LAD-YOLO: A Lightweight YOLOv5 Network for Surface Defect Detection on Aluminum Profiles

Dongxue Zhao, Shenbo Liu, Yuanhang Chen, Da Chen, Zhelun Hu, Lijun Tang
School of Physics and Electronic Science, Changsha University of Science & Technology, Changsha 410114, China

***Abstract*—In this paper, we leverage the advantages of YOLOv5 in target detection to propose a highly accurate and lightweight network, called LAD-YOLO, for surface defect detection on aluminum profiles. The LAD-YOLO addresses the issues of computational complexity, low precision, and a large number of model parameters encountered in YOLOv5 when applied to aluminum profiles defect detection. LAD-YOLO reduces the model parameters and computation while also decreasing the model size by utilizing the ShuffleNetV2 module and depthwise separable convolution in the backbone and neck networks, respectively. Meanwhile, a lightweight structure called "Ghost_SPPFCSPC_group", which combines Cross Stage Partial Network Connection Operation, Ghost Convolution, Group Convolution and Spatial Pyramid Pooling-Fast structure, is designed. This structure is incorporated into the backbone along with the Convolutional Block Attention Module (CBAM) to achieve lightweight. Simultaneously, it enhances the model's ability to extract features of weak and small targets and improves its capability to learn information at different scales. The experimental results show that the mean Average Precision (mAP) of LAD-YOLO on aluminum profiles defect datasets reaches 96.9%, model size is 6.64MB, and Giga Floating Point Operations (GFLOPs) is 5.5. Compared with YOLOv5, YOLOV5s-MobileNetv3, and other networks, LAD-YOLO proposed in this paper has higher accuracy, fewer parameters, and lower floating-point computation.**

***Keywords*—YOLOv5; ShuffleNetv2; lightweight and fast spatial pyramid pooling structure; convolutional block attention module; aluminum profiles surface defect detection**

I. INTRODUCTION

Aluminum profiles are one of the important raw materials for the manufacturing industry, widely used in industry, construction, medicine, and other industries. However, due to its complex production process and more transportation links, aluminum profiles are prone to surface defects such as scratch, dirt, pinhole, and wrinkle. These defects will directly affect the quality of aluminum profiles and even lead to distortion and deformation of aluminum profiles, which is more obvious for high-end aluminum profiles. Therefore, it is of great significance to improve the detection efficiency and accuracy of aluminum profiles surface defects to ensure the production and application of aluminum profiles.

At present, most enterprises still use traditional manual detection methods to detect defects on the surface of aluminum profiles. However, this manual inspection method is slow, subject to the influence of subjective consciousness, not only low efficiency but also poor stability, prone to misdetection, and leakage detection. Ultrasonic flaw detection, eddy current flaw detection, and other traditional non-destructive testing are also used for the detection of surface defects in aluminum profiles, but due to its slow detection speed, high cost, complex equipment operation, etc., which limits its popularity in practical applications. In 2014, Girshick et al. proposed a Regional Convolutional Neural Network (R-CNN), which broke the deadlock of slow progress in the field of target detection [1], and subsequently gave birth to Fast R-CNN [2], Faster R-CNN [3], Mask R-CNN [4], Single Shot MultiBox Detector (SSD) [5], You Only Look Once (YOLO) series [6-11], and other generalized deep learning-based target detection algorithms. As a result, deep learning-based surface defect detection is starting to develop rapidly.

For metal surface defect detection, references [12-14] combined neural networks with traditional detection algorithms to realize the detection and classification of surface defects of aluminum and other metal materials. Duan et al. [15] built a dual-stream Convolutional Neural Network (CNN) for the detection of aluminum profiles image features and gradient features, effectively realizing the classification of defect-free and multi-type defect samples. Cheng et al. [16] proposed a network DEA-RetinaNet with differential channel attention and adaptive spatial feature fusion for steel surface defect detection. The mean Average Precision (mAP) of the network on the steel surface defect dataset (NEU-DET) was 78.25%. The detection accuracy of the above methods is lower than 85%, which cannot meet the requirements of practical industrial applications.

Zeng [17] et al. proposed a data augmentation method and a migration learning technique for solving defective parts detection in steel plates. References [18-20] used Faster R-CNN to detect metal surface defects such as steel and railroad fasteners with an accuracy of more than 95%, but the detection speed is slow. Chen [21] et al. applied Convolutional Neural Networks (DCNNs) to the defect detection of fasteners and carried out experiments on high-speed railroad scenarios. ZHAO et al. [22] innovated based on YOLOv4 architecture to improve the detection accuracy of surface defects of metal materials. Wang et al. [23] proposed a structure called PE-Neck, which replaces the Neck part of the YOLOv5 network structure with a combination of scaled convolutional kernels and efficient channel attention to enhance the model's ability to extract and localize defects at different scales. However, the accuracy is only 87.4% and the strategy for generating candidate regions suffers from many flaws. Although the above methods enhance the detection accuracy by improved means, it is unable to realize the real-time detection of surface defects on

industrial aluminum profiles due to their complex network structure and large computation, and slow detection speed.

Conventional CNN inference is computationally intensive and difficult to apply in resource-constrained scenarios such as mobile and Internet of Things (IoT). Starting from SqueezeNet [24] and MobileNetV1 [25], the design of CNNs has begun to focus on efficiency in resource-constrained scenarios. The more mature lightweight networks include the MobileNet series [26-27], ShuffleNet series [28-29], GhostNet [30], etc. Li [31] et al. proposed a YOLOv3-Lite detection method, which combines a deep convolutional neural network and a feature pyramid in YOLOv3, to improve the defect detection accuracy. Xiao [32] et al. added a residual network structure to YOLOv3-Tiny, which was applied to detect obstacles in a mine, with improved accuracy compared to the original YOLOv3-Tiny, but with decreased speed. Zhang [33] et al. proposed a multi-model rail surface defect detection system based on a convolutional neural network (MRSDI-CNN). The system network uses SDD combined with YOLOv3 to improve the system's accuracy. Wang et al [34] proposed a lightweight YOLO-ACG detection algorithm that balances accuracy and speed while improving the defect detection classification error and leakage rate. Ma [35] et al improved the YOLOv4 network by replacing the backbone network with a lightweight Ghost module. At the same time, a joint attention mechanism is added to the stacked Ghost modules to ensure accuracy, so that the network is compressed and lightweight is achieved while achieving an accuracy of 94.68%. These methods have improved in detection accuracy and detection speed, but the model size is still large and memory consumption is high, which is not conducive to real-time detection on mobile, especially in devices with tight computing resources.

