AN IMPROVED RT-DETR FOR STABLE AND REAL-TIME DEFECTIVE EGG DETECTION IN EDGE COMPUTING ENVIRONMENTS
/
边缘计算环境下基于改进 RT-DETR 的稳定实时异常鸡蛋检测

Feiyu TANG, Yuhang ZHANG, Yida ZHANG, Liwei YANG*)
College of Information and Electrical Engineering, China Agricultural University, Beijing 100083/China;
Tel: +86-13810223137; E-mail: yangliwei@cau.edu.cn
Corresponding author: Liwei Yang
DOI: https://doi.org/10.35633/inmateh-78-52

Keywords: *Defective egg detection; RT-DETR-R18; Non-destructive testing; Edge computing; Inference stability*

ABSTRACT
To achieve stable abnormal egg detection under edge device deployment conditions, this paper proposes a CSDE-DETR-R18 abnormal egg detection model based on an optimized RT-DETR-R18. The CSFM module is incorporated into the feature fusion path of the RT-DETR-R18 model to aggregate multi-scale features. Subsequently, the model's ability to extract low-contrast features is enhanced by replacing the standard convolutions in the backbone network with the DEConv module, which improves its recognition capability for abnormal egg characteristics such as light-colored spots. Finally, the NWD+AIOU mixed loss function is employed to improve the model's localization accuracy for minute targets (such as fine cracks, specks, etc.). The experimental results demonstrate that the CSDE-DETR-R18 model achieved mAP@0.5, P, and R of 87.8%, 88.9%, and 87.3%, respectively, representing improvements of 3.4, 3.1, and 0.5 percentage points over RT-DETR-R18. Test results from the Raspberry Pi deployment revealed that CSDE-DETR-R18 demonstrated a significant advantage in inference stability compared with the YOLOv8 series, one of the most advanced detectors currently available. This is of great engineering significance for tasks requiring stable timing performance at equal heights, such as robotic arm sorting.

摘要
为在边缘设备部署条件下实现稳定的异常鸡蛋识别，本文提出一种基于改进 RT-DETR-R18 的异常鸡蛋检测模型 CSDE-DETR-R18。在 RT-DETR-R18 的进步特征融合路径加入 CSFM 模块聚合多尺度特征；然后使用 DEConv 模块替换骨干网络中的标准卷积提高模型的低对比度特征提取能力，增强对浅色斑点等异常鸡蛋特征的识别能力；最后用 NWD+AIOU 混合损失函数来提升模型对于微小目标（细小裂纹、斑点等）的定位精度。实验结果显示，使用 CSDE-DETR-R18 模型得到的 mAP@0.5、P 和 R 分别为 87.8%、88.9%和 87.3%，比 RT-DETR-R18 提升了 3.4、3.1 和 0.5 个百分点。树莓派部署测试结果显示，相较于目前最先进检测器之一的 YOLOv8 系列，CSDE-DETR-R18 具有明显的推理稳定性优势。这对机器臂分拣等高时序稳定性要求的任务具有重大的工程意义。

INTRODUCTION
Eggs with compromised shell quality can result from physical impact, sick hens, or malnutrition during production. This can manifest as shells that are sandy, cracked, or have surface blemishes like dirt or spots *(Gautron, 2021)*. These abnormal eggs pose a health risk to consumers and harm the economic interests of poultry farms. In actual production environments, automated abnormal egg detection enables poultry farm workers to conduct real-time monitoring, thereby efficiently controlling overall egg quality at the source and responding swiftly to potential factors such as poultry diseases. For the time being, the world's largest egg producer—cage-based egg production remains the predominant model in many countries, particularly China, which is characterized by intensive farming practices *(Yang, 2021)*. This makes rapid response to abnormal situations tremendously essential. Featured by its advantages of non-destructive and high efficiency, computer vision technology has emerged as one of the mainstream choices for automated agricultural inspection in recent years. Among them, the representative YOLO (You Only Look Once) series detection algorithms have materialized a satisfactory balance between accuracy and speed. Nonetheless, YOLO and other extensively utilized CNN detectors rely on NMS (Non-Maximum Suppression) for post-processing. When confronted with complex, dynamic recognition tasks, the fluctuating number of objects triggers striking variations in inference time, rendering them unsuitable for tasks demanding high temporal stability, such as robotic arm sorting and assembly line grading.

In an effort to address the aforementioned issues, this paper proposes a highly stable abnormal egg detection model, CSDE-DETR-R18 (Contextual Selection and Detail Enhanced Detection Transformer), grounded in an optimized RT-DETR-R18 (Real-Time Detection Transformer with ResNet-18 backbone). The RT-DETR series models leverage an end-to-end Transformer-based paradigm to circumvent multiple issues stemming from NMS dependency. The specific contributions of this research are summarized as follows:

(1) An abnormal egg detection model, CSDE-DETR-R18, is proposed based on an optimized RT-DETR-R18, further enhancing the model's accuracy in identifying abnormal eggs under cage-reared conditions. The CSFM (Contextual Selection Fusion Module) was introduced into the neck feature fusion network to mitigate feature loss in dense scenes. DEConv (Detail Enhanced Convolution) was incorporated to strengthen the model's detection capability for low-contrast defects such as minute cracks and faint spots on abnormal eggs. The NWD (Normalized Wasserstein Distance) + AIoU (Alpha-Intersection over Union) fusion loss function was adopted to improve localization accuracy for small targets.

(2) Edge Deployment Verification: To identify abnormal eggs, the CSDE-DETR-R18 model and YOLOv8n—one of the most advanced detectors currently available—were deployed on a Raspberry Pi 4B to evaluate their performance. Experimental results confirm that CSDE-DETR-R18 demonstrates superior inference stability on edge devices, providing a basis for algorithm selection in future applications such as robotic arm egg sorting and assembly line egg grading.

Traditional manual detection methods are tremendously influenced by subjective experience, resulting in unfavorable accuracy and work efficiency. As an emerging technology, computer vision (CV) has exhibited substantial potential in food and agricultural applications. For instance, convolutional neural network (CNN)–based models have been widely adopted in the domain of egg inspection. High-accuracy recognition of egg freshness or shell cracks has been achieved using architectures such as MobileNet *(Liu, 2022)*, ResNet *(Valencia, 2021)*, or hybrid models combined with BiLSTM *(Turkoglu, 2021)*. Nevertheless, the majority of these approaches not only fall into the category of image classification *(Chewprecha et al., 2023)*, but also concentrate on nothing but analyzing single cropped images. In complex scenes where multiple eggs coexist, the reliance on cumbersome pre-processing steps (e.g., region-of-interest extraction) renders such methods challenging to satisfy real-time detection requirements.

