STRAWBERRY IDENTIFICATION AND KEY POINTS DETECTION FOR PICKING BASED ON IMPROVED YOLOV8-POSE AT RED RIPE STAGE
/
基于改进 YOLOv8-pose 的红熟期草莓识别与采摘关键点检测

Jinlong WU*[1,2]，Ronghui MIAO [1,2]
[1] College of information science and engineering, Shanxi agricultural university, Taigu / China;
[2] College of agricultural engineering, Shanxi agricultural university, Taigu / China
Tel: +86-18503482797; E-mail: wujinlong8192@163.com
DOI: https://doi.org/10.35633/inmateh-76-08

Keywords: *YOLOv8-pose, strawberry identification, fruit stem, key point detection, red ripe stage*

ABSTRACT
To solve the problems of low precision in locating stem picking points and difficulty in recognizing occluded strawberry during the operation of strawberry picking robots, this paper proposed an improved YOLOv8-pose method for strawberry fruits identification and key points detection at the red ripe stage. Based on the YOLOv8-pose human posture estimation model, three categories (strawberry, stem, and picking points) were annotated. The acquired images were divided into training, validation, and test sets in an 8:1:1 ratio. In order to improve the feature extraction ability of the model for small targets, shuffle attention (SA) mechanism was added into the backbone network of YOLOv8-pose. Additionally, a comparative analysis was conducted to assess the impact of six attention mechanisms of CBAM (Convolutional block attention module), SimAM (Simple attention module), GAM (Global attention module), EMA (Efficient multi-scale attention), SK (Selective kernel attention), and SA on the detection results. Experimental results show that the proposed method can quickly and accurately detect strawberry fruits and key points for picking. The Precision (P), Recall (R), and mean average precision (mAP)50 values for both bounding boxes and key points based on SA mechanism were 99.7%, 100.0%, and 99.5% respectively, which were superior to the other attention mechanisms. Compared with YOLOv5-pose and YOLOv8-pose models, the improved model had the best P, R, and mAP50 values, and its memory usage was 6.4MB, which was also optimal. The improved method can provide crucial technical support for precise robotic strawberry picking.

摘要
为解决草莓采摘机器人作业中果梗采摘点定位精度低和遮挡草莓识别困难等问题，本文以红熟期草莓为研究对象，提出一种改进的 YOLOv8-pose 草莓果实识别及采摘关键点检测方法。以 YOLOv8-pose 人体姿态估计模型为基础，标注了草莓、果梗、采摘点 3 个类别。将采集的图像按 8:1:1 的比例划分成训练集、验证集和测试集。为了提高模型对小目标的特征提取能力，在 YOLOv8-pose 的骨干网络中添加了 shuffle 注意力机制，并对比分析了 CBAM（Convolutional block attention module），SimAM（Simple attention module），GAM（Global attention module），EMA（Efficient multi-scale attention），SK（Selective kernel attention）以及 SA 六种注意力机制对检测结果的影响。实验结果表明，本文提出的方法可以对草莓果实及采摘关键点进行快速准确检测，基于 SA 注意力机制的边界框和关键点的检测精确率（Precison，P），召回率（Recall，R），平均精度均值（Mean average precision，mAP）mAP50 均为 99.7%，100.0%和 99.5%，均优于其他注意力机制；与 YOLOv5-pose、YOLOv8-pose 模型相比，改进模型的 P，R，mAP50 值都是最优的，模型的内存占用量为 6.4MB，也是最优的，该模型可以为机器人精准采摘提供重要的技术支持。

INTRODUCTION
With various nutrients such as amino acids, vitamins, minerals, strawberry is known as the “queen of fruits” *(Hu et al., 2024)*. At present, strawberry picking still rely on manual labor, which is time-consuming and costly *(Wang et al., 2022)*. The implementation of intelligent strawberry picking is of great significance for improving picking efficiency, reducing labor costs, minimizing fruit damage, and enhancing fruit quality. The

[1] *Jinlong Wu, Lect.; Ronghui Miao, Assoc. Prof.*
[2] *Jinlong Wu, Ph.D.Stud.; Ronghui Miao, Ph.D.Stud.*

harvesting of strawberry picking robots is mainly achieved by cutting fruit stems *(Ma et al., 2025)*. However, the diameter of strawberry stem is relatively small, and affected by complex backgrounds such as lighting and obstruction from branches and leaves, making it difficult to identify and locate strawberry. Accurate positioning fruits and stems are of great significance for improving the intelligence and automation of strawberry picking robots.