The YOLOv5 algorithm is an end-to-end target detection algorithm known for its fast detection speed and high accuracy. It has found wide application in the field of surface defect detection. However, the large number of parameters in the YOLOv5 model can hinder improvements in detection speed. Additionally, its backbone layer, consisting of CSPDarknet53, faces challenges in effectively extracting features of small targets. In this paper, a lightweight aluminum profiles surface defect detection network is designed to solve this problem, which significantly improves the accuracy and detection speed. The algorithm is evaluated for its performance on aluminum profiles surface defect dataset and compared with other algorithms. The experimental results show that the LAD-YOLO proposed in this study can accurately identify aluminum profiles surface defects with excellent detection speed.

The Section I is the research purpose and significance of aluminum profiles surface defects detection and the current status of domestic and international research on target detection algorithms for metal surface defects detection. The Section II is the research on the improvement method of lightweight aluminum profiles surface defect detection model based on YOLOv5. The Section III is the experimental results and analysis of the model application. The Section IVsummarizes the research in this paper and the outlook for future research.

II. METHODOLOGY

LAD-YOLO follows the network structure of YOLOv5, which consists of four main parts: Input Layer, Backbone, Neck, and Head. The overall structure of LAD-YOLO is depicted in Fig. 1. The Input Layer takes a 640*640*3 aluminum profiles defect image as input. The Backbone network contains six ShuffleNetv2 modules, three Convolutional Block Attention Modules (CBAM), and the Spatial Pyramid Pooling Cross Stage Partial Concat structure based on ghost convolution and group convolution (Ghost_SPPFCSPC_group) for extracting surface defect features of aluminum profiles. In the neck network, depthwise separable convolution is used to extract depth features of aluminum profiles surface defects, reducing computational overhead. Simultaneously, the feature image size is doubled by using the nearest neighbor interpolation upsampling method, and the feature maps with the same size in the aluminum profiles surface defect map are connected. In the prediction layer, three different sizes of detection heads are generated to detect the aluminum profiles' surface defect image.

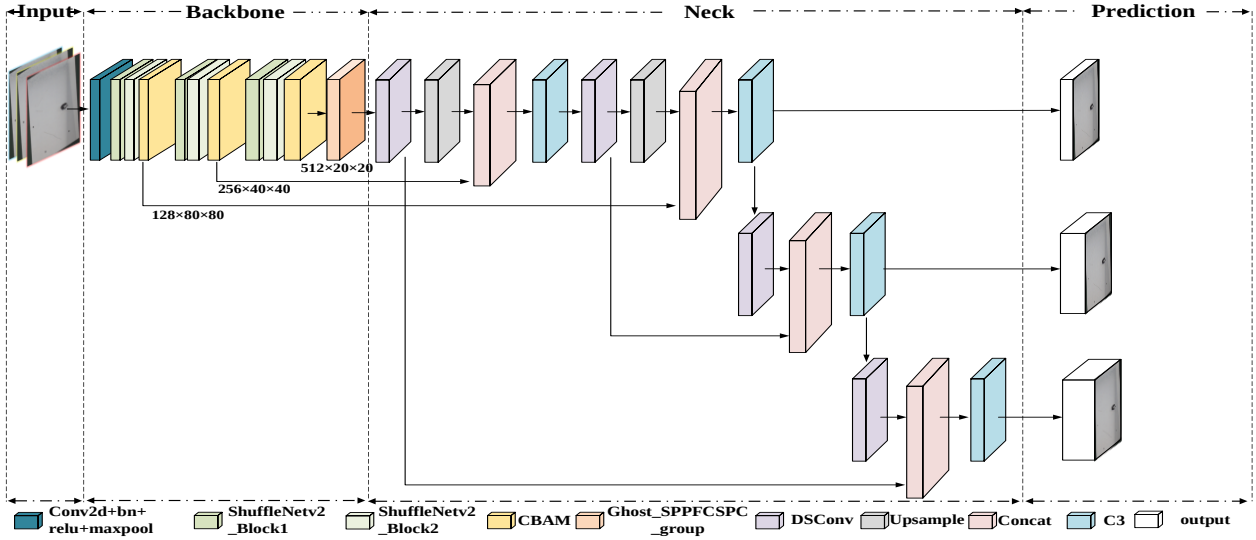

Fig. 1. LAD-YOLO network structure.

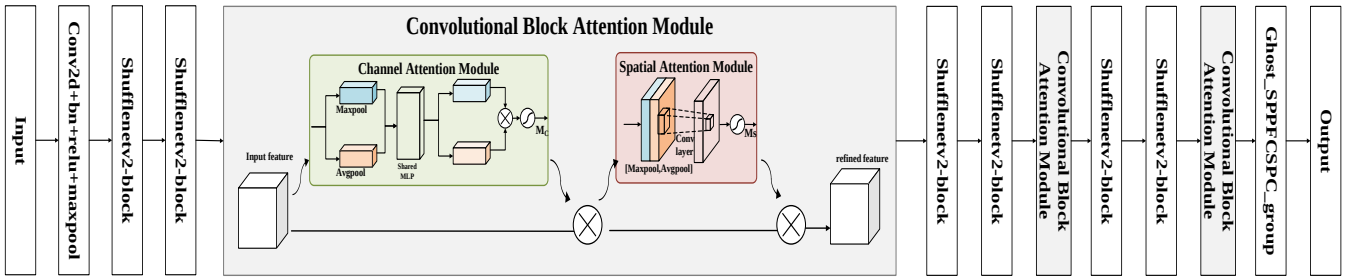

Fig. 2. CBAM and LAD-YOLO backbone layer network structure.