In recent years, deep learning–based object detection has emerged as the mainstream solution for automated inspection. Both two-stage detectors represented by Faster R-CNN (Faster Region-based Convolutional Neural Network) and one-stage detectors represented by YOLO have demonstrated strong capabilities in multi-object localization and classification under complex scenes. Among these, the YOLO series of recognition algorithms has proven to be the mainstream choice for agricultural automation in recent years, attributable to its high accuracy, lightweight nature, and fast inference speed. For instance, researchers applied it to tomato maturity detection *(Li et al., 2025)* and young apple detection *(Du et al., 2024)*. In poultry egg recognition, researchers employed enhanced versions like YOLOv5 *(Yin et al., 2024)*, YOLOv7 *(Zhao, 2023)*, and YOLOv8 *(Bambilla et al., 2025)* to achieve egg crack detection, egg size recognition, and egg grading *(Nasiri et al., 2020; Subedi et al., 2023; Yang et al., 2023)*.

Nevertheless, these CNN-based detectors all rely on non-maximum suppression (NMS) for post-processing. In high-density cage-based environments, NMS suffers from inherent limitations; for example, it may lead to missed detection of heavily overlapping objects *(Carion, 2020)*. More critically, on computation-constrained edge devices (e.g., Raspberry Pi), the serial nature of NMS causes inference latency to fluctuate sharply with variations in the number of detected targets, thereby resulting in undesirable temporal stability. In contrast, transformer-based architectures such as RT-DETR can fundamentally eliminate issues arising from NMS dependency *(Zhao et al., 2024)*. Nonetheless, few studies have investigated abnormal egg detection on edge devices using end-to-end architecture models such as RT-DETR. Therefore, this paper aims to fill this gap.

MATERIALS AND METHODS

Data Acquisition and Pre-processing

The experimental data were collected at the Bayannaoer High-Tech Breeding Farm in Inner Mongolia. The collected samples were brown-shell eggs produced by Hailan brown laying hens, including normal eggs and abnormal eggs. Abnormal eggs were classified into broken shell eggs, dirty eggs, sand shell eggs, and spotted eggs. Dirty eggs refer to eggs with feces, soil, or other external contaminants adhering to the shell surface. Broken eggs refer to eggs with cracks visible on the shell.

There were obvious abnormal color areas or stains on the spotted eggshell (not caused by foreign dirt adhesion). There were small granular or rough deposits on the surface of the sand eggshell. Refer to Fig. 1 for images of abnormal eggs.

Fig. 1 - Images of defective eggs

The Glory 90 Mobile Camera was used as a shooting tool, with a resolution of 1920×1080 pixels and a frame rate of 60 FPS. In order to build a highly representative data set and guarantee the excellent generalization performance of the model, the mobile phone was placed at a position about 30 cm horizontally and 15 cm vertically from the egg trough, and a 30°to 60°perspective was taken to optimize the effect of feature extraction. The acquisition time window was set at 08:00-09:00 a.m. and 14:00-16:00 p.m. under different lighting conditions. The dataset covered a variety of actual scenes, such as metal mesh occlusion, egg stack ghosting, and dynamic light and shadow changes. A total of 3656 eggs were captured in the video, including 2195 normal eggs, 329 sand shell eggs, 658 spotted eggs, 660 dirty eggs and 548 broken shell eggs. All egg categories were verified by professionals to ensure the accuracy of the data.

In the aftermath of acquisition, data pre-processing was performed to elevate model generalization. In the data pre-processing stage, a frame was firstly extracted every 30 frames grounded in OpenCV video sequence, and 2824 initial images were obtained. Some pictures had obvious motion blur attributable to uneven camera speed and slight jitter during shooting. On this basis, 889 high-quality egg images were retained after removing the blurred image and the image with high coincidence.

The X-AnyLabeling annotation tool was utilized to annotate 889 images, which were divided into training set, validation set and test set in line with the ratio of 6:2:2. In order to heighten the generalization ability of the model and avoid the problem of data leakage, a variety of data enhancement methods were used for and only for the training set, including flipping, random clipping, adding noise, random saturation, random contrast, and random brightness, to fit the complex environment in reality. The effect of data enhancement is shown in Fig. 2. Subsequent to enhancement, there were 745 sand shell eggs, 1776 spotted eggs, 1802 dirty eggs, 1008 broken shell eggs, and 4357 normal eggs in the training set, basically meeting the requirements of model training for the class balance of the dataset.

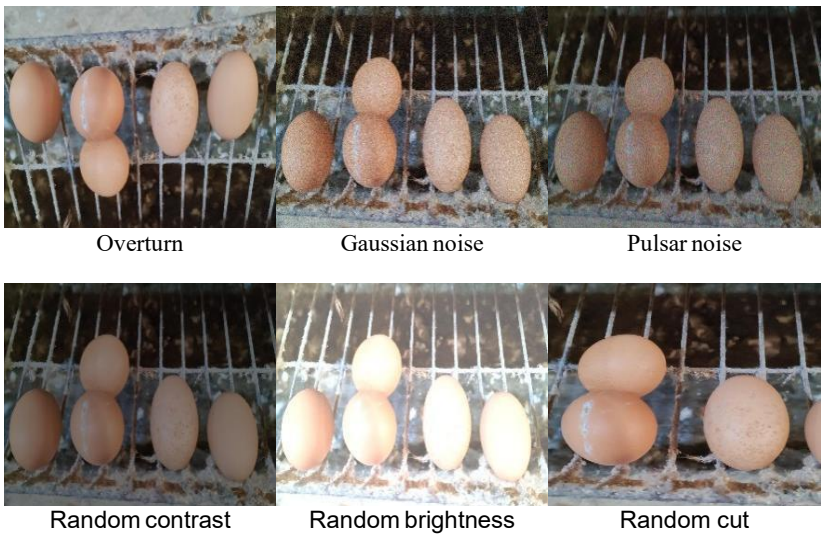

Fig. 2 - Data augmentation examples

Benchmark Model

RT-DETR is an end-to-end, NMS free real-time target detector. It uses an efficient encoder to process multi-scale features and provide higher quality initial queries to the encoder. While solving the inference delay and hyper parameter sensitivity problems spawned from the dependence of YOLO series on NMS, it avoids the difficulties of traditional DETR, which is difficult to deploy in real time on account of its high computational complexity (*Carion, 2020*).