In recent years, with the development of deep learning technology, researchers at home and abroad applied deep learning algorithms to the recognition and detection of various fruits, and have achieved certain results. *Gai et al.* (*2024*) proposed a blueberry identification algorithm named TL-YOLOv8 based on the YOLOv8 algorithm to improve the detection accuracy, by introducing an improved MPCA (Multiplexed coordinated attention) in the last layer of the backbone network, the feature extraction capability during the training process was enhanced. *Gao et al. (2025)* proposed a novel YOLOv8n-CA method for tomato maturity recognition, which defines four maturity stages: unripe, turning color, turning ripe, and fully ripe, and the experimental results showed that the YOLOv8n-CA model had a computational complexity of 6.9 GFLOPs, and a weight file size of just 4.90 MB. *Chen et al. (2024)* proposed a lightweight target detection method called LFA-YOLO to address the problem of leakage and misdetection in the detection of lychee fruit anthracnose due to the complexity of the agricultural environment, and it achieved a mAP50 of 88.61%, with precision and recall rates of 90.18% and 82.77% respectively. In the application of strawberry recognition and detection, an improved YOLOv5 object detection model was proposed, the ghost module was used to optimize the backbone network, thereby reducing the number of parameters and complexity of the model, and the experimental results indicated that this model can provide a more accurate and efficient detection method for strawberry *(Peng et al., 2025)*. An improved YOLOv5s-based method for rapid detection of strawberry ripeness in greenhouse was proposed, shuffle block was introduced as the feature extraction network in the backbone to lightweight the model, the results showed that the average precision of the improved YOLO-ODM model can reach 97.4%, indicating that this lightweight method can detect the ripeness of strawberry quickly and accurately *(Chen et al., 2023)*. A combination of improved YOLOv4 and traditional image processing technology for strawberry stem recognition in complex picking scenarios was proposed, and this model can accurately locate the picking points and demonstrate better applicability in practical picking scenarios *(Huang et al., 2023)*. However, most of the aforementioned researches mainly focused on object detection of fruits and maturities for strawberry, the detection of key points for picking were relatively few, which cannot be directly applied to strawberry picking operations. Meanwhile, the maturities of strawberry can be divided into four stages: green ripe, white ripe, color change, and red ripe. Among these, the red ripe stage is the complete ripening stage, with uniform red color on the surface, soft texture, and unique aroma. Therefore, identifying the red ripe strawberry and picking points are key steps for strawberry picking robots.

This paper takes red ripe strawberry as research object, refers on the human posture estimation model, and proposes an improved YOLOv8-pose model for detecting key points of fruit and stem. In order to improve the feature extraction ability of the model for small targets (fruit stems, picking points), SA mechanism was added to the backbone network. This study can achieve the prediction of picking point for stems, providing technical support for machine picking of strawberry.

MATERIALS AND METHODS

Data collection

The experimental strawberry images were collected from strawberry planting base in Xianggu village, Taigu District, Jinzhong city, Shanxi province in March 2024. The datasets were carried out using a Huawei P40 Pro camera. The vertical distance between the shooting device and the ground was 150 mm - 250 mm. A total of 158 images were collected, with an image resolution of 3000 pixels × 4000 pixels. The experimental processing platform was configured with Intel Core (TM) i9-9900K @ 3.60GHz, 16GB of memory, and NVIDIA GeForce RTX 2080Ti main graphics card for GPU acceleration. Figure 1 shows the collected strawberry images, which include complex environment such as single fruit, multi fruit, fruit occlusion, and branch and leaf occlusion.

(a) Multi fruit + fruit obstruction (b) Multi fruit + no fruit obstruction (c) Single fruit (d) Single fruit + leaf obstruction

Fig. 1 - Sample images of strawberry

Dataset annotation

This paper used posture estimation model to achieve strawberry identification and key points detection on fruit stems. YOLOv8 is capable of performing complex tasks such as instance segmentation *(Fan et al., 2024)*, object detection *(Kumar and Muhammad, 2023)*, image classification, posture estimation *(Ma et al., 2024)*, and key points detection. Among them, posture estimation can obtain the positions of key points in the images, and the positions are usually represented by two-dimensional coordinates (x, y) or 3D coordinates (x, y, visible), and it is mainly used to recognize various postures of the human body *(Peruzzi et al., 2025)*. In YOLOv8, there are 1 category and 17 key points, the category name is "person", and each key point represents a different part of human body. The mappings of each index to the corresponding body joint are: 0: nose; 1: left eye; 2: right eye; 3: left ear; 4: right ear; 5: left shoulder; 6: right shoulder; 7: left elbow; 8: right elbow; 9: left wrist; 10: right wrist; 11: left hip; 12: right hip; 13: left knee; 14: right knee; 15: left ankle; 16: right ankle. Figure 2 shows the key points of human body.

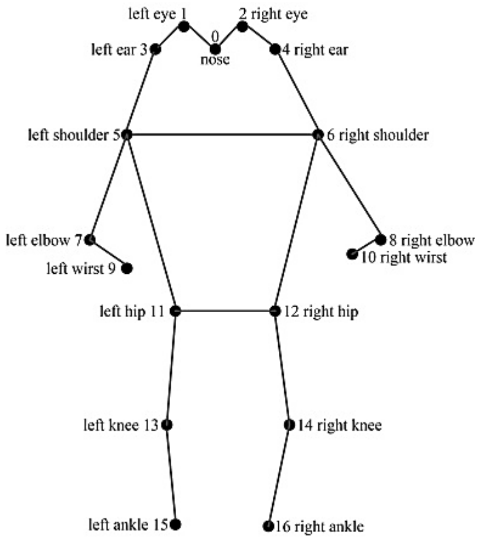


Fig. 2 – Key points of human body

Based on the posture estimation model of human body, a strawberry recognition and stem key points detection model was constructed. In this paper, the annotation tool Labelme was used to annotate the strawberry image datasets, and three types of labels were annotated: strawberry (in rectangle), stem (in point) and pick (in point). The annotation results were shown in Figure 3. The principle of labeling is: only the red ripe strawberry fruits will be labeled, and fruits that are too heavily obscured will not be labeled; the center point of the fruit stem was selected as key point; the picking point is marked at a distance of 1 cm from the stem. The format of the annotation file was JSON, in order to ensure that the datasets required for the experiment meets of COCO format, the annotated files need to be converted to TXT format.

Dataset production

After screening, a total of 98 images were selected for model training, validation, and testing. Then the image datasets were divided in an 8:1:1 ratio. Eighty percent (78 images) were randomly selected as the training sets for model training and parameter optimization. The remaining twenty percent (20 images) were used as validation and testing sets to verify the performance of the model.

