International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 14, Issue 1 [January. 2025] PP: 06-13

Chest Ring Target Segmentation Method Based on an Improved YOLOv8-seg

Yanying Lv [1], Meng Li [1,*] , Zhenwen Sun [1], Zhihang Zhang [1]
1 School of Mechanical and Vehicle Engineering, Changchun University, Changchun 130022, China,
Corresponding Author: Meng Li

ABSTRACT: *As automated target reporting technology based on chest ring targets became increasingly applied in military training, shooting competitions, and civilian shooting clubs, an improved YOLOv8-seg-based chest ring target image segmentation method was proposed to enhance the adaptability of computer vision systems in complex environments. The Coordinate Attention mechanism is integrated into the Neck to enhance the model's ability to capture spatial structure in the effective regions of chest ring targets; The U-Net network is used in the Head to replace the Proto module, improving segmentation details and better preserving edge details. Experimental results show that the model achieves an average pixel accuracy of 99.67% and an average intersection over union of 86.65%, enabling precise segmentation of chest ring target images in complex scenarios.*

Date of Submission: 27-12-2024 Date of acceptance: 06-01-2025

Ⅰ. INTRODUCTION

With the rapid development of computer technology in the field of image processing, and the continuous growth of shooting sports today[1], automatic target reporting systems based on image processing technology offer advantages such as fast statistical speed, high accuracy in identifying bullet impact points, and fairness[2].Chest ring target surface segmentation serves as the foundation for subsequent bullet hole recognition and ring value determination in automatic target reporting systems. The key technical challenge lies in achieving precise and rapid segmentation of the chest ring target surface under complex conditions. Therefore, applying computer vision and artificial intelligence technologies to the segmentation of the chest ring target surface has significant importance and practical application value.

LUO JIE[3] proposed a method for segmenting the chest ring target based on color features. This approach involves an initial extraction of the target paper region in the RGB color space, followed by a refined extraction based on the hue component in the HIS color space. This method has strict requirements for background color, limiting its general applicability. Yin Qian[4] proposed a method that extracts effective target surface information based on image grayscale and edge information, identifying the largest connected region as the target surface. This method is highly susceptible to variations in ambient lighting and cannot ensure that the largest connected region corresponds to the effective target surface. The accuracy of the above traditional recognition and segmentation methods is relatively low, and the algorithms have limited adaptability to varying environmental conditions.

In recent years, the application of deep learning in image processing has become increasingly widespread, with many researchers adopting deep learning-based methods for segmenting chest ring targets. Common models in the field of semantic segmentation include classic networks such as FCN[5], U-Net[6], FPN[7], and DeepLab[8-10]. Huang Yingqing[11] proposed a chest ring target segmentation method based on RefineNet, achieving pixel-level semantic segmentation of the effective regions of the chest ring target. Liu Hualin [12] proposed a method for chest ring target segmentation based on the PSPNet network, using mean pixel accuracy and mean intersection over union as evaluation metrics for the model. However, this method does not segment the effective regions of the chest ring target, making it less suitable for subsequent bullet hole recognition in automatic target reporting systems.

To address the above challenges, this paper proposes a model based on an improved YOLOv8-seg for segmenting the effective regions of chest ring targets. The study is conducted using a self-constructed dataset and includes a comparative analysis with various semantic segmentation algorithms. The results demonstrate that the proposed model performs well in terms of mean pixel accuracy and mean intersection over union, achieving precise segmentation of the effective regions of chest ring targets even in complex environments, thereby significantly improving the segmentation accuracy of the chest ring target surface.

Ⅱ. EXPERIMENTAL DATA

2.1Source of Experimental Data

Since there is no publicly available dataset for chest ring target surface segmentation, this study utilizes a self-constructed dataset. The sources of the chest ring target images are as follows:

1) Images of chest ring targets collected under various indoor and outdoor backgrounds.
2) Real shooting range images with background replacement using Photoshop software.
3) Various outdoor chest ring target images gathered from the internet.

