TRANSPORT PROBLEMS 2025 Volume 20 Issue 1
PROBLEMY TRANSPORTU DOI: 10.20858/tp.2025.20.1.15

Keywords: parcel sorting; lightweight network; YOLOv8; SCS module

Kailan GAO[1], Xuelin WANG[2*], Yulong XU[3], Aleksander SŁADKOWSKI[4]

INTELLIGENT LOGISTICS EXPRESS PARCEL REAL-TIME DETECTION SYSTEM BASED ON IMPROVED YOLOV8

Summary. In intelligent warehousing and logistics centers, efficient sorting and transportation of express parcels are critical tasks for daily operations. In order to handle the large number of parcels per day, existing inspection and sorting systems face challenges such as high computational costs, insufficient inspection accuracy and poor real-time response capability. These issues restrict the deployment of such systems on embedded devices and impact the overall efficiency of logistics operations. This is particularly urgent in intelligent sorting, where real-time intelligent recognition and efficient sorting transportation are crucial. In order to solve these problems, a lightweight YOLOv8-SCS-CE algorithm based on the YOLOv8n algorithm is proposed in this paper, which can quickly realize express and parcel detection. A lightweight module SCS is also used to improve the Shufflenetv2 network by utilizing the channel attention mechanism and nested structure to overcome the effect of the lightweight architecture on the detection accuracy. First, the SCS module is used to replace the backbone structure of the YOLOv8n network, which improves the Shufflenetv2 network using channel attention mechanisms and nested structures, overcoming the impact of lightweight architectures on detection accuracy. Next, to enhance the network's feature extraction capability, the ECA attention mechanism is incorporated into the forward network of C2f, forming the C2f_ECA lightweight feature extraction module, which effectively solves the issue of extracting small defect features in complex backgrounds. Experimental results show that the YOLOv8-SCS-CE model reduces the number of parameters by 59.2%, decreases the computational complexity by 54.3%, and the weight file size is only 2.8MB, a reduction of 55.5% compared to the original files. The mAP@50 of the model is only 0.8% lower than that of the original mode. The lightweight network constructed in this study recognizes fast and few parameters, which provides a key technology for the courier parcel detection industry.

1. INTRODUCTION

With the rapid development of e-commerce, the global express logistics industry has shown unprecedented growth. According to the research of Global Info Research (Global Info Research), the global courier and parcel market size is about $5203.9 billion in 2023 and according to the projections it will reach $655.37 billion by 2030[1]. In this context, how to improve the efficiency and accuracy of logistics processes has become an urgent problem for the industry [2].

[1] Qilu University of Technology (Shandong Academy of Sciences), Institute of Automation; Keyuan Road 19, Lixia, Jinan 250014, China; e-mail: 1837948379@qq.com; orcid.org/0009-0009-0072-757X
[2] Qilu University of Technology (Shandong Academy of Sciences), Institute of Automation; Keyuan Road 19, Lixia, Jinan 250014, China; e-mail: wangxuel@sdas.org; orcid.org/0000-0002-4934-5494
[3] Qilu University of Technology (Shandong Academy of Sciences), Institute of Automation; Keyuan Road 19, Lixia, Jinan 250014, China; e-mail: xylong_0406@163.com; orcid.org/0009-0007-3838-4467
[4] Silesian University of Technology; Krasiński 8, 40-019 Katowice, Poland; e-mail: Aleksander.Sladkowski@polsl.pl; orcid.org/0000-0002-1041-4309
* Corresponding author. E-mail: wangxuel@sdas.org

In order to meet this challenge, the concept of "physical internet" (PI) for smart logistics has been proposed. It improves logistics efficiency and reduces redundant resources by interconnecting physical logistics units in the global supply chain. The basic idea of PI is to enable goods to flow efficiently across the globe like packets of data by means of standardized logistics units and intelligent information management. The implementation of PI relies on a variety of new and emerging information and communication technologies (ICTs), such as the Internet of Things (IoT), big data analytics, and cloud computing, which provide real-time monitoring, dynamic optimization, and resource-sharing capabilities for logistics systems, technologies, which provide logistics systems with the ability to monitor, dynamically optimize, and share resources in real time [3]. The optimal use of resources and a high degree of process synergy are achieved through the establishment of an open, interconnected, and standardized logistics network [4], and new solutions are proposed to address the challenges faced by last-mile logistics, including the implementation of dynamic sorting, parcel tracking, and distribution path optimization [5], which provide innovative solutions to improve logistics efficiency and accuracy. In the recently published book "Solving Transportation Problems with Artificial Intelligence", new ideas for solving problems in the field of transportation using AI techniques are presented and relevant applications of AI in transportation, warehousing and logistics from all over the world are brought together [6].

In recent years, large logistics companies have been gradually adopting advanced automated sortation systems to replace the traditional manual sortation model [7]. Companies such as RightHand Robotics and Kindred Systems are enabling fulfillment centers to achieve more efficient and economical automation through a combination of mobile robots and automated forklift technology. The DoraSorter intelligent sorting robot enabled by FedEx Shanghai in 2022 is an example of an AI-powered intelligent sorting system. These advances demonstrate the potential for the application of computer vision, artificial intelligence (AI), deep learning, and robotics in automation tasks [8]. Therefore, parcel detection and sorting solutions based on deep learning technology have become an important method for intelligent logistics detection and transportation.

The main target detection methods primarily encompass traditional approaches and deep learning-based techniques, among others. The principle of traditional methods is to extract features from candidate regions of an image and detect them by a classifier. The representative methods mainly include DPM [9] and Hog+Svm, but this kind of method has a complex process of feature extraction and classification, and the selection of candidate regions of an image is slow and ineffective. Deep learning algorithms are categorized into two types: one-stage networks [10] and two-stage networks [11].