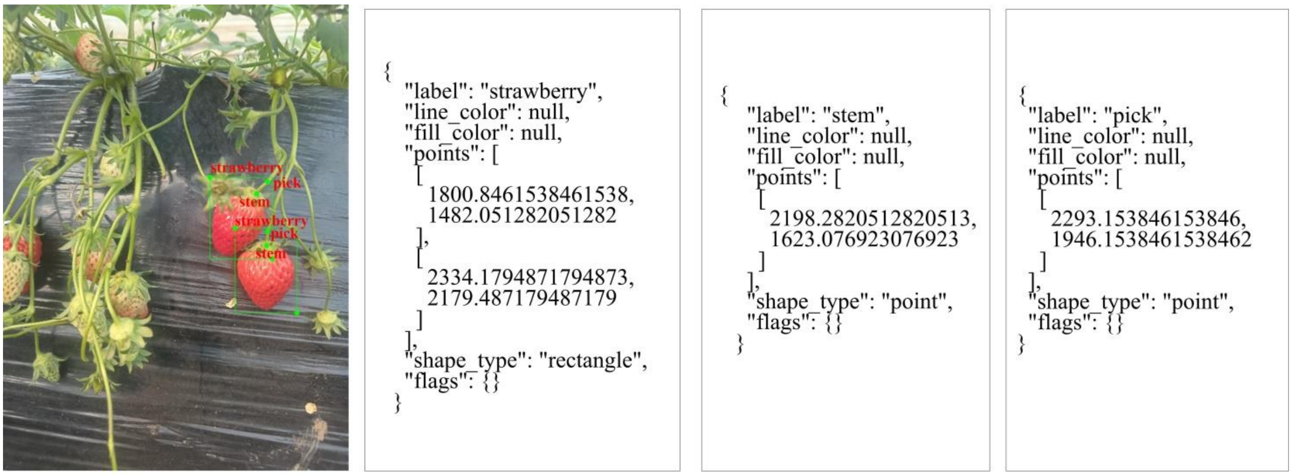


Fig. 3 – Annotation results

Dataset augmentation

YOLOv8 provides some dataset automatic augmentation methods to improve the generalization ability and detection performance of the model, and the main augment strategies include RandAugment, AutoAugment, and AugMix *(Sanat et al., 2023)*. Among them, RandAugment is a commonly used technique that enhances training data by randomly selecting a series of image transformations (such as brightness adjustment, contrast change, rotation, cropping, etc.), thus enabling the model to better learn various features and patterns in the images *(Lee et al., 2023)*. Some of the hyperparameter settings are as follows: the “hsv_h” (proportion of image hue augmentation) is 0.015; the “hsv_s” (proportion of image saturation augmentation) is 0.7; the “hsv_v” (proportion of image value augmentation) is 0.4; the “translate” (proportion of image translation) is 0.1; the “scale” (gain of image scaling) is 0.5; the “fliplr” (probability of image flip left-right) is 0.5, the “mosaic” (probability of image mosaic) is 1.0, etc. Figure 4 shows the data augmentation result, randomly applying image enhancement operations with a certain probability and intensity can generate diverse training samples. And it performs stably on different datasets and model sizes, especially suitable for resource limited scenarios.

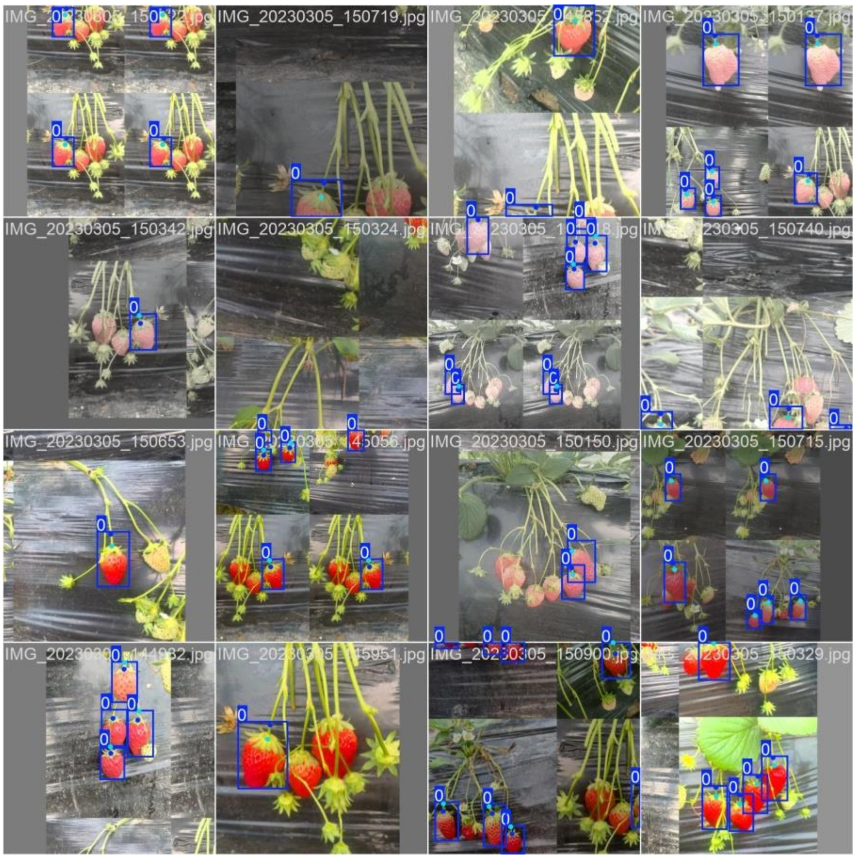


Fig. 4 – Dataset augmentation result

Strawberry identification and picking key points detection model

This paper constructed a model for strawberry identification and fruit stem key points detection based on the improved YOLOv8-pose network. In order to improve the feature expression ability, SA mechanism was added into the backbone network of YOLOv8-pose. Figure 5 shows the network structure of the improved YOLOv8-pose model.