Among the 5 categories of RT-DETR model, the RT-DETR-R18 model exhibited a lower AP value but achieved the lowest number of parameters (Params) and floating-point operations (FLOPs), while also demonstrating the fastest inference speed (FPS) for real-time object detection tasks. This makes it particularly suitable for deployment on edge devices to perform real-world abnormal egg detection tasks. This improvement enhances the model's accuracy in identifying abnormal eggs while preserving the lightweight and efficient advantages of the RT-DETR-R18 model.

Improvement of the RT-DETR-R18 model

In contrast to the current mainstream YOLO series recognition model, RT-DETR has the advantage of engineering stability as an end-to-end one-stage Transformer detection framework, but RT-DETR lacks the fine-grained representation of small targets, and the small target recognition ability is still slightly weaker than some powerful YOLO real-time detectors *(Zhao, 2024*). In the task of abnormal egg recognition, abnormal representations such as cracks, spots and rough surfaces were small-scale and similar detail features of background texture. In this regard, this paper introduced CSFM context selection module into RT-DETR-R18 neck feature fusion path to strengthen RT-DETR-R18's ability to express fine-grained abnormal targets and adaptability to complex environments; DEConv was introduced into the backbone network to replace part of the standard convolution to reinforce the extraction ability of irregular features of abnormal eggs; The NWD Loss+AIoU fusion loss function was introduced at the regression optimization level to reinforce the stability of model learning in the scene of small target scale change and fuzzy boundary.

The structure of CSDE-DETR-R18 model is depicted in Fig. 3.

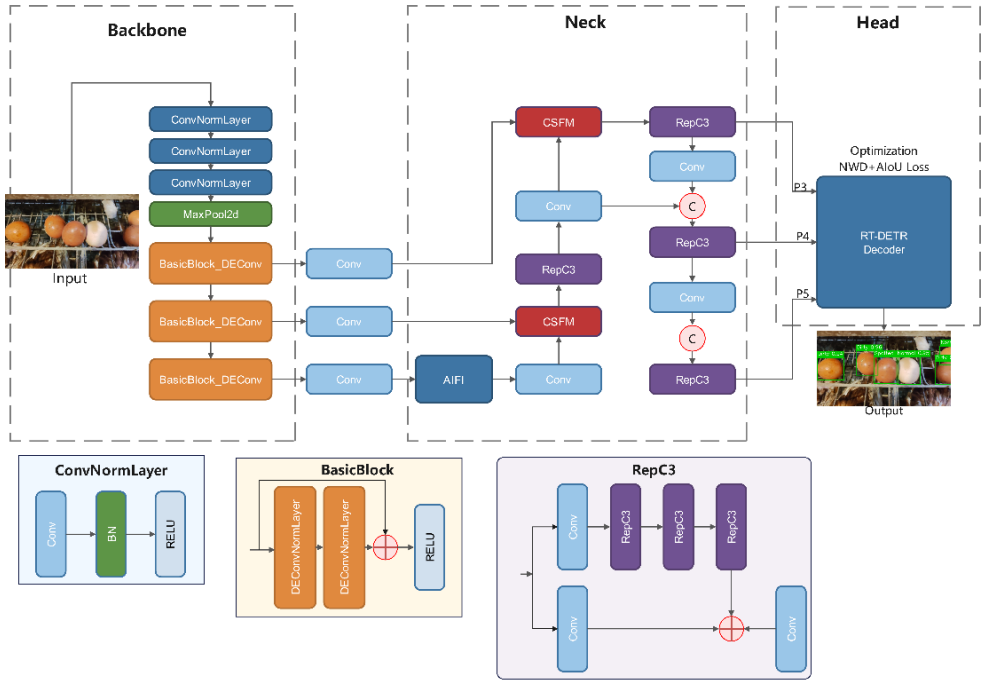

Fig. 3 - Architecture of CSDE-DETR-R18 model

Contextual Selection Fusion Module (CSFM)

The CSFM module structure is presented in Fig. 4. CSFM module (*Wang, 2025*) not only aggregates different resolution features in a step-by-step manner, but also jointly models the global information carried by deep features and the spatial details retained by shallow features with the local texture structure. On this basis, the CSFM module can avoid the dilution of the response of high-level features to small-scale defects in the fusion process, and simultaneously enhance the discriminability of shallow details in the overall feature expression. On top of that, this module can also deal well with the difficult detection of small defects on the surface of eggs.

Concurrently, CSFM introduces the channel importance evaluation and dynamic selection mechanism, calculates the channel weight (βch) through global statistics and nonlinear mapping of the fused features, enhances the texture, edge and local structure channels related to defects, and suppresses redundant or noise channels, so as to elevate the adaptability of the model to complex scenes. On this basis, the module further modulates the spatial information, guides the network to focus on the key areas such as cracks, spots and stains through the spatial weight (βsp), and improves the response strength and boundary positioning accuracy of low contrast defects.

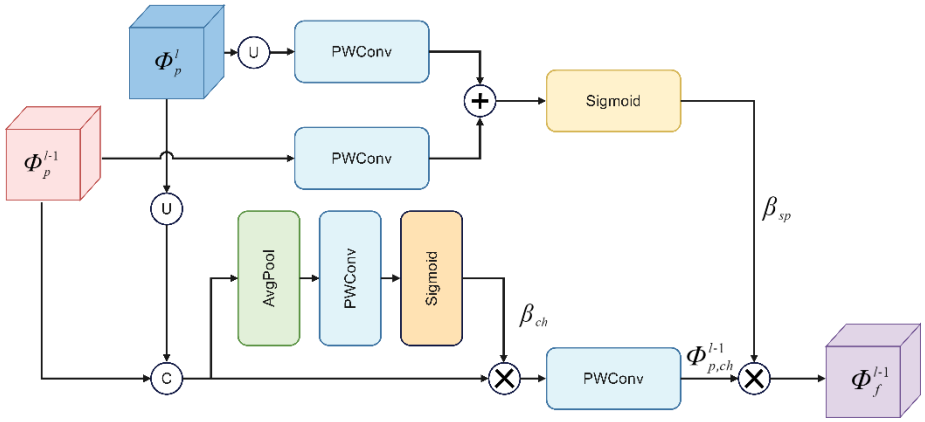

Fig. 4 - Structure diagram of CSFM module

Detail Enhanced Convolution (DEConv)

DEConv convolution module (*Chen et al., 2024*) combines conventional convolution and differential convolution (central differential convolution CDC, angular differential convolution ADC, horizontal differential convolution HDC, vertical differential convolution VDC), as presented in Fig. 5.