Since logistics detection usually needs to be conducted on edge or mobile devices, where equipment resources are limited, the classical algorithm models, which are typically large, struggle to balance speed and detection accuracy. As a result, meeting real-time detection requirements and deploying these models to hardware have become an urgent challenge. Target recognition networks improved by combining lightweight modules have a significant effect in saving computational resources, classical lightweight models include SqueezeNet, MobileNet, and GhostNet, [12] etc. The SqueezeNet is an early classical lightweight network that utilizes the Fire module for parameter compression. SqueezeNext introduces separable convolution based on SqueezeNet for optimization and improvement. MobileNet proposes depthwise separable convolution and uses pointwise convolution for inter-channel information fusion to achieve model compression; however, the network itself has fewer parameters, limited feature extraction capability, and lower recognition accuracy. GhostNet adopts a simple linear operation to produce more feature maps, which reduces the total number of parameters and computational complexity of the module, but this approach sacrifices some feature expression capability. ShuffleNet is a lightweight network that uses point-by-point group convolution and channel blending to achieve a reduction in the number of parameters and outperforms the latest lightweight networks in terms of efficiency and accuracy.

Although the aforementioned research has achieved good results, the computing power of mobile and embedded devices is still limited, and more complex deep convolutional models cannot be directly applied to embedded devices. This paper combines the advantages of current mainstream algorithms and proposes a lightweight YOLOv8n-SCS-CE model. It designs the SCS module, which performs channel shuffling operations to rearrange the channels, improving the flow of information across

different channels and enhancing feature representation. This plays an important role when the model extracts complex features. By incorporating grouped convolutions and the lightweight mechanism, the model's expressive power is increased while maintaining a low computational complexity. Finally, the C2f_ECA module is used to improve feature extraction capabilities and address small defects detection in cluttered backgrounds. It is highly beneficial for embedded devices and application scenarios requiring real-time processing.

2. YOLOV8 ALGORITHM: DESIGN, FEATURES, AND IMPROVEMENT

As an iterative version of the YOLO series, YOLOv8 redefines the realization path of real-time target detection tasks through a single-stage regression paradigm. Compared with the traditional multi-stage detection framework, the algorithm uses a deep convolutional network architecture to synchronize target localization and classification prediction in a single forward propagation.

Its core innovations can be summarized as three technical innovations: first, based on the YOLOv5 scale-adaptive concept, YOLOv8 provides a family of full-size models ranging from Nano(N) to Extra-Large(X) and provides a 640×640 and 1280×1280 dual-resolution detection backbone network. YOLOv8 provides a series of full-size models from Nano(N) to Extra-Large(X), provides a 640×640 and 1280×1280 dual-resolution detection backbone network, and integrates an improved instance segmentation branch based on YOLACT, which realizes the trade-off between accuracy and efficiency under different computational resource constraints.

Secondly, the backbone network and feature fusion layer adopts a multi-branch C2f structure instead of the original C3 module and enhances the gradient propagation path flux through splitting and connecting operations. Differentiated channel count allocation strategies are customized for different scale models instead of following a single parameter template. The final detection head structure abandons the preset anchor frame mechanism, adopts decoupled coordinate prediction based on centroids, incorporates TaskAlignedAssigner to realize the dynamic screening of positive samples, combines with Distribution Focal Loss to alleviate the category imbalance, inherits the YOLOX training strategy, and gradually turns off the Mosiaciac Loss in the final training stage to enhance the gradient propagation path flux. Inheriting the YOLOX training strategy, Mosiac enhancement is gradually turned off in the final training stage to improve the convergence stability of the model. The network structure of YOLOv8 is shown in Fig. 1.

3. LIGHTWEIGHT YOLOV8-SCS-CE CONSTRUCTED BASED ON IMPROVED YOLOV8

To address the issues of complex backgrounds, occlusions, parcel damage defects, and the challenges of having a large number of parameters and difficulties in deploying on embedded devices for realistic target detection in warehousing and logistics, YOLOv8n has been modified in terms of backbone network, feature fusion, and loss function. The lightweight YOLOv8n-SCS-CE algorithm is proposed. The parametric quantities of the algorithm are detailed in Table 1.

The enhancement of the model feature extraction capability maintains a low computational complexity. This makes it ideal for use in application scenarios that require efficient real-time processing, such as express parcel detection.

3.1. Lightweight Backbone Network Design

YOLOv8 uses deeper and more complex backbone networks to extract features. These backbone networks include more convolutional layers, Residual Blocks, Bottleneck Blocks, etc., which increase the computational effort. Therefore, it leads to high computation and slow operation speed, which makes it difficult to meet the demand for deployment in embedded or mobile devices. We propose a feature-enhanced network SCS based on a bi-dimensional attention mechanism of channels and coordinates,

which greatly compresses the number of parameters and thus realizes the balance between computational load and detection accuracy.

