IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 March 2023, accepted 27 March 2023, date of publication 17 April 2023, date of current version 26 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3267483

RESEARCH ARTICLE

An Efficient and Lightweight Pigeon Age Detection Method Based on LN-STEP-YOLO

JUN YAO, ZECHEN SHI, HANJING JIANG, QINGXIU WANG, NINGXIA CHEN, TING WU, HAIXIA XU, LING YANG, AND JUAN ZOU

Guangdong Provincial Food Safety Traceability and Control Engineering Technology Research Center, Guangzhou 510225, China
School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong 510225, China

Corresponding authors: Ling Yang (yang98613@163.com) and Juan Zou (lucy_631@163.com)

This work was supported in part by the Key Research and Development Program of Guangdong Province under Grant 2020B0202080002, Grant 2021B0202030001, and Grant 2019B020215001; in part by the Department of Education of Guangdong Province Bureau under Grant 2020ZDZX1060; in part by the Fund for Science and Technology from Guangdong Province under Grant 2020A1515010834 and Grant 2018A0303130034; in part by the Guangzhou Science and Technology Plan Project under Grant 202002030154 and Grant 201903010063; in part by the Fund for Guangzhou Rural Science and Technology Commissioner Project under Grant 20212100058; in part by the Fund from Guangzhou Science and Technology Bureau under Grant 201803020033; and in part by the Guangdong Provincial University Key Field Special Project under Grant 2022ZDZX4019.

ABSTRACT Deploying pigeon age detection model on edge equipment can solve the problem of video transmission delay and reduce the pressure of network transmission. Based on the deployment of edge devices, we made some improvements to You Only Look Once version 5 (YOLOv5) to form a new, lightweight, high-performance detector named LN-STEP-YOLO. In order to reduce the size of the model, we halved the number of channels in the YOLOv5s model. However, the decrease in the number of channels brings some problems, such as low global information acquisition, retention of image redundancy information, attention deficit, etc. To solve these problems, we did the following work. First, we proposed a new convolution structure with the effect of a large convolution kernel, StepConv. Second, a 2×2 convolution with step size 2 was used at the input to split each image into individual patches. Third, External Attention (EA) was introduced in the bottleneck structure. Fourth, a modified extremely separated convolutional block (XsepConv) was used for downsampling. Finally, we replaced the batch normalization (BN) of all non-downsampled layers with layer normalization (LN). The results showed that the improved algorithm outperformed generic lightweight networks such as Mixnet, Mobilenetv3, and Ghostnet to distinguish small-sized, overlapping pigeons, achieving 92.8% mean average precision (mAP) at about 17% of YOLOv5s parameters, 0.1% lower than that achieved by use of YOLOv5s. In addition, the improved method had 3.7G floating point operations (FLOPs) and 1.25G parameters, which allowed the detection of the growth stages of pigeons in real environments and provided a reference to guide placement of feeders in automated pigeon farming.

INDEX TERMS Pigeon, accurate feeding, overlap, greater receptive field, layer normalization.

I. INTRODUCTION

Manual feeding of pigeons is very labor intensive and time-consuming work. With the development of technology, machines can replace manual labor. Current machine feeding, usually aisle feed troughs [1], provide a fixed amount of feed by averaging the amount needed for a certain number of pigeons of a certain age, and do not allow manual on-demand feeding. However, this kind of system is problematic because the needs of the birds in a particular cage can vary due to changes in pigeon feeding patterns and growth stages, or if pigeons die or have to be transferred to a different cage [2], [3], [4]. Thus, there can be significant cage-to-cage variation in the required amount of feed. Too little food will limit the health development of the pigeons, and too much food will waste unconsumed feed, thus increasing the cost of raising pigeons. To address the inability of a feeding machine to determine how much feed should be provided, several artificial intelligence approaches have been tested.

Artificial intelligence has been increasingly applied to smart agriculture [5], including water quality prediction [2], agricultural product traceability [7], and object

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

detection [8], [9], [10], [11]. RFID [12] and spectral methods [13] can work to detect pigeons, but these methods are too expensive for routine use. Acoustic methods [14] cannot work in the noisy environment of pigeon cages. However, computer vision methods (object detection) require only one camera and can perform accurate detection.

Overall, the current mechanical equipment in pigeon farms does not consider the differentiated feeding demands of individual pigeons, while computer vision methods can efficiently detect and monitor individual pigeons. Therefore, we adopt computer vision methods for our research. Based on the small size and overlap of young pigeons, we proposed a LN-STEP-YOLO algorithm. Our contribution is as follows: First, a new convolution structure (StepConv) is proposed to obtain more global information and improve the detection performance of small targets. Second, the Focus layer is improved to reduce redundant image information. Third, introduce External Attention (EA) block to reduce the impact of complex environments. Fourth, since StepConv is not applicable in the downsampling layer, XsepConv is used instead. Fifth, replace BN layer with LN layer in non-downsampling layer to avoid excessive BN layer to reduce model performance. The result shows that this method has excellent detection performance for pigeons with small size and small overlap in the complex environment of pigeon house.

II. RELATED WORK

At present, there are few studies on pigeon age detection. Therefore, we need to refer to similar studies in other fields. Systems for pigs [8], cattle [10], and chickens [11], [15], [16], [17] detect animals as adult individuals. These systems often utilize additional modules to detect behavioral, health, and other indicators, and these additional modules may slow detection speed. In pigeon breeding, the young pigeons cannot feed themselves and need to be fed by the adults. The most economical practice is the "2+3" or "2+4" models (two adult pigeons raising three or four young pigeons in a cage) [2], [3], and often a pigeon house will have multiple cage with young pigeons at different stages, for different feed requirements. Recent research in the area of object detection has focused on plants [18], [19], [20], and there are obvious differences being plant detection and pigeon detection, including the need to distinguish the sky in the background from detecting plants in outdoor settings [18]. Additionally, plants in different growth stages may exhibit clear degree of differentiation, such as the color change of tomato [19] or the closure of flowers [20], thus allowing the detection of plant growth status. In contrast, the growth stages of pigeons reflect a slow and gradual process of change, and these stages often need to be identified by an expert.

Several commonly used lightweight networks, such as Mobilenetv3 [21], Mixnet [22], Ghostnet [23], and YOLOv5 [24], emerged as good options for edge devices. Mobilenetv3 was used to construct network structures for mobile platforms based on neural architecture search (NAS). Mixnet improved detection accuracy by combining multiple convolutional kernels of different sizes. Ghostnet reduced computation of feature maps to improve detection speed. YOLOv5 simplified the YOLOv4 model and introduced many training techniques (such as mosaic data augmentation) to achieve the performance of two-stage detectors at a small cost. In the early work, we experimented with Mobilenetv3 [21], Mixnet [22], Ghostnet [23], YOLOv5s [24] and other lightweight networks for general purpose. Experimental results show that YOLOv5s model has the highest mAP value (Section IV-C) and is the best choice for baseline model. However, the number of parameters and calculation amount of YOLOv5s model are much larger than other networks, so we will improve YOLOv5.