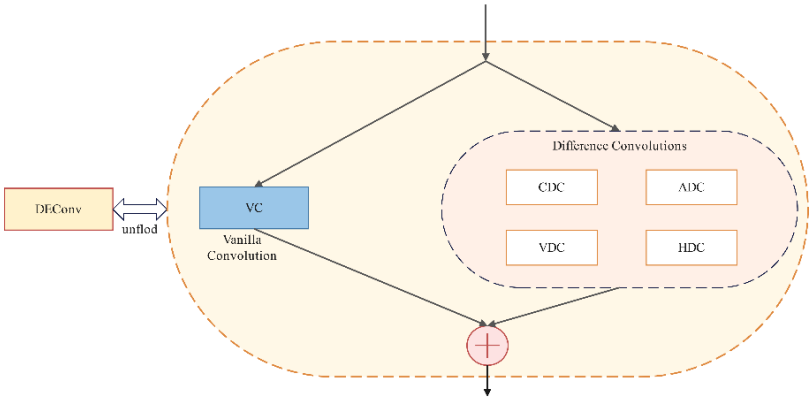

Fig. 5 - Structure diagram of DEConv convolution module

By combining traditional local descriptors with convolution operations, the DEConv module explicitly introduces prior knowledge into the convolution layer, thereby reinforcing the model's ability to understand and process image details. This feature enhances the generalization ability of the model. In addition, the DEConv module uses the reparameterization technology to simplify the multi-channel convolution operation to a standard convolution layer without adding additional computational overhead, so it can improve the richness of feature extraction while maintaining the high inference speed of RT-DETR-R18. Thanks to this explicit gradient computer system, DEConv can effectively simulate the edge detection operator and specifically capture the high-frequency details of micro cracks, thus overcoming the problem of missing detection triggered by low contrast between cracks and eggshell background.

Fusion Loss Function

NWD loss (*Xu et al., 2022*) is a loss function on the basis of Wasserstein distance, which models the bounding box as a two-dimensional Gaussian distribution, and calculates the Wasserstein distance between them to provide a measure of similarity between the bounding boxes of target detection. This measurement method can effectively deal with the box matching problem in tiny object detection, and has strong smoothness and robustness. The mathematical process of NWD loss is illustrated in equations (1) - (2).

$$L_{NWD} = exp\left(-\frac{W_2^2(N_1, N_2)}{\alpha}\right) \quad (1)$$

$$W_2(N_1, N_2) = \sqrt{\langle |b_1 - b_2|^2 \rangle} \quad (2)$$

where equation (1) is the NWD loss formula, equation (2) is the Wasserstein distance, LNWD denotes the Normalized Wasserstein Distance loss, [-], W2 symbolizes the Wasserstein distance between Gaussian distributions, [pixels], formula α is a regularized hyperparameter, and b1 and b2 represent two bounding box coordinates, [pixels].

The AIoU loss function (*He, 2021*) consists of IoU, distance loss (d/c) and shape loss, as demonstrated in equations (3) - (7)

$$L_{AIoU} = IoU + \frac{d}{c} + P_{Shape} \quad (3)$$

$$IoU = 1 - \frac{B_{gt} \cap B_{prd}}{B_{gt} \cup B_{prd}} \quad (4)$$

$$W_{dif} = \frac{\left|BoxW_{prd} - BoxW_{gt}\right|}{max\left(BoxW_{prd}, BoxW_{gt}\right)} \quad (5)$$

$$H_{dif} = \frac{\left|BoxH_{prd} - BoxH_{gt}\right|}{max\left(BoxH_{prd}, BoxH_{gt}\right)} \quad (6)$$

$$P_{Shape} = \left(1 - e^{-W_{dif}}\right)^w + \left(1 - e^{-H_{dif}}\right)^w \quad (7)$$

Among them, LAIoU denotes the Alpha-Intersection over Union loss, [-]; IoU refers to the intersection over union ratio, [-]; d represents the Euclidean distance between the center points of the boxes, [pixels]; c defines the diagonal length of the smallest enclosing box, [pixels]; PShape symbolizes the shape loss component, [-]; Bgt and Bprd are the ground truth box and predicted box respectively, [-]; Wdif and Hdif denote the normalized width and height differences, [-]. The shape loss function PShape is a paramount innovation of AIoU in contrast to traditional IoU. By introducing the difference of width height ratio to calculate the loss value, it can effectively distinguish the difference between the predicted box and the real box shape.

NWD Loss models the similarity of bounding box through Gaussian distribution to adapt to the situation of large differences in object morphology. AIoU optimizes the shape adaptation and position alignment of the box. The fusion of the above two models can not only maintain the stability of the model in the detection of a kind of tiny features in spotted eggs, but also improve the positioning and shape matching accuracy in the detection of different morphological features in broken eggs and sand eggs. Especially for small targets such as spots, the Gaussian distribution modeling mechanism of NWD minimizes the gradient disappearance problem of traditional IoU in the case of no overlap between the prediction frame and the real frame, thereby ensuring the convergence stability of the model for small defects.

Experimental Setup and Edge Deployment

In an effort to ensure the rigor of the experiment, all experiments in this article were conducted on the Windows 10 system platform. The hardware configuration included Intel Core i7-13700H (14 cores, 20 lines) and NVIDIA GeForce RTX 4060 laptop graphics card, as well as 8GB of video memory. The Python 3.10.0 programming language, PyTorch 2.5.1 deep learning framework, CUDA 12.1 and cuDNN 9.0.1 acceleration libraries were used to enhance computational efficiency, and OpenCV 4.10.0 was used as an auxiliary image processing library.

By using Raspberry Pi 4B as the deployment platform, edge deployment platform was utilized to test the real-time detection capability of edge devices. The Raspberry Pi inference environment configuration used the Ultralytics framework to run models exported in ONNX or TensorRT formats, ultimately achieving efficient performance testing in resource limited situations.

Although the performance of the original RT-DETR paper on the PaddlePaddle platform was better than that of the YOLO series (Y., Ultralytics architecture was used in deployment throughout the framework selection. This framework set a standard for real-time detection at the industrial level, ensuring consistency between RT-DETR and the underlying computational units of the comparison model, thus ensuring fairness.

On top of that, to support the rapid deployment of abnormal egg recognition on construction sites, this framework could be efficiently converted into edge-compatible formats such as TensorRT and ONNX. This compatibility not only simplified the integration of the algorithm with agricultural machinery but also underscored its strong potential for practical engineering application.