YOLOv8-pose model

With advantages of higher detection accuracy and speed, new backbone network structure, anchor free detection head, and new loss function *(Jo et al, 2024)*, YOLOV8 model can achieve tasks such as object detection, instance segmentation, and key points detection, and it has been widely used in security monitoring, autonomous driving, smart homes and industrial automation. The structure of YOLOv8 mainly consists of three parts: Backbone, Neck, and Head.

a. Backbone: responsible for feature extraction. A series of convolutional and deconvolution layers are adopted to realize deep feature extraction. The introduction of residual connections and bottleneck structures can improve the performance of the network *(Zhai et al., 2024)*. Compared with YOLOv5, C2f module is introduced to replace the original C3 module, which can achieve light weighting with fewer parameters. Moreover, the depth wise separable convolution and dilated convolution techniques can further enhance feature extraction capabilities.

b. Neck: located between the backbone and head parts, responsible for feature fusion enhancement. The spatial pyramid pooling fusion (SPPF) can fuse features maps from different stages of backbone to enhance feature representation capabilities *(Dong et al., 2024)*.

c. Head: responsible for generating the final detect results. This section uses simple convolutional and upsampling layers, combined with the feature maps from the neck part to achieve object detection.

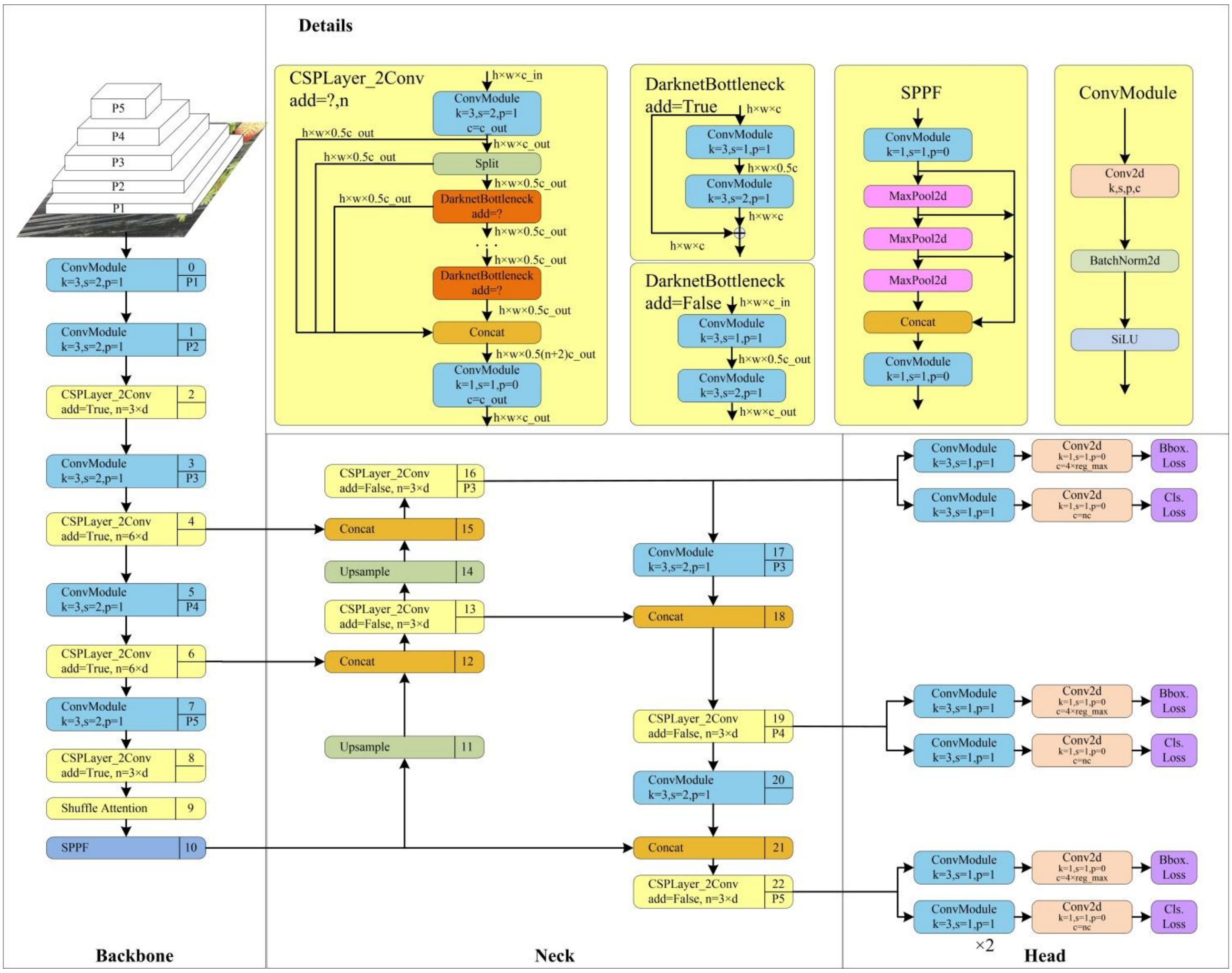


Fig. 5 – The network structure of the improved YOLOv8-pose model

SA mechanism

Attention module can focus on useful feature information and suppress unimportant feature. SA mechanism is capable of selectively weighting the information within the feature maps *(Zhang and Yang, 2021)*. By introducing this mechanism, YOLOv8 model can be made more stable and accurate in finding the position and classifying the tested object. The advantages of SA mechanism are: it efficiently combines spatial and channel attention mechanisms, thus improving the performance of the model; by dividing the channel dimensions of the input feature map into multiple groups, the computational cost is effectively reduced; by using channel rearrangement technique, the expression ability of channel features is enhanced.

