

Received 11 June 2023, accepted 27 June 2023, date of publication 3 July 2023, date of current version 18 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3291598

## RESEARCH ARTICLE

# MA-YOLO: A Method for Detecting Surface Defects of Aluminum Profiles With Attention Guidance

LINGJIE JIANG<sup>1,2,3</sup>, BAOXI YUAN<sup>1,2,3</sup>, YUQIAN WANG<sup>4</sup>, YAHONG MA<sup>1,2,3</sup>, JIAWEI DU<sup>1,2,3</sup>, FENG WANG<sup>1,2,3</sup>, AND JIANXIN GUO<sup>1,2,3</sup><sup>1</sup>School of Electronic Information, Xijing University, Xi'an, Shaanxi 710123, China<sup>2</sup>Shaanxi Key Laboratory of Integrated and Intelligent Navigation, Xi'an, Shaanxi 710068, China<sup>3</sup>Xi'an Key Laboratory of High Precision Industrial Intelligent Vision Measurement Technology, Xijing University, Xi'an, Shaanxi 710123, China<sup>4</sup>Graduate Office, Xijing University, Xi'an, Shaanxi 710123, China<sup>5</sup>School of Computer Science, Xijing University, Xi'an, Shaanxi 710123, China

Corresponding author: Baoxi Yuan (yxbupt@163.com)

This work was supported in part by the Foundation of Shaanxi Key Laboratory of Integrated and Intelligent Navigation under Grant SKLIIN-20190102; in part by the Natural Science Foundation of Shaanxi Province under Grant 2021JM-537, Grant 2019JQ-936, and Grant 2021GY-341; and in part by the Research Foundation for Talented Scholars of Xijing University under Grant XJ20B01, Grant XJ19B01, and Grant XJ17B06.

**ABSTRACT** Aluminum Profiles (APs) are aluminum materials obtained by hot melting and extruding aluminum rods. It has the characteristics of low cost, strong plasticity, easy processing, and recyclability, and therefore plays an important role in industrial production. However, defects such as Non-Conductive (NC), Scratch, Orange Peel (OP), and Dirty Point (DP) often occur during the production and processing of APs, which can seriously affect the quality of APs. In addition, surface defects of APs also have problems such as fuzzy regional definition, large-scale variation, imbalance of aspect ratio, and high inter-class defect similarity, making defect detection more challenging. To solve these problems, this paper proposes an attention-guided object detection algorithm called MA-YOLO, specifically for Surface Defect Detection (SDD) of APs. The algorithm is based on YOLOv5s. Firstly, the K-Means++ clustering algorithm is used to optimize the anchor boxes, which alleviates the problem of aspect ratio imbalance. Secondly, by improving the multi-scale Feature Fusion Network (FFN), the detection performance of the model to detect the defects with unbalanced aspect ratio is improved, and the adaptability of the model to defects of different scales is enhanced. Finally, a novel Max Pooling Average Pooling (MA) attention module is proposed to improve the overall detection performance of the model, especially for small-scale defects. Experimental results on the aluminum profile surface defect dataset show that MA-YOLO has better detection performance and superiority than the current mainstream object detection algorithms, and compared with the baseline YOLOv5s, the mAP<sub>50</sub> and F1 score are increased by 2.9% and 2.2%, respectively, while keeping the model lightweight. This indicates that MA-YOLO has broad application prospects in the surface defect detection of APs.

**INDEX TERMS** Attention mechanism, computer vision, aluminum profiles, surface defect detection, YOLOv5, K-Means++.

**NOMENCLATURE** $\omega_i$  Different learnable weights. $\varepsilon$  Non-zero factor.

C Dimension of the feature map.

H

Height of the feature map.

W

Width of the feature map.

AL\_Dataset

Aluminum profile surface defect dataset.

APs

Aluminum profiles.

BiFPN

Bidirectional feature pyramid network.

CBAM

Convolutional block attention module.

IoU

Intersection over union.

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar<sup>1</sup>.

MA	Max pooling average pooling.
MLP	Multilayer perceptron.
MV	Machine vision.
NC	Non-conductive.
NMS	Non-maximum suppression.
OP	Orange peel.
PANet	Path aggregation network.
PB	Paint bubble.
PC	Parti-colour.
PCT	Principal component thermography.
CL	Corner leak.
CNN	Convolutional neural networks.
DL	Deep learning.
DP	Dirty point.
Faster R-CNN	Faster region-based convolutional neural network.
FFN	Feature fusion network.
FPN	Feature pyramid network.
GOCM	Gradient-only co-occurrence matrix.
PPT	Pulsed phase thermography.
R-CNN	Region-based convolutional neural network.
RPN	Region proposal network.
SDD	Surface defect detection.
SE	Squeeze and excitation.
SSD	Single shot multibox detector.
TSR	Thermal signal reconstruction.
YOLO	You only look once.

## I. INTRODUCTION

Aluminum Profiles (APs) are an important raw material in industrial production, with good corrosion resistance, thermal conductivity, plasticity, and recyclability. They are widely used in high-end manufacturing industries such as automobile manufacturing, equipment manufacturing, and rail transportation [1], [2]. Furthermore, the low cost and processability characteristics of APs make them a preferred alternative to expensive raw materials such as copper. In recent years, with the continuous development and upgrading of the industrial manufacturing industry, there have been higher requirements for the overall quality of APs [3], [4]. However, improper human operations, uneven production equipment, and low-quality raw materials in the production process of APs can lead to various types of defects on the surface of APs, which affect aluminum profiles' overall quality and cause economic losses [5]. Moreover, some APs with severe defects can pose serious hidden dangers to the performance and quality of the product. The detection of surface defects is a crucial step in guaranteeing the quality of industrial products [6]. Therefore, to improve the overall quality of APs and meet the production requirements of the modern manufacturing industry, a highly efficient detection method needs to be designed to achieve precise detection of surface defects on APs.

