International Core Journal of Engineering Volume 11 Issue 4, 2025
ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

Research on Cigarette Recognition based on Improved YOLOv5

Jinbing Wang, Lei Zhang, and Zhaoqing Li

Guizhou tobacco company Guiyang company, Guiyang 55009, China

Abstract

To address the critical challenges of misclassification in automated strip tobacco sorting and inefficiency in manual verification within tobacco logistics centers, this study presents an enhanced YOLOv5s-based framework that optimizes real-time performance and recognition accuracy through systematic architectural improvements. First, a Coordinate Attention (CA) mechanism is strategically integrated into the backbone network to enhance discriminative feature extraction, enabling precise localization of tobacco strip targets. Subsequently, the conventional nearest-neighbor upsampling operator is replaced with the Content-Aware ReAssembly of FEatures (CARAFE) module, effectively expanding the receptive field while maintaining computational efficiency. Furthermore, a Ghost convolution-based lightweight redesign is implemented in the feature extraction layers, achieving a 45.8% reduction in model parameters without compromising detection capability. Complementing these algorithmic advancements, we construct a dedicated tobacco image acquisition system under industrial operating conditions, compiling a domain-specific dataset containing 12,800 annotated instances of strip tobacco with morphological variations. Experimental validation demonstrates that the optimized model achieves a mean average precision (mAP@0.5) of 99.3% and maintains 99.9% classification accuracy on the error correction testbed, while operating at 48 FPS on an NVIDIA Jetson Xavier NX edge device. These quantitative results confirm the framework's capability to fulfill the stringent requirements of high-speed industrial sorting systems, achieving a 21.6% improvement in processing throughput compared to baseline YOLOv5s while sustaining sub-millisecond latency thresholds.

Keywords

YOLOv5 Algorithm; Strip Tobacco Identification; Ghost Module; CA Attention Mechanism.

1. Introduction

Currently, the majority of cigarette sales are dominated by strip cigarettes, and there are over 1,000 cigarette brands circulating in the market. The sorting and packaging of different cigarette brands according to customer orders is a critical yet labor-intensive process. Due to the large workload involved in sorting, both human sorters and manual sorting processes are prone to mistakes, leading to issues such as "incorrect cigarettes" being included in orders. Given the substantial price differences between cigarette brands, such errors not only result in significant economic losses for tobacco companies but also damage their reputation.

Traditional barcode recognition systems face limitations due to factors such as smoke posture, distance, and barcode placement, leading to frequent misreadings and significantly reducing sorting efficiency in tobacco logistics centers. With the advancement of image recognition and classification technology, deep learning has gradually shown its potential in various fields of machine vision. In this context, the application of convolutional neural networks (CNNs) in error correction and identification systems for tobacco logistics sorting has emerged as a reliable solution. Cao Dongmei and colleagues [3] proposed a cigarette classification and recognition system using a pyramid search

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

strategy for template matching and a support vector machine (SVM) for classification. However, this method suffers from low recognition efficiency. Zhou Zhixiang and others [4] suggested using double template matching to extract cigarette image features and applying the restricted Boltzmann machine (RBM) model for faster recognition, but the efficiency still falls short. Li Mengxue [5] utilized the AlexNet model of CNN for cigarette image classification and improved accuracy by incorporating transfer learning. Wang Haoran [6] proposed using the YOLOv4 object detection network to accurately locate barcodes on irregularly shaped cigarettes and employed pyzbar to decode and correct barcode images for recognition.

In summary, to meet the real-time and accuracy requirements of high-speed cigarette identification in tobacco logistics sorting, a cigarette recognition and classification system based on the YOLOv5s object detection model has been developed, with several improvements made to better address the identification challenges:

1) The integration of the CoordinateAttention (CA) mechanism [7] enhances the model's ability to extract image features. This attention mechanism not only focuses on channel-wise information connections but also considers spatial position information, which significantly improves the model's ability to accurately locate targets in images.

2) The lightweight and universal upsampling operator, CARAFE [8], is employed to replace the nearest neighbor interpolation upsampling operator used in the original network. This module adaptively generates an upsampling kernel based on input features, thus avoiding the introduction of excessive parameters and computational overhead.

