from the perspective of post-processing the prediction results, which can be integrated after any detection network for post-processing to improve the detection performance of the model. When the detected target meets the following conditions, the detection of small targets can be extended to the joint detection and verification of large

and small targets, i.e., condition 1: there are large targets in the detection picture that can be associated with small targets, and the intersection over union (IoU) of the two is greater than 50%; condition 2: the detection accuracy of large targets in the target group is higher than that of small targets.

Based on the idea of extending the detection target, we propose the joint feature verification model, which extracts the category information and location information after detection. As shown in Fig.3(a), ideally to confirm whether the small target is within the large target range, only the four vertices of the small target need to be verified whether they are within the large target range, but most of the image boxes detected by the model have some deviation from the real target box, as shown in Fig.3(b), so the method of verifying the four vertices is not appropriate. To this end, we decided to construct a joint verification function $H(AreaO, AreaL)$ to verify the authenticity of the small target, which is shown in (1).

$$H(AreaO, AreaL) = V(S_{AreaO\cap AreaL} - \omega S_{AreaO})$$
$$V(x) = \begin{cases} 0, x < 0 \\ 1, x > 0 \end{cases} \quad (1)$$

where V(x) is the judgment function, $AreaO$ and $AreaL$ represent the area occupied by the small target and the large target respectively, which contains the location information of the corresponding large and small targets; S_{AreaO} is the area occupied by the small target O; S_{AreaL} is the area occupied by the large target L; $S_{AreaO\cap AreaL}$ is the area of the overlap between the two; ω is the deviation threshold coefficient, which is used to set the permissible deviation range of the small target to prevent the false target from being recognized as the true target with deviation, thus leading to misjudgment. As shown in Fig.3(c), when setting $\omega = 025$, then $S_{AreaO\cap AreaL} < \omega$, then$H(AreaO, AreaL) = 0$, the target is identified as a false target and its detection result is deleted. Also as shown in Fig.3(d), when $\omega = 085$ is set, then $S_{AreaO\cap AreaL} > \omega$, then $H(AreaO, AreaL) = 1$, the target is identified as a true target and its detection result is retained.

The detailed algorithm of the model is shown in Table 1. Firstly, the following parameters are defined: $AreaO$ is the area occupied by the small target, its upper left corner coordinates, width, and height information is stored in the set

$$O = \{O_1 (x_{o1}, y_{o1}, w_{o1}, h_{o1}), O_2 (x_{o2}, y_{o2}, w_{o2}, h_{o1}), \ldots, O_m (x_{om}, y_{om}, w_{om}, h_{om})\}$$
$$O = \{O_1 (x_{o1}, y_{o1}, w_{o1}, h_{o1}), O_2 (x_{o2}, y_{o2}, w_{o2}, h_{o1}), \ldots, O_m (x_{om}, y_{om}, w_{om}, h_{om})\},$$

the corresponding center point is C_o, and its coordinate information is stored in the set

$$C_o = \{C_{o1} (x_{o1}, y_{o1}), C_{o2} (x_{o2}, y_{o2}), \ldots, C_{om} (x_{om}, y_{om})\}.$$

$AreaL$ is the area occupied by the large target, and its upper left corner coordinates and width and height information are included in the set

$$L = \{L_1 (x_{l1}, y_{l1}, w_{l1}, h_{l1}), L_2 (x_{l2}, y_{l2}, w_{l2}, h_{l2}), \ldots, L_n (x_{ln}, y_{ln}, w_{ln}, h_{ln})\};$$

the corresponding center point coordinates are C_L, and its coordinate information is included in the set

$$C_L = \{C_{L1} (x_{L1}, y_{L1}), C_{L2} (x_{L2}, y_{L2}), \ldots, C_{Ln} (x_{Ln}, y_{Ln})\}.$$

As shown in Fig.1, the joint verification model is placed at the output of the network to verify the detection results.

TABLE 1. Joint verification algorithm.

