2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)

ALC: Automated Lung Cancer Detection Framework in Thoracic CT Scans Based on Deep learning

Mingzi Wu[1,a]
[1]School of Computer Science and Software Engineering
Chengdu Neusoft University, Chengdu, China
[a]e-mail: wcassis0@gmail.com

Yuheng Li[2,*]
[2]School of Cyberspace Security (School of Cryptology)
Hainan University, Haikou, China
*e-mail: liyuheng@hainanu.edu.cn

Kaiyi Feng[3,b]
[3]School of Computer Science and Software Engineering
Chengdu Neusoft University, Chengdu, China
[b]e-mail: 3090086203@qq.com

2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA) | 979-8-3315-0976-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/AIITA65135.2025.11047899

***Abstract*—With the continuous increase in the incidence and mortality rates of lung cancer, the demand for automated analysis of chest CT scans, as an important means of early detection, is becoming increasingly urgent. Traditional object detection models struggle to balance accuracy and efficiency, especially when dealing with CT images that have low contrast and complex backgrounds, where their performance is limited. To address this issue, this paper proposes a novel high-precision lightweight lung cancer detection framework named ALC-Net, aiming to improve the accuracy and computational efficiency of lung cancer detection in CT scans. The model incorporates the ELA module and combines multi-scale feature fusion technology to effectively capture the subtle features of lung lesions. At the same time, the Dynamic block is adopted to adaptively adjust the feature weights, significantly reducing the computational complexity and ensuring real-time performance and deployment feasibility. The experiments are verified based on the self-built Lung-Dataset, which contains 4536 chest CT images. ALC-Net achieves 98.2% and 79.3% in mAP50 and mAP50-95 respectively, showing a significant improvement compared to YOLOv8n. Moreover, the number of parameters (293.1M) and computational cost (10.8GFLOPs) remain low, meeting the requirements of clinical real-time detection and providing an efficient solution for the early automated screening of lung cancer.**

Keywords-YOLOv8; Lung Cancer Detection; Deep Learning; Lightweight Model; Multi-scale Feature Fusion

I. Introduction

Lung cancer is one of the malignant tumors with the highest incidence and mortality rates globally. Early detection is of vital importance for improving the survival rate of patients. Chest CT scans have become the gold standard for lung cancer screening due to their advantages of high resolution and non-overlapping imaging. However, the massive data and complex structures of CT images pose great challenges to manual interpretation, and there is an urgent need for automated detection technologies to assist doctors in improving the diagnostic efficiency and accuracy.

In recent years, deep learning has made remarkable progress in the field of medical image detection. However, existing models still face multiple challenges in the application of chest CT. Traditional object detection models such as Faster R-CNN have a large computational load and are difficult to meet the real-time requirements in clinical settings. While lightweight models improve the speed, they often sacrifice accuracy, especially in the detection of low-contrast and tiny lesions. In addition, the complex anatomical structures of the lungs, respiratory artifacts, and the diversity of lesion morphologies further increase the difficulty of detection.

Aiming at the above problems, this paper proposes a high-precision lightweight lung cancer detection framework, ALC-Net. Taking YOLOv8n as the baseline, this framework enhances the extraction of lesion edge features by introducing the edge learning attention (ELA) module, combines the high-resolution feature pyramid block（HGBlock）, a multi-scale feature pyramid block, to fuse context information at different scales, and adopts a dynamic block to adaptively optimize the feature weights, significantly reducing the computational cost while ensuring the detection accuracy. The experimental results show that ALC-Net achieves an mAP50 of 98.2% on the self-built chest CT dataset, an increase of 4.2% compared with the baseline model. Moreover, the number of parameters and the computational amount are only 293.1M and 10.8G FLOPs, meeting the requirements for clinical deployment.

The contributions of this paper are as follows:

(1) We proposed an ALC-Net network framework in thoracic CT scans. This architecture combines the ELA module, the HGBlock, and the dynamic block, which significantly improves the detection accuracy and computational efficiency of lung cancer lesions in low-contrast CT images.

(2) We constructed a lung cancer detection dataset, Lung-Dataset, containing 4536 thoracic CT images. This dataset covers three types of lesions: nodules, cancers, and adenocarcinomas, and is annotated by medical experts, ensuring data diversity and clinical representativeness.