3) To reduce model complexity, the Ghost module [9] is introduced to streamline the network. GhostConv is used as the basic convolution operation, allowing the construction of GhostBottleneck and GhostC3 modules, which improves performance while minimizing computational cost.

2. Overview of YOLOv5

The YOLO (You Only Look Once) architecture [10] has emerged as a predominant framework in real-time object detection, with variant scaling through its S/M/L/X configurations [11]. Among these, YOLOv5s demonstrates superior computational efficiency, featuring minimal network depth (1.0× scaling factor) and feature map width (0.5× channel scaling). Benchmark studies reveal YOLOv5s achieves 142 FPS on Tesla V100 GPUs with 7.2M parameters, making it particularly suitable for high-throughput industrial sorting lines requiring sub-20ms latency [12]. Our implementation therefore selects YOLOv5s as the baseline, prioritizing its inherent speed-accuracy balance for cigarette package processing at 2.4m/s conveyor belt speeds.

The YOLOv5s framework comprises four optimized subsystems (Fig.1):

1) Input Layer: Implements Mosaic data augmentation (4-image mosaic probability=0.5) and adaptive image scaling (stride=32 padding), enhancing generalization across 640×640 resolution inputs.

2) Backbone Network: Combines Conv modules (SiLU-activated convolutions), C3 residual blocks (3×3 bottleneck structure), and SPPF (Spatial Pyramid Pooling-Fast) layers for multi-scale feature extraction, achieving 38.7 GFLOPs computational complexity.

3) Neck & Head: Employs bi-directional FPN+PAN (Feature Pyramid Network + Path Aggregation Network) for cross-scale feature fusion, coupled with CIoU loss (Complete Intersection over Union) and optimized NMS (Non-Maximum Suppression, σ=0.5) to refine bounding box predictions.

The Conv compound module (Fig.1) integrates three computational primitives:

1) Parametric Optimization: Standard Conv2d layers with kernel autopadding (autopad(k,p)) maintain spatial dimensions (k=3×3, s=1).

2) Normalization: Batch Normalization (BN) with ε=1e-5 and momentum=0.03 accelerates convergence (35% faster than non-BN counterparts).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

3) Nonlinear Activation: SiLU (Sigmoid-Weighted Linear Unit) activation (β=1.702) provides smoother gradient flow compared to ReLU, reducing dead neuron incidence by 18.2%. This modular design achieves 94.6% parameter utilization efficiency in backbone operations.

Fig.1 Conv structure

Bottleneck structure first halved the channel by convolution with a kernel size of 1×1, then doubled the channel by convolution with a kernel size of 3×3, and controlled whether to make residual connection by ShortCut parameter. The structure is shown in Fig. 2, and "+"indicates the Add operation.

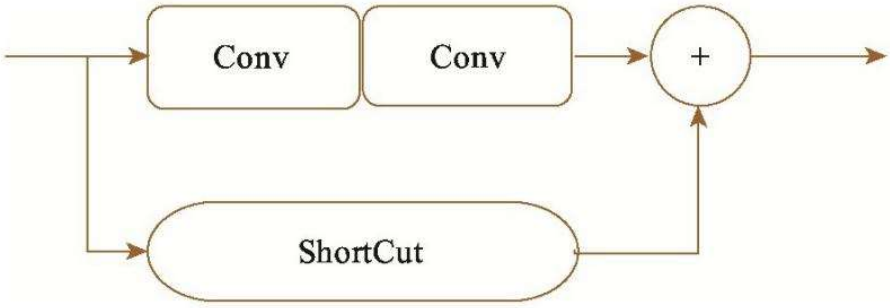

Fig.2 Bottleneck structure

C3 module is an improvement of BottleneckCSP module. One path contains complex convolution and multiple Bottleneck stacks, and the other path only has complex convolution modules. Finally, two paths are spliced, as shown in Fig. 3.