In general, the lack of a suitable algorithm to guide the feeding machine, the complex environment of a pigeon house (including interference from light, feces, and other disturbances), the difficult in assessment of pigeon growth stages, and the difficulty in detecting young pigeons have limited used of smart feeding systems for pigeon care. In this work, we modified the YOLOV5n network as LN-STEP-YOLO, a method that can identify the growth stage of each pigeon in each cage with fast and high accuracy in an edge device environment, providing data that can be used to guide a feeding machine to feed on demand.

III. LN-STEP-YOLO MODEL

A. THE STRUCTURE OF LN-STEP-YOLO MODEL

YOLOv5 is a mainstream algorithm developed for object detection in a single stage, with advantages of high speed, accuracy, and easy to use. We achieved good results in pigeon age recognition using YOLOv5s, but this method was not modified based on the edge deployment environment, making it difficult to detect small-sized pigeons and distinguish overlapping pigeons. Based on our previous experience with YOLOv5s, we first halved the number of channels of YOLOv5s to YOLOv5n, which reduced the parameters by nearly 75%, but the mean average precision (mAP) of the model only decreased by 1.6%, suggesting use of YOLOv5n as the baseline. Our subsequent work was based on a smaller YOLOv5n. The overall structure of YOLOv5 [Figure 1 (a-c)] includes the backbone (CSPDarknet) for extracting features, neck [Feature Pyramid Network (FPN) and Path Aggregation Network (PAN)] for fusing features, and head for output.

Figure 2 shows the improved network structure (LN-STEP-YOLO). To distinguish it from the standard network structure, we renamed C3 as C3s (C3 with StepConv). The details of the improvements are described in subsections III-B-III-F.

This study used the loss functions L, Intersection over Union (IOU) loss L_B, Category loss L_C, Object Loss L_O, and three constants (β, γ, and ζ), to define the loss function of YOLOv5, as follows:

$$L = \beta L_B + \gamma L_c + \zeta L_O \tag{1}$$

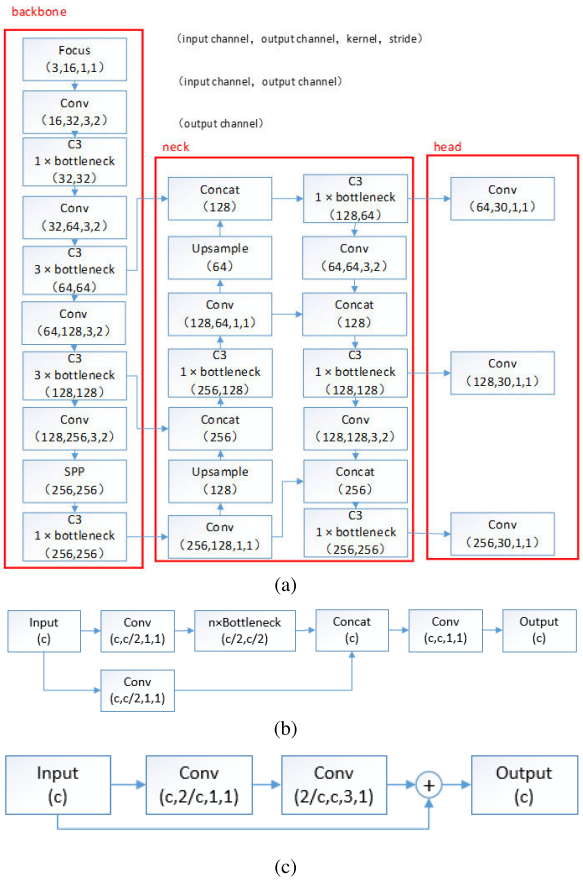


FIGURE 1. YOLOv5 network structure and its modules. (a) YOLOv5 network structure. (b) C3 block. (c) Bottleneck block. Number of input channels "c" and number "n" are derived from (a).

Using the network output x_n, Ground truth y_n, and using L_C as an example,L_C was defined as:

$$L_c = -\sum \left[y_n \times \log x_n + (1 - y_n) \times \log (1 - x_n) \right] \quad (2)$$

IOU loss is calculated using CIOU [25]. Using the intersection between prediction box and Ground truth IOU, the Euclidean distance between the prediction box and Ground truth center coordinates $\rho^2(b, b^{gt})$, the diagonal distance of the minimum closure area between prediction box and Ground truth c, the balance parameters α, and the prediction box and Ground truth's aspect ratio v, the L_B was defined as follows:

$$L_B = 1 - CIOU \quad (3)$$

$$CIOU = IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (4)$$

$$\alpha = \frac{v}{1 - IOU + v} \quad (5)$$

$$v = \frac{4}{\pi^2}(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2 \quad (6)$$

B. StepConv

Many neural networks [such as Inception [26] and Mixnet [22]] improved the network field of view by using

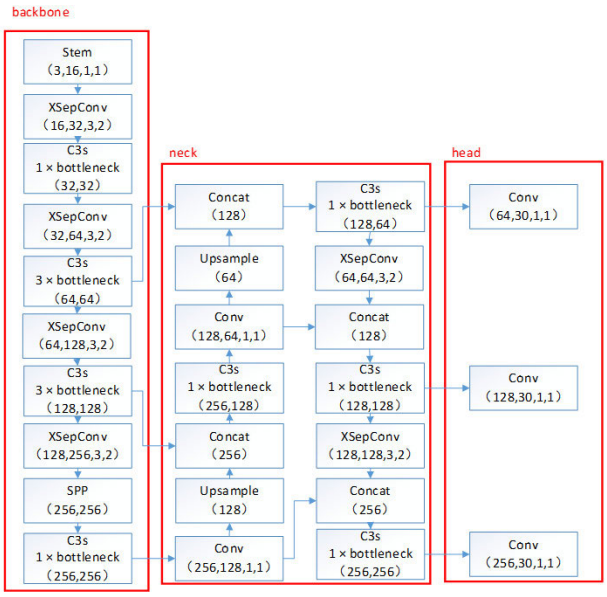


FIGURE 2. LN-STEP-YOLO structure. The meanings of the numbers in parentheses are the same as in Figure 1.

multiple larger convolutional kernels for effectively improved network accuracy. Large convolution kernels expanded the receptive field, but also increased the computational load, so Ultralytics [24] did not use large convolution kernels in YOLOv5 to maintain a lightweight model. Assuming a parallel expansion like for InceptionConv [Figure 3 (a)], 3 × 3, 5 × 5, and 7 × 7 convolutional operations were performed. The computational overhead incurred was huge compared to the use of only 3 × 3 convolutional kernels. MixConv [Figure 3 (b)] [22] was optimized on this basis by scaling the internal channels to perform 3 × 3, 5 × 5, and 7 × 7 convolution operations, but this still increased the computational overhead by using large convolution kernels. To address the extra computational overhead caused by large convolutional kernels, Vgg [27] used multiple consecutive 3 × 3 convolutional kernels to replace one large convolutional kernel. In this way, multiple 3 × 3 convolution kernels can be directly used to eliminate the large convolution kernels used in Mixnet to reduce computation, but the computational overhead was still higher than a standard 3 × 3 convolution. Inspired by Ghostconv's [23] [Figure 3 (c)] multiplexed channels, we next proposed a new convolution structure "StepConv" [Figure 3 (d)], in which a 3 × 3 convolution was decomposed into three 3 × 3 convolutions in a step-like progression. By passing the receptive field in this recursive manner, the receptive field could be expanded with reduced parameters. By superposition of these convolution, the receptive field of StepConv was equivalent to the effect of different convolution kernels (3 × 3, 9 × 9, and 27 × 27).