Algorithm JointVerification(AreaO,AreaL)

Input:

$C_o = \{C_{o1}(x_{o1}, y_{o1}), C_{o2}(x_{o2}, y_{o2}), \ldots, C_{om}(x_{om}, y_{om})\}$;

// Small target center point coordinates

$O = \{O_1(x_{o1}, y_{o1}, w_{o1}, h_{o1}), O_2(x_{o2}, y_{o2}, w_{o2}, h_{o1}), \ldots, O_m(x_{om}, y_{om}, w_{om}, h_{om})\}$

// Small target bounding box information

$C_L = \{C_{L1}(x_{L1}, y_{L1}), C_{L2}(x_{L2}, y_{L2}), \ldots, C_{Ln}(x_{Ln}, y_{Ln})\}$ //

Large target center point coordinates

$L = \{L_1(x_{l1}, y_{l1}, w_{l1}, h_{l1}), L_2(x_{l2}, y_{l2}, w_{l2}, h_{l2}), \ldots, L_n(x_{ln}, y_{ln}, w_{ln}, h_{ln})\}$

// Large target bounding box information

Output: $O_{new[h]}$

1:h=0;
2:for i=0 to m
3: j←Find the coordinates $C_{L[j]}(x_{L[j]}, y_{L[j]})$ of the nearest center point to point $C_{o[i]}(x_{o[i]}, y_{o[i]})$ in the set C_L and take its serial number j;
4: $AreaL = [x_l, y_l, w_l, h_l]$ ←Take out the large target bounding box information $L_{[j]}(x_{l[j]}, y_{l[j]}, w_{l[j]}, h_{l[j]})$ corresponding to the serial number j in the set L ;
5: if Point $C_{o[i]}(x_{o[i]}, y_{o[i]})$ is inside the bounding box L
6: $AreaO = [x_o, y_o, w_o, h_o]$ ←Take out the small target bounding box information $O_{[i]}(x_{o[i]}, y_{o[i]}, w_{o[i]}, h_{o[i]})$ corresponding to the serial number i in the set O;
7: if H(AreaO, AreaL) // H(AreaO, AreaL) for 1 verification pass, for 0 verification fail
8: $O_{new[h]}$ ←The information of the small target corresponding to serial number i //Calibration passed, and information corresponding to small targets is stored in format for network output
9: h++;
10: end if;
11: i++;
12: end if;
13:end for;

IV. EXPERIMENT RESULTS

A. CAGED CHICKEN DATA COLLECTION AND PROCESSING

The construction of dataset is an important step to perform deep learning based target detection. In this paper, broilers in cage breeding mode in modern standard broiler farms are

FIGURE 4. Examples of datasets in different lighting environments.

FIGURE 5. Examples of datasets in different hues.

FIGURE 6. Examples of datasets in different contexts.

used as the research object, and photos of caged chickens in real caged chicken farms are collected, using cameras with different pixels as the acquisition equipment, and fixed machine positions are placed at multiple angles and locations in the chicken house for acquisition, and a total of 1700 images are collected. In order to improve the richness of images in the dataset and enhance the detection performance of the trained model, the following points should be considered in the acquisition process:

(1) Acquire images of caged chickens under different lighting intensities (exemplified in Fig.4): the cages in the coop are placed in a stacked manner with sufficient lighting on the upper layer and less light on the lower layer. Photographs of the upper and lower cages were collected to improve the richness of the illuminated environment of the caged chicken images.

TABLE 2. Number of labels in each category.

Category	close	open	head
Quantity	4942	3144	5762

(2) Acquisition of caged chicken pictures with different tones (example is Fig.5): the imaging effect will be different due to the influence of image acquisition equipment and the surrounding environment. Different equipment is used to capture caged chickens in different environments separately to improve the richness of the tones of caged chicken pictures.

(3) Get pictures of caged chickens in different backgrounds (example is Fig.6): the background of the chicken cages will change with different positions in the chicken coop, and the caged chickens in different positions will be photographed to

FIGURE 7. Example of chicken beak labeling. The labels are open, close, and head.

TABLE 3. Environment configuration.

Configuration Name	Configuration Information
Operating System	Ubuntu 20.04.3LTS
Memory	64G
CPU	Intel®Core™i9-12900KF×24
GPU	NVIDIA GeForce RTX 3070
Programming Languages	Python 3.8
GPU Driver	CUDA11.3 CuDNN8.2.0

improve the richness of the background of the caged chicken pictures.

The image data of the captured caged chickens were filtered to remove the low-quality images such as blurred and no valid information, and 1500 valid images were obtained. Labelimg software was used to label the images into three categories: chicken shut up (close), chicken open mouth (open), and chicken head (head). The number of targets in each category is shown in Table 2, and the image labeling examples are shown in Fig.7.

B. EXPERIMENTAL ENVIRONMENT AND PARAMETER CONFIGURATION

The following experimental environment is used in this paper as shown in Table 3. In order to make up for the shortcoming of insufficient number of data sets and strengthen the model training effect, this paper adopts the migration training approach to train the model. The optimizer is Adam, the localization loss function is CIoU Loss, and the Classification loss and Confidence loss are BCE Loss. The training parameters are set as follows: the initial learning rate is 0.01; the momentum parameter is 0.937; the weight decay coefficient is 0.0005; the number of iterations is 300, the batch size is 8, and the input image size is 640∗640.