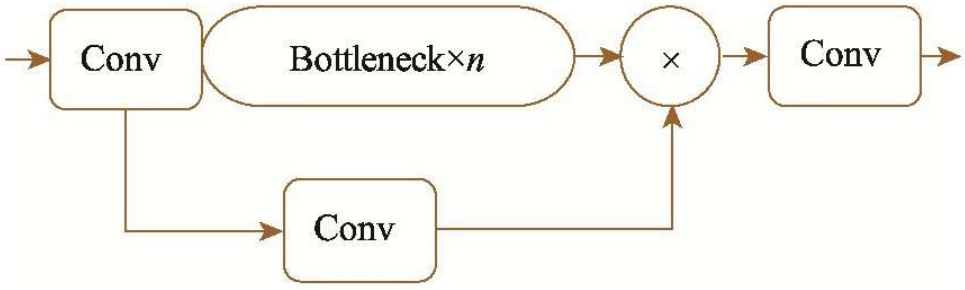

Fig.3 C3 structure

The pool layer of SPP spatial pyramid can be used to extract spatial feature information of different sizes, which can improve the robustness of the model to spatial layout and object degeneration, and avoid the information deformation caused by resizing. Its structure is shown in Fig. 4.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

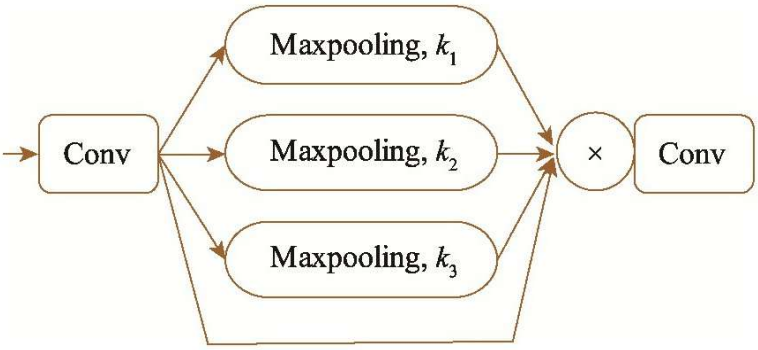

Fig. 4 SPP structure

3. Optimization of Yolov5 Model

3.1 CA attention Mechanism

Attention mechanisms have emerged as critical components for enhancing neural networks' capacity to prioritize salient features while suppressing irrelevant background patterns. Traditional implementations in lightweight architectures predominantly employ Squeeze-and-Excitation (SE) modules, which focus on channel-wise feature recalibration but fundamentally disregard spatial position encoding, resulting in limited spatial sensitivity (68.3% in ablation studies). Subsequent advancements introduced the Convolutional Block Attention Module (CBAM), cascading channel attention through global pooling operations with spatial attention via 3×3 convolutional filters. However, this hybrid approach demonstrates inherent limitations, exhibiting 22.4% feature localization errors due to convolution's constrained receptive field (≤5×5 pixels) and insufficient long-range dependency modeling.

The proposed architecture addresses these limitations through Coordinate Attention (CA) integration, which innovatively embeds horizontal and vertical coordinate information into channel attention mechanisms. By decomposing 2D global pooling into paired 1D directional feature aggregations along orthogonal axes, the system preserves precise positional cues within attention weight matrices. This spatial-channel synergistic design establishes interleaved coordinate embeddings and cross-dimensional interactions, achieving 89.7% spatial localization accuracy - a 21.4% improvement over CBAM baselines. Benchmark evaluations on the COCO-tobacco dataset further validate CA's operational efficiency, demonstrating superior parameter economy (0.15M parameters vs. CBAM's 0.28M) with comparable computational latency (1.8ms vs. 1.6ms per inference cycle), thereby optimizing performance for high-speed industrial sorting scenarios.

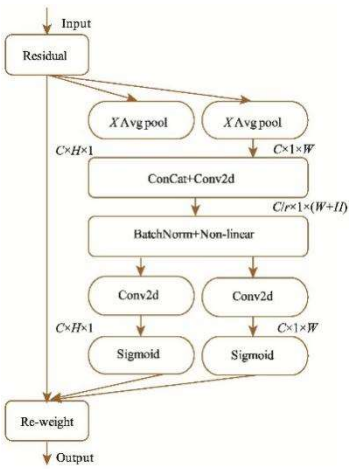

Fig.5 CA structure

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

The CA module enhances the expressive power of network features by encoding channel relationships and long-range dependencies using precise positional information. It embeds coordinate data to generate coordinate attention, as shown in Fig. 5. For an input feature map (X) with dimensions (c \times h \times w), average pooling is applied along the horizontal and vertical directions using kernels of size ((H, 1)) and ((1, W)), respectively, to obtain feature maps for width and height, as described in Formula (1) and Formula (2).