The structure of SA is shown in Figure 6. The processing procedure of SA module mainly includes: (1) Making the input feature map into G groups to obtain sub features of each group; (2) Use spatial attention and channel attention to obtain information of the sub features; (3) Channel shuffle operation is used to fuse the features from different groups. The blue parts represent the spatial attention mechanism, and the green ones represent the channel attention mechanism. In Figure 6, X is the feature map; c , h and w represent the number of channels, spatial height and width respectively; the number of channels for each group is c/g ; the number of channels for each attention branch is $c/2g$; GN is group normalization.

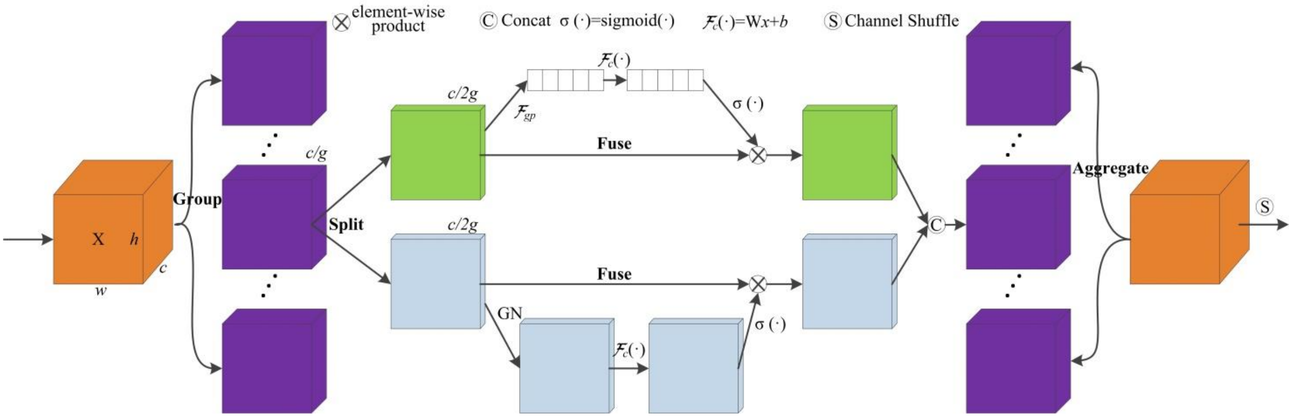


Fig. 6 – SA mechanism structure diagram.

Model evaluation

In order to evaluate the performance of identification and detection, mAP, P and R were selected as evaluation indicators. Among them, mAP refers to the average precision at different recall rates. P represents the proportion of samples predicted as positive cases that are actually positive cases. R represents the proportion of samples correctly identified as positive cases by the model among all actual positive cases. The calculation formulas for evaluation indicators are shown in equations (1) to (3).

$$mAP = \frac{\sum_{i=1}^{C} \int_0^1 P_i dR_i}{C} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

In which:

C is the number of categories; TP (True positives) is the number of correctly identified positive samples; TN (True negatives) is the number of correctly identified negative samples; FP (False positives) is the number of samples incorrectly identified as positive; FN (False negatives) is the number of incorrectly identified negative samples.

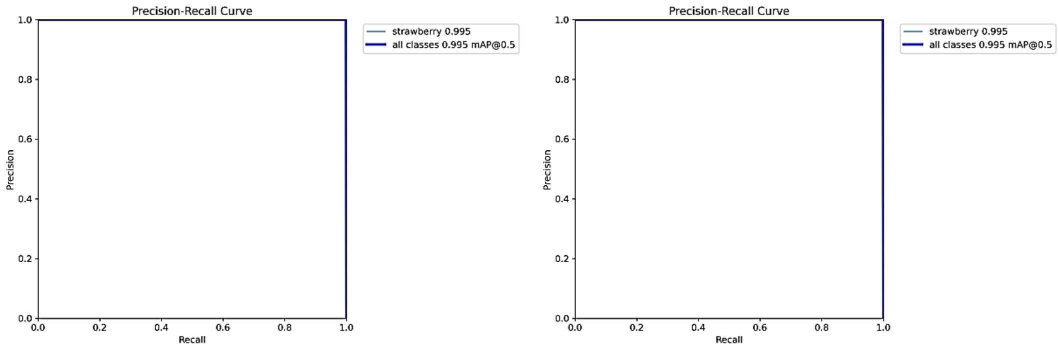
Model parameters

In the improved YOLOv8-pose model, the batch size is 16; the epoch is 200; the initial learning rate is 0.01; the momentum is 0.937. Batch size represents the number of training set images input for each batch; epoch represents the number of iterations of model training; learning rate is a parameter that controls the size of the gradient descent step in each iteration update; momentum can accelerate the update of model parameters.

RESULTS AND DISCUSSION

Identification and detection results of the improved YOLOv8-pose

Figure 7 shows the PR curves of box and pose on the training sets. PR curve is a common method used to evaluate the performance of the classification models. The horizontal axis of the PR curve is precision, and the vertical axis is recall. The closer the PR curve is to the upper right corner, the better the performance of the model. It can be seen that the mAP50 values for both box and pose were 0.995, which were close to 1, indicated that the model had a better balance between precision and recall, and its performance was good.