A. Backbone Structure

The Backbone Structure shown in Fig. 2 consists of three parts: ShuffleNetv2, Convolutional Block Attention Module (CBAM), and the Spatial Pyramid Pooling-Fast Cross Stage Partial Concat structure combining Ghost Convolution and Group Convolution (Ghost_SPPFCSPC_group). ShuffleNetv2 uses Channel Split, 1*1 convolution, depthwise separable convolution, and mixing and washing of channels to accomplish the detection of input aluminum profiles defect information, which reduces the memory access time, reduces the number of model parameters, and improves the detection speed. The adoption of the ShuffleNetv2 structure drastically reduces the number of parameters of the model, but also brings a certain loss of accuracy. To compensate for the loss of accuracy, CBAM is used to embed into the backbone, as shown in Fig. 2.

CBAM mainly consists of two key modules: the Channel Attention Module and the Spatial Attention Module. The Channel Attention Module captures the importance of each feature channel by calculating global statistics and applies attention weights to each channel. The Spatial Attention Module highlights important spatial regions in the feature map by computing global statistics and applying attention weights to different spatial positions. Combining the above two modules, CBAM enables the network to adaptively focus on significant channels and spatial areas, improving feature representation for aluminum profiles surface defect detection tasks.

B. Ghost_SPPFCSPC_Group Structure

The Spatial Pyramid Pooling (SPP) structure can effectively capture target features at different scales by stacking pyramid layers of different sizes together, improving the model's detection ability for targets of different sizes. The Spatial Pyramid Pooling-Fast (SPPF) structure is a faster structure proposed based on the SPP structure. The Cross Stage Partial structure consists of two parts, the convolution, and the complex structure, in parallel to increase the speed of the network.

In 2023, wang et al. [11] first proposed the Spatial Pyramid Pooling Cross Stage Partial Concat (SPPCSPC) structure in YOLOv7, as shown in Fig. 3, which uses the SPP and CSP modules for better handling of multi-scale targets.

The use of the SPPCSPC structure can effectively improve the model detection accuracy, but it will increase the amount of computation and the number of model parameters. To improve the speed, this paper replaces the SPP in the SPPCSPC structure with the SPPF structure to obtain the Spatial Pyramid Pooling-Fast Cross Stage Partial Concat (SPPFCSPC) structure. To reduce the amount of computation and parameters, Ghost convolution and group convolution are used to replace standard convolution in the SPPFCSPC structure.

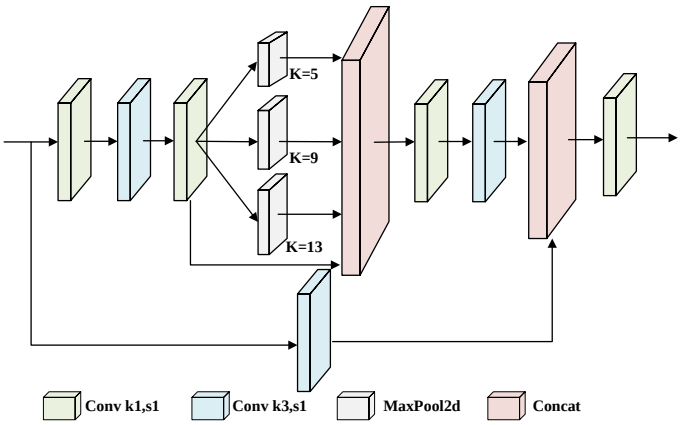

Fig. 3. Spatial pyramid pooling cross stage partial concat structure.

The standard convolution, Ghost convolution, and group convolution are compared and analyzed below. Fig. 4 shows the operation process of standard convolution, Eq. (1) is the standard convolution parameters, where c is the number of input channels, n is the number of output channels, and the size of the convolution kernel is $k*k$.

$$P_{std} = c \cdot n \cdot k \cdot k \quad (1)$$

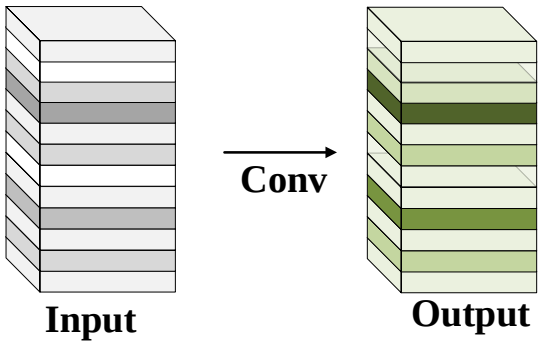

Fig. 4. Standard convolution operation.

Fig. 5 shows the group convolution operation process, which divides the input channels and output channels into the same number of groups, and then allows the input channels and output channels in the same group number to be fully

connected. Eq. (2) is the parameters of the group convolution. Where *g* is the number of groups divided into output channels.

$$P_{Group} = \frac{c}{g} \cdot \frac{n}{g} \cdot g \cdot k \cdot k \quad (2)$$

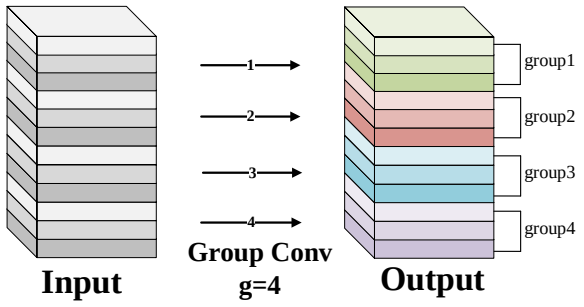

Fig. 5. Group convolution operation.