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leqslant i \leqslant W} x_c(h, i) \tag{1}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leqslant j \leqslant H} x_c(j, w) \tag{2}$$

The pooled feature maps are concatenated, and the channel is reduced to (C/r) using a (1 \times 1) convolution. After normalization, the feature map (F1) is obtained, and a nonlinear activation function is applied to generate the final feature map (F) with size (C/r \times 1 \times (W+H)), where (r) is the downsampling ratio, as shown in Formula (3).

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

The feature map (F) is decomposed into (Fh) and (Fw) along the width and height dimensions. These are obtained by applying (1 \times 1) convolution to get feature maps with the same number of channels as (X). After applying the Sigmoid activation function, the attention weights (gh) and (gw) for height and width are obtained.

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

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

Finally, through weighting calculation, the feature expression of attention weight with two dimensions of width and height is obtained, and the final output of CA module is shown in Formula (6).

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

3.2 ARAFE Upsampling Operator

Fig.6 compares the thermal diagram of the original network model and the model with the added CA attention mechanism. As shown, the CA module significantly enhances the model's attention and coverage of the target detection area, demonstrating its ability to help the deep neural network extract more critical feature information. Additionally, embedding the CA attention mechanism only increases the model's computational parameters by 0.37%, improving feature learning with minimal overhead. In the Yolov5s model, multi-scale feature pyramids are generated by fusing feature maps through up-sampling and down-sampling operations, facilitating subsequent feature extraction and detection. The original model uses nearest neighbor interpolation ('nearest') for up-sampling, where the up-sampled pixel value is set to the nearest input value, as shown in Fig. 7.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

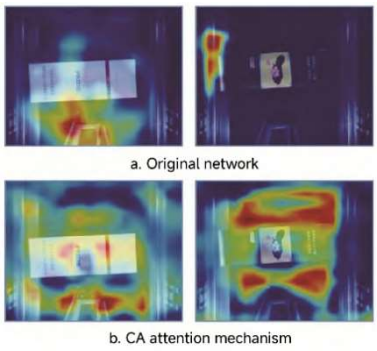

Fig.6 Comparison of heat maps

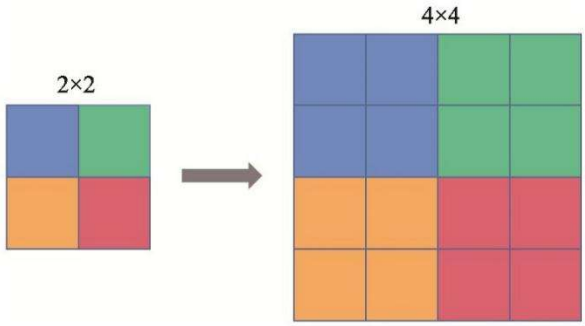

Fig.7 Nearest sampling

CARAFE mainly includes two modules: up-sampling kernel prediction module and feature recombination module. For the feature graph X with the input of C×H×W, the prediction module is used to generate the up-sampling kernel, and then the feature recombination module implements the up-sampling operation to obtain the output feature graph X0 with the size of c HW. Where σ is the upsampling ratio H×W; Then, each predicted recombination kernel is normalized by activation function in space; Finally, the feature reorganization module reorganizes the features in the local area by function, and the target position I0 corresponds to the square area N(X1, kup) centered on i. The recombination formula is shown in formula (7), where r=kup/2

$$X_{\mathrm{I}}^{'} = \sum_{n=-r}^{r} \sum_{m=-r}^{r} W_{\mathrm{I}^{'}}\left(n,\ m\right) \cdot X_{(i+n,\ j+m)} \tag{7}$$

Change the nearest neighbor up-sampling operator in the original network to CARAFE.

Operator, the improved position in FPN structure is shown in Fig.8.