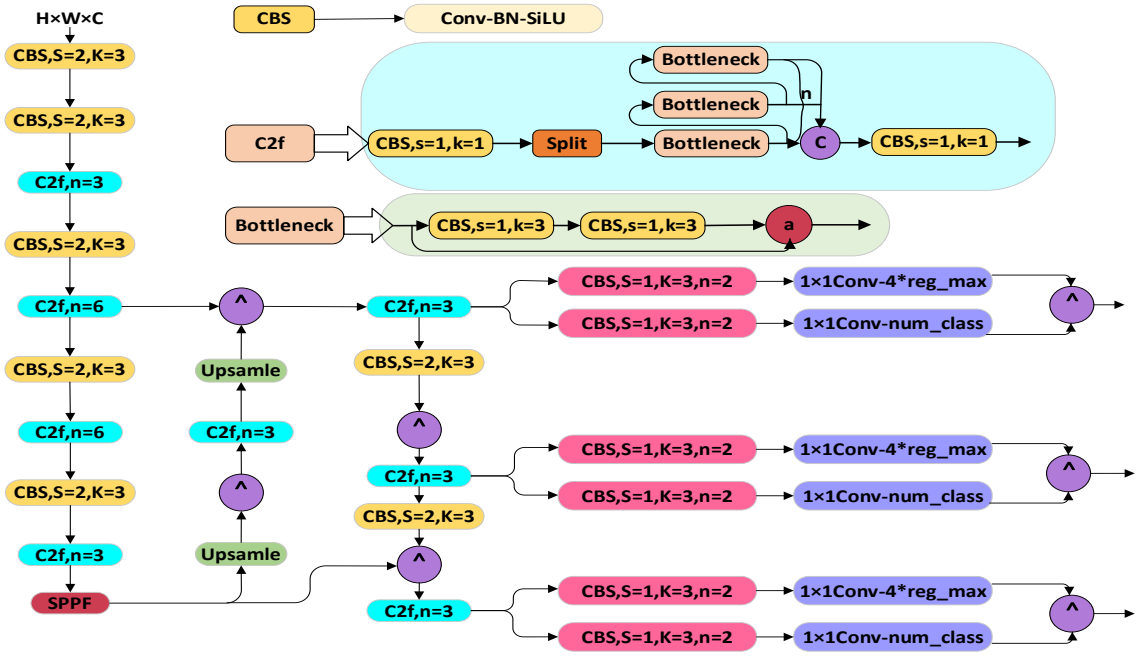

Fig. 1. YOLOv8n's network architecture

Table 1

Model Structural Parameters

Type	Input channel number	Output channel number	Output feature map	Parameter	Convolution kernel size	Stride
Conv	3	16	320*320*16	464	3	2
SCS	16	24	160*160*24	936		2
SCS	24	24	160*160*24	456		
SCS	24	48	80*80*48	2160		2
SCS	48	48	80*80*48	2400		
SCS	48	96	40*40*96	6624		2
SCS	96	96	40*40*96	7104		
SCS	96	192	20*20*192	22464		2
SCS	192	192	20*20*192	11712		
SPPF	192	192	20*20*192	92736		
Upsamle	192	192	40*40*192	0		
Concat			40*40*288	0		
C2f_ECA	288	96	40*40*96	83523		
Upsamle	96	96	80*80*96	0		
Concat			80*80*144	0		
C2f_ECA	144	48	80*80*48	21027		
Conv	48	48	40*40*48	20832	3	2
Concat			40*40*144	0		
C2f_ECA	144	96	40*40*96	69699		
Conv	96	96	20*20*96	83136	3	2
Concat			20*20*288	0		
C2f	288	192	20*20*192	277632		
Detect	48 96 192			525622		

Propose a feature-enhanced network SCS based on a bi-dimensional attention mechanism of channels and coordinates, which greatly compresses the number of parameters and thus realizes the balance between computational load and detection accuracy. As shown in Fig. 2.

In the backbone network part, the initial spatial features are first captured using conventional convolution to reduce the size of the feature map and reduce the subsequent computation. Then several SCS networks are used for lightweight feature extraction, which improves ability to characterize and computational efficiency of the network through various ways such as attention mechanism, lightweight design, multi-scale feature extraction and feature fusion.

The neck network part replaces the C2f module with the C2f_ECA module and adds an efficient channel attention mechanism to obtain a certain performance gain without basically changing the number of parameters.

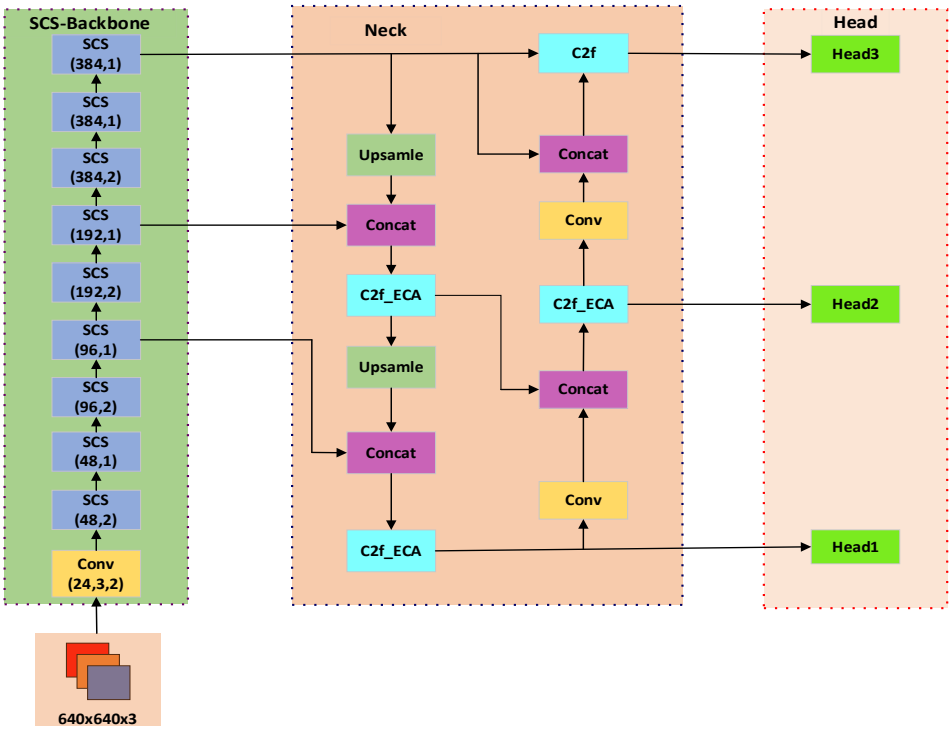

Fig. 2. YOLOv8n-SCS-CE: network architecture optimized with lightweight modules

3.2. C2f_ECA: Lightweight Feature Extraction Module

To enhance inter-pixel dependencies, attention mechanisms (CBAM, SE) have been introduced into convolutional neural networks, demonstrating great potential in the integration of cross-dimensional attention weights with input features. However, these attentional modules involve a large number of pooling operations that significantly increase the computational requirements. To overcome the limitation of practical factors, an efficient channel attention module, ECA, is used here, and the network structure is shown in Fig. 3. ECA captures the local cross-channel interaction information by considering each channel and its KNN (k-Nearest Neighbors). It can efficiently be implemented by a fast 1D convolution with size k to accomplish a non-dimensionality reducing local cross-channel interaction strategy. This mechanism helps to perform inter-channel interactions more efficiently while maintaining inter-channel correlation, thus improving the expressiveness and performance of the network.

The ECA mechanism enhances channel feature selection and modeling capability by introducing efficient channel attention, while maintaining computational efficiency. This compensates for the lack of inter-channel correlation and fine-grained feature selection in the C2f module. As a result, this paper adds the ECA attention mechanism to the C2f forward propagation process and constructs the C2f_ECA module. The new structure is shown in Fig. 4.