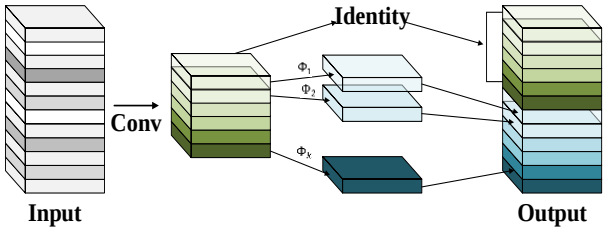

Fig. 6. Ghost convolution operation.

Fig. 6 shows the Ghost convolution operation process. First, a standard convolution operation with *m* (*m*≤*n*) output channels (where *n* is the number of final output channels) is performed on the feature map with input channel *c* to obtain a feature map with m channels. Second, a new feature map is obtained by *s-1* linear operations. Finally, the two feature maps are connected to obtain an output feature map with *n* channels (*n*=*m***s*). Eq. (3) is the parameters of Ghost convolution. Where the convolution kernel size of linear operations in Ghost Module is *d*×*d*.

$$P_{Ghost} = \frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d \quad (3)$$

By comparing the above three convolution parameters, Eq. (4) shows the parameters compression ratio (R_1) for standard and Ghost convolution, and Eq. (5) shows the parameters compression ratio (R_2) for standard and group convolution. Where let *k*=*d*.

$$R_1 = \frac{n \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s + c - 1} \approx s \quad (4)$$

$$R_2 = \frac{n \cdot c \cdot k \cdot k}{\frac{c}{g} \cdot \frac{n}{g} \cdot g \cdot k \cdot k} = g \quad (5)$$

From the above computation results, it is shown that the number of parameters of standard convolution is g times more than that of group convolution and s times more than that of Ghost convolution, so the number of parameters can be drastically reduced by choosing group convolution and Ghost convolution compared to standard convolution.

Therefore, in this paper, a lightweight SPPFCSPC structure (Ghost_SPPFCSPC_group) is designed by combining SPPFCSPC, Ghost convolution, and group convolution, as shown in Fig. 7. The structure utilizes the smaller number of parameters of Ghost convolution and group convolution to achieve lightweight. The Ghost_SPPFCSPC_group structure uses a smaller computational cost to enable the fusion of multi-scale features and improve feature representation.

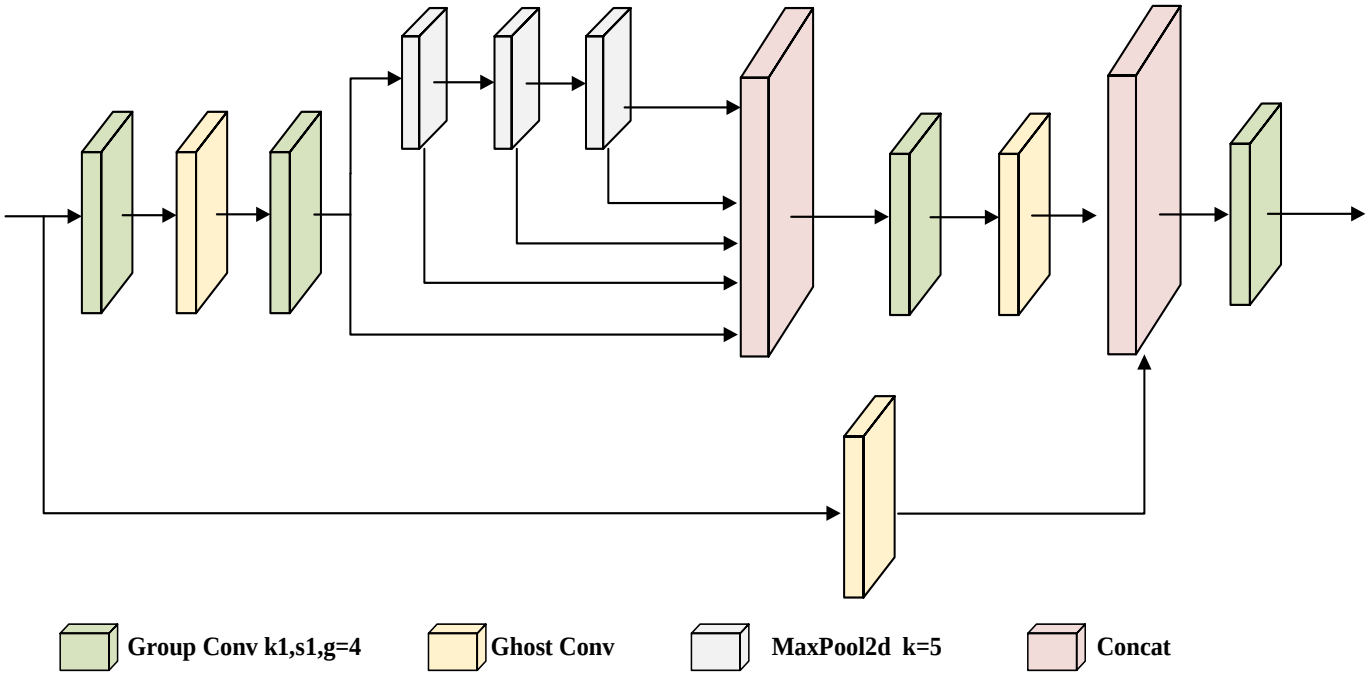

Fig. 7. Ghost_SPPFCSPC_group structure.

C. Depthwise Separable Convolution

Depthwise Separable Convolution (DSConv) contains two parts, Depthwise Convolution and Pointwise Convolution. As shown in Fig. 8, Depthwise Convolution computes the convolution of each channel separately to extract the features of each channel; Pointwise Convolution computes the feature map generated by Depthwise Convolution and adopts a convolution kernel with the size of 1×1×M convolution kernel, weighted combination in the depth direction, to realize the fusion of features between the channels, to generate a new feature map.