The parameters and FLOPs of the model were calculated based on the number of channels identified in Figure 3. In terms of parameters, StepConv was $\frac{11}{16}$ of the standard convolution, and in terms of FLOPs, StepConv was slightly less than $\frac{11}{16}$ of the standard convolution.

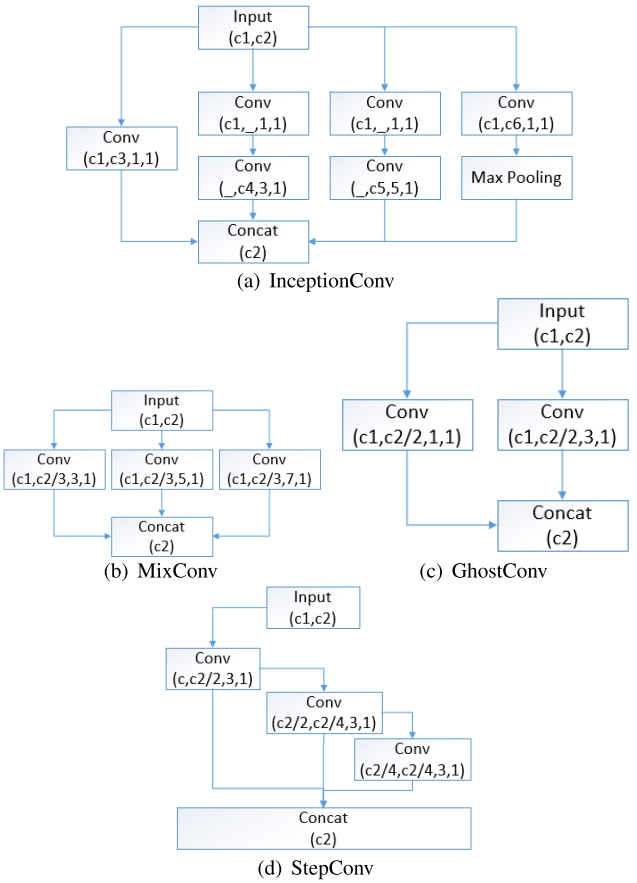


FIGURE 3. Different Conv structures. (a) InceptionConv, (b) MixConv, (c) GhostConv, (d) StepConv. The numbers in parentheses mean the same as in Figure1. Here, 'c1', 'c2' and '_' represent the number of inputs, outputs, and the hidden channels of the whole structure, respectively.

TABLE 1. Comparison of improvement.

Experimental	mAP	Parameters (K)
STEP-YOLO	0.917	1570810
STEP-YOLO (with stem layer)	0.921	1571626

C. STEM LAYER

To deal with the redundancy inherent in natural images [28], we utilized the basis of Transformer [29], MLP-Mixer [30], and Convnet [28] to segment input images into a relatively small 2×2 convolution, where the input image could be partitioned into individual patches, i.e., stem layers [Figure4 (b)]. The standard YOLOv5 was a slice of the original image [Focus, Figure4 (b)], and the stem layer deepened the convolution depth of the model. Since the input image had only three channels, the cost of splitting patches by convolution was extremely small. From the results shown in Table 1, the stem layer increased the parameters only by 816k, but increased the mAP by 0.4.

D. EXTERNAL ATTENTION (EA)

People can rapidly detect and understand complex images, but it is a complicated process to determine the most

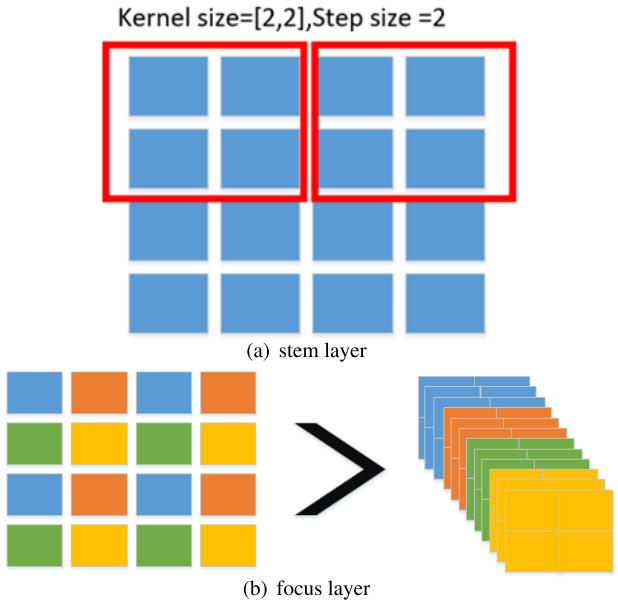


FIGURE 4. Network first layer structure. (a) stem layer, (b) focus layer.

efficient strategies for machines to effectively and quickly detect and make decisions. Self-attention is an exciting direction in attention-related research [31], but the computational complexity of self-attention was squared and ignores potential connection between different samples. EA [32] is proposed to solve this problem, and contains only two linear and two normalization layers and has linear computational complexity.

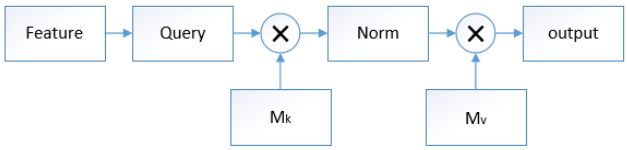


FIGURE 5. The structure of external attention (EA).

Figure5 shows the structure of EA. Query is a 1×1 convolution, M_k, M_v are linear layers and are learnable parameters independent of the input, serving as memory for the entire training data set. EA is formulated as follows:

$$F_{out} = Norm(FM_k^T)M_v \tag{7}$$

where input feature $F \in R^{(N \times d)}$, N is the number of image pixels, d is the number of channels in the input feature map, memory cell $M_k M_v \in R^{(S \times d)}$, S is the hyperparameter with a value of 64 [32], and Norm is the double-normalization method [32].