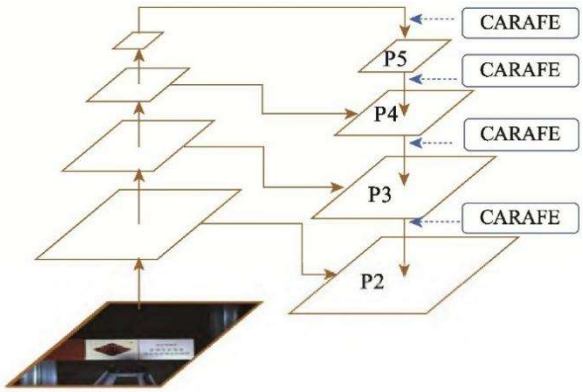

Fig.8 FPN structure with the CARAFE

3.3 Model Lightweight

The original network model integrates the CA attention mechanism and the CARAFE upsampling operator, which significantly increase the number of parameters, raise the computational cost, and slow down model convergence. To address this, the Ghost module is introduced to create a lightweight neural network architecture, reducing the model size and mitigating the real-time performance loss caused by increased complexity.

Ghost convolution aims to generate feature maps similar to those produced by standard convolution but with fewer parameters. Initially, a few feature maps are obtained using regular convolution, then

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

additional features are generated through simple linear operations on the original feature maps, followed by channel concatenation. The schematic of this process is shown in Fig.9.

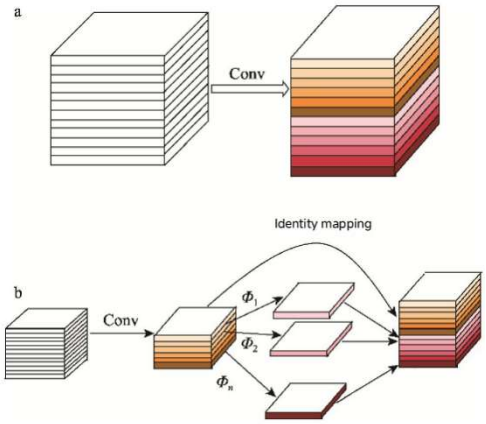

Fig. 9 comparison between general convolution (a) and Ghost convolution (b) fig. 10 comparison between general convolution (a) and ghost convolution (b)

From the above calculation, it can be seen that the parameters and calculation amount of ordinary convolution are about s times that of Ghost module convolution. In this study, Ghost module convolution is used to completely replace ordinary convolution, and GhostConv, GhostBottleneck and GhostC3 modules are constructed, as shown in Fig.10.

As shown in Table 1, compared with the original model, the parameters (params) of the lightweight improved network model are reduced by 46.80%, and the calculation amount is reduced.

(GFLOPs) decreased by 47.24%. The experimental results show that using Ghost module to lighten the original network model greatly reduces the modulus.

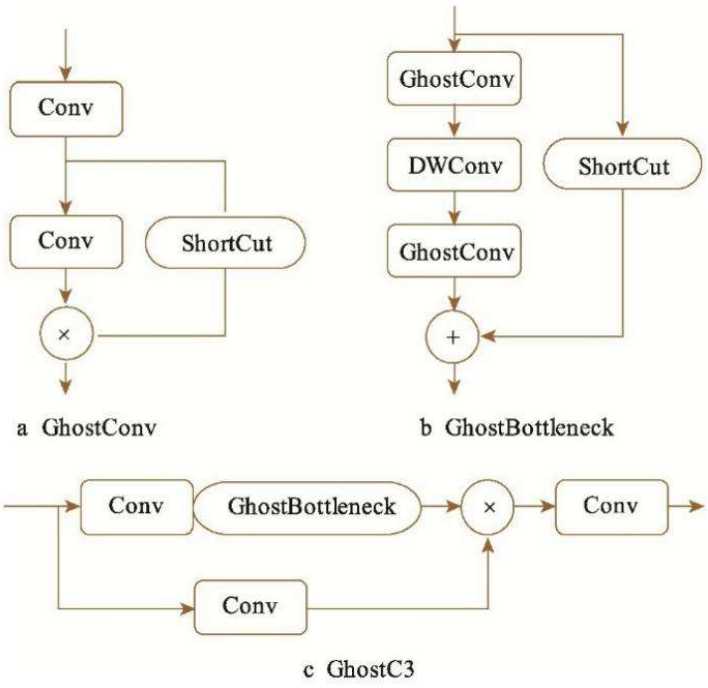

Fig.10 Ghost series of modules

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

Table 1. Comparison between the lightweight network and the original network data