The development of Surface Defect Detection (SDD) technology is closely related to the progress of science and technology. It can be divided into three stages of overall history: the manual detection stage, the machine device detection stage, and the Machine Vision (MV) detection stage [7]. In areas where production is not highly automated, most factories still rely on traditional manual detection methods for detecting metal surface defects. However, manual detection methods are affected by subjective factors, missed detection and false detection occur frequently, and the learning and training costs of manual labor are huge. Along with the popularity of automated production, manual detection methods are gradually being eliminated [8]. Machine device detection methods mainly use signal collection devices to collect specific optical, electrical, or magnetic signals or ultrasound to detect defects on metal surfaces [9]. Guo et al [10]. using eddy current testing on Inconel 738LC alloy, derived the relationship between the influencing factors such as excitation frequency, lift-off distance, defect depth and size, residual heat, surface roughness, and the defect EC signal. D'Accardi et al [11] carried out a study on the pulsed thermography technique, comparing the performance of Pulsed Phase Thermography (PPT), Thermal Signal Reconstruction (TSR), and Principal Component Thermography (PCT) for the detection of surface defects in APs, and systematically listing the advantages, disadvantages, and sensitivity of various thermographic algorithms. Lou et al. [12] proposed a non-destructive detection method based on low-frequency (20-50HZ) electromagnetic technology to detect internal defects in steel plates. The results show that the proposed method can detect internal defects buried to a depth of 6mm in a 12-mm thick 20# steel plate or pipeline, achieving the best results compared with the method of the same period. Although, these methods have been proposed to solve the problem of metal SDD to a certain extent. However, machine device-based detection methods suffer from high detection costs, high learning costs, and limited detection accuracy, which are not conducive to large-scale use in industrial production.

Currently, with the rapid development of MV technology, SDD methods based on MV are beginning to be widely used in various industrial sectors, including automotive parts [13], solar panels [14], printed circuit boards [15], electronic displays [16], steel and APs [17]. In the traditional sense, MV detection methods first acquire defect images with an industrial camera and preprocess the images, then manually design feature extraction methods based on specific defect features, and finally, perform data dimensionality reduction on the feature information and input it into a classifier for classification. Compared to manual detection methods and machine detection methods, MV detection methods offer advantages such as reliability, convenience, and efficiency. Wang et al. [18]. proposed a guiding template-based SDD method for steel strips. The core idea is to detect surface defects on steel strips through a template-matching method. Experimental results show that a detection accuracy of 96.2%

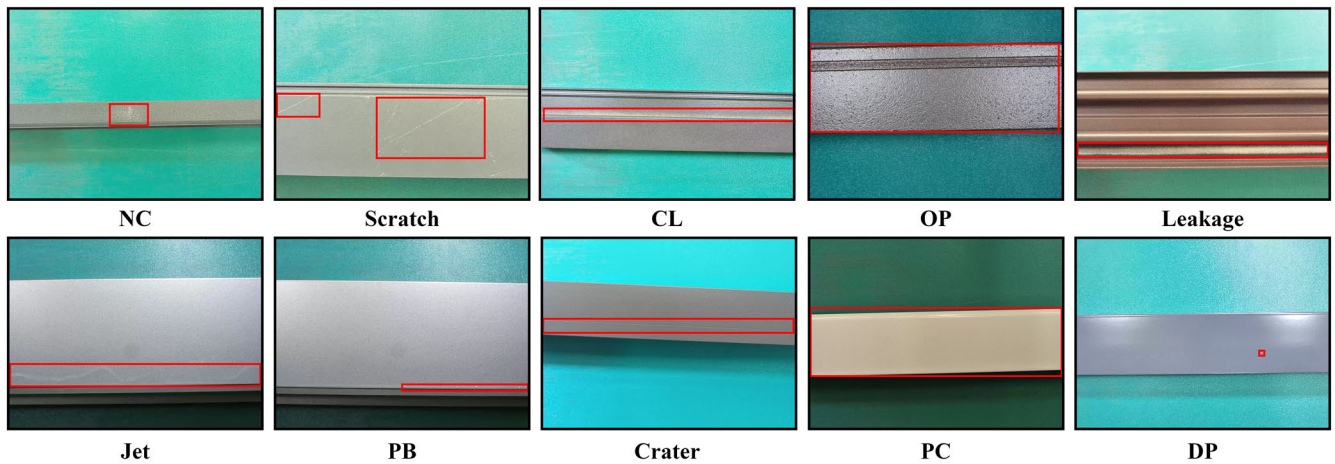
was achieved on 1500 test images, but the implementation process is complex, and the detection speed needs to be improved. Shi et al. [19] improved the Sobel edge detection operator by adding 45°, 135°, 180°, 225°, 270°, and 315° orientation templates to enrich the object edge information and thus improve the detection accuracy of the model. Jayaweera et al. [20], developed a system for detecting surface defects in APs based on the canny operator, which achieves edge extraction of defects, but their research is not complete and in-depth enough. Chondronasios et al. [21], by using the Sobel edge detector to obtain the gradient magnitude of the image and proposed the Gradient-only Co-occurrence Matrix (GOCM). Classification and detection of defect-free, blistered, and scratched APs surfaces were achieved based on the GOCM, with a test accuracy of 98.6% on a self-built dataset. All of the above are traditional MV-based defect detection methods. Although basically automated detection of surface defects has been achieved, the feature extraction methods used for defect detection require manual design and are not universally applicable and robust, and even require “one method for one scene”. Therefore, it is difficult to promote its use.

Deep Learning (DL), a major branch of machine learning, has made breakthroughs in recent years, especially Convolutional Neural Networks (CNN), which have been gradually applied to various object detection scenarios by their powerful feature extraction and non-linear representation capabilities [22]. Depending on the processing approach, DL-based object detection algorithms can be broadly classified into single-stage object detection algorithms and two-stage object detection algorithms. The more typical of the two-stage object detection algorithms include Region-CNN (R-CNN) [23], Faster Region-based CNN (Faster R-CNN) [24], Mask R-CNN [25], and so on. This class of algorithms divides the object detection process into two stages, with the first stage generating multiple proposal boxes in the image via a Region Proposal Network (RPN) and the second stage fine-tuning the proposal boxes. Due to the fact that the detection process is divided into two stages, good detection accuracy can be obtained on the one hand, but on the other hand, the detection speed is slow. Unlike the two-stage object detection algorithm, the single-stage object detection algorithm discards the RPN and performs regression detection directly on the object, so the detection process is a holistic one. The representative algorithms are Single Shot MultiBox Detector (SSD) [26] and You Only Look Once (YOLO) [27], and the detection speed of this class of algorithms has been greatly improved while guaranteeing a certain accuracy rate. In industrial production, real-time is an important metric. Therefore, YOLO is widely used as a typical single-stage object detection algorithm for defect detection in industrial production [28].