The double-normalization process was performed as follows. First, Equation 7 was established. Second, Softmax (Equation 8) was used on the second dimension of $\tilde{a}_{(i,j)}$. Finally, L1 Normalization was performed on the third dimension of $\tilde{a}_{(i,j)}$ (Equation 9). After this calculation, the output of

double-normalization $a_{(i,j)}$ was obtained.

$$\tilde{a}_{(i,j)(i,j)} = FM_k^T \tag{8}$$

$$\hat{a}_{(i,j)} = \frac{exp(\tilde{a}_{(i,j)})}{\sum_{i=0}^{k} exp(\tilde{a}_{(k,j)})} \tag{9}$$

$$a_(i,j) = \hat{a}_{(i,j)} \sum_{i=0}^{k} exp(\hat{a}_{(i,j)}) \tag{10}$$

Figure6 (a - f) show that with EA, the model better extracted the effective features (the region where the pigeons are located).

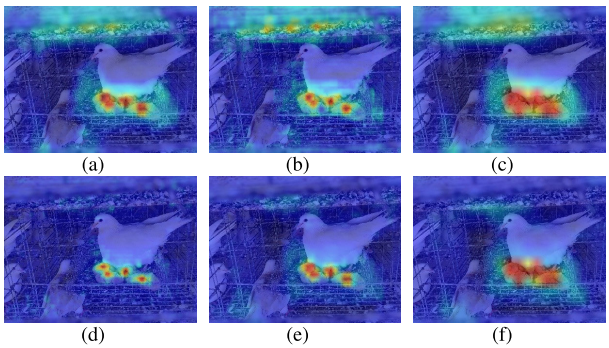


FIGURE 6. Input feature maps of the three output layers superimposed on the original image. (a - c) without EA, (d - f) with EA. The color shifts from blue to red, and the closer the color is to red, the higher the attention of the region.

E. DOWNSAMPLING IMPROVEMENT

Downsampling results in information loss, and it is common practice [33], [34], [35] to expand the output channel of the convolution operation to compensate for this information loss. If output channel expansion is used on top of StepConv, the first convolution of StepConv for channel expansion poses two problems. First, the computation of this module increases exponentially. Second, with multiple downsampling layers and multiplying the number of channels, the computation of the whole model explodes, which is not consistent with the design concept of stepwise convolution. Therefore, we chose to optimize the downsampling layer in the modified, extremely separated convolutional block [XsepConv [36]].

The standard XSepConv [Figure7 (a)] was concatenated in the order of 2×2 (step=1), 1×3 (step=2), 3×1 (step=1) DW convolution, SE layer, and 1×1 standard convolution. The improved XSepConv [Figure7 (b)] separated the 2×2 convolution (stem) independently, in parallel with the subsequent structure. The mAP values were equal before and after this modification, but the improved XSepConv brought more "Patchify" layers to the model, and the computational overhead of the improved XSepConv was slightly reduced compared to the standard XSepConv due to the increased step length.

F. LAYER NORMALIZATION (LN)

After incorporating the improvements described in Section III-E (Table 2), we observed a substantial reduction

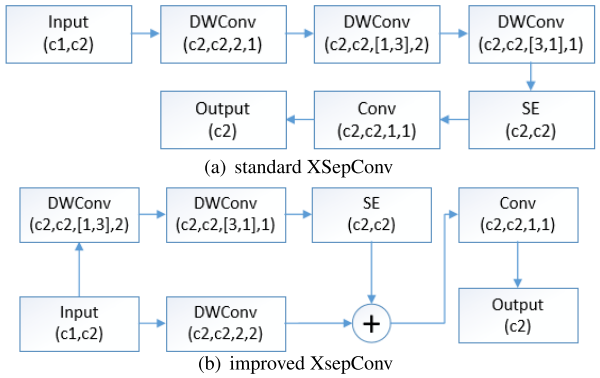


FIGURE 7. Comparison of XSepConv structures. (a) standard XSepConv, (b) improved XsepConv.

TABLE 2. Comparison of improvements.

Experiment	mAP	FLOPs (G)	Parameters (G)
YOLOv5n	0.913	5.0	1.78
STEP-YOLO (Section III.B-E)	0.907	4.2	1.25

TABLE 3. Experimental comparison of LN replacement of BN.

Experiment	mAP
Standard	0.907
In all network layers (I)	0.913
Only the downsampling layer (II)	0.915
Only the non-downsampling layer (III)	0.928

in the computational effort and parameters of the model. However, the mAP of the improved model was even lower than the baseline (YOLOv5n) by 0.04. This was not the result we wanted, so we revisited the structure of the model and found that our proposed network included a large number of BN layers (a StepConv module and an XsepConv module with three and four times the volume of BN layers as a standard convolution, respectively). The BN was still the preferred solution for most vision tasks, despite the fact that too many BNs can have a bad effect on the model. A simpler LN is often used in Transformers, and given that the previous improvements significantly changed the model structure, we next evaluated the LN replacement of BN [28]. The formula for LN is as follows:

$$y = \frac{x - E(x)}{\sqrt{Var(x) + \varepsilon}} \times \gamma + \beta \tag{11}$$

where $E(x)$ represents the mean, $Var(x)$ is the standard deviation, γ and β are the training parameters, and ε is $1e-5$ to avoid $Var(x)$ equal to 0.

Downsampling can reduce the resolution, so we did four experiments (Table 3) differentiated by the downsampling layer. We tested the standard model (the model in Section III-E), keeping the same BN; Experiment I, with BN replaced with LN in all network layers; Experiment II, BN replaced with LN only in the downsampling layer; and Experiment III, replacing BN with LN only in the non-downsampling layer.

The experiments in Table 3 showed that LN replacement of BN was feasible, with improvements for all three variations. Surprisingly, Experiment III exhibited better performance and even made up for the reduced mAP in Section III-E. Therefore, we chose to use LN in the non-downsampling layer (Experiment III).

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental environment used was Windows 10 64-bit operating system, Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz 2.50GHz, 64GB RAM, NVIDIA GeForce RTX 2080 Ti graphics card, and the same training parameters as YOLOv5 [24] were used in the experiment. To evaluate the model, we selected AP, mAP, FLOPs, and Parameters. AP and mAP reflect precision and recall, and the higher the values of AP and mAP, the better the model's classification performance. FLOPs and Parameters reflect the model's parameters and computational complexity, and the smaller the values of FLOPs and Parameters, the less complex the model is. Although the neck part of YOLOv5 introduces additional computational complexity [37], it is beneficial to the learning of the model. Therefore, we retained the neck part of YOLOv5.

A. DATASET

Images obtained from multiple perspectives allowed the model to learn more effective features. The data collection times were 8:00 am to 11:00 am, 2:00 pm to 5:00 pm, and 8:00 pm to 10:00 pm every day. The location of the pigeon farm was Meizhou, Guangdong, China. A total of 988 images (Figure8) were obtained. Of these, 788 were randomly selected as the training set, and the remaining 200 formed the test set (approximate 4:1 ratio). Each image contained 1 to 5 young pigeons, and the total number of samples was about 2700.