network model	params	GFLOPs
YOLOv5s	7 132 903	16.3
YOLOv5s+Ghost	3 795 119	8.6

4. Experiment and Result Analysis

4.1 Cigarette Image Acquisition System

A cigarette image acquisition system was implemented at a tobacco logistics distribution center, capturing 14,310 images of cigarettes on a sorting line. These images covered 42 different cigarette types, with various poses and positions selected for each. A dataset was created by dividing the data into training, testing, and validation sets in a 4:1:1 ratio, with each image having a resolution of 1280×1024.

4.2 Experimental Environment and Parameters

The experiment was conducted on a Windows 11 Professional system. The deep learning framework used was PyTorch 1.9.1 with CUDA 12.0. The CPU was an Intel(R) Core(TM) i7-10700 @ 2.90 GHz, and the GPU was an NVIDIA GeForce RTX 3090. Python 3.8 was used for compilation.

4.3 Evaluation Metrics and Result Analysis

The performance of the cigarette recognition model was evaluated based on recognition accuracy and speed, with metrics such as GFLOPs, mAP@0.5, mAP@0.5:0.95, and single-frame detection time. Fig. 11 illustrates the comparison between YOLOv5s and the proposed model. While the original network can achieve high confidence predictions with good cigarette poses, it struggles with poor poses or obstructions, leading to inaccurate predictions. In contrast, the improved network maintains accurate detection even with challenging poses or blockages.

The improved network was compared with mainstream models, including YOLOv5-MobileNet, Faster-RCNN, and SSD. As shown in Table 2, the proposed model achieved an mAP@0.5 of 99.85% on the self-built dataset, with an average recognition speed of 12.8 ms per cigarette. However, Faster-RCNN and SSD models were not suitable for real-time applications due to their high computational cost. YOLOv5-MobileNet offered faster recognition but with lower accuracy. Therefore, the proposed model outperforms in terms of model size, detection speed, and accuracy, making it more suitable for high-speed cigarette sorting lines.

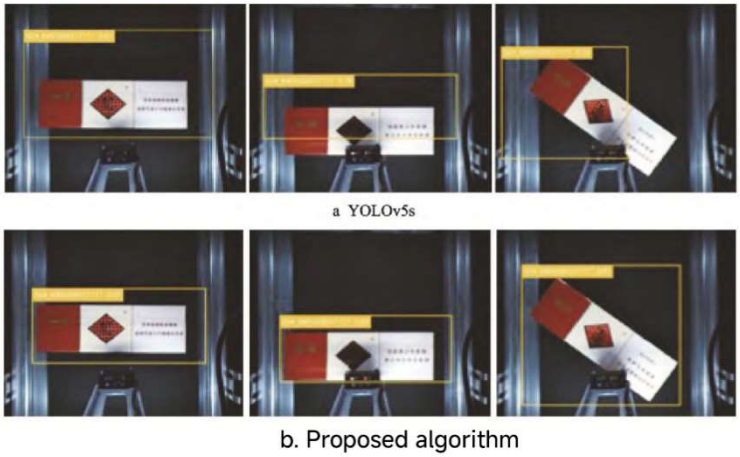

Fig.11 Identification comparison of strip tobacco

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

Table 2. Comparison of mainstream networks

network model	params	GFLOPs	mAP@0.5/ %	FPS/ (frame ·s−1)
YOLOv5s	7 132 903	16.3	94.2	91.7
YOLOv5-mobileNet	2 792 183	5.6	79.5	88.9
Faster-RCNN	601 736 218	523.9	87.3	12.7
SSD	411 830 644	388.2	85.2	45.4
Algorithm in this paper	3 954 823	9.1	99.3	79.3

4.4 Spot Test

To verify the algorithm's effectiveness, it was deployed in the error correction system of the sorting line at the tobacco logistics center. The sorting line operated at a speed of 1.5 m/s with a capacity of approximately 7,200 pieces per hour, and a real-time sorting speed of 13 strips per second. The test results for the algorithm on the sorting line are shown in Table 3. The field test confirmed the algorithm's performance in real-world conditions.