After screening, collecting, and field acquisition, a dataset of 4,598 images was constructed for this study. The collected samples are shown in Figure 1.

Fig.1 Display of datasets

2.2 Experimental Dataset Processing

In this study, Labelme software is used to annotate the effective regions of the chest ring targets, classifying them as "target" and generating corresponding .json label files. The annotation process utilizes polygonal labeling, with some annotated samples shown in Figure 2. To ensure that the model can effectively learn the data features without overfitting or underfitting, the dataset is randomly divided into training, validation, and test sets in a ratio of 8:1:1 for the semantic segmentation experiments.

Fig2. Labeling results

Ⅲ. RESEARCH METHOD

3.1 YOLOv8-seg Network Structure

The YOLOv8 segmentation model is an instance segmentation model based on YOLACT[13-15], and it classifies models based on the network's depth and width, including YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x[16]. These variations are suitable for different application scenarios depending on the model size and

computational complexity.

YOLOv8-seg is the semantic segmentation model of YOLOv8, which extends the object detection capabilities by adding pixel-level object classification. It identifies objects in the input image and generates masks based on the dataset's labels, enabling real-time pixel-level segmentation. YOLOv8n is the model with the fewest parameters and computational requirements within the YOLOv8 series. Given the high real-time performance demands of automatic target reporting systems and the need for deployment on embedded devices, this study selects YOLOv8n-seg as the foundational network model.

The model primarily consists of three components: the Backbone, the Neck, and the Head, as shown in Figure 3. The Backbone layer is composed of SPPF (Spatial Pyramid Pooling - Fast) and multiple CBS and C2f modules, and it is primarily responsible for extracting features from the image. The SPPF is a spatial pyramid pooling layer that aggregates image features at different scales, capturing multi-scale contextual information.

The Neck layer primarily performs feature fusion on the feature maps output by the Backbone. It integrates features extracted at three different scales using upsampling and then applies downsampling through CBS modules. The resulting feature maps are then passed to the Head layer.

The Head section generates detection boxes and segmentation masks through operations such as convolution and upsampling, producing the semantic segmentation results. The Proto module, which is unique to the YOLOv8-seg model, processes the features extracted by the Backbone using upsampling operations to generate prototype masks. This serves as the foundation for the subsequent fine segmentation branch.

3.2 Improved YOLOv8-seg model

The Backbone and Neck parts of YOLOv8 have been optimized compared to previous YOLO series models. However, the chest ring target scenarios are complex and variable, requiring the model to maintain high segmentation accuracy even under challenging conditions such as complex backgrounds and significant lighting variations.The YOLOv8-seg model still has limitations in capturing image details and feature extraction capabilities. To enhance the model's feature extraction ability during segmentation tasks, this study incorporates the Coordinate Attention (CA) mechanism [17]after the feature layer output of the Neck. The CA mechanism applies pooling and encoding to the obtained feature maps along both horizontal and vertical directions, effectively embedding spatial information into the feature maps. This mechanism captures spatial details while compressing channel dimensions and extracting specific coordinate information, allowing the model to focus on critical features. It enhances attention to target coordinate information while reducing focus on background and other irrelevant regions. Compared to other attention mechanisms, CA is better at capturing positional and channel information, improving the model's generalization ability and detection accuracy without significantly increasing complexity.