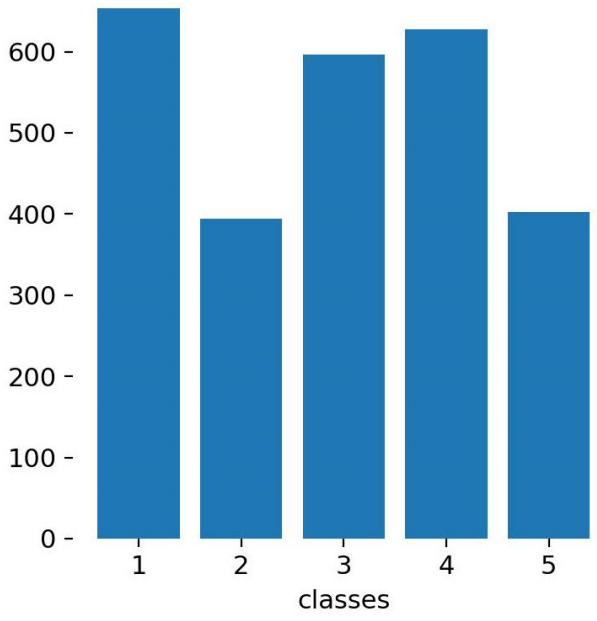


FIGURE 8. Number of pigeons in the five stages. Classes 1-5 correspond to stages 1-5, respectively, and the y-axis shows the number of pigeons in each class.

The dataset was divided according to experienced workers, pigeon hatching record and the guidance of the food intake of young pigeons [38], [39]. Pigeon hatching records refers to the artificial records of the breeding workers, including the hatching time of pigeons and the number of days of growth. The five stages of the pigeons were designated (Figure9) as follows: 1 (birth, 1-4 days), 2 (5–8 days), 3 (9–14 days), 4 (15–20 days), and 5 (after 21 days).

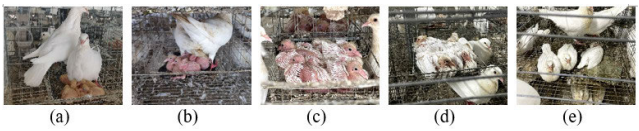


FIGURE 9. The five stages of the pigeon. (a) stage 1, (b) stage 2, (c) stage 3, (d) stage 4, (e) stage 5.

Stage 1 is the pigeon spawn stage. The pigeons at this stage were easily obscured and were difficult to observe, so we added some photos of pigeons in stage 1. Stage 5 corresponds to the slaughtering stage [40] and the sample size was smaller. We also performed data enhancement (Figure10) of the images by flipping, changing the saturation or color, and stitching four randomly cropped images (mosaic) [24].

FIGURE 10. Data enhancement (flipping, changing the saturation or color, and mosaic).

B. NETWORK PERFORMANCE EVALUATION

Given that true negatives (TN) are not utilized in object detection frameworks, it is recommended that object detection algorithms steer clear of TN-based metrics like TPR, FPR, and ROC curves [41]. Instead, the assessment of object detection algorithms should be based on precision (P) and recall (R), which are defined as average precision (AP) and mean average precision (mAP), respectively, as follows [42]:

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

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

$$AP_{11} = \frac{1}{11} \sum_{R \in (0, 0.1, \cdots, 0.9, 1)} P(R) \tag{14}$$

$$P(R) = \max_{\tilde{R}:\, \tilde{R} \geq R} P(R) \tag{15}$$

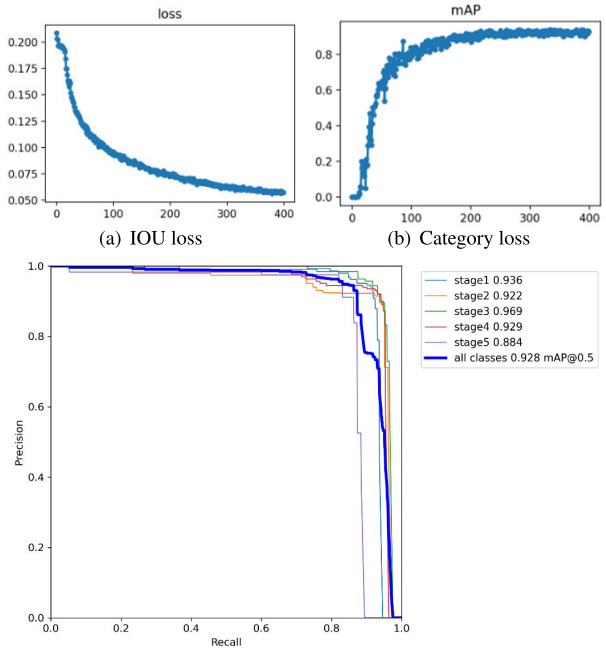
where TP represents true positives, FP represents false positives, and FN represents false negatives. The AP is calculated

using the 11-point interpolation method [42], and P(R) represents the highest precision achieved when the recall value R surpasses a predefined recall threshold R.

C. ANALYSIS AND EVALUATION

Figures 11 (a - c) show three plots of the LN-STEP-YOLO in the validation set obtained for the loss [Figure11 (a)], mAP [Figure11 (b)], and P-R curves [Figures 11 (c)]. As shown in Figure11, the curves behave well. The loss starts to converge at about 350 epochs. mAP was stable at about 200 epochs. The AP values for the five stages of the model were 0.936, 0.922, 0.969, 0.929, and 0.884, respectively.

Figure12 (a) shows the labels we made (ground truth) and Figure12 (b - g) show the results of different network detections. Figures 12 (a) - (e) were analyzed in combination with Tables 4-6.