With the emergence of the concept of Industry 4.0, industrial production is gradually upgrading from automation to intelligence. MV combined with DL provides a new

solution for intelligent production. More and more scholars are beginning to apply DL technology to the SDD of industrial products [29], [30]. Compared to traditional MV detection methods, DL-based detection methods have autonomous feature learning capabilities, thus eliminating the need for manual feature design and offering good robustness and generalization, as well as requiring larger datasets. He et al. [31], designed an end-to-end detection network with multiscale feature fusion for steel plate defects detection and proposed the NEU-DET dataset. The experimental results show that the proposed method achieves up to 82.3% mAP on the NEU-DET dataset and a detection speed of 20 ft/s. Bachmann et al [32], through an experimental study, propose data augmentation and transfer learning as the key ingredients for training small sample datasets. The proposed algorithm achieves a detection accuracy of 47% mAP and a detection speed of 10 FPS on a self-built dataset of surface defects in APs. Dong et al. [33], proposed a novel PGA-Net for pixel-wise detection of surface defects by designing a pyramidal Feature Fusion Network (FFN) and a global context attention network. The mAPs of 82.15%, 74.78%, 71.31%, and 79.54% were achieved on the NEU-Seg, DAGM 2007, MT\_defect, and Road\_defect datasets, respectively. Wei and Bi [34], conducted a study on surface defects in APs and improved on the Faster R-CNN by multi-scale feature fusion, ultimately achieving a detection accuracy of 75.8% mAP, an improvement of 12.5%. Due to the two-stage object detection algorithm used, the detection speed was slow, and the FPS was only 1.2. Duan and Zhang [35], proposed a two-branch gradient image-based CNN for aluminum SDD, in which the original RGB image and the gradient image are input in two branches, and finally, the feature information of the two branches is fused using the concat operation. However, the two-branch network structure design may slow down the detection speed and have an impact on real-time performance. Ma et al. [36], Improvement of YOLOv4 using depth-separable convolution and a parallel dual-channel attention module for surface defects detection on APs. Li et al. [37] proposed a method for classifying surface defects in APs based on RepVGG and Convolutional Block Attention Module (CBAM). The classification experiments on ten APs surface defects showed that the classification accuracy was as high as 99.41%. However, the study only carried out the classification of defect images and did not explore defect detection in depth. Wang et al. [38], proposed MS-YOLOv5 based on YOLOv5, improved with PE-Neck and multi-stream networks. Tests on a dataset of surface defects of APs with seven defects showed that the mAP could reach 87.4%, and the FPS was 19.1, which basically met the real-time requirements. However, the dataset has fewer defect types and does not fully reflect actual industrial production conditions.

In summary, it is shown that DL-based object detection methods are feasible for detecting surface defects in industrial products. However, the existing research on the detection of



**FIGURE 1.** Some APs surface defect images in AL\_Dataset and the defect areas are labeled with red boxes.

surface defects in APs suffers from several deficiencies as a whole. Firstly, the single pursuit of higher detection accuracy leads to a slower detection speed of the algorithm, ignoring the real-time requirements in industrial inspection. Secondly, only the classification of defect images has been studied without delving into the detection of defects in images, where it is also crucial to clarify the specific defect location in industrial production. Finally, the study covers fewer defect categories, as it is inevitable that a wide variety of defects will be encountered in a complex production environment. On the other hand, there are problems such as a fuzzy definition of the defect area, large variation of defect scale, an imbalance in defect aspect ratio, and high similarity between inter-class defects for APs surface defects. Therefore, combining the shortcomings of current research and the characteristics of aluminum surface defects, this paper improves upon YOLOv5 and proposes an attention-guidance MA-YOLO for APs surface defect detection. Experiments show that MA-YOLO can detect APs surface defects more efficiently while satisfying the real-time requirements of industrial inspection. The specific work in this paper is as follows.

- (1) To address the problem of the randomness of the initial values taken by the K-Means clustering algorithm. We adopted the K-Means++ clustering algorithm to optimize the original anchor boxes, making the clustering results more globally optimal. This not only alleviates the problem of defect high aspect ratio imbalance but also improves the detection accuracy and convergence speed of the network.
- (2) To address the problem of large variation in defect scale and an imbalance in defect aspect ratio in the aluminum surface. This paper optimized a multi-scale FFN, which improved the detection performance of high aspect ratio defects and enhanced its adaptability to different scale defects.
- (3) For the problem of fuzzy definition and high inter-class defect similarity for APs surface defects, this paper

proposes a novel Max Pooling Average Pooling (MA) attention module to enable the network to focus on defect feature information and suppress useless background and noise interference, thus improving the detection accuracy of the network.

- (4) Extensive experiments on the surface defects dataset of APs have shown that MA-YOLO has a significant improvement in precision while meeting real-time requirements of industrial production, proving the efficiency of the proposed method. The effectiveness of the proposed improvement strategy is also verified by ablation experiments.

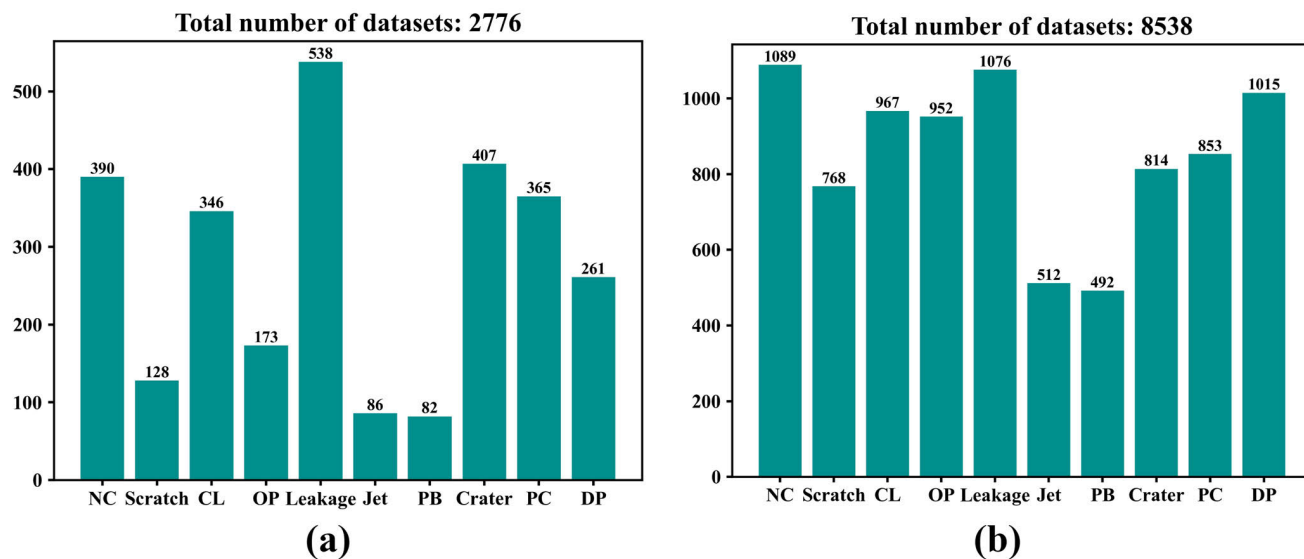
The rest of the paper is organized as follows: Section II introduces the dataset used in this paper, as well as the data augmentation methods, and outlines the YOLOv5 algorithm. Section III presents MA-YOLO and details specific improvement strategies. Section IV describes the experimental environment, parameter settings, and evaluation metrics for this paper. Section V gives detailed experimental results, performs a visual analysis, and verifies the effectiveness of the improvement strategy and the superiority of the proposed method. Section VI summarizes the work of this paper and indicates future research directions.