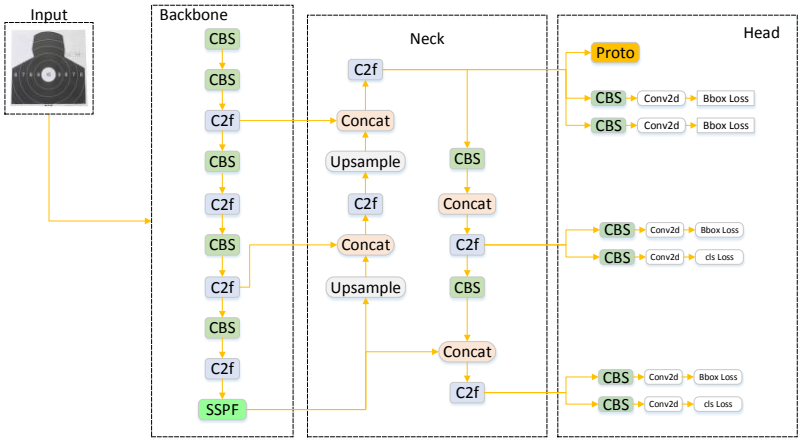

Fig3. YOLOv8-seg model network structure

In the YOLOv8-seg model, the Head layer is responsible for outputting bounding boxes, object classes, and segmentation masks. This study focuses on segmenting the effective regions of the chest ring target, which is crucial for accurate bullet hole recognition in automatic target reporting systems. Therefore, precise localization is highly important. Since the YOLOv8-seg model adds semantic segmentation functionality based on object detection, its segmentation head is relatively simple. To enhance its segmentation capabilities, this study integrates a U-Net network into the Head layer, replacing the Proto module. U-Net is a deep learning network specifically designed for image segmentation tasks[18], featuring a symmetric U-shaped structure composed of an encoder and a decoder. The encoder part of the U-Net employs downsampling to extract deep image information. After each layer's output, it captures deep feature information while simultaneously reducing the image

dimensions. The decoder part uses transposed convolution for upsampling, restoring the image resolution. During the upsampling process, it concatenates feature maps, generating high-precision segmentation masks and achieving the fusion of detailed image features. U-Net introduces skip connection[19], which allow for training an efficient segmentation model even on relatively small datasets. This makes it well-suited for chest ring target segmentation, where no public dataset is available. The improved network structure is shown in Figure 4.

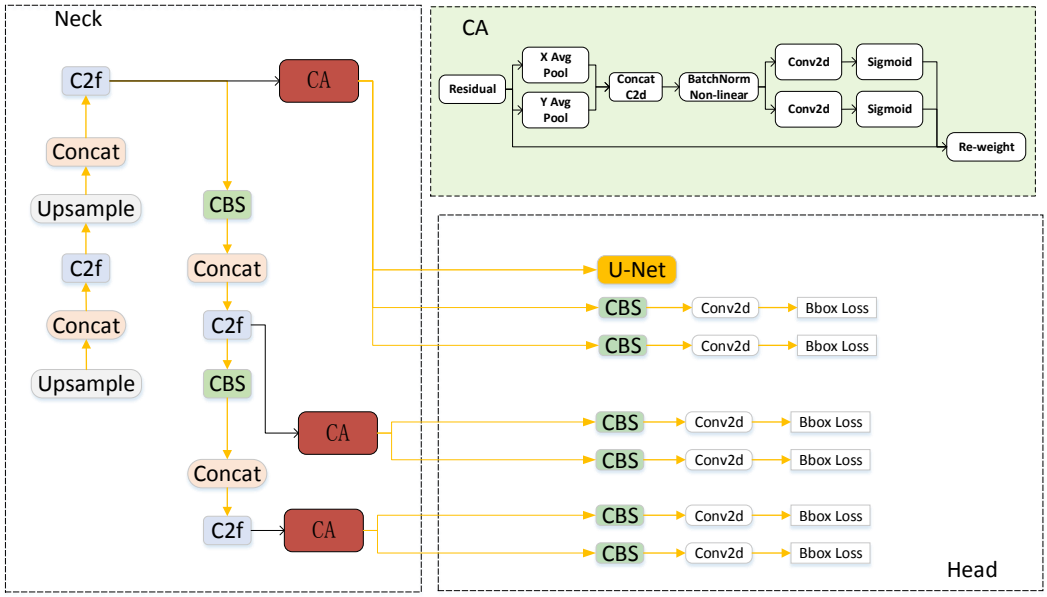

Fig4. Improved network structure of YOLOv8-seg model

Ⅳ. Experimental Results and Analyses

4.1 Experimental Evaluation Indicators

In semantic segmentation tasks, Mean Intersection over Union (MIoU) and Mean Pixel Accuracy (MPA) [20]are important metrics for assessing model performance. This study uses MPA and MIoU as evaluation metrics for the proposed model.