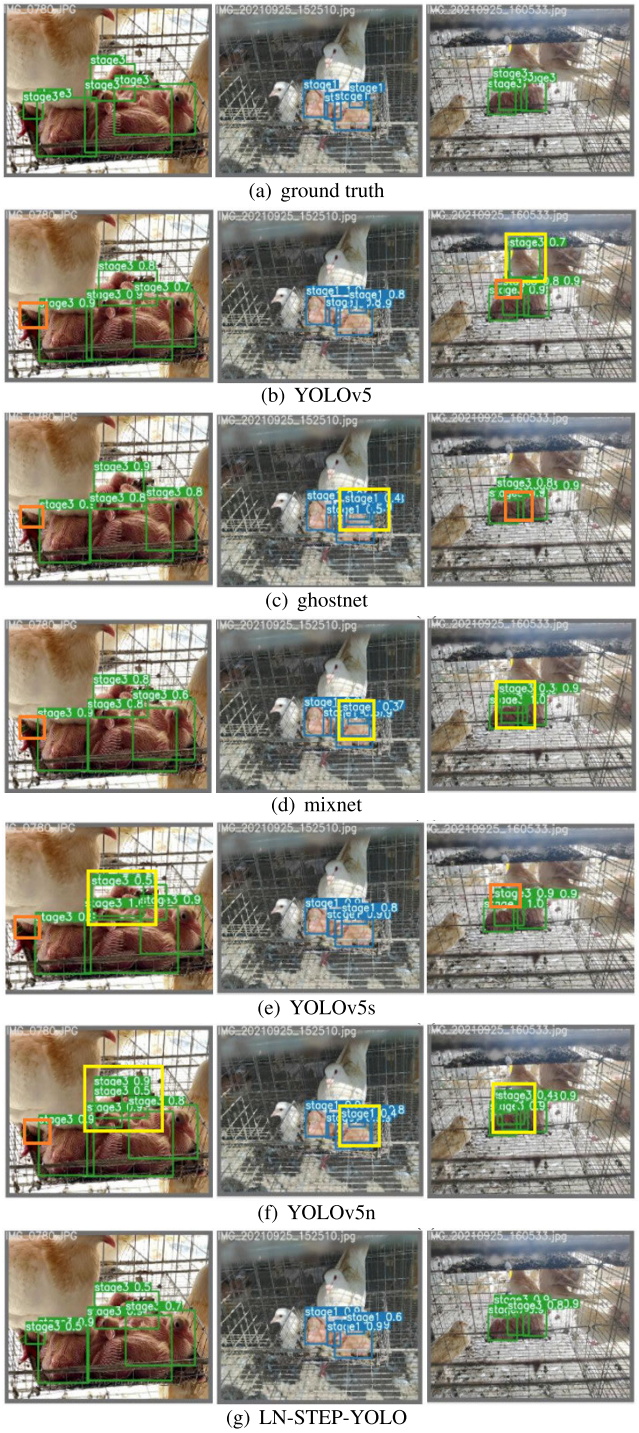
(a) IOU loss (b) Category loss

(c) Precision vs. Recall curve

FIGURE 11. Cross-validation curves. (a) loss, (b) mAP, (c) P-R curves.

To analyze the impact of the improved method proposed here on the performance of the YOLOv5 algorithm, we designed six sets of experiments (Table 4) according to the order of problem solving. The experimental environment used was Windows 10 64-bit operating system, Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz 2.50GHz, 64GB RAM, NVIDIA GeForce RTX 2080 Ti graphics card, and the same training parameters used in the experiment.

YOLOv5n was obtained by halving the number of channels of YOLOv5s. In terms of mAP values (Tables 4 and 5), the accuracy of YOLOv5n was only 1.6% lower than that of YOLOv5s (Experimental group *1*), which still had a high mAP. However, in terms of practical results, YOLOv5n easily produced multiple detection frames on one object when the pigeon overlap was high [as shown in the 1st and 3rd pictures of Figure12 (f)], making it not suitable for practical application. This was also a drawback of simply halving the number of YOLOv5s channels, which made the YOLOv5n model less able to recognize overlapping pigeons, resulting in multi-detection. However, some heavily obscured pigeons, such as



(a) ground truth

(b) YOLOv5

(c) ghostnet

(d) mixnet

(e) YOLOv5s

(f) YOLOv5n

(g) LN-STEP-YOLO

FIGURE 12. Detection effects of different models. (a) ground truth, (b) ghostnet, (c) mobilenetv3, (d) mixnet, (e) YOLOv5s, (f) YOLOv5n, (g) LN-STEP-YOLO. In the yellow box, the number of detections is greater than the ground truth, and in the orange box, the number of detections is less than the ground truth.

TABLE 4. Effect of improved methods on model performance.

Experimental group	Improvement points					mAP	FLOPs (G)	Parameters (G)
	StepConv	Stem	EA	XSepConv	LN			
I	×	×	×	×	×	0.913	5.0	1.78
II	✓	×	×	×	×	0.917	3.7	1.57
III	✓	✓	×	×	×	0.921	3.9	1.57
IV	✓	✓	✓	×	×	0.926	4.4	1.78
V	✓	✓	✓	✓	×	0.907	4.2	1.25
VI	✓	✓	✓	✓	✓	0.928	3.7	1.24

TABLE 5. Performance of different models.

Model	mAP	FLOPs(G)	Parameters(G)	Memory(M)
Mobilenetv3	0.89	8.4	3.55	**135**
Mixnet	0.910	8.8	3.54	332
Ghostnet	0.911	5.4	3.03	149
YOLOv5s	**0.929**	19.0	7.06	144
Ours	*0.928*	***3.7***	***1.25***	*154*

the leftmost pigeon in the first panel of Figure12 (a), were only correctly identified by the LN-STEP-YOLO method [Figure12 (g)].

By observing Table 4 and Fig. 12 (f) and (g), it can be seen that the performance of these experimental groups gradually improved. In Experiment II, StepConv was introduced, resulting in a 0.004 increase in mAP, a 1.3G decrease in FLOPs, and a 0.21G increase in parameters. In Experiment III, Stem was introduced, resulting in a 0.004 increase in mAP, a 0.2G increase in FLOPs, and no change in parameters. In Experiment IV, EA was introduced, resulting in a 0.005 increase in mAP, a 0.5G increase in FLOPs, and a 0.21G increase in parameters. The introduction of an improved XSepConv in Experiment V resulted in a 0.019 decrease in mAP, a 0.2G decrease in FLOPs, and a 0.53G decrease in parameters. In Experiment VI, LN was introduced in all non-downsampling layers, resulting in a 0.021 increase in mAP, a 0.5G decrease in FLOPs, and a 0.01G decrease in parameters. LN-STEP-YOLO had 1.5% more mAP, 1.3G less computation, and 0.54G less parameters than YOLOv5n, with better actual results, especially for the recognition of smaller, overlapping pigeons, than YOLOv5n.

To analyze the proposed LN-STEP-YOLO, we compared the performance of the model with that of other mainstream lightweight networks (Tables 5 and 6). In Tables 5 and 6, the Intersection of Union (IOU) of average precision (AP) is 0.45 and the same training parameters were used for all experiments. Bolded numbers indicate that the item is the best, and italicized numbers are the result of LN-STEP-YOLO.

As shown in Tables 5 and 6, the mAP of LN-STEP-YOLO was 0.928, the FLOPs was only 3.7G, and the number of parameters was only 1.25G. In terms of training memory usage, Mixnet with large convolution kernels had the highest memory consumption, while Mobilenetv3, Ghostnet, YOLOv5s, and LN-STEP-YOLO had similar memory usage. The AP of LN-STEP-YOLO outperformed that of the other models (including YOLOv5s) in Stages 1-3 where the overlap was higher and the body size was relatively small. In addition, the AP of LN-STEP-YOLO was second to Mixnet and

TABLE 6. Classification performance of different models.