## II. RELATED WORK

### A. DATA INTRODUCTION

To facilitate readers' follow-up research, the open data set from the APs surface defect detection contest held by Ali Tianchi is used in the experimental part of this paper [39]. For the sake of presentation, we will name this data set AL\_Dataset in this article. AL\_Dataset contains a total of ten defects that are more common in production, including Non-Conductive (NC), Scratch, Corner Leak (CL), Orange Peel (OP), Leakage, Jet, Paint Bubble (PB), Crater, Parti-Colour (PC), and Dirty Point (DP). AL\_Dataset contains 2776 defect images, all set to a resolution of  $2560 \times 1920$ . Some of the surface defect images of APs are shown in Figure 1.





**FIGURE 2.** Changes in the number of defective images in AL\_Dataset for each category. (a) Original number distribution of defect images of each category in AL\_Dataset; (b) The number distribution of defect images of each category in AL\_Dataset after data augmentation.

**TABLE 1.** Various data augmentation methods adopted in this paper.

Data Augmentation Methods	Reasons for Use	Expected Results
Gaussian Noise	The differences between the captured and original images can be caused by factors such as image acquisition equipment and production environment.	By simulating noise in the images, the detection algorithm has strong robustness and adaptability to images collected in actual production environments.
Rotational Transformation	The unfixed placement of APs and the uncertainty of the location of various defects in the image increase the difficulty and complexity of defect detection.	By using image rotation transformation technology, the detection algorithm's perception ability for defects in different regional positions has been improved.
Brightness Transformation	In actual production environments, the brightness of the surrounding environment often changes, and a single brightness is not conducive to testing the generalization of the detection model.	By using image brightness transformation technology, diverse brightness scenes in actual production environments are simulated to improve the adaptability of detection models in complex environments.
Contrast Transformation	In actual production, the defects on the surface of APs are sometimes obvious and sometimes blurred, which increases the difficulty for the model to learn various types of defects.	By using image contrast transformation technology, simulate diverse defect situations in actual production to enhance the robustness of detection models in complex environments.

**B. DATA AUGMENTATION PROCESSING**

Compared with other computer vision tasks, SDD does not have a large and unified data set such as ImageNet [40], PASCAL-VOC [41], and COCO [42]. Defect detection mainly studies specific applications in different detection objects and scenarios. Compared with more than 14 million sample data in the ImageNet dataset, the most critical problem faced in SDD is the small sample size problem. Even in many real industrial scenes, there are only a few or dozens of defect images.

In this paper, the number of defective images of each category in AL\_Dataset was statistically analyzed and the results are shown in Figure 2a. It was observed that, on the one hand, AL\_Dataset has the problem of small data volume compared to large datasets such as ImageNet and COCO.

On the other hand, the number of Leakage defective images in AL\_Dataset is 538 (the most), and the number of PB defective images is 82 (the least), which has a severe problem of sample unevenness. DL is a strong data-driven discipline, the quality and quantity of data often significantly impact the detection model's performance [43]. When the training data is insufficient, the detection model is more prone to overfitting, which can affect the model's detection accuracy and generalization ability. When the training samples are unbalanced, data classes with larger samples will affect the model's learning of data classes with smaller samples, which can also interfere with the model's accuracy. Regarding the above two issues, and combined with the characteristics of surface defects in APs and specific production environmental factors, this article uses Gaussian noise, rotation transformation, brightness