Throughout the parameter settings, the training parameters were set as follows: the input resolution was 640×640, the number of iteration rounds was 300 epochs, the batch size was set to 8, and the optimizer used AdamW. The number of decoder layers was 3, and the number of query vectors was set to 300.

RESULTS

Evaluation Metrics

In order to show the advantages of RT-DETR as an end-to-end detector over the recognition model of NMS dependency (represented by YOLOv8), and the improvement effect of RT-DETR-R18 model, Mean Average Precision (mAP), Precision (P), Recall (R), Post processing, and NMS Latency were selected as evaluation indicators. The specific mathematical definitions for the accuracy metrics are formulated as follows:

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$AP = \int_0^1 P(R)dR \tag{10}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{11}$$

Here, P denotes the precision rate, [%]; R defines the recall rate, [%]; TP (True Positive) denotes the number of defective eggs correctly detected, [-]; FP (False Positive) symbolizes the number of background objects or normal eggs incorrectly classified as defective, [-]; and FN (False Negative) refers to the number of defective eggs that are missed by the model, [-]; AP denotes the average precision of the defect category (e.g., broken, dirty, or spotted eggs), mAP refers to the mean average precision across all categories, [%], and N represents the total number of defect categories considered in this study.

Ablation experiment

To further validate the effectiveness of the elevated CSDE-DETR-R18 model and its adaptability to abnormal egg recognition tasks, this study conducted six ablation experiments, with results presented in Table 1.

Table 1

Ablation study results of the proposed improvements on RT-DETR-R18

Test Number	CSFM	DEConv	NWD+AIoU	mAP @0.5/%	P/%	R/%	FLOPs/ $\times 10^9$	Params/ $\times 10^6$
1	-	-	-	84.4	85.8	86.8	57.0	19.8
2	✓	-	-	86.8	86.2	87.6	62.3	21.2
3	-	✓	-	86.4	89.2	86.4	41.9	18.7
4	-	-	✓	85.3	87.0	87.1	57.0	19.8
5	✓	✓	-	87.2	88.7	86.9	47.2	20.1
6	✓		✓	86.7	86.8	88.1	62.3	21.2
7	✓	✓	✓	87.8	88.9	87.3	56.4	20.1

Note: "–" demonstrates that the corresponding improvement strategy is not used, while "✓" illustrates that it is used.

As illustrated in Table 1, CSDE-DETR-R18 demonstrated a pronounced improvement over RT-DETR-R18:

- +3.4 percentage points improvement in mAP@0.5
- +3.1 percentage points improvement in Precision (P)
- +0.5 percentage points improvement in Recall (R)

The FLOPs and parameters of CSDE-DETR-R18 were essentially on par with those of RT-DETR-R18. It could also be observed that both single and dual improvements could further enhance model performance, demonstrating the validity of the fusion improvement strategy.

Visualization Analysis

The partial detection results of RT-DETR-R18 and the heightened CSDE-DETR-R18 were visualized in Fig.6.

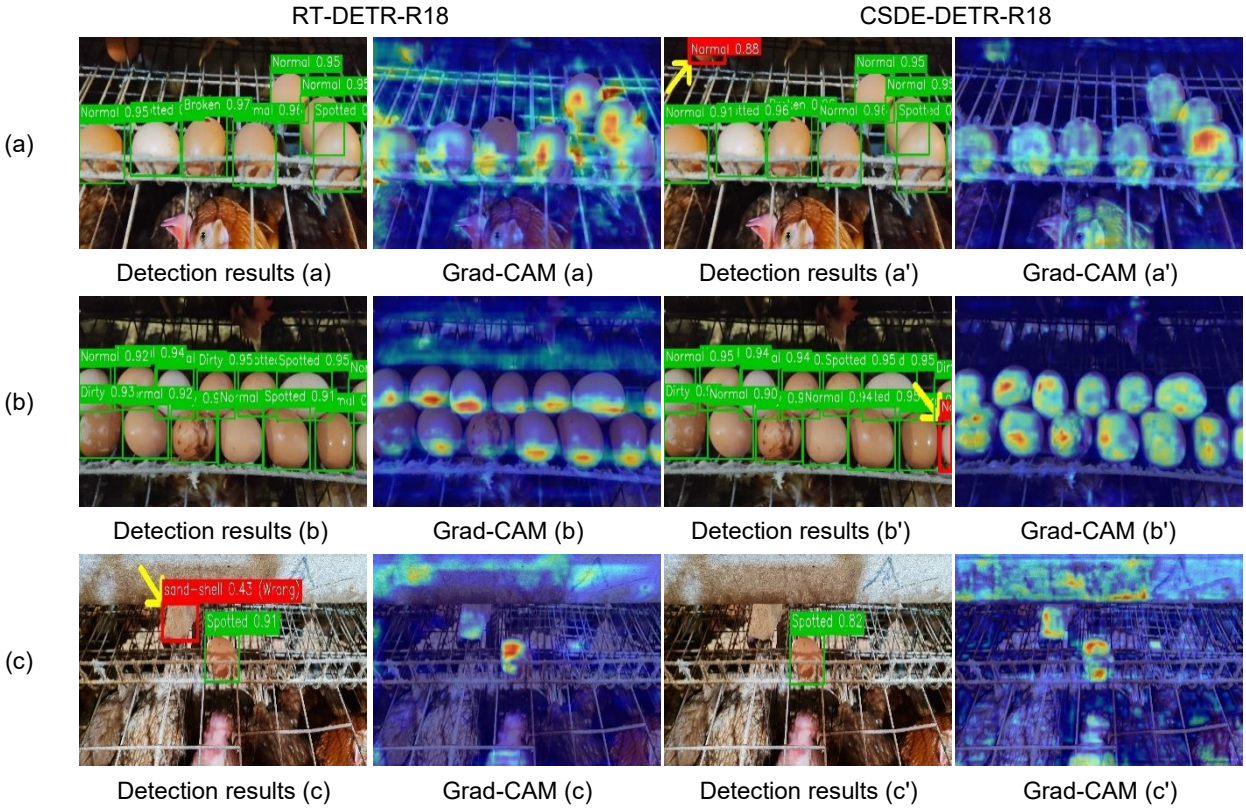

Fig. 6 - Visualization of model detection result

Through the above three image groups, a, b, and c in Fig. 6, the advantages of CSDE-DETR-R18 can be clearly observed. In group (a), for the egg in the upper left corner under dim lighting, RT-DETR-R18 demonstrated activation but exhibited diffuse attention, ultimately leading to a missed detection, while CSDE-DETR-R18 correctly focused its activation on the egg itself, achieving successful detection. In group (b), CSDE-DETR-R18 exhibited a more complete receptive field with its attention region covering the entire egg, and so successfully detected partially obscured eggs that RT-DETR-R18 failed to detect. In the images of group (c), both models activated on environmental objects. In contrast to RT-DETR-R18, which demonstrated weak activation (light color) and failed to extract sufficient information, CSDE-DETR-R18 exhibited more conspicuous activation on this environmental object, demonstrating that the model actively recognized sufficient detailed features and successfully classified it as a negative sample.