(3) The experimental results show that ALC-Net achieves 98.2% mAP50 and 79.3% mAP50 - 95 on the self - built dataset, with an increase of 4.2% and 3.2% respectively compared to YOLOv8n. Meanwhile, it maintains a low number of parameters(293.1M) and computational cost(10.8GFLOPs), meeting the requirements of clinical real - time detection.

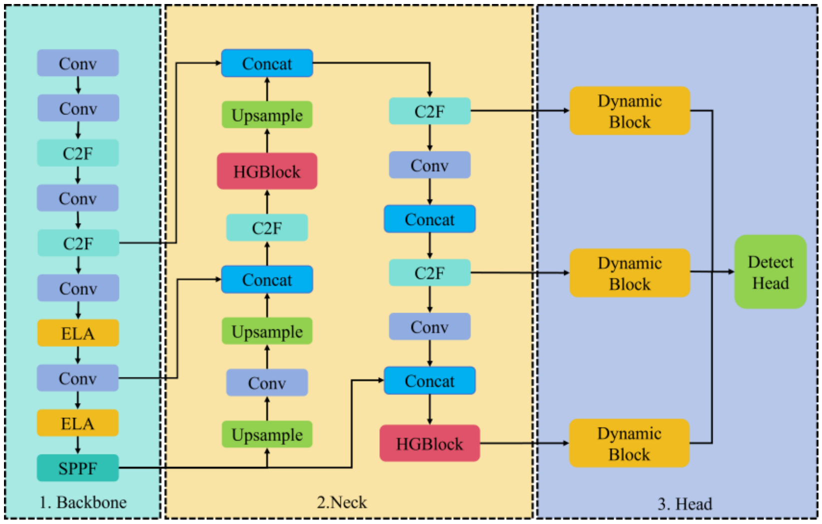

Figure 1. Network structure of improved ALC-Net.

II. RELATED WORK

Early detection of lung cancer mainly relies on traditional image processing and machine learning techniques. These methods use manually designed features, such as grayscale features and texture features[1][2], and combine them with classifiers like support vector machines and random forests for lesion recognition[3][4]. However, manually designed features are difficult to comprehensively represent the complex characteristics of lung cancer lesions. When faced with diverse lesion morphologies and complex lung backgrounds, the detection accuracy is limited.

With the rise of convolutional neural networks, their powerful feature learning ability has brought new breakthroughs to lung cancer detection. Classic network architectures, such as VGGNet[5], ResNet[6], etc., have been widely applied in medical image analysis. Meanwhile, the success of U-Net[7] and its variants in medical image segmentation tasks has also inspired researchers to apply their ideas to lung cancer detection, assisting lesion detection through the accurate segmentation of the lung area.

In recent years, object detection algorithms have been widely applied in lung cancer detection. The YOLO series[8] models, with their high detection speed and good performance, have demonstrated excellent performance in industrial detection, been widely applied[9], and gradually introduced into medical image detection. Its continuously iterated versions, such as YOLOv5 and YOLOv8, have continuously improved in detection accuracy and speed through the improvement of network architectures and the optimization of training strategies. However, due to the low contrast and noise interference of chest CT images, as well as the complexity of the pulmonary anatomical structure, existing object detection models still face problems such as insufficient accuracy and missed detection of tiny lesions in lung cancer detection. In addition, in the context of complex medical images, how to more accurately guide attention remains a research challenge. There is still a need for a lung cancer detection method that can ensure high accuracy while having a low computational complexity to meet the actual clinical needs.

III. RESEARCH METHOD

A. *ALC-Net*

In this paper, we proposed a novel network named ALC-Net for lung cancer detection to address the specific challenges in lung CT scan image analysis. Based on the YOLOv8n baseline, the model introduces the HGBlock[10] to tackle issues such as limited resolution and insufficient feature extraction in lung CT images. By cascading convolutional kernels of varying sizes, this module effectively captures multi-scale features, comprehensively integrating local and global lesion information to enhance the model's feature representation capabilities in complex pulmonary environments and ensure accurate capture of features of various lung cancer lesions.