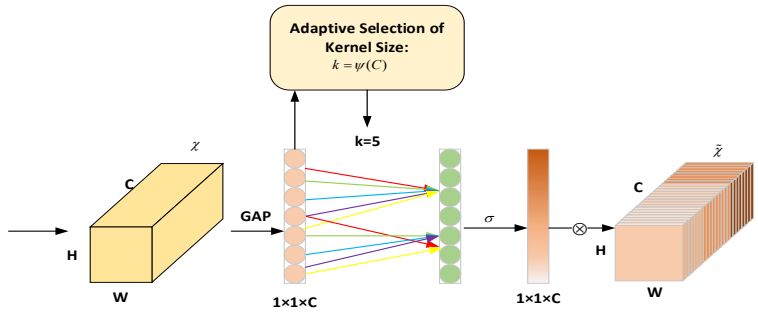

Fig. 3. Network architecture with ECA attention mechanism

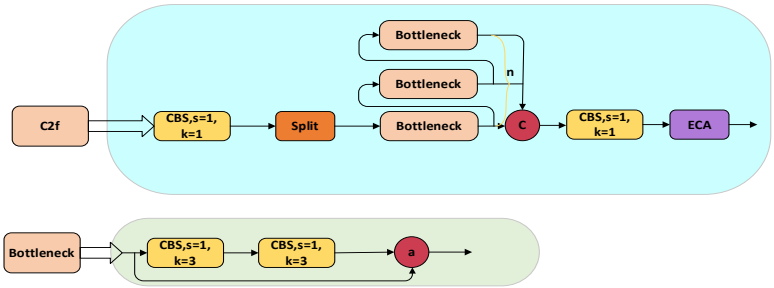

Fig. 4. Architecture of C2f_ECA

4. LIGHTWEIGHT MOUDULE SCS

This section proposes a lightweight feature extraction module, SCS, which realizes feature capture and balanced scheduling of computational resources by means of a multi-branch collaborative architecture and a dynamic attention fusion strategy. In this section, its core design will be elaborated in turn: first, to address the three-level topology, feature decoupling, two-branch distillation, and attention recalibration of the module; second, to explore the cascaded extended convolution and gated-time attention mechanisms introduced in branch expansion; and finally, to analyze the adaptive optimization effect of cross-dimensional coordinate attention (CA) in complex logistics scenarios.

4.1. SCS: Design of The Lightweight Module

The lightweight feature extraction module SCS proposed in this paper incorporates the CoordAttention attention mechanism based on Shufflenetv2 [13], which enhances feature extraction for ShuffleNetV2, and improves the model's performance in handling complex scenes and detailed features while maintaining lightweight and efficiency. ShuffleNetV2 is designed to be lightweight through a streamlined architecture and techniques such as group convolution. It introduces the channel shuffle operation, which significantly reduces the computational load and parameter count of the model, facilitates the exchange and fusion of information between different channels, and enhances the model's efficiency on mobile and embedded devices. The expression ability of the model is improved.

The improved SCS module first goes through the initial processing of the input image by the conv_maxpool module, which extracts the base features and downsamples them. Then the SC module is introduced to process the feature map in depth. Through the stacking of multiple layers of SC modules, feature extraction and fusion at different scales are realized to enhance the classification performance of the model. The fc layer pools extracts features globally on average and then inputs them to the fully connected layer to realize the final classification. Through the stacking of multi-layer SC modules, step-by-step downsampling, and feature extraction, combined with the CoordAttention mechanism, the selectivity and expressive ability of features are enhanced, and finally, classification is achieved through the fully connected layer.

As shown in Fig. 5, when the step size is 1, the input features are firstly separated by channel decoupling, the left channel completes multilevel feature abstraction and feature interaction by connecting SC enhancement modules in series, and the right channel adopts the jump connection to interact the original features with the transformed features across layers of information. Finally, the fused features are reorganized by channel dimension substitution, which significantly reduces the statistical dependence between feature mappings, and thus improves the compactness and effectiveness of feature expression.When the step size stride is 2, the input channel is divided into two parts. In the left half, a 3×3 convolutional layer is inserted between the SC module for downsampling. In the right half, a downsampling operation is performed first, followed by a 1×1 convolutional layer, and then the spatial attention mechanism (CoordAttention) is applied to enhance the correlation between different channels. Then the two channels perform a feature fusion with the information obtained. The two channels then fuse the obtained information into features and finally go through the channel shuffle module. In this way, more local information can be extracted while retaining the global information. It not only effectively improves the performance of ShuffleNetV2 in feature extraction, but also maintains the lightweight and efficiency.