Pixel Accuracy (PA) represents the ratio of the number of correctly classified pixels to the total number of pixels in an image, reflecting the model's prediction accuracy for each class's pixel precision. MPA (Mean Pixel Accuracy) is the average of the accuracy for each class, calculated as the sum of the correctly classified proportions for all classes. The calculation formula is shown in Equation 1.

$$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \qquad (1)$$

Where k is the number of kinds of target objects in the image segmentation task except the background, i denotes the real class, j denotes the predicted class, is the number of pixels that are correctly classified, is the number of pixels that predict the real class i as the other class j, and is the number of pixels that predict the other class as the i class.

In this formula, k represents the number of target object classes in the image segmentation task, excluding the background. The variable i denotes the true class, and j represents the predicted class. p_{ii} refers to the number of pixels correctly classified as class i, p_{ij} denotes the number of pixels where the true class i is predicted as class j , and p_{ji} represents the number of pixels where other classes are incorrectly predicted as class i .

In semantic segmentation, Intersection over Union (IoU) is used to represent the similarity between the predicted region of an object and its true label. In this study, IoU indicates the degree of alignment between the model's extracted effective region of the chest ring target and the actual effective region. MIoU is calculated as the average IoU across all classes, with values ranging from 0 to 1. A higher MIoU value indicates more accurate segmentation by the model. In this study, image pixels are classified into two categories: the effective region of the chest ring target surface and the background. The calculation formula is shown in Equation 2.

$$MIoU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \quad (2)$$

4.2 Environment Configuration

The experimental environment for this study is configured as follows: The hardware setup includes a Windows 11 system, an Intel Core i7-14700HX CPU, and an NVIDIA GeForce RTX 4070 GPU. The software environment consists of Python 3.10.1, the deep learning framework PyTorch, and the development software PyCharm. A self-constructed dataset is used for this study.

The training parameters of the proposed model are shown in Table 1.

Table 1.Training parameters

Parameter	Value
Training Set Size	3678 images
Validation Set Size	459 images
Input Image Size	640×640 pixels
Number of Classes	2 (Effective region, Background)
Epoch	100
Batchsize	4
Initial Learning Rate	0.0025
Optimizer	Adam

4.3 Model Training Results and Analysis

The improved YOLOv8-seg-based model was used to perform semantic segmentation on chest ring target images, with the effective region of the chest ring target treated as the target instance. After training, the model is able to accurately identify and segment the target instance while ignoring the background. Additionally, the model outputs the class label and confidence score for each detected object.

The model's loss curve is shown in Figure 5. At the beginning of training, the loss is relatively high. As the model iterates, the loss values for both the training set and validation set decrease sharply within the first 10 epochs, indicating that the model is learning from the data and significantly reducing the error. After 50 epochs, the various loss values stabilize, with no significant signs of overfitting or underfitting, indicating that the model's learning has gradually saturated. The image shows that the smoothed curve aligns closely with the original loss curve, suggesting that the model is relatively stable during training, with no major fluctuations observed. The losses of the validation set and training set are closely aligned, indicating that the model has good generalization performance on the validation set. The results demonstrate that the model performs well in the semantic segmentation tasks presented in this study.