In addition, to overcome the difficulties of low recognition accuracy and weak robustness in detecting lung cancer lesions against complex pulmonary backgrounds, this paper employed an efficient ELA[11]. This method focused on pulmonary lesion regions, strengthens the model's recognition ability for lung cancer lesions, enables precise differentiation between lesions and normal tissues, and reduces false positive rates. To meet the requirements of clinical real-time detection and convenient deployment on terminal devices, the dynamic block[12] is incorporated into the existing framework. By leveraging multiple dynamic experts, this block significantly expands model parameters with minimal additional computational overhead, enhancing the model's learning capacity. This allows the model to adaptively adjust convolution parameters according to input lung CT images, further improving detection accuracy and efficiency. The overall model architecture is illustrated in Figure 1.

B. *ELA model*

The Edge Learning Attention(ELA) module is an efficient local attention mechanism designed to enhance model performance in visual tasks by focusing on edge information. Its structure is illustrated in Figure 2. The core idea of this module is that the model should prioritize key edge regions in images or feature maps, rather than background or irrelevant information. This focus enables the model to better capture fine details, particularly edges and structural features, thereby improving task accuracy and efficiency. The workflow of the ELA module primarily consists of three steps:

(1) Edge detection: If the input image is I, the edge graph E is obtained through the Sobel filter. The formula is as follows:

$$E = \text{Sobel}(I) \tag{1}$$

where $\text{Sobel}(I)$ denotes the Sobel operator is used for edge detection and the edge graph E is obtained, which dimension is $M \times N$. Here, M represents the number of features, corresponding to the types of key features extracted from chest CT images. N represents the number of samples, that is, the number of chest CT image slices involved in feature extraction.

(2) Attention weight calculation: based on limbic information E, calculate the attention weight A for each position. The formula is as follows:

$$A = \frac{1}{1+\exp(-\gamma E)} \tag{2}$$

where γ denotes adjust parameters, which can control the degree of influence of the edges. A larger γ will make the weight of the edge more prominent, and a smaller γ will reduce the influence of the edge.

(3) Weighted feature plots: Finally, the attention weights A are multiplied by the input feature map F to obtain the weighted feature map F_{ELA}, and the formula is as follows:

$$F_{ELA} = F \cdot A \tag{3}$$

where F denotes enter feature map, A is the corresponding attention weight, which indicates multiplication by elements.

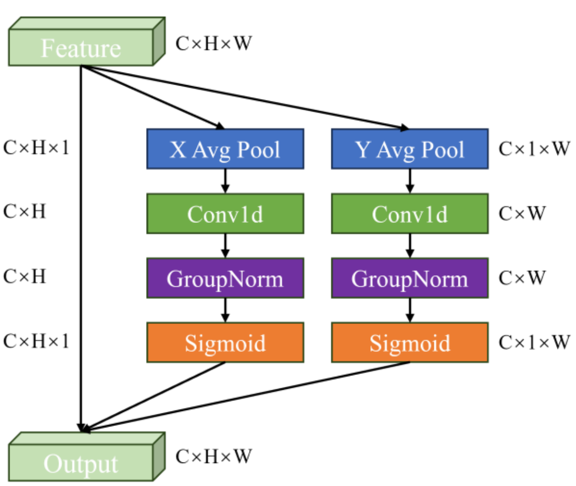

Figure 2. ELA network structure.

C. HGBlock

The High-Resolution Feature Pyramid Block (HGBlock) module extracts multi-scale features by cascading convolutional kernels of different sizes. This allows the module to extract features through multiple convolutional layers and optimize them via 1x1 convolutions and an Effective Channel Attention (ESE) module. By focusing on edge information in the image, the module enhances its feature extraction capabilities. Its structure is shown in Figure 3. Compared to traditional convolutional networks, the ESEBlock is specifically designed with an edge-sensitive mechanism that adaptively identifies and weights edge regions, improving the model's performance in handling detail-rich or structured information.

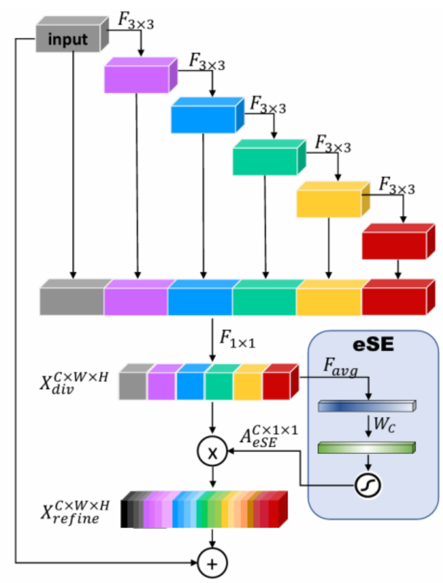

Figure 3. HGBlock network structure.