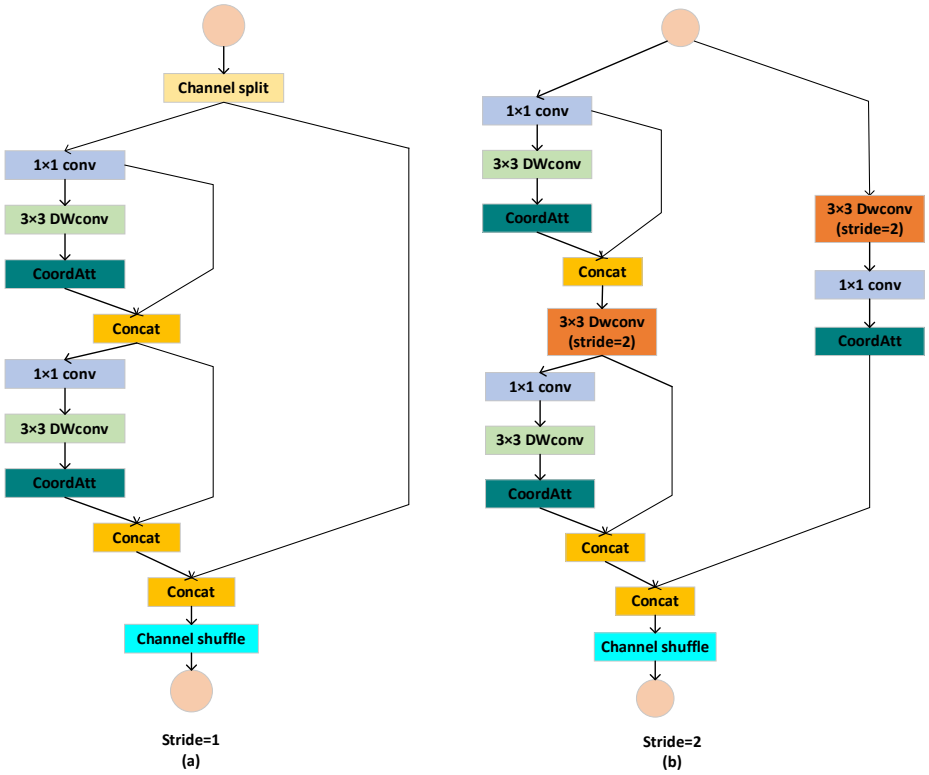

Fig. 5. Lightweight module SCS

4.2. Use of Multiple SC Module

SC module is improved on the basis of the original Shufflenetv2 model, and the network structure is shown in Fig. 6. Shufflenetv2 reduces the number of parameters of the network structure while improving the inference speed on the basis of retaining the channel blending and group convolution. However, considering the original model is prone to noise and background interference in image processing.The spatial attention mechanism, CoordAttention, is added to the original architecture to help the network pay more attention to the important feature regions and improve the expression ability of the features by applying spatial attention processing to the features. Meanwhile, the SC Module is nested several times in the network structure, and each time this module is used, the input features are processed to a certain extent, and the channel attention is applied several times, which helps the network to pay better attention to the important feature channels and improve the expression ability of the features, so as to increase the network's perception of the input image. Each module contains nonlinear

transformations such as ReLU [14] activation functions, which increase the nonlinear expressive ability of the network through repeated combinations to improve the diversity and effectiveness of features. The model can effectively enhance several tasks such as image classification, object detection and other different tasks by enhancing the depth and representation ability of the network.

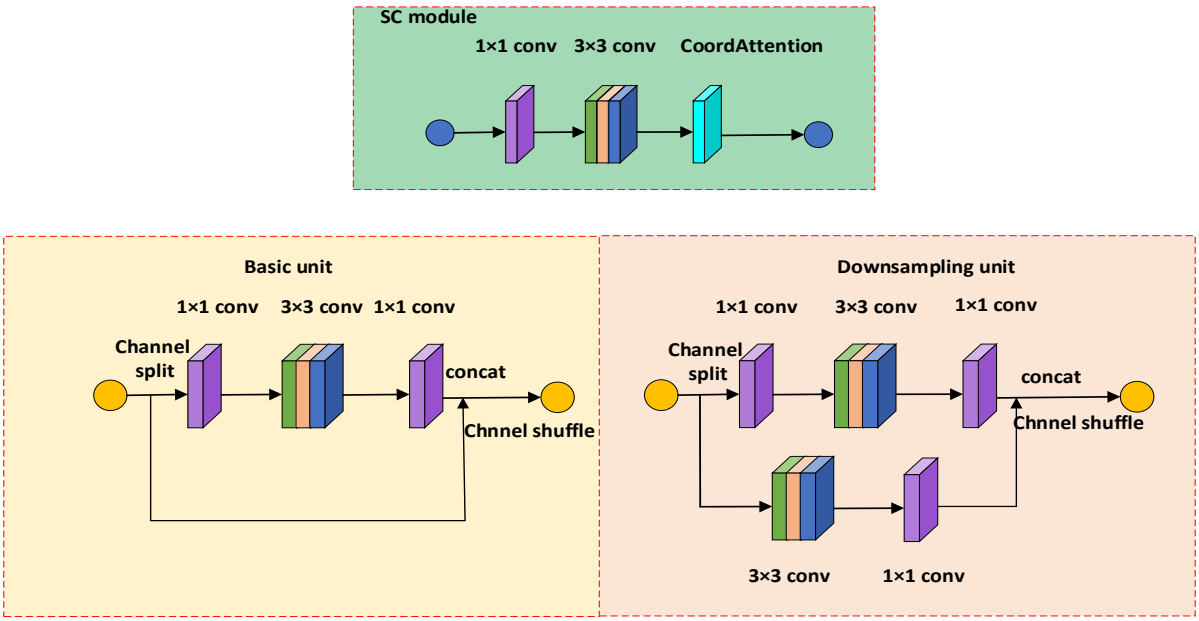

Fig. 6. SC module

4.3. CoordAttention Mechanism

CoordAttention Attention Mechanism is a positional attention mechanism that encodes spatial information in a neural network using transform [15] so as to model channel relationships and their connectivity, and fuses spatial information by means of channel weighting. Its module is shown in Fig. 7.By virtue of its flexibility, the mechanism can be embedded into the ShuffleNetV2 structure to enhance the target localization and recognition accuracy of the model.

For the *c-th* channel of height *h* the output can be described as:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} X_c(h,i) \tag{1}$$

The output for the *c-th* channel of width w is described by the following equation:

$$z_c^w = \frac{1}{H} \sum_{0 \le j < H} X_c(j,w) \tag{2}$$

The coordinate attention generation operation is proposed, which first cascades the two feature maps generated by the previous module and then transforms the F_1 by a shared 1×1 convolution, which is expressed in the following equation:

$$f = \delta(F_1(\left[z^h \; z^w \right])) \tag{3}$$

Next, Eq (3) is sliced into two independent tensors along the spatial dimension, and then two 1x1 convolutions F_h and F_w are utilized to convert the feature maps f^h and f^w to the same number of channels as the input X, yielding the final attention vectors g^h and g^w.

$$g^h = \sigma(F_h(f^h)) \tag{4}$$

$$g^w = \sigma(F_w(f^w)) \tag{5}$$

Then expanding on g^h and g^w, the final output of this module can be expressed as follows.

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{6}$$

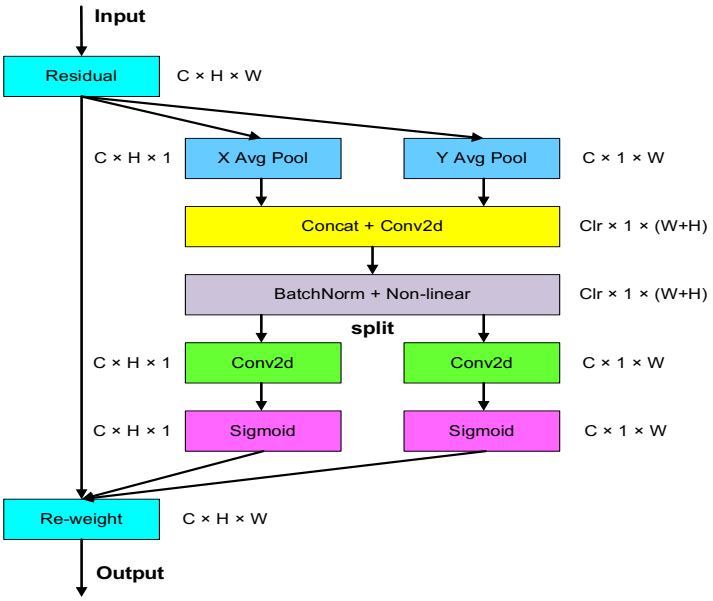

Fig. 7. Network architecture with CoordAttention

5. EXPERIMENTAL RESULTS AND ANALYSIS

5.1. Experimental Platform

The experiment platform is based on the deep learning framework built on the GPU of Intel(R) Lris(R) Xe Graphics, the CPU of 12th Gen Intel(R) Core (TM) i7-12700H, Python3.10, PyTorch2.0.1, and Cuda11.8, and the experiment-related parameters are shown in Table 2.