C. EVALUATION METRICS

In this paper, the Precision(P), Average Precision(AP), mean Average Precision(mAP), and detection rate FPS of the generic target detection evaluation metrics are used as the evaluation metrics of the model to assess the target detection accuracy and the computational speed of the model. The Precision is the proportion of all targets predicted by the model that are correctly predicted, also called the accuracy rate [23], as shown in (3).

$$\text{Precision}=\frac{\text{TP}}{(\text{TP}+\text{FP})} \tag{2}$$

where TP is the true case, the true category is positive and the predicted category is positive; FP is the false positive case, the true category is negative and the predicted category is positive.

The P-R curve, or Precision-Recall curve, reflects the predictive ability of the model. The mAP is the area under the P-R curve, reflecting the average accuracy of each category. The mAP is the average of the AP of each category, reflecting the accuracy of category recognition globally [21]. The mAP is shown in (4).

$$\text{m}AP = \frac{1}{k}\sum_{i=1}^{k} AP(i) \tag{3}$$

D. ABLATION EXPERIMENT

To verify the validity of the three improvements proposed in this paper, we give the mAP, number of parameters, computational volume, and FPS of all models by ablation experiments. Ablation experiments are designed as follows. Experiment 1 represents the detection results of the YOLOv5 network; Experiments 2-4 represent only one improved method based on the original YOLOv5 network, respectively, and Experiments 5-7 represent only one improved method eliminated based on the RJ-YOLOv5 network proposed in this paper, respectively. The improved methods are: reconstruction of the backbone network with the improved ResC module, construction of a four-layer pyramid with fused Transformer, and integration of the joint feature verification model at the output; Experiment 8 indicates the model proposed in this paper.

In this experiment, the chicken head as an auxiliary detection label was not included in the experimental requirements, so the mAP in the ablation experimental results only represents the mean average precision of the detection of the opening and closing status of the chicken beak (the results are rounded to one decimal place). The experimental results are shown in Table 4.

From the above Table, When IoU takes 0.5,we can see that: the mAP of the YOLOv5 algorithm represented in Experiment 1 is 78.4%, which is less accurate. Experiment 2 and Experiment 5 show that the backbone network reconstructed with the Resc module has an improved effect on the detection of chicken beaks. From the results of Experiment 3 and Experiment 6, it can be seen that the detection accuracy is improved to some extent by adding a four-layer pyramid of fusion Transformer to the neck network. Experiments 4 and 7 show that the addition of the joint feature verification model leads to a significant improvement in the detection precision of the model. Experiment 8 shows the detection results of the

TABLE 4. Results of ablation experiments.

Experiment	ResC	Transformer -Four-layer	JFV	mAP@0.5	mAP@0.5:0.95	Parames	FPS
1				78.4	40.3	70.3M	82
2	√			79.1	41.3	69.9M	77
3		√		79.6	42.1	102.1M	68
4			√	81.4	42.6	70.3M	79
5		√	√	83.3	44.8	102.2M	74
6	√		√	83.7	44.6	70.9M	80
7	√	√		81.3	44.1	101.9M	72
8	√	√	√	85.6	48.5	101.9M	69

FIGURE 8. Examples of images used to test the model.

FIGURE 9. Examples of YOLOv5 model test results.

FIGURE 10. Examples of RJ-YOLOv5 model test results.

final proposed network model in this paper, which improves the accuracy by 7.2% compared to the original YOLOv5 network. When the IoU is taken as 0.5:0.95, the mAP of the original YOLOv5 is 40.3%, and the mAP of the algorithm proposed in this paper is 48.5%, which has obvious effect in the improvement of small target detection accuracy.

Fig.8 shows the comparison of YOLOv5 and RJ-YOLOv5 test results, where Fig.8 shows the original figure, Fig.9 shows the YOLOv5 test results, and Fig.10 shows the test results of the RJ-YOLOv5 model proposed in this paper. From the above experimental data and image comparison, it can be seen that the detection results of YOLOv5 for

TABLE 5. Results of ablation experiments.