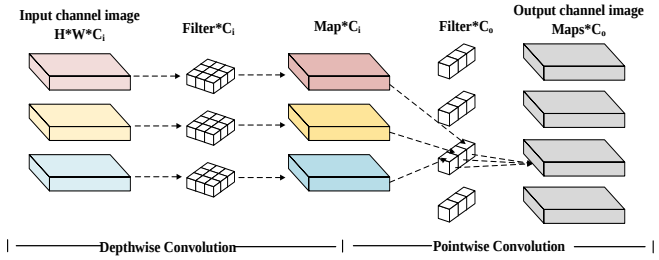

Fig. 8. Depthwise separable convolution module structure.

Eq. (6) is the FLOPs for standard convolution. Eq. (7) is the FLOPs Operations for Depthwise Separable Convolution. Eq. (8) is the ratio R of Depthwise Separable Convolution to the standard convolutional computation. The input feature map size is $H \times W \times C_i$, the output feature map size is $H \times W \times C_o$, H, W, C_i and C_o denote the height, width, number of input channels, and number of output channels of the feature map respectively, and the size of the convolution kernel in the standard convolution is $K_1 \times K_2$.

$$C_{std} = K_1 \cdot K_2 \cdot H \cdot W \cdot C_i \tag{6}$$

$$\begin{aligned} C_{\text{separable}} &= C_{depthwise} + C_{\text{pointwise}} \\ &= K_1 \cdot K_2 \cdot H \cdot W \cdot C_i + H \cdot W \cdot C_i \cdot C_o \end{aligned} \tag{7}$$

$$\begin{aligned} R = \frac{C_{\text{separable}}}{C_{std}} &= \frac{K_1 \cdot K_2 \cdot H \cdot W \cdot C_i + H \cdot W \cdot C_i \cdot C_o}{K_1 \cdot K_2 \cdot H \cdot W \cdot C_i \cdot C_o} \\ &= \frac{1}{C_o} + \frac{1}{K_1 \cdot K_2} \end{aligned} \tag{8}$$

From Eq. (8), it can be seen that the floating-point computation of Depthwise Separable Convolution is only $\frac{1}{C_o} + \frac{1}{K_1 \times K_2}$ of the standard convolution. Assuming the convolution kernel size of 3×3 for Depthwise Convolution, the computation of the standard convolution is about eight to nine times that of Depthwise Separable Convolution. Replacing the standard convolution with the Depthwise Separable Convolution reduces the floating-point computation.

III. Results And Discussion

A. Datasets Introduction

In this paper, for aluminum profiles defect detection, the Hikvision high-definition industrial camera model MV-CS050-10GC-PRO is used to collect the sample images of aluminum profiles and make the aluminum profiles defect datasets, and some of the data in the datasets are shown in Fig. 9.

The labeling categories are pinhole, scratch, dirt, and wrinkle. Since the original image samples are too few, panning, rotating, changing brightness, shearing, mirroring, and other means of expanding the datasets are chosen to expand the original aluminum profiles surface defect datasets. After the expansion, the total number of defect images is 5013, including 6325 pinholes, 3042 dirt, 5863 scratches, and 2415 wrinkles. The datasets are categorized into 60% training set, 20% validation set, and 20% test set containing 3008, 1002, and 1003 images, respectively.

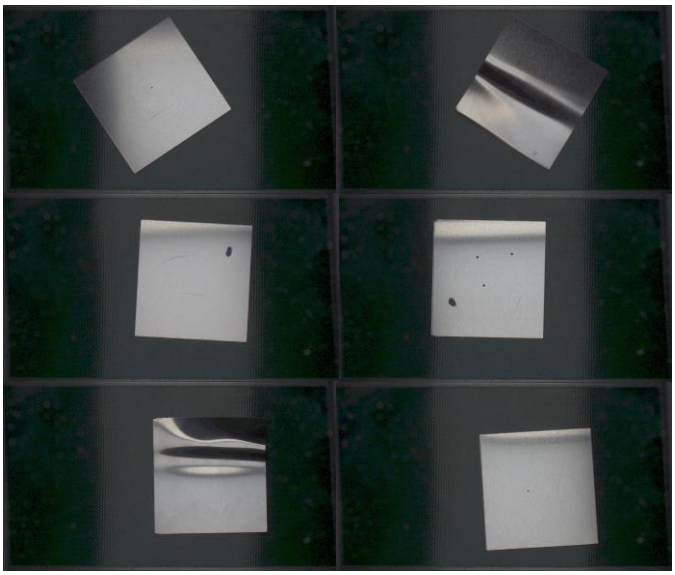

Fig. 9. Part of the datasets.

B. Evaluation Metrics

Precision (P), Recall (R), Average Precision (AP), and Mean Average Precision (mAP) are used as the evaluation metrics of detection effectiveness. The mAP is the average value of AP for all defect categories, which is used as a comprehensive index for evaluating precision. The higher the values of AP and mAP, the better the algorithm is for detecting the target defects. P, R, AP, and mAP are calculated as follows: (9), (10), (11), (12). *TP* is the number of defects in the positive samples that were detected as correct, *FP* is the number of defects in the negative samples that were incorrectly detected as correct, *FN* is the number of defects in the positive samples that were not detected, and *m* is the number of defect categories. In addition, Floating Point Operations (FLOPs), parameters, and Model Size are used to evaluate the lightness of the model.

$$P = \frac{TP}{TP+FP} \tag{9}$$

$$R = \frac{TP}{TP+FN} \tag{10}$$

$$AP_i = \int_0^1 PR\mathrm{d}r \tag{11}$$

$$mAP = \frac{1}{m}\sum_{i=1}^{m} AP_i \tag{12}$$

C. Experimental Process

The GPU used for defect detection training and testing is NVIDIA TITAN RTX, and the specific configuration of the experimental platform is shown in Table I. During the training

experiments, the optimizer chooses the stochastic gradient descent with momentum (SGD) with a momentum factor of 0.937. The weight attenuation coefficient is set to 5×10−4. The learning rate is initially set to 10−3, while the Cosine Annealing is used to reduce the learning rate to 10−5. The batch size is set to 64, and the epochs are set to 500.