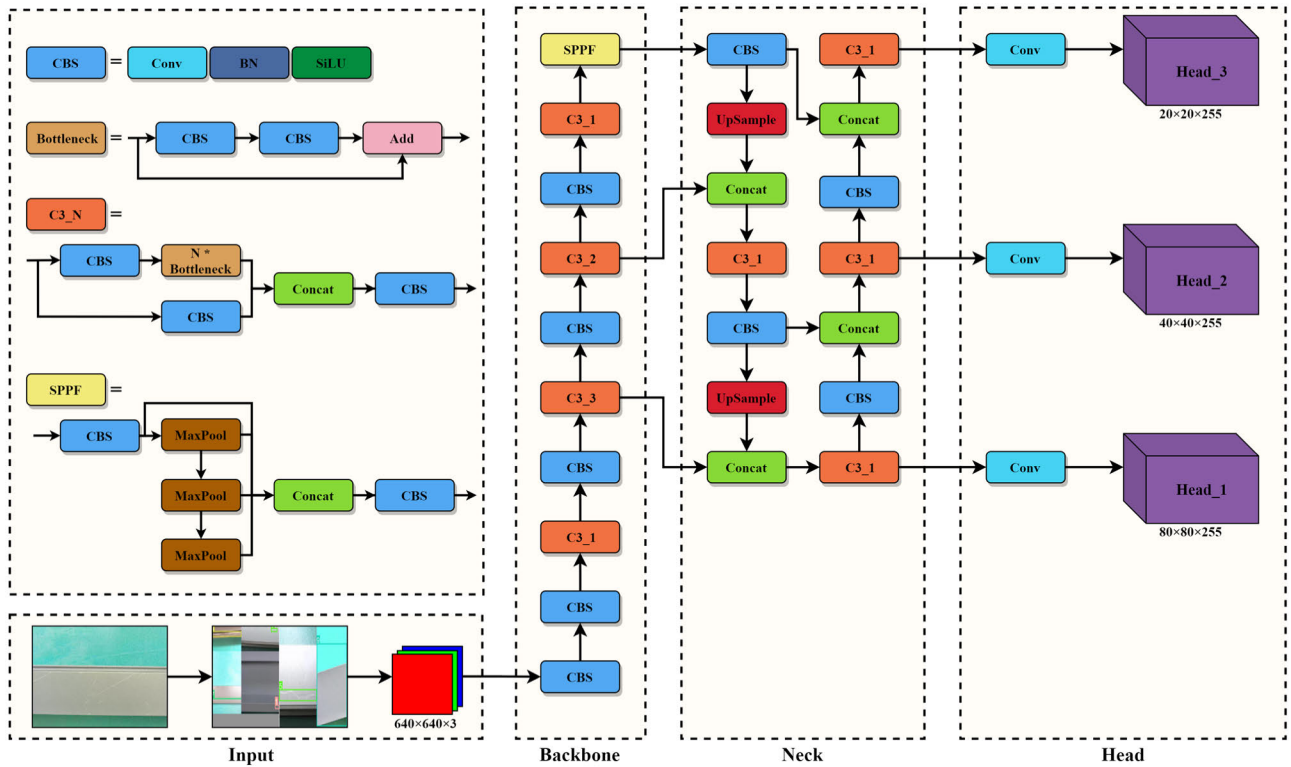


FIGURE 3. The overall architecture of YOLOv5s.

transformation, and contrast transformation to enhance and expand AL\_Dataset through data augmentation. In Table 1. This paper details the reasons for using the above data augmentation methods and the expected results. After data augmentation, AL\_Dataset contains 8538 defect images, and each class contains at least about 500 images. The results are shown in Figure 2b, thus effectively improving the problem of an insufficient dataset and an unbalanced distribution of sample data. Finally, we divided AL\_Dataset into training set, validation set, and test set in a ratio of 8:1:1 for the research on SDD of Aluminum Profiles in this paper.

C. INTRODUCTION TO THE YOLOv5 MODEL

YOLO is a representative single-stage object detection algorithm, which cleverly transforms object detection into a regression problem in its design. Firstly, the offset between the ground truth box and the predicted box is calculated through the loss function, then an optimizer is used to regress the offset of predicted boxes, finally achieving accurate detection with this algorithm. Moreover, compared to two-stage object detection algorithms such as Faster R-CNN and Mask R-CNN, YOLO has a significant advantage in detection speed, making it more favored by the industry. YOLOv5 is the most widely used and technically mature algorithm in the YOLO family and is widely used in industrial production [44]. Compared to previous versions of YOLO, YOLOv5 combines many of the best designs from advanced networks

and therefore offers enhanced detection performance, faster detection, and smaller size. Depending on the depth and width of the network, YOLOv5 can be further divided into YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. From left to right, the detection performance of each network increases in turn, increasing in size but getting slower in detection speed. YOLOv5s was chosen as the baseline algorithm for this paper due to real-time considerations for detecting surface defects in APs. The overall architecture of YOLOv5s is shown in Figure 3.

On the whole, YOLOv5s can be divided into four parts, namely Input, Backbone, Neck, and Head. Firstly, there is the input part of the model, which is responsible for pre-processing the input images, including image enhancement and adaptive image scaling. Next is the Backbone, also called the feature extraction network, which consists of the CBS module, the C3 module, and the SPPF module. The CBS module replaces the traditional Pooling operation and is responsible for the feature map’s scale and dimensional transformation. The C3 module is a modified version of the Cross Stage Partial module [45], which adopts a two-branch structure and stacks the residual structure to enhance the learning capability of the network and maintain a good lightweight, which can effectively extract the shallow texture features and in-depth semantic features of the image. SPPF module, on the other hand, uses different scales of the Max Pooling operation, which effectively expands the receptive

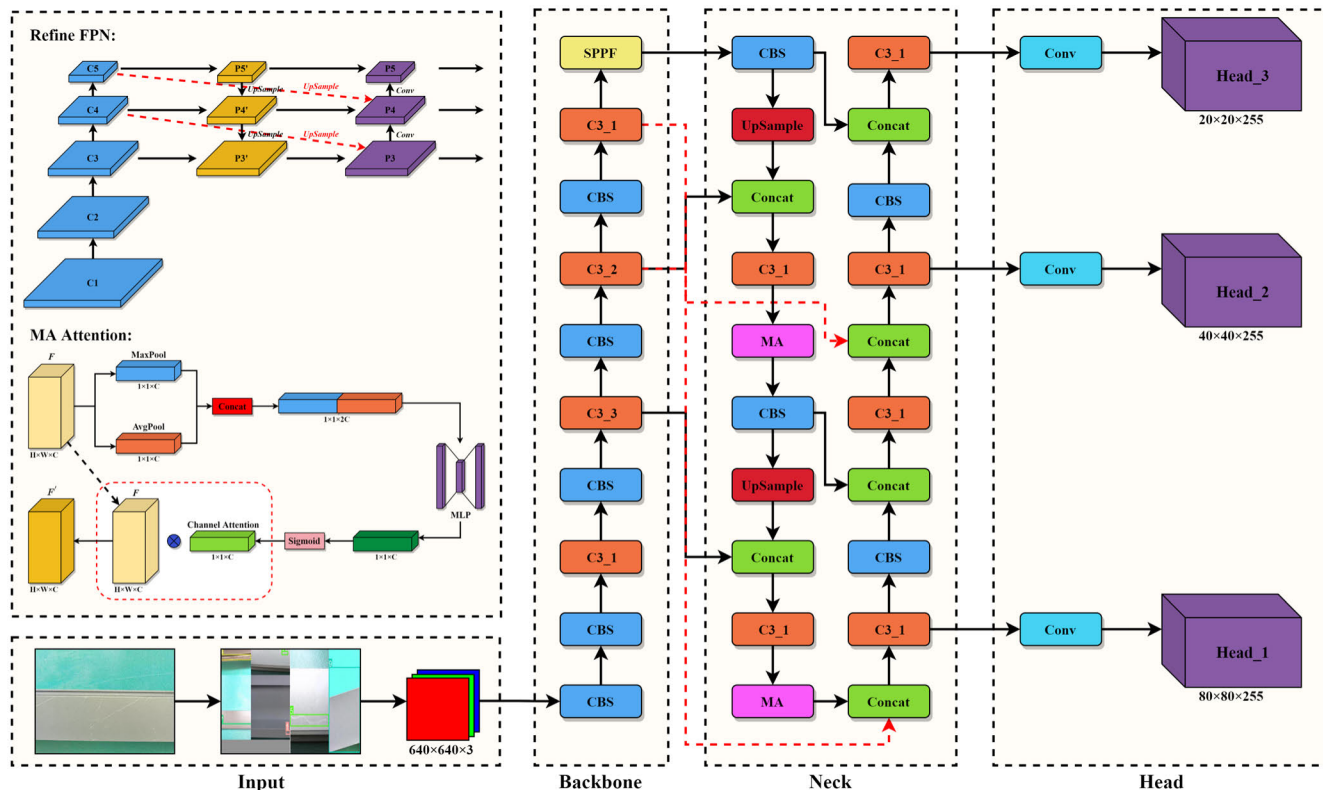


FIGURE 4. The overall architecture of MA-YOLO.