(a) Box PR curve *(b) Pose PR curve*

Fig. 7 - PR curves on the training sets

Figure 8 shows the loss curves on the training sets. The loss functions of YOLOv8-pose mainly includes five types of losses: bounding box loss (box_loss), key point loss (pose_loss), key point confidence loss (kobj_loss), classification loss (cls_loss), and distribution focal loss (dfl_loss). Bounding box loss represents the deviation between the predicted bounding box and the actual bounding box. Key point loss represents the deviation between the predicted key point positions and the actual positions. Key point confidence loss is used to measure the confidence level of the model in predicting key points, and it is adjusted by comparing it with the true confidence level. Classification loss represents the deviation between the predicted category and the true category. Distribution focal loss is used to optimize the predicted distribution of bounding boxes. As shown in figure 8, with the increase of training epochs, the loss value continuously decreased and tended to stabilize, indicated that the performance of the model was gradually improving in the classification and prediction tasks.

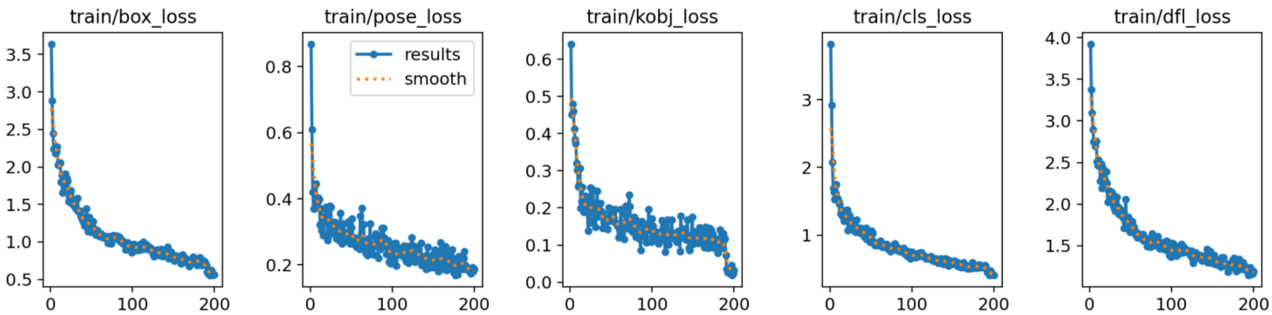


Fig. 8 - Loss curves on the training sets

Figure 9 shows the identification and detection results of strawberry. In the recognition results, the blue rectangle represents the identified results of strawberry at red ripe stage; the red and yellow points present the picking and stem key points. It can be seen that the improved model can accurately identify the strawberry at red ripe stage, because it only has one target in this paper, the detection and recognition rate is relatively high. There is a certain overlap between the stem and picking points, due to the small color difference of key points, the detection results are not ideal.

Fig. 9 – The prediction results of the improved YOLOv8-pose model

Identification and detection results of different attention mechanisms

In order to verify the impact of attention mechanisms on the performance of the model, six attention mechanisms of CBAM *(Woo, et al., 2018)*, SimAM *(Yang, et al., 2021)*, GAM *(Liu et al., 2021)*, EMA *(Zhang et al., 2024)*, SK *(Peng et al., 2020)* and SA were set for comparative experiments.

Table 1 shows the detection results of different attention mechanisms. Here, mAP50 represents the average accuracy of the model when the IoU (Intersection over union) threshold is 0.5, while mAP50-95 represents the average accuracy when the IoU ranging from 0.5 to 0.95. And mAP50-95 provides a more comprehensive evaluation of the performance under different degrees of overlap.

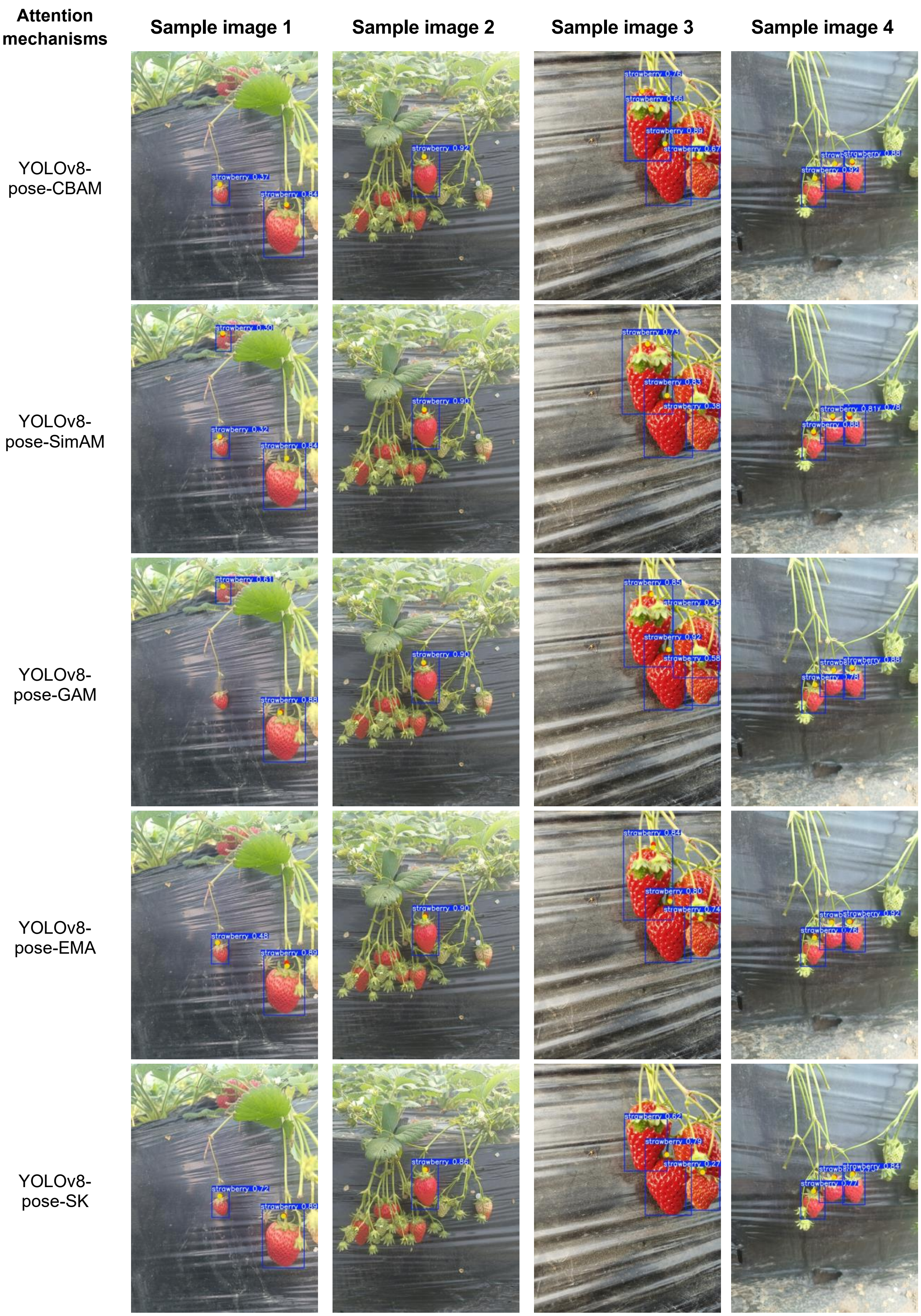
From table 1, it can be seen that the indicators of all the attention mechanisms can reach 81% or above, and the detection results were good; the detection results of YOLOv8-pose-EMA and YOLOv8-pose-SK were relatively poor; the memory usage of YOLOv8-pose-SK was 17.5MB, which was the highest among all the attention mechanisms; the box and pose P, R, mAP50 of SA were all 99.7%, 100.0% and 99.5%, and the memory usage of YOLOv8-pose-SA was 6.4MB, these values were all optimal.