TABLE I. THE SPECIFIC CONFIGURATIONS OF THE EXPERIMENTAL PLATFORM

Name	Version
CPU	Intel(R) Xeon(R) Gold 5218R CPU @ 2.10GHz
Memory Bank	32GB
GPU	NVIDIA TITAN RTX
GPU Memory	24GB
Operating System	Windows10
Software environment	CUDA11.6
Python Version	Python 3.8
Deep learning framework	PyTorch 1.12

D. Comparative Experiment and Analysis

Comparison experiments are conducted by LAD-YOLO with SSD, YOLOv3, YOLOv3-tiny, YOLOv4-tiny, YOLOv5, and YOLOv5s-MobileNetv3, and the results are shown in Table II. As can be seen in Table II, LAD-YOLO achieved 96.9% mAP, 97.4% Precision, and 95.7% Recall on the aluminum profiles surface defects dataset, and the model size is 6.64MB. Compared with the YOLOv5s algorithm, mAP increases by 2.8% and model size decreases by 58%. Compared with YOLOv5s-MobileNetv3, the precision is improved by 7.2% and the model size is reduced by 8.8%. The experimental results show that the LAD-YOLO network improves the precision and recall rate of defect detection, and reduces the model parameters and size.

E. Ablation Experiment

To further verify the role of each improvement in enhancing the performance of the algorithm, ablation experiments are conducted. The results are shown in Table III.

From Table III, the mAP of the baseline model YOLOv5s is 94.1%, the model size is 14,08MB, the number of parameters is 7.03×106 and the GFLOPs is 15.8. It can be seen that after using ShuffleNetV2 and CBAM, the mAP is improved by 1.7% compared to YOLOv5s, the model size is reduced from 14.08MB to 6.69MB, and the GFLOPs are reduced from 15.8 to 5.8; after using the Ghost_SPPFCSPC_group structure, the mAP is again improved by 1.5%, with a slight increase in Model Size and GFLOPs; with the use of deep separable convolution, the mAP is 96.9%, Model Size is again reduced to 6.64M, and GFLOPs are 5.5. compared to the original network. mAP is improved by 2.8%, Model Size is reduced by 52.8%, and GFLOPs are reduced by 65.6%. The results show that improvements to YOLOv5 are necessary everywhere.

F. Test Results of Defect Detection

Fig. 10 shows the schematic diagram of four kinds of defect detection in aluminum profiles. The non-maximum suppression (NMS) is used in the prediction as the post-processing method. The confidence was set to 0.5 and the IoU was set to 0.6.

TABLE II. COMPARATIVE RESULTS OF EVALUATION METRICS FOR DIFFERENT METHODOLOGIES

Methods	Precision (%)	Recall (%)	mAP (%)	Model Size (MB)
SSD	68.4	70.2	70.6	90.13
YOLOv3	91.9	87.7	91.2	120.67
YOLOv3-tiny	87.6	84.7	86.9	33.79
YOLOv4-tiny	90.9	89.6	90.8	23.03
YOLOv5s	95.3	94.0	94.1	14.07
YOLOv5s-MobileNetv3	89.9	88.4	89.7	7.28
LAD-YOLO（OURS）	97.0	95.7	96.9	6.64

TABLE III. RESULTS OF THE ABLATION EXPERIMENT

ShuffleNetV2	CBAM	SPPFCSPC	Ghost_SPPFCSPC_group	DSConv	mAP	Parameters (10^6)	Model Size（MB）	GFLOPs
--	--	--	--	--	94.1%	7.03	14.08	15.8
√					92.2%	3.23	6.66	5.8
√	√				95.8%	3.25	6.69	5.9
√	√	√			97.3%	9.77	19.46	11.1
√	√		√		97.0%	3.85	7.91	6.3
√	√		√	√	96.9%	3.19	6.64	5.5