field of the model and increases the richness of the extracted features. Once again, Neck, also called the FFN, has the main purpose of effectively fusing feature map information from different scales in the feature extraction network, making the network more capable of feature representation. In YOLOv5, a Path Aggregation Network (PANet) is used [46], which improves the Feature Pyramid Network (FPN) and adds a new bottom-up feature transfer path to enhance the interactive fusion of shallow and profound information. Finally, there is the Head, also known as the detection head. In YOLOv5, there are three detection heads, one for large objects ( $20 \times 20$ ), one for medium objects ( $40 \times 40$ ), and one for small objects ( $80 \times 80$ ). The detection head generates multiple predictor boxes in the detection image, first filtering those below the confidence threshold and then removing the redundant boxes by Non-Maximum Suppression (NMS) to obtain the final detection result of the network.

III. METHODS

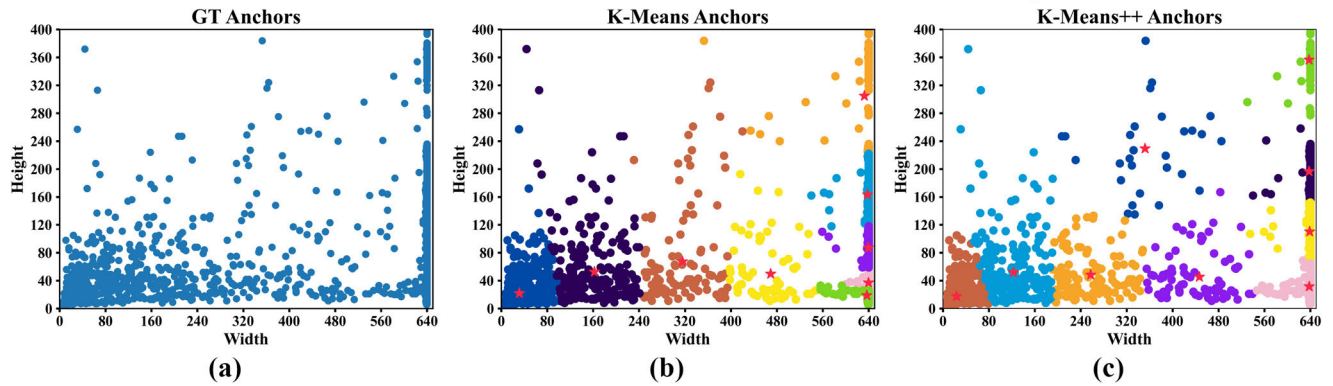
A. MA-YOLO OVERVIEW

Although, in natural scenes, YOLOv5 exhibits strong object detection performance and real-time object detection speed. However, in complex aluminum SDD scenarios, there are problems such as a fuzzy definition of defect areas, large variation of defect scale, imbalance of defect aspect ratio, and high similarity between inter-class defects, so further optimization of YOLOv5 needs to be taken. To address the above

problems, this paper firstly adopts the K-Means++ clustering algorithm to optimize the original anchor boxes [47], which can make the initial anchor box settings more relevant to the AL\_Dataset, alleviate the problem of extreme defect aspect ratio, and help improve the detection accuracy and convergence speed of the network. Next, by cross-scale feature linking and weighted feature fusion, PANet is improved to make in-depth feature information more effectively transmitted, alleviating the problem of large variations in defect scale and imbalance of defect aspect ratio. Moreover, the effective fusion of deep and shallow information also enhances the adaptability of the network to different scales of features. Finally, this paper introduces a novel MA attention module, which enables the network to focus on the defect features on the surface of APs and improves the detection accuracy of the network. With the above-mentioned improvement strategies, this paper proposes a SDD method with attention guidance for APs, called MA-YOLO. The overall architecture of MA-YOLO is shown in Figure 4.

B. K-MEANS++ CLUSTERING ALGORITHM

YOLOv5 is a single-stage object detection algorithm based on the anchor boxes. The setting of the initial anchor box affects the algorithm’s detection accuracy and convergence speed to a certain extent. In order to make the initial anchor box size more fitting to the training dataset, the ground truth box in the training dataset is usually clustered using the



**FIGURE 5.** The visualization results of the K-Means clustering algorithm and the K-Means++ clustering algorithm on the AL\_Dataset. (a) Distribution of ground truth boxes in AL\_Dataset; (b) Division of ground truth boxes in AL\_Dataset after K-Means clustering; (c) Division of ground truth boxes in AL\_Dataset after K-Means++ clustering.

**TABLE 2.** The Comparison of Default Anchor Box Sizes, K-Means Clustering Anchor Box Sizes, and K-Means++ Clustering Anchor Box Sizes.