The principle of ESE can be briefly described as follows: First, assume the input feature map is F, and the edge map is E, where the edge map is obtained through an edge detection algorithm. Next, an edge enhancement coefficient α is calculated to adjust the influence of edge features. Finally, the ESEBlock obtains the weighted feature map F_{ESE} by multiplying the edge enhancement coefficient with the original feature map F, as expressed by the following formula:

$$F_{ESE} = F \cdot \alpha \tag{4}$$

where · denotes element-wise multiplication, α is the edge weighting coefficient, and $F_E SE$ represents the enhanced feature map.

D. Dynamic Block

The Dynamic block is a neural network module based on dynamic feature adjustment, designed to enhance network performance in complex tasks by adaptively adjusting the weights and activation values of features within the network. The core idea of this module is to dynamically learn the importance of features and flexibly adjust the network's computation process based on the varying characteristics of input data and task requirements. Its workflow can be described through the following steps:

(1) Feature Analysis and Weight Calculation: Assume the input feature map is F. The Dynamic block first computes a dynamic weight matrix W, which is used to weight each feature. This weight matrix W is dynamically generated from the contextual information in the feature map and can be computed using an attention mechanism or other adaptive methods. Specifically, the calculation formula for the weight matrix is as follows:

$$\mathrm{W} = \mathrm{f}\left(\mathrm{F}\right) \tag{5}$$

where f represents a learning function (such as convolution, fully connected layers, or attention mechanisms), used to compute the weight matrix W based on the input feature map F.

(2) Dynamic Weighting: After obtaining the weight matrix W, the Dynamic block performs element-wise multiplication with the input feature map F to adjust the importance of each feature. This process can be expressed by the following formula:

$$\mathrm{F}_{\text{dynamic}} = \mathrm{F} \cdot \mathrm{W} \tag{6}$$

where $\mathrm{F}_{\text{dynamic}}$ is the dynamically weighted feature map, and $\cdot$ denotes element-wise multiplication. Through this operation, important features in the feature map are enhanced, while less important features are suppressed.

(3) Dynamic Activation Adjustment: Building on the dynamically weighted features, the Dynamic block further enhances the network's expressive power by adjusting the parameters of the activation function. Specifically, the dynamically adjusted activation values can be expressed as:

$$\mathrm{A}_{\text{dynamic}} = \sigma\left(\mathrm{F}_{\text{dynamic}} + \Delta\right) \tag{7}$$

where $\mathrm{A}_{\text{dynamic}}$ is the output after applying the activation function (e.g., ReLU or Sigmoid) with dynamic adjustments, and Δ is a bias term learned by the network to further fine-tune the range of activation values.

(4) Output: The dynamically adjusted feature map $\mathrm{F}_{\text{dynamic}}$ can be passed to subsequent network layers for task processing. The output of the Dynamic block helps improve the network's adaptability in complex scenarios, enabling it to respond flexibly based on the characteristics of the input data.

IV. Experiment and Results

A. Dataset

The Lung-Dataset of lung CT scan images in this study is sourced from two hospitals: Wuhan Third People's Hospital and Jingzhou Central Hospital. The data from these two hospitals ensures the diversity and representativeness of the dataset. We collaborated with the respiratory medicine departments, thoracic surgery departments of Wuhan Third People's Hospital and the lung cancer diagnosis and treatment center of Jingzhou Central Hospital. Through cooperation with their imaging departments, we obtained a large number of lung CT scan images from different patients, enhancing the dataset's applicability. The dataset contains 4536 images in total, covering three categories: nodule, cancer, and adenocarcinoma, which were annotated and labeled by domain experts.

B. Experimental configuration

All experiments in this study were conducted on two servers, each equipped with two NVIDIA GeForce RTX 3090 GPUs. Each GPU has 24 GB of memory, and the servers run on the Ubuntu 22.04 operating system. All models used in the experiments were implemented in a Python environment and built using the PyTorch framework. The proposed segmentation model was trained on the constructed Coal-Dataset, which was divided into training, validation, and test sets in an 8:1:1 ratio.