For this reason, it was concluded that the optimized CSD-DETR-R18 model was improved in three aspects, namely an expanded perception range, more focused attention, and more refined feature extraction. By employing this improved CSDE-DETR-R18 model, it can better adapt to conditions such as egg blockage, changes in lighting conditions, and strong background noise in cage environments.

Comparative Experiments

In order to assess the effectiveness of CSDE-DETR-R18 at a deeper level, several models with similar scales were selected for comparative analysis. The main algorithm paradigms in the field of automated detection within the agricultural and food industries were outlined below.

These included the classic two-stage detector, Faster R-CNN with a ResNet-18 backbone; followed by the traditional single-stage detector, SSD300 (Single Shot MultiBox Detector with 300×300 input); and concluding with the high-performance single-stage detectors, YOLOv8m and YOLOv8n. All experiments were conducted using the same experimental dataset and hardware platform.

Experimental results were revealed in Table 2 and Fig. 7.

Table 2

Performance comparison of different detection paradigms in cage-rearing environments

Detector Paradigm	Model	mAP@0.5(%)	Parameters (M)	FLOPs (G)	Inference Speed (FPS)
NMS-dependent Two-stage	Faster R-CNN	91.3	28.2	215.0	2.67
Legacy One-stage	SSD300	86.7	26.3	61.3	17.42
NMS-dependent One-stage	YOLOv8m	90.5	25.9	78.9	49.01
NMS-dependent One-stage	YOLOv8n	90.2	3.2	8.7	176.08
End-to-End One-stage	CSDE-DETR-R18	87.8	20.1	56.4	68.49

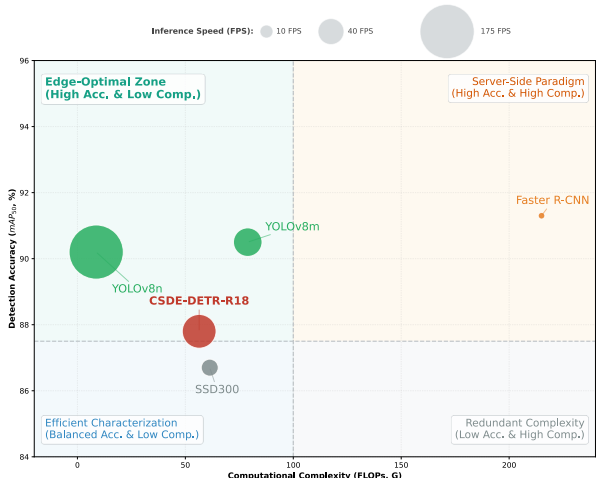

Fig. 7 - Accuracy-Efficiency Trade-off Analysis

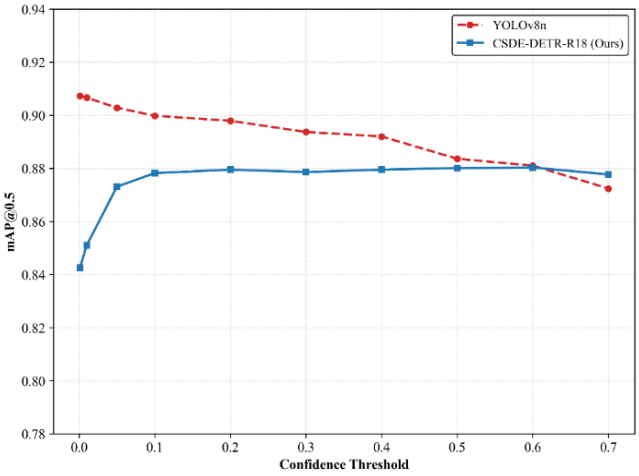

Fig. 8 - Comparison of detection accuracy under different confidence thresholds

The specific experimental data were presented in Table 2. For intuitive comparison, a precision-complexity-FPS trade-off bubble chart was plotted on the basis of the data in Table 2. As depicted in Fig. 7, SSD300 offered a balanced performance across accuracy, model size, and FPS, though none of these metrics were exceptional. Meanwhile, Faster-R-CNN achieved the highest accuracy but was limited by high FLOPs and low FPS, failing to meet the requirements for edge deployment and real-time detection applications. The YOLOv8n, YOLOv8m, and CSDE-DETR-R18 models all achieved high accuracy, moderate FLOPs, and high FPS that met application requirements. Among the three, YOLOv8n offered the second-highest accuracy, noticeably higher FPS, and the lowest FLOPs, appearing to be the optimal choice for edge deployment in abnormal egg detection. Nonetheless, as a consequence of its reliance on NMS, YOLOv8n suffered from threshold sensitivity and significant fluctuations in inference time during practical applications. This made it unsuitable for meeting the stability requirements of high-synchronization tasks like robotic arm sorting in the future. On that account, this paper conducted a further comparative analysis between CSDE-DETR-R18 and YOLOv8n in Section 4.5.

Advantages of End-to-End Detection Paradigm

Universally acknowledged for its end-to-end architecture, CSDE-DETR-R18 minimized a diverse spectrum of issues triggered by NMS post-processing, making it a more robust engineering choice. This paper analyzed and studied it from the perspectives of threshold sensitivity and inference stability.

Threshold Sensitivity

Traditional NMS dependency models, such as YOLOv8, adopted a "one-to-many" positive sample allocation strategy, and their performance was highly dependent on the fine coupling of confidence threshold and NMS threshold in the post-processing link, thereby making the model more sensitive to changes in super parameters. By virtue of the one-to-one set prediction mechanism in the end-to-end architecture, RT-DETR and others eliminated the generation of redundant prediction boxes from the underlying architecture, and could rely on the NMS post-processing. Fig. 8 demonstrated the sensitivity curve after a comprehensive comparison and

profound analysis of how different confidence thresholds affected the mAP value of abnormal egg recognition on the validation set.