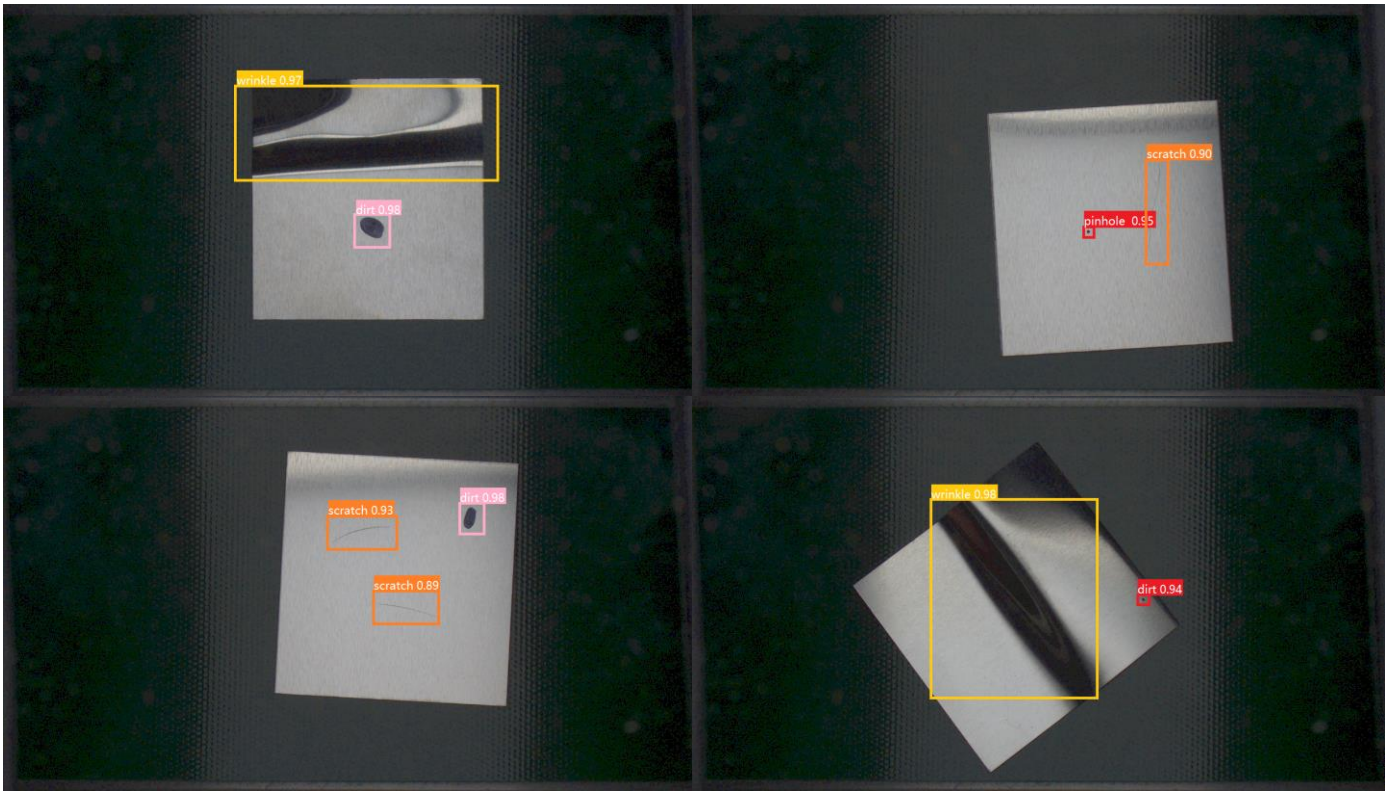

Fig. 10. Partial detection images for aluminum profiles defect detection.

The obtained LAD-YOLO P-R curve for defect detection is shown in Fig. 11, and it can be seen that the AP of pinholes is 98.8%, the AP of dirt is 99.3%, the AP of scratches is 91.1%, and the AP of wrinkles is 98.2%. Except for scratches, all other types of defects have an AP of 98% or more. The poor detection of scratches is due to its high defect precision requirements and susceptibility to environmental influences.

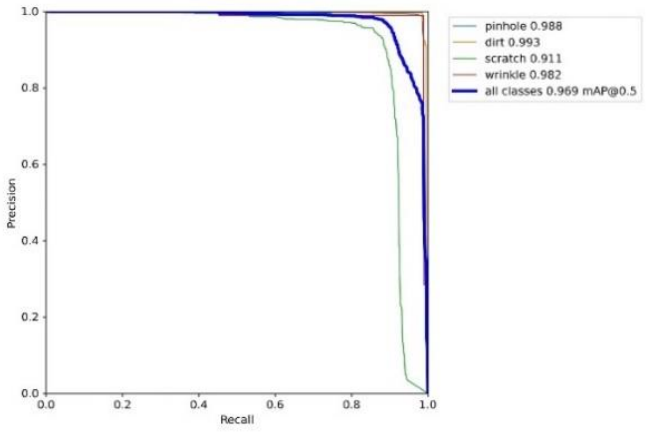

Fig. 11. LAD-YOLO P-R curve.

The accuracy of detecting various defects under different methods is plotted in Fig. 12. It is evident from the figure that LAD-YOLO exhibits improvements in accuracy for different defect types compared to other methods. Specifically, LAD-YOLO shows an increase in accuracy of 4.6% for pinhole detection, 2.3% for dirt detection, 3.2% for scratch detection, and 0.9% for wrinkle detection when compared to YOLOv5s.

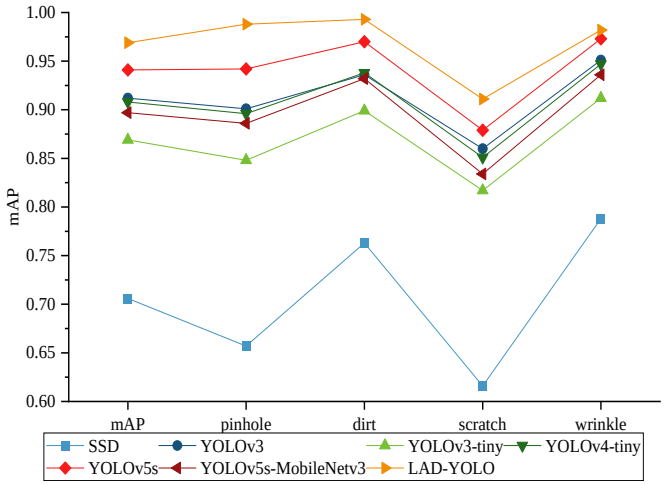

Fig. 12. Accuracy of defects in each category under different methods.

The results demonstrate that LAD-YOLO achieves enhanced accuracy across all defect categories, with particularly notable improvements in the detection of small targets, such as pinholes and scratches.

IV. Conclusion

We propose a lightweight aluminum profiles surface defect detection network, which involves improvements to the backbone and neck layers of YOLOv5. By designing the Ghost_SPPFCSPC_group structure with low floating-point

operation and combining it with the ShuffleNetV2 module and the Convolutional Block Attention Module (CBAM) to construct the backbone network, we reduce the model parameters and computation amount while obtaining richer feature information, thus improving the network's ability to detect defects on small-sized targets. By using depthwise separable convolution to replace the standard convolution in the neck layer, the number and size of model parameters are further reduced to improve the network operation speed. The specific experimental results are as follows:

(1) The Model Size of LAD-YOLO is only 6.64MB, which is 52.84% less compared with YOLOv5s; its GFLOPs are only 5.5, which is 65.19% less compared with YOLOv5s. It shows that LAD-YOLO occupies fewer memory resources, which is more helpful to be applied to platforms with scarce computational resources to achieve low-cost aluminum profiles surface defect detection.