Figure 10 shows the partial identification and detection results of different attention mechanisms. It can be seen that all models were able to identify strawberry fruits and key points of picking with high accuracy; the model of YOLOv8-pose-GAM had a certain of miss detection. Due to the fact that both the stem and picking point are green, there was a certain overlap in the results.

Table 1

Detection results of different attention mechanisms

Attention mechanisms	Box				Pose				Memory usage (MB)
	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	
YOLOv8-pose-CBAM	99.7	99.9	99.5	82.0	99.7	99.9	99.5	97.5	6.4
YOLOv8-pose-SimAM	99.7	100.0	99.5	82.1	99.7	100.0	99.5	97.3	6.4
YOLOv8-pose-GAM	99.0	96.0	99.1	81.3	100.0	99.5	99.5	97.5	7.3
YOLOv8-pose-EMA	90.9	100.0	98.6	83.7	90.9	100.0	98.6	95.0	7.7
YOLOv8-pose-SK	93.2	96.0	98.7	81.1	100.0	95.1	99.2	96.6	17.5
YOLOv8-pose-SA	**99.7**	**100.0**	**99.5**	**83.5**	**99.7**	**100.0**	**99.5**	**97.3**	**6.4**


Attention mechanisms
Sample image 1
Sample image 2
Sample image 3
Sample image 4
YOLOv8-pose-CBAM
YOLOv8-pose-SimAM
YOLOv8-pose-GAM
YOLOv8-pose-EMA
YOLOv8-pose-SK

YOLOv8-pose-SA

Fig. 10 – The prediction results of different attention mechanisms

Identification and detection results of different models

In order to verify the detection performance of different posture estimation models, this paper compared the results of YOLOv5-pose, YOLOv8-pose, and the improved YOLOv8-pose method. Table 2 shows the detection results of different models. From table 2, it can be seen that the results of our method were optimal, and superior to the other models.

Table 2

Detection results of different models

Models	Box				Pose				Memory usage (MB)
	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	
YOLOv5-pose	85.6	82.6	83.7	71.2	82.1	83.8	84.1	73.4	7.5
YOLOv8-pose	91.3	83.5	95.7	74.7	81.7	96.0	96.3	93.7	6.4
Improved YOLOv8-pose of our method	**99.7**	**100.0**	**99.5**	**83.5**	**99.7**	**100.0**	**99.5**	**97.3**	**6.4**

CONCLUSIONS

Faced with the characteristics of small stem diameter and easy damage to the flesh of strawberry, in this paper, an improved YOLOv8-pose method was proposed for strawberry fruit identification and key points detection method at red ripe stage. Based on the YOLOv8-pose human posture estimation model, three categories of strawberry, stem, and picking points were annotated. In order to improve the feature extraction ability of the model for small targets, SA mechanism was added into the backbone network of YOLOv8-pose. Experimental results show that the proposed method can quickly and accurately detect strawberry fruits and key points of picking, and can provide crucial technical support for precise robotic strawberry picking.

At present, there are varieties of strawberry, and the method in this paper is only suitable for detecting red strawberry. It may have certain limitations in identifying and locating strawberry in other colors. Subsequent research will introduce white strawberry or pink strawberry to enrich the images, so that improving the generalization ability and adaptability to different strawberry. Generally, intelligent fruit and vegetable picking robots are composed of vision, control, execution and walking systems. Among them, the visual system plays a crucial role in harvesting robots. As the “eyes” of the picking robots, it can accurately identify the position, size, and maturity of the target fruit through image recognition and object detection technology, and provide precise guidance for the execution system. The future research direction is to combine the method proposed in this paper with RGBD depth cameras to further obtain three-dimensional information of fruits, and guide the robotic arm to perform picking operations.