Model	mAP@0.5(%)	GFLOPs	FPS(f/s)
SSD	55.8	60.38	48
Faster RCNN	63.2	12.9	23
YOLOv4	76.8	119.8	76
YOLOv5	78.4	205.7	82
YOLOv7	79.8	189.9	77
YOLOv8	81.6	267.8	88
Ours	85.6	228.7	69

targets with small feature scales in images have the problem of missed detection, whereas the RJ-YOLOv5 model reconstructs the backbone network by the ResC module, incorporates the four-layer pyramid structure of Transformer, and adds the joint feature verification module to enhance the detection capability of the network for small targets. This enables the RJ-YOLOv5 network to accurately identify targets with small feature scales and partial occlusions in images, which has significantly improved the detection effect compared with the original network.

E. COMPARISON OF MAINSTREAM ALGORITHMS

The algorithm proposed in this paper is compared with the current mainstream target detection algorithms, and the experimental results are shown in Table 5. From the table, the mAP of SSD [24] and Faster RCNN [25] is 55.8% and 63.2% respectively, and the detection accuracy of these two algorithms is around 60%, therefore, they are not suitable for detecting the opening and closing of chicken beaks. Higher detection accuracy and faster detection speed of YOLOv4 [26], YOLOv5, and YOLOv7 [27] compared to the low detection accuracy of SSD and Faster RCNN. However, there is still room for its detection accuracy to rise. The detection accuracy of the RJ-YOLOv5 network proposed in this paper is 85.6%, which is a significant improvement compared with YOLOv4, YOLOv5, and YOLOv7. Although the detection speed is slightly reduced, it still meets the demand for real-time monitoring of the opening and closing of chicken beaks.

V. CONCLUSION

The physiological sign of the opening and closing state of the chicken's mouth can be used as one of the reference indicators to assist in measuring the feeding quality of the chicken.

To study the problem of detecting the opening and closing status of chicken beaks in the complex context of chicken coops, this paper proposes the RJ-YOLOv5 target detection algorithm based on the YOLOv5 network structure. This paper reconstructs the backbone of YOLOv5 and proposes to use the improved ResC module to extract chicken beak features; proposes to fuse Transformer's four-layer feature pyramid layer to fully fuse feature information; meanwhile, a joint feature verification (JFV) method is proposed to solve the problem of dense distribution and occlusion among chickens and improve the detection accuracy of small targets. Experimental results show that the improved RJ-YOLOv5 algorithm can achieve effective detection of occluded targets and small targets in complex backgrounds.

The joint feature verification model proposed in this paper can be applied to a variety of scenarios where there are large and small targets that can be jointly verified, substantially improving the detection effectiveness of the model. In addition, although this paper takes caged chickens as the research background, the method can also be applied to the feeding management system of other livestock and poultry to assist environmental controllers in monitoring their physiological signs, thus improving livestock and poultry production.

The contribution of this work is twofold. One is the application of object detection in new areas, and the other is the improved detection of small objects and complex backgrounds by the YOLOv5 model.