Table 2

Experimental Parameters

Experimental parameters	Parameter's quantity
Epoch	200
Batch size	16
Optimizer	SGD
Learning rate	0.01

5.2. Experimental Dateset

There are few public datasets about express parcels, so the dataset in this paper collects a large number of images for autonomous labeling on the basis of a small number of public datasets. Realistic scenarios are simulated by implementing probabilistic horizontal flipping, angular random rotation, and exposure adjustment. This dataset contains a total of 10,330 labeled images, including the training set and the validation set, which are divided into 8:2 ratio. There are two categories in the dataset, where "0" represents parcel and "1" represents box. The prediction results are shown in Fig. 8. The classification can be detected more accurately by training the model, but there is still a certain false detection rate in complex environments. Fig. 8 shows the prediction results for the selected test cases in the dataset. These examples demonstrate the bounding boxes of the model predictions, illustrating its performance after training. Although classification accuracy is significantly improved by model optimization, a certain level of false detection still occurs in complex environments. The selected cases highlight typical successful and challenging scenarios, providing insights into the model's current

strengths and areas for improvement. This figure illustrates only a subset of the results and does not represent the complete evaluation dataset.

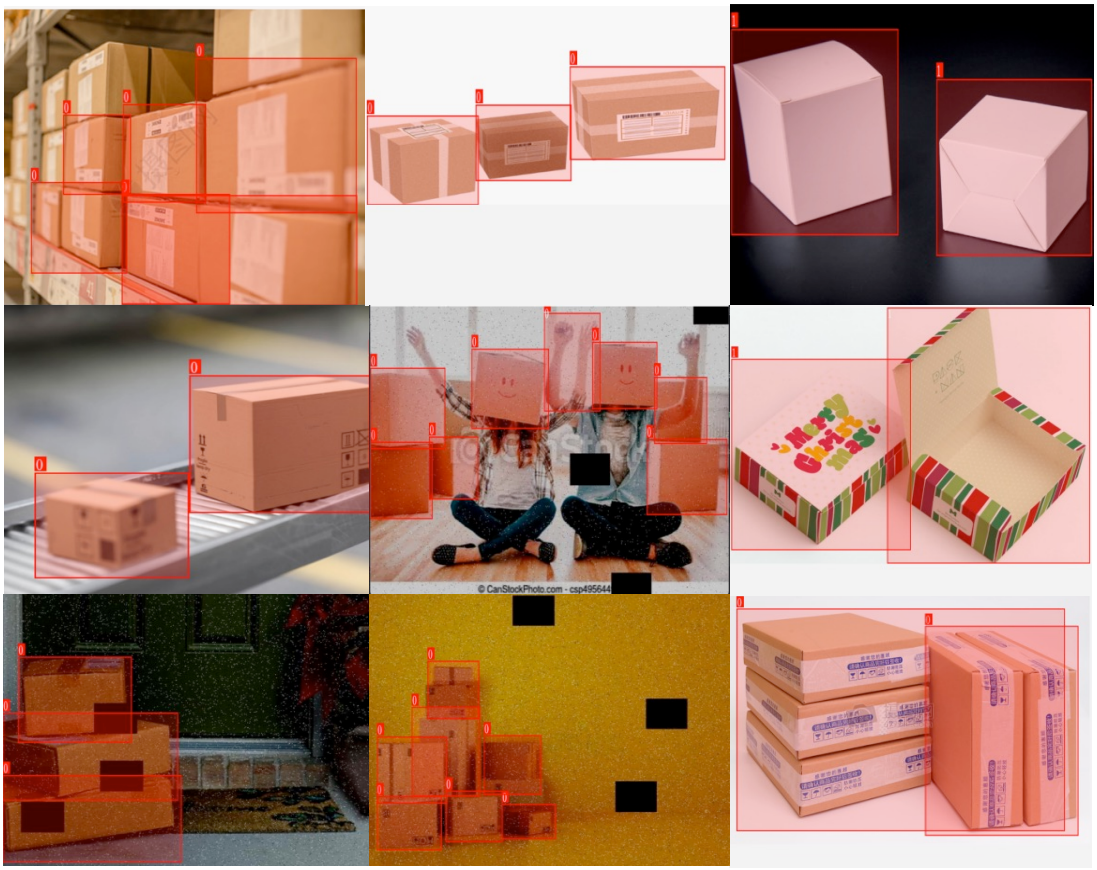

Fig. 8. Predicted effect diagram

5.3. Performance Comparison Experiment

By replacing the backbone network of the benchmark model YOLOv8n with a lighter feature extraction network, model complexity is reduced and real-time detection performance is improved. The experimental results show (see Table 3) that the depth and complexity of the model can be greatly improved without compromising the accuracy of the model by making lightweight improvements to the backbone and neck-level networks. This approach achieves a more efficient model structure while maintaining performance. As can be seen from the Table 3, using SCS-CE to improve the network structure, SCS-CE has a parameter count of 1.22×10^6M, GFlOPs of 3.7, and a model size of 2.8MB Compared to the benchmark model YOLOv8n, the parameter count is reduced by 60%, the computational complexity is reduced by 54%, and the model size is reduced by 55%. This reduction substantially improves the computational efficiency and inference speed of the model. It is favorable for deployment in resource-constrained environments, such as embedded devices and mobile devices. However, the performance is still outstanding in terms of precision (mAP@0.5), recall (R) and precision (P). The mAP@0.5 of SCS-CE was 91.4%, and its recall and precision were 85.9% and 83.6%, respectively, which were only 2.6% and 0.4% lower. In the number of network layers, depth of the model reaches 292 layers, although the number of layers is more, the role of lightweight modules not only does not increase the computational overhead but also reduces the complexity of the model.