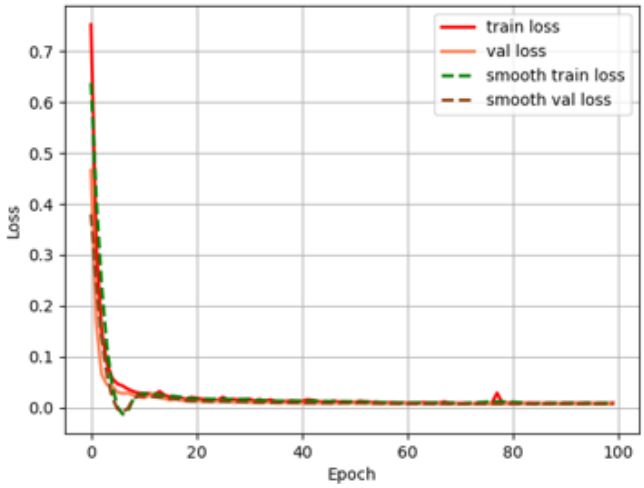

Fig5. Model training loss diagram

Model Precision refers to the proportion of true positive instances among those predicted as positive by the model[21]. Recall indicates the proportion of instances correctly identified as positive out of all actual positive instances [22]. A high precision indicates that the model has a high accuracy in its positive predictions. However, if the recall is low at the same time, it means that the model misses actual positive instances. In such cases, the F1

score is introduced to evaluate the model. The F1 score is the harmonic mean of precision and recall, providing a comprehensive assessment of the model's performance. As shown in Figure 6, the X-axis represents the confidence score, where a higher confidence indicates a greater likelihood that the detected object belongs to the correct class. The Y-axis represents the F1 score, with values closer to 1 indicating better model performance. From Figure 6, it can be observed that at a confidence score of 0.884, the F1 score reaches 1.00, indicating that the model achieves optimal precision and recall at this threshold. At this point, the model is able to accurately detect the target without missing any actual instances of the target.

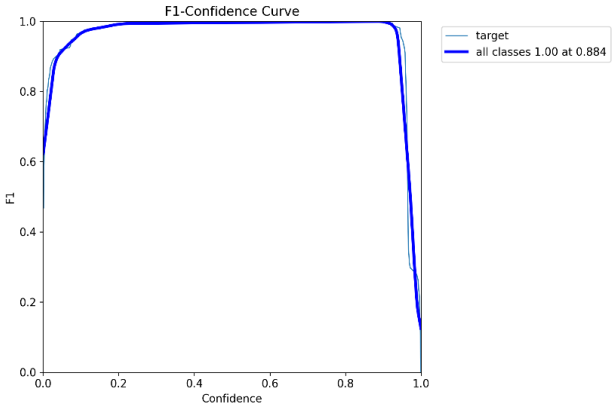

Fig6. The F1-condience curve

4.4 Performance Comparison With Other Semantic Segmentation Models

To evaluate the performance of the proposed model in the chest ring target segmentation task, a comparative analysis was conducted through experiments against other semantic segmentation models, including PSPNet[23]and DeepLabV3+[24]. The comparative experiments were conducted using the same experimental environment and dataset as the proposed model. The visualization of the results is shown in Figure 7. In Figure 7, image (a) is the sample image, and image (b) represents the test result of the PSPNet model. The results show that the segmented image has uneven edges, and the head details of the chest ring target are incomplete. There is a significant discrepancy between the segmented area and the actual effective region, indicating that the model's representation capability is relatively weak. Image (c) shows the results of the DeepLabV3+ model. It can be seen that this model is able to accurately extract the effective area of the chest ring target surface. However, there are issues with rough detail extraction along the edges of the segmented image, and some parts of the head of the chest ring target are missed during segmentation. Image (d) shows the results of the YOLOv8-seg model. The model effectively extracts the chest ring target's valid area, with a target detection confidence of 0.96. Compared to the DeepLabV3+ and PSPNet models, the segmentation performance of the YOLOv8-seg model shows noticeable improvement. Image (e) shows the results of the proposed model, with a target detection confidence of 0.97, which is higher than that of the YOLOv8-seg model, indicating improved segmentation performance.It can be seen that the proposed algorithm provides accurate segmentation results at a high confidence level, achieving good segmentation performance even with a limited dataset.