C. Results of comparative experiment

In order to verify the performance of different models in the task of lung cancer detection, this study selected widely used deep learning-based object detection algorithms for experiments. As can be seen from Table 1, compared with classic networks such as YOLOv5n-p6 and Faster-RCNN, ACL-Net achieved improvements of 11.5% and 8.9% in the mAP50, and increases of 5.0% and 3.9% in mAP50-95. The ACL-Net model outperforms classic network algorithms such as the YOLO series and Mask-RCNN in terms of both mAP and Recall, and the detection results are shown in Figure 4.

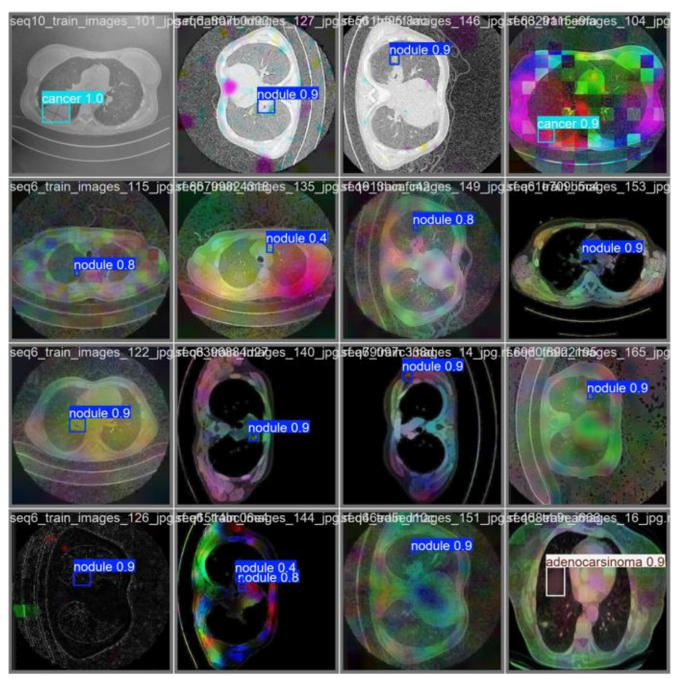

Figure 4. Detection result diagram.

TABLE I. COMPARATIVE EXPERIMENT WITH OTHER SOTA MODEL.

Method	mAP50 (%)	mAP50-95(%)	Precision (%)	Recall (%)	F1(%)	FLOPs	Param
YOLOv5n-p6	86.7%	74.3%	89.6%	88.1%	91.8%	11.2G	445.3M
Faster-RCNN	89.3%	75.4%	90.4%	91.0%	94.2%	15.3G	440.8M
YOLOv7-u7	92.4%	75.0%	93.7%	90.1%	93.4%	142.G	3788.7 M
YOLOv8n	94.0%	76.1%	93.9%	91.6%	92.7%	12.1G	326.4M
YOLOv8s	95.1%	78.6%	95.4%	93.8%	94.6%	42.7G	1179.2 M
YOLOv9c	97.4%	81.1%	95.8%	95.3%	95.5%	159.1G	2783.9 M
YOLOv10n	97.3%	77.4%	96.1%	94.2%	95.1%	11.8G	285.0M
Mask-RCNN	95.9%	78.1%	96.3%	80.8%	87.9%	344.1G	62.9M
Cascade Mask R-CNN	97.1%	80.2%	96.7%	79.8%	87.4%	1822.7 G	77.0M
YOLACT	95.6%	72.0%	96.0%	78.5%	86.4%	61.8G	34.7M
SOLOv2	96.7%	72.3%	96.8%	75.2%	84.6%	250.9G	46.2M
Mask2Former	98.1%	79.1%	96.5%	94.5%	95.5%	262.1G	44.0M
ALC-Net	98.2%	79.3%	96.4%	94.7%	95.5%	10.8G	293.1M

D. *Results of ablation experiment*

In order to evaluate the real-time detection effect of ALC-Net in the scenario of chest CT scans, this paper conducts ablation experiments. The experiments gradually add the ELA, HGBlock, and Dynamic block on the basis of YOLOv8n. As shown in Table 2, the results show that after adding the ELA, the model's mAP50 increases from 94.0% to 96.8%, and the mAP50-95 increases from 76.1% to 77.9%, which improves the feature aggregation ability. After further adding the HGBlock, the mAP50 increases by 1.1%, and the mAP50-95 increases by 0.8%. Finally, after adding the Dynamic block, ALC-Net reaches 98.2% in mAP50 and 79.3% in mAP50-95, showing the best performance. At the same time, the computational amount does not increase, making it convenient for terminal deployment.In summary, ELA, HGBlock, and Dynamic Block enhance ALC-Net's accuracy for real-time lung cancer detection.