REFERENCES

[1] M. Tixier-Boichard, "From the jungle fowl to highly performing chickens: Are we reaching limits?" *World's Poultry Sci. J.*, vol. 76, no. 1, pp. 2–17, Jan. 2020.
[2] G. Abbas, "Prospects and challenges of adopting and implementing smart technologies in poultry production," *Pakistan J. Sci.*, vol. 74, no. 2, p. 108, 2022.
[3] M. G. L. Cândido, I. F. F. Tinôco, L. F. T. Albino, L. C. S. R. Freitas, T. C. Santos, P. R. Cecon, and R. S. Gates, "Effects of heat stress on pullet cloacal and body temperature," *Poultry Sci.*, vol. 99, no. 5, pp. 2469–2477, May 2020.
[4] B. Liu, "Temperature management of cage broilers during brooding period," (in Chinese), *Animals Breeding Feed*, vol. 20, no. 3, pp. 46–47, 2021.
[5] S. S. Niu and F. Liu, "Research on summer ventilation technology of 5-row, 4-layer layer cages for laying hens," (in Chinese), *Shandong J. Animal Sci. Veterinary Med.*, vol. 43, no. 3, pp. 11–14, 2022.
[6] C. Lamping, M. Derks, P. G. Koerkamp, and G. Kootstra, "ChickenNet—An end-to-end approach for plumage condition assessment of laying hens in commercial farms using computer vision," *Comput. Electron. Agricult.*, vol. 194, Mar. 2022, Art. no. 106695.
[7] D. Wu, D. Cui, M. Zhou, and Y. Ying, "Information perception in modern poultry farming: A review," *Comput. Electron. Agricult.*, vol. 199, Aug. 2022, Art. no. 107131.
[8] T. Leroy, E. Vranken, E. Struelens, B. Sonck, and D. Berckmans, "Computer vision based recognition of behavior phenotypes of laying hens," in *Proc. ASAE Annu. Meeting*, 2005, p. 1.
[9] M. Sozzi, G. Pillan, C. Ciarelli, F. Marinello, F. Pirrone, F. Bordignon, A. Bordignon, G. Xiccato, and A. Trocino, "Measuring comfort behaviours in laying hens using deep-learning tools," *Animals*, vol. 13, no. 1, p. 33, Dec. 2022.
[10] W. Zhao, M. Syafrudin, and N. L. Fitriyani, "CRAS-YOLO: A novel multi-category vessel detection and classification model based on YOLOv5s algorithm," *IEEE Access*, vol. 11, pp. 11463–11478, 2023.
[11] H. Wang, Y. Xu, Y. He, Y. Cai, L. Chen, Y. Li, M. A. Sotelo, and Z. Li, "YOLOv5-fog: A multiobjective visual detection algorithm for fog driving scenes based on improved YOLOv5," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
[12] G. Dai, L. Hu, J. Fan, S. Yan, and R. Li, "A deep learning-based object detection scheme by improving YOLOv5 for sprouted potatoes datasets," *IEEE Access*, vol. 10, pp. 85416–85428, 2022.
[13] Z. Liu, Y. Gao, Q. Du, M. Chen, and W. Lv, "YOLO-extract: Improved YOLOv5 for aircraft object detection in remote sensing images," *IEEE Access*, vol. 11, pp. 1742–1751, 2023.
[14] J. Zhao, X. Zhang, J. Yan, X. Qiu, X. Yao, Y. Tian, Y. Zhu, and W. Cao, "A wheat spike detection method in UAV images based on improved YOLOv5," *Remote Sens.*, vol. 13, no. 16, p. 3095, Aug. 2021.
[15] L. Wang, Y. Cao, S. Wang, X. Song, S. Zhang, J. Zhang, and J. Niu, "Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN," *IEEE Access*, vol. 10, pp. 60622–60632, 2022.

[16] S. Li, Y. Li, Y. Li, M. Li, and X. Xu, "YOLO-FIRI: Improved YOLOv5 for infrared image object detection," *IEEE Access*, vol. 9, pp. 141861–141875, 2021.
[17] S. Gao, M. Cheng, K. Zhao, X. Zhang, M. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.
[18] W. Zhou, Y. Chen, C. Liu, and L. Yu, "GFNet: Gate fusion network with Res2Net for detecting salient objects in RGB-D images," *IEEE Signal Process. Lett.*, vol. 27, pp. 800–804, 2020.
[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.
[20] Y. Sun, W. Liu, Y. Gao, X. Hou, and F. Bi, "A dense feature pyramid network for remote sensing object detection," *Appl. Sci.*, vol. 12, no. 10, p. 4997, May 2022.
[21] Z. Xue, H. Lin, and F. Wang, "A small target forest fire detection model based on YOLOv5 improvement," *Forests*, vol. 13, no. 8, p. 1332, Aug. 2022.
[22] Y. Yu, J. Zhao, Q. Gong, C. Huang, G. Zheng, and J. Ma, "Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5," *Remote Sens.*, vol. 13, no. 18, p. 3555, Sep. 2021.
[23] J. Hu, G. Li, H. Mo, Y. Lv, T. Qian, M. Chen, and S. Lu, "Crop node detection and internode length estimation using an improved YOLOv5 model," *Agriculture*, vol. 13, no. 2, p. 473, Feb. 2023.
[24] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
[25] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[26] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

LINAN ZU received the Ph.D. degree. She is currently an Associate Professor and a Master's Tutor. Her research interests include autonomous navigation and machine vision.

XIAOYU CHU is currently pursuing the M.S. degree in electronic information with the Qingdao University of Science and Technology, Qingdao, China. Her research interests include computer vision, artificial intelligence, and image processing.

QIAOMEI WANG is currently pursuing the M.S. degree in electronic information with the Qingdao University of Science and Technology, Qingdao, China. Her research interests include computer vision, image processing, robot positioning, and navigation.

YUNPENG JU received the Ph.D. degree. He is currently a Master's Tutor. His research interests include robot control and navigation, and analysis of mechanical system dynamics model.

MINGYUE ZHANG received the M.S. degree from the Huazhong University of Science and Technology, in 2011, and the Ph.D. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2014. She is currently an Associate Professor and a Master's Tutor with the Qingdao University of Science and Technology. Her research interests include intelligent control of nonlinear systems and machine vision.