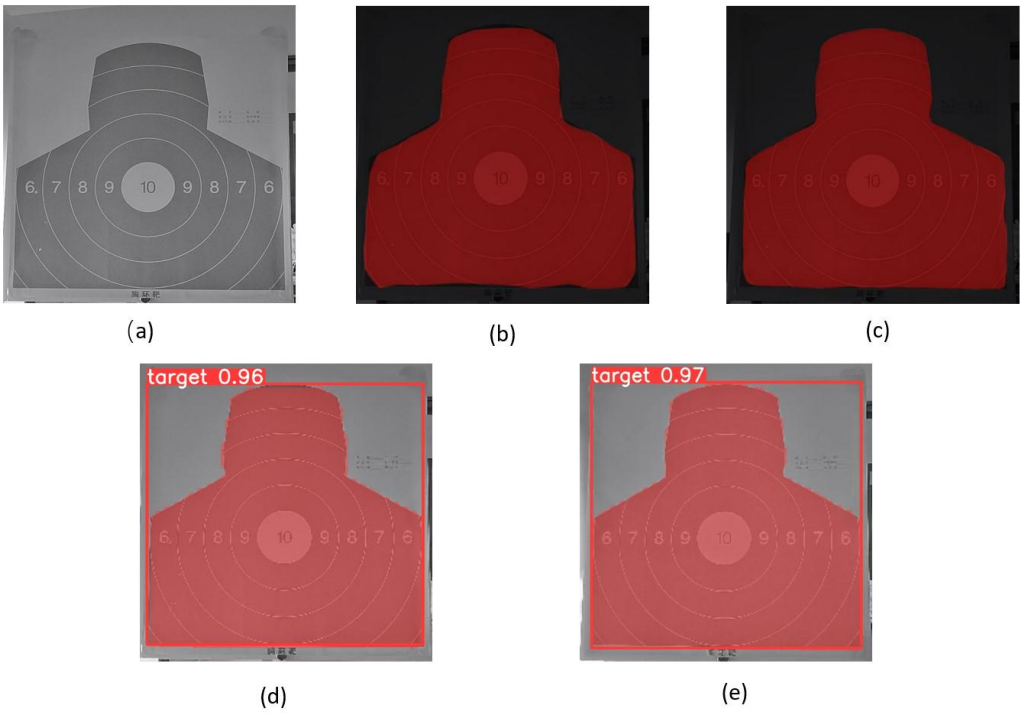

Fig7. Comparison of thoracic ring target segmentation results from different semantic segmentation models(a)sample(b)YOLOv8n-seg(c)YOLOv8n-seg+CA+U-Net(d)DeepLabv3+(e)PSPNet

The quantitative comparison of the segmentation results from the four methods is shown in Table 2. As observed from Table 2, the experimental results of the proposed model demonstrate a significant advantage over the other three models. In the segmentation task of this study, the proposed model achieved MPA of 99.67% and MIoU of 86.65%, demonstrating the best extraction performance. This aligns with the visual analysis of the segmentation results shown in Figure 7. The MPA of the PSPNet model is 97.86%, which is significantly lower than that of the YOLOv8-seg and DeepLabV3+ models, with the proposed model achieving the highest MPA of 99.67%. The MIoU of the proposed model reaches 86.65%, which is higher than that of the other three models. This indicates that the proposed model exhibits a high degree of overlap between the segmented area and the ground truth, resulting in more accurate segmentation of the edge information of the valid region. Additionally, the model demonstrates strong generalization capability, allowing it to accurately segment the effective area of the chest ring target.

Table 2.Comparison of segmentation performance results from different models

Model	MPA%	MIoU%
YOLOv8-seg	98.65	74.26
YOLOv8-seg+CA+U-Net	99.67	86.65
DeepLabv3+	98.51	66.04
PSPNet	97.86	65.33

The proposed CA attention mechanism and U-Net network improvements based on YOLOv8-seg have enhanced the model's segmentation performance, enabling the model to better focus on important features in the image and adapt to the segmentation tasks of chest ring targets in complex environments.