Fig. 9 shows that the improved lightweight model in this paper is slightly lower mAP@0.5 than YOLOv8n, with a reduction of only 0.8%, but a small improvement compared to the SCS model and significantly better than other models. It is important to note that this slight drop in accuracy comes at the cost of significantly reducing model complexity and improving the applicability of embedded

systems. The SCS-CE network structure mAP@0.5 to 91.4% with the same number of parameters, while ShuffleV2, Mobilenetv3-CE and Mobilenetv3 have a relatively low mAP@0.5 with the same number of parameters. It can be seen that the YOLOv8n-SCS-CE can still maintain high accuracy at a lower number of parameters.

Table 3

Comparison of Detection Performance of Different Models

Backbone	Parameters/$\times 10^6$ M	GFlOPs	Model size/MB	mAP@0.5/%	R/%	P/%	Layers
Mobilenetv3	1.19	2.8	2.7	87.4	83.2	79.1	286
ShufflenetV2	0.45	1.7	1.2	80.4	78.3	71.1	238
Mobilenetv3-CE	1.18	2.8	2.7	88.8	84.7	79.2	298
SCS	1.22	3.7	2.7	91.1	85.7	81.5	280
SCS-CE (ours)	1.22	3.7	2.8	91.4	85.9	83.6	292
YOLOv8n	3.0	8.1	6.3	92.2	88.5	84	168

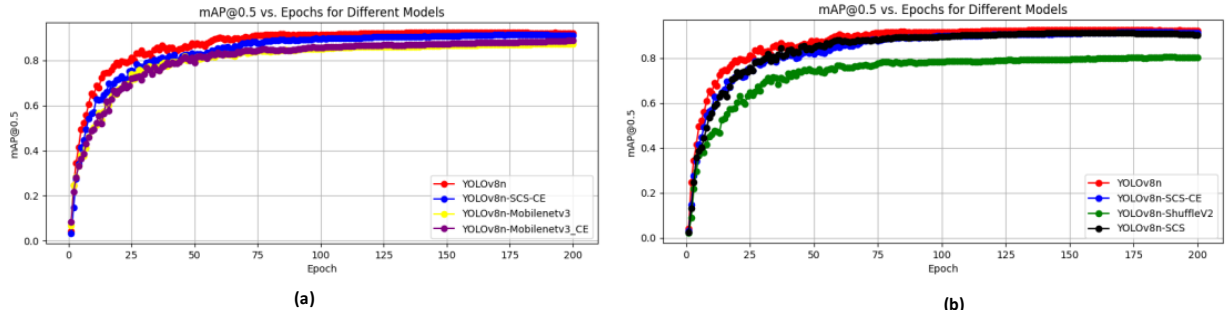

Fig. 9. Precision (mAP@0.5) comparison for different models

Fig. 10. Comparison of training effect of different models

SCS-CE is mainly characterized by the fact that the method can significantly reduce the number of parameters under the premise of guaranteeing high accuracy, with smaller scale, which is suitable for resource-constrained occasions. Therefore, the research results of this paper can not only meet the real-time requirements, but also provide practical solutions for large-scale industrial applications, which is of great practical significance.

5.4. Ablation Experiments

The effectiveness of the model proposed in this paper is verified by designing a set of ablation experiments, which keep the training environment and training parameters consistent. The results are shown in Table 4, where √ represents the use of the module and × represents none. Among them, the base model uses YOLOv8n, firstly, Shufflenetv2 is used to replace the backbone network in the network architecture, and the lightweight effect shows excellent performance, and the number of parameters is optimized significantly in comparison with YOLOv8n, so the SCS module is proposed on this basis. Replacing the Shufflenetv2 module with the SCS module, it can be found that the number of parameters increases a small amount but the Map@50 increases to 91.1%, which is only 1.1% down compared with the original model. To further optimize the model, the ECA attention mechanism is added to the C2f forward network to form the CE module, replacing the previous three C2f modules of the neck network, mAP@50 increases by 0.3% on the basis of YOLOv8n-SCS, and the accuracy reaches 91.3%, indicating that the model achieves lightweighting while maintaining the detection performance.

As can be seen from the Table 4, the parameter counts of YOLOv8n, YOLOv8n-Shufflenetv2, YOLOv8n-SCS, and YOLOv8n-SCS-CE are all lower than the benchmark model, and above the lightweighting, we can see that the optimization of Shufflenetv2 is obvious. Considering the problem of detection accuracy, by comparison we know that the mAP@50 of YOLOv8n-SCS-CE is significantly better than the YOLOv8n-Shufflenetv2 model, which basically takes into account both lightweight and detection accuracy.

With the same number of training rounds and the training platform, comparing metrics such as edge loss, classification loss, precision, recall, average precision (mAP@0.5), and average accuracy (mAP@0.5-0.95) reveals that YOLOv8n-SCS-CE performs well across several performance metrics, as shown in Fig. 11. It can be seen that YOLOv8n-SCS-CE maintains a low border loss and classification loss throughout the training process, indicating that the SCS-CE module can accurately locate the target. The precision curve indicates that the model's performance is comparable to the benchmark model, as evidenced by other tests. This demonstrates that the SCS-CE module effectively reduces false detections while decreasing the number and complexity of parameters. In the recall curves, YOLOv8n-SCS-CE and YOLOv8n performed similarly. The mAP@0.5-0.95 and mAP@0.5 curves coincide with the benchmark model, which further verifies that the SCS-CE module balances the detection accuracy and parameter quantity.