When the confidence threshold was immensely low, CSDE-DETR-R18 contained background noise with considerably low classification score in a large number of boxes, and the interference accuracy was calculated accordingly. At this time, YOLOv8n used one-to-many matching to generate massive overlapping prediction boxes, which ensured the recall rate by means of a large-scale search, and the initial value of mAP maintained a high level. After the confidence threshold entered the reasonable interval (>0.1), the CSDE-DETR-R18 sensitivity curve suggested excellent consistency, and YOLOv8n performance demonstrated a significant downward trend with the increase of the threshold, within the confidence interval of 0.6-0.7 mAP@0.5, the value was already lower than CSDE-DETR-R18. As demonstrated by the above analyses, CSDE-DETR-R18's end-to-end architecture had better parameter robustness in abnormal egg recognition.

Inference Performance Evaluation of Edge Devices in High Density Scenarios

Edge devices generally have limited computing power. For this reason, it holds practical significance to carry out model deployment research for embedded platforms to reduce hardware costs and promote the engineering implementation of technology. Raspberry Pi is a common edge device for agriculture and food industry with low cost and high deployment flexibility.

In the abnormal egg recognition task, eggs may be dense. CSDE-DETR-R18 achieved considerable recognition accuracy in the data set containing a multitude of dense egg scenes in this paper. In practical production applications, not only the accuracy of the identification model is required, but also the time stability of the model is tested. In the dense target scenario, the NMS depends on the detector to process multiple redundant prediction boxes, and the post-processing time is strikingly prolonged. This problem is more prominent when the computing power of edge devices is limited. To be more specific, the CPU single-core performance of edge devices is difficult to support the exponential growth of serial logic operations of the NMS algorithm in the high-density scenario. In comparison, DETR series models use a fixed number of query mechanisms, which can maintain a constant inference frame rate.

To compare the inference stability of YOLOv8n and CSDE-DETR-R18 under a unified metric, the normalized delay ratio Lnorm was introduced as an evaluation metric.

$$L_{\text{norm}} = \frac{T_{\text{base}} + T_{\text{post}}}{T_{\text{base}}} \tag{12}$$

In Equation (12), T_{base} represents the normalization constant, [ms]. It was obtained by recording the time interval between the start of the model's post-processing phase and the initial point in the aftermath of image loading, subsequently taking the average across all experiments. T_{post} defines the recorded time interval of the model's post-processing phase, [ms].

The closer L_{norm} is to 1, the stronger the stability of the model's inference time. Using CSDE-DETR-R18 and YOLOv8n on a Raspberry Pi to process 50 images of abnormal eggs, box plots were generated by employing the median and quartiles, with scatter plots overlaid. The results are depicted in Fig. 10.

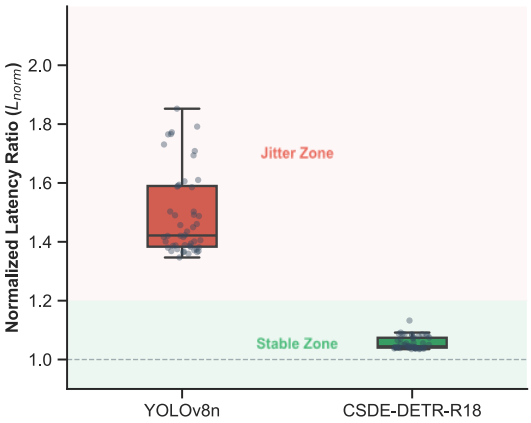

Fig. 9 - Normalized Latency Distribution

In Fig. 9, the median line of YOLOv8n fell within the Jitter Zone where L_{norm} exceeded 1.2. The inter-quartile range was exceptionally wide, with sample points exhibiting a highly dispersed distribution. Frame-to-frame inference latency demonstrated significant fluctuation.

The median of CSDE-DETR-R18 fell within the Stable Zone and was close to 1. Simultaneously, the inter-quartile range was considerably small and the sample points were highly concentrated, demonstrating that the post-processing overhead of CSDE-DETR-R18 was nearly zero and exhibited strong timing determinism.

As adequately demonstrated by the above results, CSDE-DETR-R18 fundamentally can eliminate the need for NMS through its end-to-end architecture, thereby avoiding the inference stability issues associated with it. CSDE-DETR-R18 employed the fixed-query mechanism of the DETR series models. In comparison with the YOLO series, which required serial logical comparisons of numerous candidate bounding boxes during post-processing —leading to fluctuating inference latency and non-linear growth—it achieved stable computation cycles on edge devices like the Raspberry Pi. In contrast to YOLOv8n, CSDE-DETR-R18 sacrificed the marginal precision gap and the extreme speed in sparse scenes, but gained the all-important temporal stability of the whole scene in the identification of abnormal eggs in the agricultural and food industries.

CONCLUSIONS

(1) The CSDE-DETR-R18 abnormal egg recognition model was proposed: on the basis of RT-DETR-R18, the context selection fusion module CSFM was introduced into the neck fusion path, the standard convolution in the backbone network was replaced by the detail enhanced convolution DEConv, and the NWD Loss+AIoU fusion loss function was adopted to replace the GIoU loss function at the regression optimization level, mAP@0.5, P and R increased by 3.4, 3.1, and 0.5 percentage points, respectively, while the floating-point operations and parameters remained basically the same.

(2) In comparison with YOLOv8n, CSDE-DETR-R18 overcomes the delay fluctuation and long tail effect stemmed from NMS in NMS dependent detectors. Although YOLOv8n still has a slight advantage in absolute accuracy indicators, CSDE-DETR-R18 demonstrates near-zero fluctuation in inference time in high-density scenarios, making it a more robust engineering choice. This research confirms that in the resource-constrained edge computing scenario, the end-to-end architecture has greater system stability value and can make up for the current minor accuracy disadvantage. In the long run, the end-to-end architecture eliminates the post-processing link of NMS, making the model have a better global optimization potential. With the lightweight improvement of relevant pre-training technology and attention mechanism, the detection accuracy of this architecture still has huge room for improvement, which is expected to further solve the dilemma of having both accuracy and stability in the edge computing scene in the future, providing a new perspective for the selection of recognition algorithms in agriculture and food industry.

ACKNOWLEDGEMENT

This work was supported by the Research on National Standards for Radio Frequency Identification (RFID) Tags (Project No. 69193060).