V. CONCLUSIONS

1.The semantic segmentation model used in this study, based on the improved YOLOv8-seg, incorporates a CA attention mechanism and replaces the Proto module with a U-Net network. This model achieves a precision of 99.67% and MIoU of 86.65%. Compared to the original model, the proposed model captures both global and local details more effectively. Compared to other semantic segmentation models, it demonstrates higher segmentation accuracy and stronger edge information extraction capability.

2.This study adopts an improved Yolov8n-seg model, which shows good performance on the chest ring target dataset. This lightweight model maintains high accuracy while achieving optimal segmentation efficiency,

making it well-suited for chest ring target segmentation in specific scenarios. It enables the image processing-based automatic scoring system to be more widely applied in environments with limited computational resources.

REFRENCES

[1]. YU Zi, Research on a Bullet Hole Localization System Based on Image Processing Technology[D].Harbin University of Science and Technology, 2017.

[2]. LI Changjun,Design of an Automatic Target Reporting System Based on DM642[D]. Southwest Jiaotong University,2010.

[3]. LUO Jie,ZHANG Zhiming. Chest Silhouette Segmentation Algorithm Based on Color Feature[J].Science Technology and Engineering, 2016,16(23):239-243

[4]. YIN Qian,LIAO Qiang.An Effective Target Extraction Algorithm Based on Gray Characteristics of the Image of a Chest Silhouette[J].Science Technology and Engineering, 2017,17(22):260-264

[5]. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[6]. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.

[7]. Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[8]. Chen L C. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. arXiv preprint arXiv:1412.7062, 2014. DEEPLABv1

[9]. Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848. DEEPLABv2

[10]. Chen L C. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017. DEEPLABv3

[11]. HUANG Yingqing,CHEN Xiaoming,XIE Zhihong.Chest Ring Target Image Segmentation Method Basedon RefineNet.[J].Journal of Ordnance Equipment Engineering,2021,42(09):231-236.

[12]. LIU Hualin,ZHANG Yi,WANG Haipeng. Chest RingTarget Image Segmentation Method in Complex Environment[J]. Journal of Ordnance Equipment Engineering, 2020,41(05):107-112.

[13]. Bolya D, Zhou C, Xiao F, et al. Yolact: Real-time instance segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9157-9166.

[14]. Qadri S A A, Huang N F, Wani T M, et al. Plant Disease Detection and Segmentation using End-to-End YOLOv8: A Comprehensive Approach[C]//2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE). IEEE, 2023: 155-160.

[15]. Yue X, Qi K, Na X, et al. Improved YOLOv8-Seg network for instance segmentation of healthy and diseased tomato plants in the growth stage[J]. Agriculture, 2023, 13(8): 1643.

[16]. LU Zice,LIU Xiaofang,WANG Dewei.Semantic Segmentation Method for PCB Solder Joint Based on Improved YOLOv8[J].Radio Engineering, 2024,54(07):1614-1621.

[17]. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[18]. Minaee S, Boykov Y, Porikli F, et al. Image segmentation using deep learning: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(7): 3523-3542.

[19]. GAO Kangzhe,WANG Fengyan,LIU Ziwei,WANG Mingchang. Semantic Segmentation of Remote Sensing Images Based on Improved U-Net Network[J/OL].Journal of Jilin University(Earth Science Edition):1-11[2024-09-12].

[20]. YUAN Wei,ZHOU Tian,XI Zongshun,ZHOU Xingyi.MUNet:A Multi-branch Adaptive Deep Learing Network for Remote Sensing Image Semantic Segmentation[J].Journal of Geomatics Science and Technology,2020,37(06):581-588.

[21]. CHEN Menghua,ZHANG Tongyun,ZHOU Ziyang,LU Min.High-precision Batch Automatic Extraction of Building Based on DeepLabV3+[J/OL].Journal ofGeomatics Science and Technology: 1-6[2024-09-13].

[22]. WU Yongjun,CUI Can,HE Yongfu.Rail Surface Defect Detection Based on Semantic Augmentation and YOLOv8[J/OL]. Journal of Railway Science and Engineering: 1-12.

[23]. Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[24]. Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.