Method	Anchor Boxes
The preset sizes of YOLOv5	Small object detection layer: (10,13); (16,30); (33,23)
	Medium object detection layer: (30,61); (62,45); (59,119)
	Large object detection layer: (116,90); (156,198); (373,326)
The K-Means clustering algorithm	Small object detection layer: (27,19); (156,52); (305,62)
	Medium object detection layer: (466,51); (630,18); (639,41)
	Large object detection layer: (638,83); (638,162); (638,303)
The K-Means++ clustering algorithm	Small object detection layer: (24,18); (123,52); (257,48)
	Medium object detection layer: (352,229); (447,45); (637,32)
	Large object detection layer: (638,110); (638,197); (638,357)

K-Means algorithm and fine-tuned using a genetic algorithm before the YOLOv5 training. However, the convergence in the K-Means clustering process is heavily dependent on the initial value of the cluster centers, which may lead to a large difference between the randomly initialized cluster centers and the optimal ones. To address this problem, this paper uses the K-means ++ clustering algorithm to improve it. The core idea is to make the centers as far away from each other as possible when initializing the cluster centers. Thus, the clustering results are closer to the optimal global solution. Unlike natural scenes, the surface defects of APs often present problems of large variation in defect scale and imbalance of defect aspect ratio. Therefore, this paper proposes to use the K-Means++ algorithm to cluster the ground truth boxes in AL\_Dataset to get more suitable anchor boxes for model training. The step-by-step process of the K-Means++ algorithm is as follows:

1. randomly selecting a sample as the first initialized cluster center;

2. calculate the distances of all samples to the known initialized clustering centers, dividing them by probability and preferring to select the distant sample as the new clustering center;

3. Repeat step 2 until the required number of cluster centers has been selected;

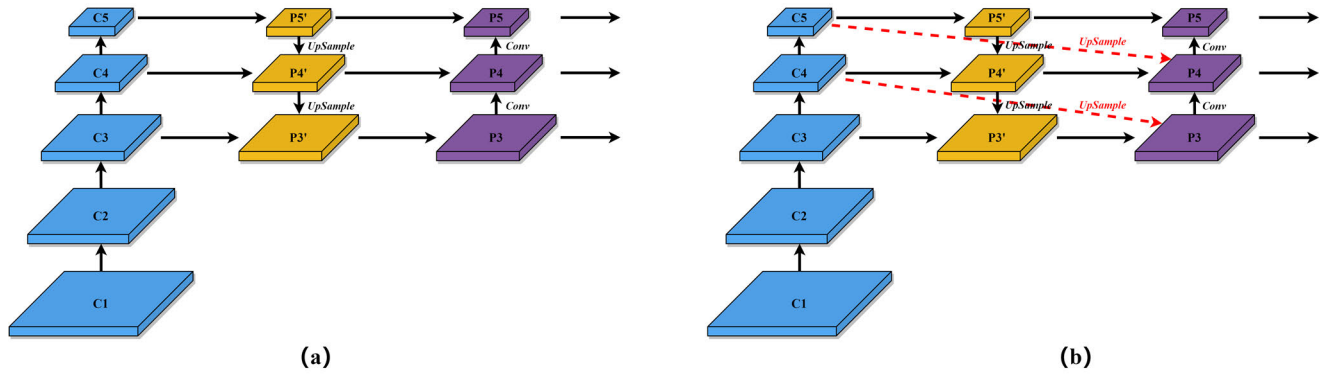
4. calculate the distance of each sample to each cluster center and reclassify to the cluster with the closest distance;

5. recalculate the average height and width of all samples in each cluster and obtain the new cluster centers;

6. repeat steps 4 and 5 until the distance traveled by the cluster centers is less than a threshold, or the set upper limit of the calculation is reached.

In this paper, we set 9 cluster centers based on the number of detection layers and the preset number of anchor boxes per layer. Specifically, we preset 3 scales of anchor boxes for each detection layer, and the entire model contains 3 detection layers. Before clustering, we adaptively scaled the image from a size of  $2560 \times 1920$  to  $640 \times 480$ . In Figure 5, we present the clustering visualization results of the K-Means algorithm and the K-Means++ algorithm. Figure 5a shows the size distribution of all real boxes in the AL\_Dataset training set, where the x-axis represents the width of the real box and the y-axis represents the height of the real box. It can be observed from the image that although there is some dispersion in the size of the real boxes, they are mainly concentrated in the lower left corner of the image and on the vertical line with a width





**FIGURE 6.** Improvement of Multi-scale FFN. (a) The original multi-scale FFN in YOLOv5; (b) the improved multi-scale FFN.

value of 640. Figures 5b and 5c show the clustering results of the K-Means algorithm and the K-Means++ algorithm on the AL\_Dataset training set, respectively. In the figure, different colors represent different clusters, and red stars are used to represent cluster centers. By observing the figure, it can be found that compared with the K-Means algorithm, the K-Means++ algorithm has a better partitioning effect, and the cluster centers of each cluster are relatively dispersed. At the same time, we also found that there are problems with large defect scale variation and imbalanced defect aspect ratios in the AL\_Dataset. Table 2 records the preset sizes of the original anchor boxes in YOLOv5, the anchor box sizes after K-Means clustering, and the anchor box sizes after K-Means++ clustering.

### C. MULTI-SCALE FEATURE FUSION NETWORKS

The multi-scale FFN is mainly proposed to solve the problem of multi-scale differences of objects in the detection image [48]. The basic idea is to fuse the image edge texture features extracted by the shallow network with the image semantic features extracted by the deep network for output so that the detection network can achieve better localization and regression [49]. FPN [50] is the first one to propose the idea of multi-size feature fusion, which achieves an interactive fusion of information by building a top-down network structure to transfer feature information from deep layers to shallow layers, enhancing the network's detection of small objects. PANet adds a bottom-up feature information transfer path to the FPN to enhance the information fusion between feature maps, thus improving the overall detection performance of the network. The Bidirectional Feature Pyramid Network (BiFPN) [51] optimizes the bidirectional fusion structure of PANet by proposing cross-scale connections and weighted feature fusion, further enhancing the multi-scale feature sensing capability of the detection network. In this paper, the PANet is improved concerning the ideas of cross-scale connections and weighted feature fusion in BiFPN to address the problem of the large variation of defect scale and an imbalance in defect aspect ratio in APs. The weighted cross-scale feature fusion is achieved by up-sampling the deep feature maps in the backbone network, which enables the deep

feature information to be transferred more effectively, thus improving the network's performance in detecting multi-scale defects. In Figure 6, a represents the original FFN in YOLOv5, and b represents the improved FFN. Taking the P4 detection layer in Figure 6b as an example, the fusion process of its feature information can be represented by formulas (1) and (2). Where  $\omega_i$  represents different learnable weights,  $\varepsilon = 0.0001$  is used to avoid numerical instability. *Conv* represents a convolutional operation, and *UpSample* represents the UpSampling operation.

$$P4' = \text{Conv}\left(\frac{\omega_1^* C4 + \omega_2^* \text{UpSample}(P5')}{\omega_1 + \omega_2 + \varepsilon}\right) \quad (1)$$

$$P4 = \text{Conv}\left(\frac{\omega_3^* \text{Conv}(P3) + \omega_4^* P4' + \omega_5^* \text{UpSample}(C5)}{\omega_3 + \omega_4 + \omega_5 + \varepsilon}\right) \quad (2)$$