ACKNOWLEDGEMENT

This research was funded by Shanxi Province Applied Basic Research Youth Project, grant number 202203021212428; Shanxi Province Applied Basic Research Youth Project, grant number 202203021212414.

REFERENCES

[1] Chen, J. S., Liu, M., & Li J. C. J. (2024) LFA-YOLO: Detection of lychee fruit anthracnose based on UAV images and deep learning [J]. *Applied Engineering in Agriculture*, *40*(5), 515-523.

[2] Chen, R. F., Xie, Z., & Lin, C. (2023) YOLO-ODM based rapid detection of strawberry[J]. *Journal of Huazhong Agricultural University, 42*(04), 262-269.

[3] Dong, X. H., Li, S. X., Zhang, J. X. (2024) YOLOV5s object detection based on Sim SPPF hybrid pooling[J]. *Optoelectronics Letters, 20*(6), 367-371.

[4] Fan, Y. H., Zhang, L. X., Zheng, C. X., Wang X. Y, Zhu J. H, & W, L. (2024) Instance segmentation of faces and mouth-opening degrees based on improved YOLOv8 method[J]. *Multimedia Systems, 30*(5), 269-278.

[5] Gao, X., Ding, J. Y., Zhang, R. H., & Xi, X. B. (2025) YOLOv8n-CA: Improved YOLOv8n Model for Tomato Fruit Recognition at Different Stages of Ripeness [J]. *Agronomy, 15*(1), 188.

[6] Gai, R., Liu, Y., & Xu, G. (2024) TL-YOLOv8: A Blueberry Fruit Detection Algorithm Based on Improved YOLOv8 and Transfer Learning[J]. *IEEE Access, 12*(000), 13.

[7] Hu, H. P., Chen, M. X., Huang, L. B., & Guo, C. (2024). BHI-YOLO: A Lightweight Instance Segmentation Model for Strawberry Diseases[J]. *Applied Sciences, 14*(21), 9819.

[8] Huang, J. C., Wang, T., & Zhang, D. (2023) Recognition and localization of strawberry stalk leaf based on improve YOLOv4[J]. *Journal of Chinese Agricultural Mechanization, 44*(9), 205-213.

[9] Jo, S., Woo, J., Kang C. H., & Kim, S.Y. (2024) Damage detection and segmentation in disaster environments using combined YOLO and Deeplab [J]. *Remote sensing, 16*(22), 4267.

[10] Kumar, D., & Muhammad, N. (2023) Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8 [J]. *Sensors, 23*(20), 8471.

[11] Lee, G., Yonrith, P., Yeo, D., & Hong, A. (2023) Enhancing detection performance for robotic harvesting systems through RandAugment [J]. *Engineering Applications of Artificial Intelligence, 123*(PartC), 11.

[12] Liu, Y., Shao, Z., & Hoffmann, N. (2021) Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions [C]. *arXiv preprint arXiv: 2112.05561*.

[13] Ma, M. X., Ma, H., & Song, H.W. (2024) Pose estimation algorithm for small target pedestrians in urban street view based on YOLO-pose [J]. *Computer engineering, 50*(4), 177-186.

[14] Ma, Z. H., Dong, N. S., Gu J. Y., Cheng H. C., Meng Z. C, & Du X. Q. (2025) STRAW-YOLO: A detection method for strawberry fruits targets and key points [J]. *Computers and Electronics in Agriculture, 230*: 109853.

[15] Peng, Y., Qiao, Y. H., Zhang, C. Y., Yao, J. & Bao, D. L. (2025) Research of strawberry detection method with improved YOLOv5 model [J]. *Journal of Chinese Agricultural Mechanization, 46*(1), 213-219.

[16] Peng, Z. H., Li, Q., & Liang, S. (2020) Multi-Scale Dense Selective Kernel Spatial Attention Network for Single Image De-raining [C]. *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), IEEE*, 341-347.

[17] Peruzzi, G., Galli, A., Giorgi G., & Pozzebon, A. (2025) Sleep posture detection via embedded machine learning on reduced set of pressure sensors [J]. *Sensors, 25*(2), 458.

[18] Sanat, R., Diego, D., & Cristians, G. (2023) TRandAugment: temporal random augmentation strategy for surgical activity recognition from videos [J]. *International journal of computer assisted radiology and surgery, 18*(9), 1665-1672.

[19] Wang, Y., Yan, G., Meng, Q. L., Yao, T., Han, J. F., & Zhang, B. (2022) DSE-YOLO: Detail semantics enhancement YOLO for multi-stage strawberry detection [J]. *Computers and Electronics in Agriculture, 198*, 107057.

[20] Woo, S., Park, J. C., Lee J. Y., & Kweon, I. S. (2018) CBAM: Convolutional Block Attention Module[C]. *Proceeding of the European conference on computer vision (ECCV)*, 3-19.

[21] Yang, L., Zhang, R. Y., Li, L., & Xie, X. (2021) SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks [C]. *International Conference on Machine Learning, Proceedings of Machine Learning Research (PMLR), 11863-11874.*

[22] Zhai, Z. L., Niu N. W. J, Feng B. M., Xu, S. Y., Qu C. Y., & Zong C. (2024) An improved YOLOv8 model enhanced with detail and global features for underwater object detection [J]. *Physica Scripta: An International Journal for Experimental and Theoretical Physics,* (9), 99.

[23] Zhang, Q. L., & Yang, Y. B. (2021) SA-Net: Shuffle Attention for Deep Convolutional Neural Networks [C]. *ICASSP 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2235-2239*.

[24] Zhang, X., Li, L. Z., & Han, L. (2024) A cross-spatial network based on efficient multi-scale attention for landslide recognition [J]. *Landslides, 21*(12), 2913-2925.