TABLE II. ABLATION EXPERIMENT RESULTS.

Method	mAP50(%)	mAP50-95(%)	Precision (%)	Recall (%)	F1(%)	FLOPs	Param
YOLOv8n	94.0%	76.1%	93.9%	91.6%	92.7%	12.1G	326.4M
YOLOv8n+ELA	96.8%	77.9%	94.6%	93.5%	94.0%	12.1G	337.9M
YOLOv8n+ELA +HGBlock	97.9%	77.1%	95.2%	93.6%	94.4%	12.5G	340.4M
YOLOv8n+ELA +Hgblock+Dynamic	98.2%	79.3%	96.4%	94.7%	95.5%	10.8G	293.1M

E. *Discussion*

In this study, ALC-Net achieved a balance between high accuracy and high efficiency in the detection of lung cancer in low-contrast CT images by introducing the ELA module, HGBlock, and Dynamic block. However, although the model performed excellently on the self-built dataset, further optimization is still required for practical applications. For example, the model's ability to detect rare lung cancer subtypes may be limited by the diversity of the dataset and the number of samples. In addition, the robustness of ALC-Net to image noise and artifacts still needs to be improved, especially in complex clinical scenarios. Future research will explore more efficient data augmentation and noise suppression methods to further improve the generalization ability and robustness of the model.

V. CONCLUSION

The ALC-Net proposed in this paper brings truly remarkable new breakthroughs to the detection of lung cancer in chest CT scans. By introducing the ELA module, which cleverly captures key features, combining the multi - scale feature fusion technology of HGBlock, and adopting the Dynamic block, it effectively solves the problem that traditional models struggle to balance accuracy and efficiency in lung cancer detection. The experimental results show that ALC-Net performs excellently on the self-built Lung-Dataset carefully constructed with 3240 chest CT images. It achieves outstanding results in terms of mAP50 and mAP50 - 95, showing a significant improvement compared with the baseline model YOLOv8n. At the same time, it maintains a low number of parameters and computational cost, meeting the requirements of clinical real-time detection.

REFERENCES

[1] Rashid U, Jaffar A, Rashid M, et al. Nodule Detection Using Local Binary Pattern Features to Enhance Diagnostic Decisions[J]. Computers, Materials & Continua, 2024, 78(3).

[2] Mehrnia S S, Safahi Z, Mousavi A, et al. Landscape of 2D Deep Learning Segmentation Networks Applied to CT Scan from Lung Cancer Patients: A Systematic Review[J]. Journal of Imaging Informatics in Medicine, 2025: 1-30.

[3] Sabry A H, Bashi O I D, Ali N H N, et al. Lung disease recognition methods using audio-based analysis with machine learning[J]. Heliyon, 2024, 10(4).

[4] El-Askary N S, Salem M A M, Roushdy M I. Features processing for random forest optimization in lung nodule localization[J]. Expert Systems with Applications, 2022, 193: 116489.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[6] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[7] Neha F, Bhati D, Shukla D K, et al. U-Net in Medical Image Segmentation: A Review of Its Applications Across Modalities[J]. arXiv preprint arXiv:2412.02242, 2024.

[8] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[9] Li Y, Chen Q, Zhu J, et al. PM-YOLO: A Powdery Mildew Automatic Grading Detection Model for Rubber Tree[J]. Insects, 2024, 15(12): 937.

[10] Lv H, Xiang Q, Xiao J. Object detection based on improved RT-DETR for human-robot collaboration manufacturing system[C]//Fourth International Conference on Mechanical, Electronics, and Electrical and Automation Control (METMS 2024). SPIE, 2024, 13163: 2144-2148.

[11] Xu W, Wan Y. ELA: Efficient Local Attention for Deep Convolutional Neural Networks[J]. arXiv preprint arXiv:2403.01123, 2024.

[12] Han K, Wang Y, Guo J, et al. ParameterNet: Parameters Are All You Need for Large-scale Visual Pretraining of Mobile Networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 15751-15761