REFERENCES

[1] Bambilla R.L.M., Lee M.J.B., Valiente F.L., (2025). Freshness Condition Detection, Size Classification with Weight Prediction and Sortation of Philippine Egg Produce Using YOLOv8 and OpenCV, *2025 17th International Conference on Computer and Automation Engineering (ICCAE)*, IEEE, pp.158-164, Piscataway, NJ/USA. DOI: https://doi.org/10.1109/ICCAE64891.2025.10980499

[2] Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S., (2020). End-to-End object detection with transformers, *Proceedings of the European Conference on Computer Vision (ECCV)*, vol.12346, Springer, pp.213-229, Cham/Switzerland. DOI: https://doi.org/10.1007/978-3-030-58452-8_13

[3] Chen Z., He Z., Lu Z.M., (2024). DEA-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention, *IEEE Transactions on Image Processing*, vol.33, IEEE, pp.1002-1015, Piscataway, NJ/USA. DOI: https://doi.org/10.1109/TIP.2024.3354108

[4] Chewprecha K., Songserm O., Gertphol S., (2023). Comparison of SSD Models for Fertile Egg Image Classification, *2023 7th International Conference on Information Technology (InCIT)*, IEEE, pp.259-264, Chiang Rai/Thailand. DOI: https://doi.org/10.1109/InCIT60207.2023.10412958

[5] Du Y., Gao A., Song Y., Guo J., Ma W., Ren L., (2024). Young apple fruits detection method based on improved YOLOv5, *INMATEH - Agricultural Engineering*, vol.73, no.2, pp.84-93, Bucharest/Romania. DOI: https://doi.org/10.35633/inmateh-73-07

[6] Gautron J., Stapane L., Le Roy N., Nys Y., Rodriguez-Navarro A.B., Hincke M.T., (2021). Avian eggshell biomineralization: an update on its structure, mineralogy and protein tool kit, *BMC Molecular and Cell Biology*, vol.22, BioMed Central, pp.11, London/UK. DOI: https://doi.org/10.1186/s12860-021-00350-0

[7] He J., Erfani S., Ma X., Bailey J., Chi Y., Hua X.S., (2021). α-IoU: A Family of Power Intersection over Union Losses for Bounding Box Regression, *Advances in Neural Information Processing Systems (NeurIPS)*, vol.34, Curran Associates, Inc., pp.20230-20242, Red Hook, NY/USA. URL: https://proceedings.neurips.cc/paper/2021/hash/a8f15eda80c50adb0e71943adc8015cf-Abstract.html

[8] Li J.M., Huang Z.L., Xia L.Q., Sun H., (2025). Tomato maturity detection based on improved YOLOv8n, *INMATEH - Agricultural Engineering*, vol.75, no.1, pp.619-629, Bucharest/Romania. DOI: https://doi.org/10.35633/inmateh-75-53

[9] Liu X., Shen C.Y., Lyu X.Z., Dong M.P., Bao Q.H., Zhang Y.Z., (2022). Egg freshness recognition model based on improved MobileNetV3-Large (基于改进 MobileNetV3-Large 的鸡蛋新鲜度识别模型), *Transactions of the Chinese Society of Agricultural Engineering*, vol.38, no.17, pp.196-204, Beijing/China. DOI: https://doi.org/10.11975/j.issn.1002-6819.2022.17.021

[10] Nasiri A., Omid M., Taheri-Garavand A., (2020). An automatic sorting system for unwashed eggs using deep learning, *Journal of Food Engineering*, vol.283, Elsevier, pp.110036, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.jfoodeng.2020.110036

[11] Subedi S., Bist R., Yang X., Chai L., (2023). Tracking floor eggs with machine vision in cage-free hen houses, *Poultry Science*, vol.102, no.6, Elsevier, pp.102637, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.psj.2023.102637

[12] Turkoglu M., (2021). Defective egg detection based on deep features and Bidirectional Long-Short-Term-Memory, *Computers and Electronics in Agriculture*, vol.185, Elsevier, pp.106152, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2021.106152

[13] Valencia Y.M., Majin J.J., Taveira V.B., Salazar J.D., Stivanello M.E., Ferreira L.C., Stemmer M.R., (2021). A novel method for inspection defects in commercial eggs using computer vision, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol.XLIII-B2-2021, Copernicus Publications, pp.809-816, Göttingen/Germany. DOI: https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-809-2021

[14] Wang Z.S., Wang C.F., Li X.S., Xia C.Q., Xu J.W., (2025). MLP-Net: Multilayer Perceptron Fusion Network for Infrared Small Target Detection, *IEEE Transactions on Geoscience and Remote Sensing*, vol.63, IEEE, pp.1-13, Piscataway, NJ/USA. DOI: https://doi.org/10.1109/TGRS.2024.3515648

[15] Xu C., Wang J.W., Yang W., Yu H., Yu L., Xia G.S., (2022). Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.190, Elsevier, pp.79-93, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.isprsjprs.2022.06.002

[16] Yang N., (2021). Egg production in China: current status and outlook, *Frontiers of Agricultural Science and Engineering*, vol.8, no.1, Higher Education Press, pp.25-34, Beijing/China. DOI: https://doi.org/10.15302/J-FASE-2020363

[17] Yang X., Bist R.B., Subedi S., Chai L., (2023). A Computer Vision-Based Automatic System for Egg Grading and Defect Detection, *Animals*, vol.13, no.14, MDPI, pp.2354, Basel/Switzerland. DOI: https://doi.org/10.3390/ani13142354

[18] Yin J.J., Kang J.Q., Xiao D.Q., (2024). Lightweight detection algorithm for duck egg cracks based on improved YOLOv5l (基于改进 YOLOv5l 的轻量化鸭蛋裂纹检测算法), *Transactions of the Chinese Society of Agricultural Engineering*, vol.40, no.5, pp.216-223, Beijing/China. DOI: https://doi.org/10.11975/j.issn.1002-6819.202307259

[19] Zhao Y., Lv W., Xu S., Wei J., Wang G., Dang Q., Liu Y., Chen J., (2024). DETRs Beat YOLOs on Real-time Object Detection, *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp.16965-16974, Seattle, WA/USA. DOI: https://doi.org/10.1109/CVPR52733.2024.01605

[20] Zhao Z.X., Wei H.F., Huang Y., Huang X.P., Mi Y.L., Luo Y.F., (2023). Online real-time detection system for broken shell eggs based on improved YOLOv7 (基于改进 YOLOv7 的破壳鸡蛋在线实时检测系统), *Transactions of the Chinese Society of Agricultural Engineering*, vol.39, no.20, pp.255-265, Beijing/China. DOI: https://doi.org/10.11975/j.issn.1002-6819.202305228