Model	AP (IOU=0.45)				
	Stage1	Stage2	Stage3	Stage4	Stage5
Mobilenetv3	0.878	0.861	0.93	0.916	0.864
Mixnet	0.894	0.909	0.933	0.966	**0.946**
Ghostnet	0.913	0.904	0.967	0.905	0.863
YOLOv5s	0.919	0.916	0.964	**0.967**	0.879
Ours	***0.936***	***0.922***	***0.969***	*0.929*	*0.884*

YOLOv5s in Stage 4 and second only to Mixnet in stage 5. YOLOv5s contains a much higher number of channels than other networks and has a richer feature combination relationship, resulting in far greater computation. Mobilenetv3 and Ghostnet showed AP values in Stages 4-5 that were lower than Mixnet, indicating that large convolutional kernels allowed improved detection of targets with larger body size. The inferiority of LN-STEP-YOLO to Mixnet for the detection of large-sized pigeons may be a drawback of splitting the large convolutional kernel into multiple small convolutional kernels. However, the combination of multiple small convolutional kernels (StepConv) still retained some of the advantages of large convolutional kernels for the detection of larger targets, so LN-STEP-YOLO still outperformed Mobilenetv3 and Ghostnet in detection during Stages 4-5. The detection of smaller, overlapping pigeons is more important, because larger pigeons are usually not hidden at another angle and were generally easy to detect, while smaller pigeons were more likely to be hidden by their parents or other young pigeons, making them difficult to detect.

Overall, compared to other commonly used lightweight algorithms, LN-STEP-YOLO boasts the lowest computational and parameter requirements while demonstrating the best detection performance in the more crucial Stage1-3 pigeon detection. Additionally, its mAP score ranked second, only 0.1% lower than YOLOv5s, with LN-STEP-YOLO having approximately 17% of YOLOv5s' parameters and 19% of its Flops. These results suggest that LN-STEP-YOLO is a highly efficient and accurate object detection model, particularly in detecting pigeons in practical applications.

V. CONCLUSION

In this work, an LN-STEP-YOLO pigeon age detection algorithm was developed to address the limitations of existing general-purpose target detection algorithms for recognition of the growth stages of pigeon members in cages in an environment of edge devices. In order to cater to edge devices with limited computing power, this method first reduces the channel count of YOLOv5s by half, resulting in lower feature relationship complexity and computational cost as compared to YOLOv5s. While low computational cost is an advantage, it also brings about the drawback of low feature relationship complexity, which may affect the model's ability to express features. To tackle the problem of low complexity in feature relationships, we have developed a new method called LN-STEP-YOLO. Our approach boasts several innovations:

firstly, we have introduced a convolution structure that has a large convolution kernel efficiency while requiring less computational complexity. Secondly, we have improved the input module and XsepConv. Finally, we have integrated EA and LN to further enhance the model's performance. Experimental results indicate that the LN-STEP-YOLO method has a mAP that is only 0.1% lower than that of YOLOv5s, and has lower computational cost, and outperforms YOLOv5s in Stage1-3 of pigeon detection. Finally, the developed method in this paper reached mAP of 0.928, parameter of 3.7G, and FLOPs of 1.25G in the validation set, indicating this method shows excellent performance for the detection of pigeons with small size and overlap and meets the requirements of determining pigeons' growth stage in a real environment. The pigeon farm is situated in a semi-exposed environment, and specific weather conditions, such as heavy fog, can cause interference, such as rain streaks appearing in the image, resulting in a decrease in image quality and an increase in noise levels. Low-quality and high-noise level images may negatively impact the model's performance. Therefore, in the future, we will take into account the effects of specific weather conditions and further optimize the LN-STEP-YOLO model to effectively address these challenges.

REFERENCES

[1] D. Shan, Y. An, Y. Jiang, X. Fei, and S. Yang, "Design and application of the automatic feeding machine running behind the pigeon cage to be used for the large-scale pigeon breeding," *Feed Ind.*, vol. 39, no. 9, pp. 14–19, 2018, doi: 10.13302/j.cnki.fi.2018.09.003.

[2] X. Wang, S. Li, X. Zhan, G. Suo, C. Rao, and S. Liu, "Research on the crude protein needs of breeding pigeons under the production model of '2+4' for meat pigeon breeding," *Feed Ind.*, vol. 30, no. 17, pp. 59–60, 2009.

[3] H. Hou, X. Wang, C. Yang, O. Bao, W. Lv, Y. Tu, W. Zhao, and J. Yao, "Research on the weight growth pattern of European pigeon II strain at 27 days of age under the '2+3' model," *Shanghai J. Animal Husbandry Veterinary Med.*, no. 6, pp. 20–21, 2020, doi: 10.14170/j.cnki.cn31-1278/s.2020.06.007.

[4] M. Asaduzzaman, M. Mahiuddin, M. Howlider, M. Hossain, and T. Yeasmin, "Pigeon farming in Gouripur Upazila of Mymensingh district," *Bangladesh J. Animal Sci.*, vol. 38, nos. 1–2, pp. 142–150, Jan. 1970.

[5] R. Hu and W. Liu, "Technological revolution, disruptive technology and smart agriculture," Smart Agricult., Tech. Rep., pp. 1–6. [Online]. Available: https://kns.cnki.net/kcms/detail/10.1681.S.20220720.2005.002.html

[6] T. Fu, G. Liu, Q. Wan, T. Wu, L. Zhao, L. Lin, and L. Yang, "Establishment of a water nitrite nitrogen concentration prediction model based on stacked autoencoder-BP neural network," *J. Fisheries China*, vol. 43, no. 4, pp. 958–967, 2019.

[7] J. Zou, H. Jiang, Q. Wang, N. Chen, T. Wu, and L. Yang, "Accurate identification of agricultural inputs based on sensor monitoring platform and SSDA-HELM-SOFTMAX model," *J. Sensors*, vol. 2021, pp. 1–12, Nov. 2021.

[8] 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.

[9] L. Zhang, H. Gray, X. Ye, L. Collins, and N. Allinson, "Automatic individual pig detection and tracking in pig farms," *Sensors*, vol. 19, no. 5, p. 1188, Mar. 2019.

[10] M. Saar, Y. Edan, A. Godo, J. Lepar, Y. Parmet, and I. Halachmi, "A machine vision system to predict individual cow feed intake of different feeds in a cowshed," *Animal*, vol. 16, no. 1, Jan. 2022, Art. no. 100432.

[11] N. Li, H. Ren, and Z. Ren, "Research of behavior monitoring method of flock hens based on deep learning," *J. Hebei Agricult. Univ.*, vol. 44, no. 2, pp. 117–121, 2021, doi: 10.13320/j.cnki.jauh.2021.0035.

[12] Y. Li, "Construction of cold chain logistics of aquatic products based on Internet of Things technology," *Logistics Technol.*, vol. 41, no. 6, pp. 105–109, 2022.

[13] L. Yang, Q. Wang, H. Yang, Z. Shi, L. Su, T. Wu, L. Lin, and J. Zou, "Muscle crispness of crispy grass carp (*Ctenopharyngodon idella*) based on Raman spectroscopy," *J. Fisheries China*, vol. 46, no. 7, pp. 1235–1245, 2022.

[14] H. Jiang, "Research on target recognition method based on spectrum analysis," Zhejiang Ocean Univ., Zhoushan, China, Tech. Rep., 2020.