(2) The detection accuracy of LAD-YOLO is much higher than that of current detection methods, including SSD, YOLOv3, YOLOv3-tiny, YOLOv4-tiny, YOLOv5s, and YOLOv5s-MobileNetv3, etc. Compared with YOLOv5s, the mAP of LAD-YOLO is 96.9%, an improvement of 2.8%; compared with YOLOv5s-MobileNetv3, the accuracy is improved by 7.2%. The results indicate that the LAD-YOLO network not only achieves model lightweight but also shows an improvement in accuracy.

In the forthcoming phases, we intend to augment the variety of defects within our dataset, thereby enhancing its diversity. Furthermore, to bolster the model's resilience and versatility in real-world scenarios, we will acquire images portraying authentic situations characterized by uneven lighting, occlusions, and other intricacies. This approach aims to further elevate the model's overall performance.

ACKNOWLEDGMENT

This research was funded by the Postgraduate Scientific Research Innovation Project of Changsha University of Science & Technology, Grant Number CXCLY2022141, the Open Research Fund of Hunan Provincial Key Laboratory of Flexible Electronic Materials Genome Engineering, Grant Number 202019, the Open Research Fund of the Hunan Province Higher Education Key Laboratory of Modeling and Monitoring on the Near-Earth Electromagnetic Environments, Grant Number N202107, and the Postgraduate Scientific Research Innovation Project of Hunan Province, Grant Number CX20200896.

REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell, J. Malik; "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[2] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.

[5] Wei, Liu. "SSD: Single shot multibox detector In Computer vision ECCV2016 14th European conference proceedings Part I (eds Leibe, B., Matas, J., Sebe, N. & Welling, M). 21–37." (2016).

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2016, pp. 779-788.

[7] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263-7271.

[8] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[9] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[10] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," arXiv preprint arXiv:2107.08430, 2021.

[11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464-7475.

[12] Li Q H, Liu D. Aluminum plate surface defects classification based on the BP neural network [J].Applied Mechanics and Materials, 2015.734:543-547.

[13] Ferguson M K, Ronay A K, Lee Y T, et al. Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning [J]. Smart and sustainable manufacturing systems. 2018, 2.

[14] Song L, Lin W,Yang Y, et al. Weak micro-scratch detection based on deep convolutional neural network[J].IEEE Access.2019,7: 27547-27554.

[15] Duan C, Zhang T. Two-stream convolutional neural network based on gradient image for aluminum profile surface defects classification and recognition[J]. IEEE Access,2020, 8:172152.172165.

[16] X. Cheng and J. Yu, "RetinaNet With Difference Channel Attention and Adaptively Spatial Feature Fusion for Steel Surface Defect Detection," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021.

[17] W. Zeng, Z. You, M. Huang, Z. Kong, Y. Yu and X. Le, "Steel sheet defect detection based on deep learning method", Proc. 10th Int. Conf. Intell. Control Inf. Process. (ICICIP), pp. 152-157, Dec. 2019.

[18] X. Jin et al., "DM-RIS: Deep multimodel rail inspection system with improved MRF-GMM and CNN", IEEE Trans. Instrum. Meas., vol. 69, no. 4, pp. 1051-1065, Apr. 2020.

[19] X. Chen and H. Zhang, "Rail Surface Defects Detection Based on Faster R-CNN," 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), Tianjin, China, 2020, pp. 819-822.

[20] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, Y. Li. Railway track fastener defect detection based on image processing and deep learning techniques: a comparative study.Eng. Appl. Artif. Intell., 80 (2019), pp. 66-81.

[21] J. Chen, Z. Liu, H. Wang, A. Núñez, Z. Han Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network.IEEE Trans. Instrum. Meas., 67 (2) (2018), pp. 257-269.

[22] Zhao H L, Yang Z F, LI J. Detection of metal surface defects based on YOLOV4 algorithm [J]. Journal of Physics: Conference Series, 2021, 1907 (1):12-43.

[23] Wang, T.; Su, J.; Xu, C.; Zhang, Y. An Intelligent Method for Detecting Surface Defects in Aluminium Profiles Based on the Improved YOLOv5 Algorithm. Electronics 2022, 11, 2304.

[24] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[25] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv: 1704.04861, 2017.

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2018, pp. 4510-4520.

[27] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, et al., "Searching for mobilenetv3," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1314-1324.

[28] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2018, pp. 6848-6856.

[29] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in Proceedings of the European Conference on computer vision (ECCV), 2018, pp. 116-131.

[30] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 1580-1589.

[31] Y. Li, Z. Han, H. Xu, L. Liu, X. Li, K. Zhang.Yolov3-lite: a lightweight crack detection network for aircraft structure based on depthwise separable convolutions Appl. Sci.-Basel, 9 (18) (2019)

[32] D. Xiao, F. Shan, Z. Li, B. T. Le, X. Liu, and X. Li, "A Target Detection Model Based on Improved Tiny-Yolov3 Under the Environment of Mining Truck," in IEEE Access, vol. 7, pp. 123757-123764, 2019.

[33] H. Zhang, Y. Song, H. Zhong, L. Liu, et al., "MRSDI-CNN: Multi-Model Rail Surface Defect Inspection System Based on Convolutional Neural Networks," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 11162-11177, Aug. 2022.

[34] C. Wang, M. Sun, Y. Cao, K. He, et al. Lightweight Network-Based Surface Defect Detection Method for Steel Plates[J]. Sustainability, 2023, 15(4): 3733.

[35] Z. Ma, Y. Li, M. Huang, Q. Huang, J. Cheng, et al. Automated real-time detection of surface defects in manufacturing processes of aluminum alloy strip using a lightweight network architecture. J Intell Manuf 34, 2431–2447 (2023).