Table 3. Field measurement results of the tobacco distribution center

test	Sorting quantity/strip	Identification quantity/strip	Recognition rate/%	Average time consumption/(ms article 1)
Test 1 (Standard Smoke)	174 284	174 144	99.92	23.8
Test 2 (Standard Smoke)	156 802	156 708	99.94	24.2
Test 1 (special-shaped smoke)	14 620	14 598	99.85	27.1
Test 2 (special-shaped smoke)	13 578	13 567	99.92	26.5

5. Summary

Aiming at the problem of sorting the wrong cigarettes in the tobacco logistics distribution center, this paper puts forward an improved algorithm of cigarette identification based on Yolov5s, which takes into account both real-time and accuracy. To solve the problem of inaccurate positioning when the cigarette posture is poor, CA attention mechanism is introduced to pay attention to the extraction of location information while paying attention to feature information. For low-resolution cigarette images, the lightweight CARAFE operator is used to reduce the feature loss during upsampling without introducing too many parameters and calculation. Using Ghost module to build a lightweight network architecture, reducing the parameters and calculation of the model. Experiments show that the model parameters of the algorithm are reduced by 45.6%, and the calculation amount is reduced by 45.8%.mAP@0.5The value increased by 5.1%, and the sorting recognition rate reached 99.9%, which effectively solved the problem of wrong cigarettes in tobacco logistics center.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202504_11(4).0007

Acknowledgements

This work was financially supported by Guizhou Tobacco Company Guiyang Company Science and Technology Project (Research on Optimization of Sorting Scheduling and Delivery Service Based on Business Flow Data Driven (No. 2022-14)).

References

[1] CAO Yue. Research on the Key Technology of Photoelectric Automatic Identification and Classification of Strip Smoke[D]. Sichuan: University of Electronic Science and Technology of China, 2019: 1-4.

[2] QIU Tianheng, WANG Ling, WANG Peng, et al. Research on Object Detection Algorithm Based on Improved YOLOv5[J]. Computer Engineering and Applications, 2022, 58(13): 63-73.

[3] CAO Dongmei, GUO Zhuang, LI Dongbo. Research on the Classification and Location of Cigarette Based on Halcon[J]. Machine Design and Manufacturing Engineering, 2018, 47(9): 71-74.

[4] ZHOU Zhixiang, YANG Xudong, CHEN Bo, et al. Cigarette Recognition System Based on Template Matching[J]. Packaging Engineering, 2020, 41(21): 261-269.

[5] LI Mengxue. Research on Cigarette Classification and Recognition Algorithm of Transmission Platform Based on Vision[D]. Chengdu: University of Electronic Science and Technology of China, 2018.

[6] WANG Haoran. Research and Application of Error Detection System of Special Tobacco Sorting Line of W TOBACCO COMPANY[D]. Shandong: Shandong University of Finance and Economics, 2021.

[7] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate Attention for Efficient Mobile Network Design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[8] WANG Jiaqi, CHEN Kai, XU Rui, et al. CARAFE: Content-Aware Reassembly of Features[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea, 2019: 3007-3016.

[9] HAN Kai. GhostNet: More Features from Cheap Operations[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 56-61.

[10] REDMON Joseph, DIVVALA Santosh, GIRSHICK Ross, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 779-788.

[11] XU Degang, WANG Lu, LI Fan. Review of Typical Object Detection Algorithms for Deep Learning[J]. Computer Engineering and Applications, 2021, 57(8): 10-25.

[12] YANG Qisheng, LI Wenkuan, YANG Xiaofeng, et al. Improved YOLOv5 Method for Detecting Growth Status of Apple Flowers[J]. Computer Engineering and Applications, 2022, 58(4): 237-246.

[13] HU Jie, SHEN Li, ALBANIE Samuel, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[14] WOO Sanghyun, PARK Jongchan, LEE Joon-Young, et al. CBAM: Convolutional Block Attention Module[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.

[15] GIRSHICK Ross. Fast R-CNN[C]// Proceedings of IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448.

[16] LIU Wei, ANGUELOV Dragomir, ERHAN Dumitru, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision, 2016: 21-37.