Table 4

Improved Model Ablation Test Results

Model	Shufflenetv2	SCS	CE	mAP@0.5/%	Model size/MB	Parameters	GFlOPs
YOLOv8n	×	×	×	92.2	6.3	3×10^6	8.1
YOLOv8n-Shufflenetv2	√	×	×	80.4	1.2	4.56×10^5	1.7
YOLOv8n-SCS	√	√	×	91.1	2.7	1.22×10^6	3.7
YOLOv8n-SCS-CE (ours)	√	√	√	91.4	2.8	1.22×10^6	3.7

6. CONCLUSIONS

To achieve lightweight, high-precision, and rapid detection of express parcels, this paper proposes the YOLOv8n-SCS-CE detection model, which ensures lightweight operation while maintaining detection accuracy. Ablation experiments show that integrating the SCS module into the backbone network can effectively reduce the model parameters, while appropriate cross-channel interactions not only reduce the network complexity but also keep the performance stable. In the neck network, replacing C2f with C2f_ECA improves the feature extraction capability while reducing the number of parameters, thus improving the detection accuracy. The experimental results show that compared with the original

network, the parameter quantity of YOLOv8n-SCS-CE is reduced by 59.2%, the weight file is reduced to 2.8M, which is 55.5% less compared with the source code, and the mAP@50 decreases by only 0.8%, while the detection speed is improved by more than double. The experimental results show that the proposed YOLOv8n-SCS-CE model realizes a lightweight learning network and improves the detection speed while maintaining the detection accuracy, which is suitable for deployment in embedded device systems. This research can provide lightweight detection and identification technology for the intelligent warehousing and logistics industry and solve the problem of unmanned intelligent sorting and transportation.

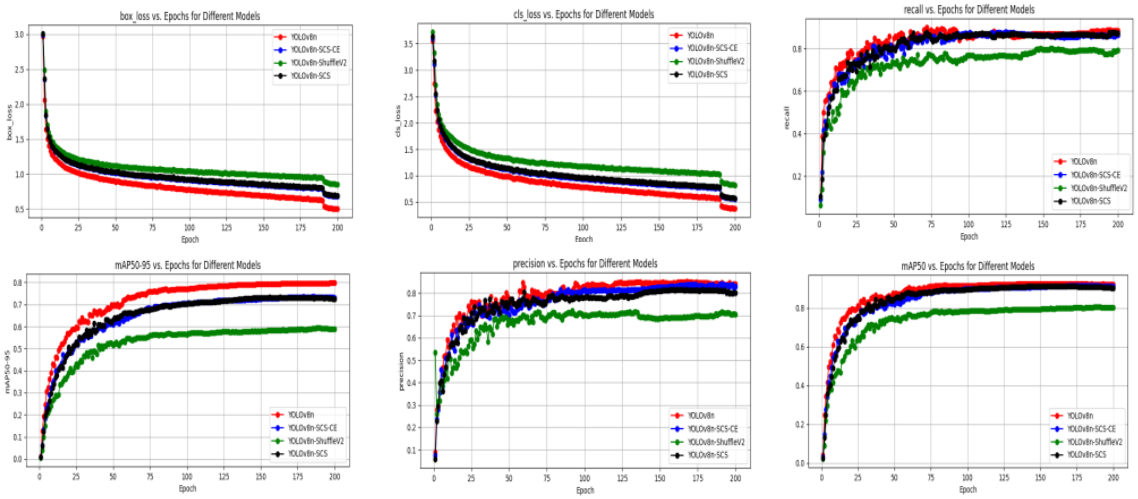

Fig. 11. Comparison of performance indicators

Acknowledgments

Project Supported by Shandong Provincial Natural Science Foundation of China (ZR2023MF077).

References

1. Wang, J. & Lim, M.K. & Zhan, Y. & et al. An intelligent logistics service system for enhancing dispatching operations in an IoT environment. *Transportation Research Part E: Logistics and Transportation Review.* 2020. Vol. 135. P. 101886.
2. Matusiewicz, M. Logistics of the future – Physical internet and its practicality. *Transportation Journal.* 2020. Vol. 59(2). P. 200-214.
3. Tran-Dang, H. & Krommenacker, N. & Charpentier, P. & et al. Toward the internet of things for physical internet: Perspectives and challenges. *IEEE internet of things journal.* 2020. Vol. 7(6). P. 4711-4736.
4. Ha, N.T. & Akbari, M. & Au, B. Last mile delivery in logistics and supply chain management: a bibliometric analysis and future directions. *Benchmarking: An International Journal.* 2023. 30(4): Vol. 113. P. 1170.
5. Ciprés, C. & de la Cruz, M.T. The physical internet from shippers perspective. *Towards User-Centric Transport in Europe: Challenges, Solutions and Collaborations.* 2019. P. 203-221.
6. Sładkowski, A. *Using Artificial Intelligence to Solve Transportation Problems.* First Edition. Cham: Springer. 2024. XII. 565 p. DOI: 10.1007/978-3-031-69487-5.
7. Moshood, T.D. & Sorooshian, S. The Physical Internet: A means towards achieving global logistics sustainability. *Open Engineering.* 2021. Vol. 11(1). P. 815-829.

8. Treiblmaier, H. & Mirkovski, K. & Lowry, P.B., et al. The physical internet as a new supply chain paradigm: a systematic literature review and a comprehensive framework. *The International Journal of Logistics Management.* 2020. Vol. 31(2). P. 239-287.
9. Sirignano, J. & MacArt, J.F. & Freund, J.B. DPM: A deep learning PDE augmentation method with application to large-eddy simulation. *Journal of Computational Physics.* 2020. Vol. 423. P. 109811.
10. Yeh, J.F. & Lin, K. M. & Lin, C.Y. et al. Intelligent mango fruit grade classification using alexnet-spp with mask r-cnn-based segmentation algorithm. *IEEE Transactions on AgriFood Electronics.* 2023. Vol. 1(1). P. 41-49.
11. Patnaik, S.K. & Babu, C.N. & Bhave, M. Intelligent and adaptive web data extraction system using convolutional and long short-term memory deep learning networks. *Big Data Mining and Analytics.* 2021. Vol. 4(4). P. 279-297.
12. Chen, F. & Li, S. & Han, J. & et al. Review of lightweight deep convolutional neural networks. *Archives of Computational Methods in Engineering.* 2024. Vol. 31(4). P. 1915-1937.
13. Han, J. & Yang, Y. & L-Net: lightweight and fast object detector-based ShuffleNetV2. *Journal of Real-Time Image Processing.* 2021. Vol. 18(6). P. 2527-2538.
14. Daubechies, I. & DeVore, R. & Foucart, S., et al. Nonlinear approximation and (deep) ReLU networks. *Constructive Approximation.* 2022. Vol. 55(1). P. 127-172.
15. Yu, G. & Yu, M. & Xu, C. Synchroextracting transform. *IEEE Transactions on Industrial Electronics.* 2017. Vol. 64(10). P. 8042-8054.

Received 21.01.2024; accepted in revised form 13.03.2025