[15] F. Zhang, Y. Wang, M. Lv, J. Wang, Z. Xu, and X. Chen, "Design and experiment of auto-precision feeding system for piglets," *Trans. Chin. Soc. Agricult. Machinery*, vol. 49, no. 7, pp. 39–45, 2018.

[16] K. Gao, Y. Huang, Z. Li, and A. Duan, "Design and development of integrated multi-functional calf feeding equipment," *Xinjiang Agricult. Mechanization*, no. 6, pp. 29–30 and 46, 2021, doi: 10.13620/j.cnki.issn1007-7782.2021.06.007.

[17] I. Adewumi and I. Atanda, "Development of semi-automated portable box poultry brooder," *Development*, vol. 2, no. 1, pp. 1–6, 2016.

[18] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Comput. Electron. Agricult.*, vol. 157, pp. 417–426, Feb. 2019.

[19] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3," *Sensors*, vol. 20, no. 7, p. 2145, Apr. 2020.

[20] Q. Yang, W. Li, X. Yang, L. Yue, and H. Li, "Improved YOLOv5 method for detecting growth status of apple flowers," *Comput. Eng. Appl.*, vol. 58, no. 4, pp. 237–246, 2022. [Online]. Available: https://kns.cnki.net/kcms/detail/11.2127.TP.20211012.1636.012.html

[21] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.

[22] M. Tan and Q. V. Le, "MixConv: Mixed depthwise convolutional kernels," 2019, *arXiv:1907.09595*.

[23] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1580–1589.

[24] *Ultralytics*. Accessed: Nov. 5, 2021. [Online]. Available: https://github.com/ultralytics/yolov5

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

[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[28] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11976–11986.

[29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[30] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-mixer: An all-MLP architecture for vision," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 24261–24272.

[31] Z. L. Zhu, Y. Rao, Y. Wu, J. N. Qi, and Y. Zhang, "Research progress of attention mechanism in deep learning," *J. Chin. Inf. Process.*, vol. 33, no. 6, pp. 1–11, 2019.

[32] M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu, "Beyond self-attention: External attention using two linear layers for visual tasks," 2021, *arXiv:2105.02358*.

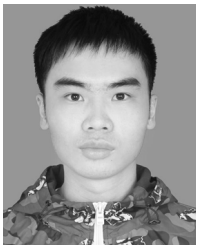
[33] *Huawei-Noah*. Accessed: Nov. 3, 2022. [Online]. Available: https://github.com/iamhankai/ghostnet.pytorch

[34] *Chinhsuanwu*. Accessed: Nov. 9, 2022. [Online]. Available: https://github.com/chinhsuanwu/coatnet-pytorch

[35] S.-Y. Zhou and C.-Y. Su, "A novel lightweight convolutional neural network, ExquisiteNetV2," 2021, *arXiv:2105.09008*.

[36] J. Chen, Z. Lu, J.-H. Xue, and Q. Liao, "XSepConv: Extremely separated convolution," 2020, *arXiv:2002.12046*.

[37] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[38] W. Dong, W. Shen, H. Zhang, S. Zhuang, and H. Hao, "Preliminary research on egg weight loss during hatching and daily gain of white king pigeon," *J. Yangling Vocational Tech. College*, no. 4, pp. 1–3, 2003.
[39] M. Chen, "The growth and nutrition regulation of domestic pigeon," *Poultry Sci.*, no. 8, pp. 49–52, 2019.
[40] D. Li, Y. Dai, and X. Su, "Processing status and development prospect of pigeon," *Meat Ind.*, no. 3, pp. 52–53, 2015.
[41] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[42] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jul. 2020, pp. 237–242.



JUN YAO is currently pursuing the master's degree in food safety and intelligent control with the Zhongkai University of Agricultural Engineering. His research interests include computer vision, intelligent agriculture, and agricultural informatization.

ZECHEN SHI is currently pursuing the master's degree in agricultural engineering and information technology with the Zhongkai University of Agricultural Engineering. His research interests include computer vision, plant and animal phenotypic information methods, and natural language processing.

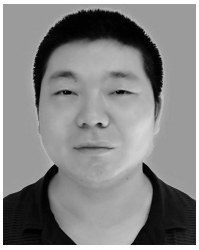
HANJING JIANG received the B.S. degree in digital media technology from the Guilin University of Electronic Technology, Guilin, China, in 2018, and the M.S. degree in agricultural engineering and information technology from the Zhongkai University of Agriculture and Engineering, Guangzhou, China, in 2020. Since 2018, she has been a Research Assistant with Guangdong Provincial Food Quality Safety Traceability and Control Engineering Technology Research Center, Guangzhou. Her research interests include rapid detection of agricultural inputs/food quality and safety traceability.

QINGXIU WANG received the M.S. degree in agricultural engineering and information technology from the Zhongkai University of Agriculture and Engineering, Guangzhou, China, in 2021. Her research interests include hyperspectral technology, Raman spectroscopy technology, and crispy grass carp brittleness detection method research.

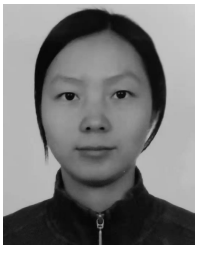
NINGXIA CHEN received the B.S. and M.S. degrees from the Beijing University of Post and Telecommunication, in 2004 and 2007, respectively. She is currently a Lecturer with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, China. Her research interests include machine learning, intelligent agriculture, and mobile communications.

TING WU received the Ph.D. degree in agricultural electrification and automation with South China Agricultural University, in 2018. He is currently an Associate Professor with the Zhongkai University of Agriculture and Engineering, China. His research interests include intelligent agriculture, agricultural informatization, and nondestructive testing technology for agricultural products.

HAIXIA XU received the Ph.D. degree in optics from Sun Yat-sen University, Guangzhou, China, in 2010. Since 2010, she is currently a Lecturer with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, China. Her research interests include photo-communication and the optical properties of optical devices.



LING YANG received the B.S. degree in computer application from the Shenyang University of Technology, Shenyang, China, in 2002, and the M.S. degree in computer technology from Sun Yat-sen University, Guangzhou, China, in 2009. Since 2003, he has been a Professor with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, and the Guangdong Provincial Food Quality Safety Traceability and Control Engineering Technology Research Center, Guangzhou. His research interests include agricultural artificial intelligence/rapid detection of agricultural inputs/food quality and safety traceability.



JUAN ZOU received the B.S. degree in computer application from the Shenyang University of Technology, Shenyang, China, in 2002, and the M.S. degree in computer software and theory from Sun Yat-sen University, Guangzhou, China, in 2010. Since 2003, she has been a Lecturer with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, and a Research Assistant with Guangdong Provincial Food Quality Safety Traceability and Control Engineering Technology Research Center, Guangzhou. Her research interests include agricultural artificial intelligence/rapid detection of agricultural inputs/food quality and safety traceability.