### D. MA ATTENTION MODULE

The attention mechanism originates from the human visual system's perception of different things. Its basic idea is to make the detection network focus on useful feature information in the image, and suppress useless background and noise interference, thereby improving the network's detection performance. In recent years, research based on attention mechanisms has been continuously proposed. Squeeze and Excitation (SE) Attention is a typical channel attention proposed by Hu et al. [52]. It mainly learns the correlation between different channels in the feature map by performing SE operations on the input feature map, giving higher weights to feature maps with better performance and lower weights to those with poorer performance. CBAM is an efficient attention module proposed by Woo et al. [53]. Based on the SE attention module, it adds a spatial attention module and combines channel attention and spatial attention to enable the network to perceive the correlation between different channels in feature maps and focus on the position information of objects in feature maps. Currently, the mainstream method mainly learns the correlation between different channels in the feature map through Max Pooling and Average Pooling. In addition, there are mixed Pooling, Stochastic Pooling, and

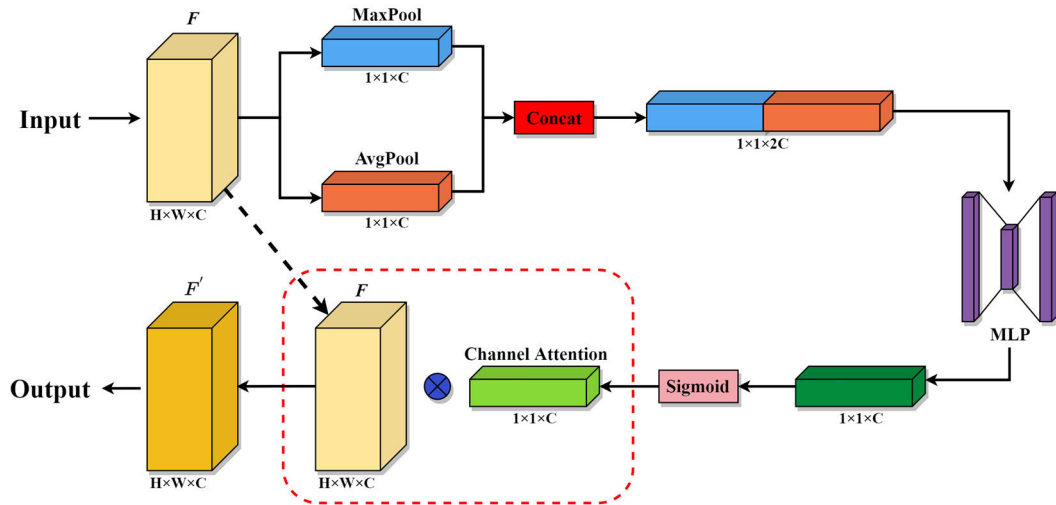


FIGURE 7. Schematic diagram of the MA attention module.

Random Pooling operations. However, they have problems such as high computational complexity, low accuracy, and difficult deployment. Therefore, they are not considered in this research [54]. Based on the research of SE and CBAM, this paper proposes a novel MA attention module for surface defect detection in APs. The specific implementation details are shown in Figure 7. We set the original input feature map as  $F$  ( $H \times W \times C$ ), and perform Max Pooling and Average Pooling respectively. On the one hand, Max Pooling can better preserve important spatial information in the  $F$  and reduce noise interference to some extent. On the other hand, Average Pooling can better preserve overall spatial information in the  $F$  and is relatively smooth. The combination of both can effectively learn spatial information from the  $F$ . At this moment,  $F$  undergoes Max Pooling and Average Pooling to obtain two channel attention maps with a size of  $1 \times 1 \times C$  each. Then, the two channel attention maps are concatenated along the channel dimension, resulting in a size of  $1 \times 1 \times 2C$ . Then, by building a Multilayer Perceptron (MLP) to perform SE operations on the channel attention map, further highlighting more valuable channel information. At this moment, the size of the channel attention map becomes  $1 \times 1 \times C$ . Then, non-linear mapping is achieved through the Sigmoid function to improve the adaptability and expressive ability of the channel attention map. Finally, multiply the obtained channel attention with the original input feature map  $F$  to obtain MA Attention Module's final output feature map  $F'$  ( $H \times W \times C$ ). The experiment shows that YOLOv5s, by combining the MA attention module, not only improves the problem of defining fuzzy defect areas but also enhances the network's ability to distinguish between similar defects in different classes.

## IV. EXPERIMENT

### A. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTING

All experiments in this paper were conducted on a high-performance DL workstation with the following overall

configuration: CPU: Intel Xeon Silver 4210 CPU; GPU: NVIDIA TITAN RTX 24GB; Memory: 64GB; DL framework: Python 3.8.10, Cuda 10.2, Pytorch 1.8.1, Torchvision 0.9.1; Operating system: Windows 10. In the experiment, we set the ratio of training set, validation set, and test set to 8:1:1. We used a unified training parameter setting, including input image size of  $640 \times 640$ , BatchSize of 16, an initial learning rate of 0.01, the minimum learning rate of 0.0001 and optimized with SGD optimizer with default weight decay of 0.0005 and momentum of 0.937. To ensure the stability and reliability of the model, we trained each model for 200 epochs.

### B. EVALUATION METRICS

In DL-based object detection algorithms, Precision (P), Recall (R), Average Precision (AP), mean Average Precision (mAP), and F1 score are important metrics for evaluating the merits of detection models. Where precision represents the proportion of true labeled samples among the number of correct samples predicted by the model, and recall refers to the proportion of all labeled positive samples that are correctly predicted as positive by the model. Next, AP refers to the average precision of a certain category in the dataset. It can be represented by the area enclosed by the PR curve with R as the horizontal axis and P as the vertical axis. Meanwhile, mAP represents the average precision of all categories in the dataset and is usually represented by mAP at an Intersection over Union (IOU) threshold of 0.5, also called mAP<sub>50</sub>. Finally, the F1 score is the best balance between accuracy and recall, which gives a more comprehensive picture of the overall performance of the model. Their calculation formulas are given in (3)-(7). Where TP is True Positive, FP is False Positive, FN is False Negative, TN is True Negative,  $i$  is the number of the category, and N is the total number of categories.

$$P = \frac{TP}{TP + FP} \quad (3)$$

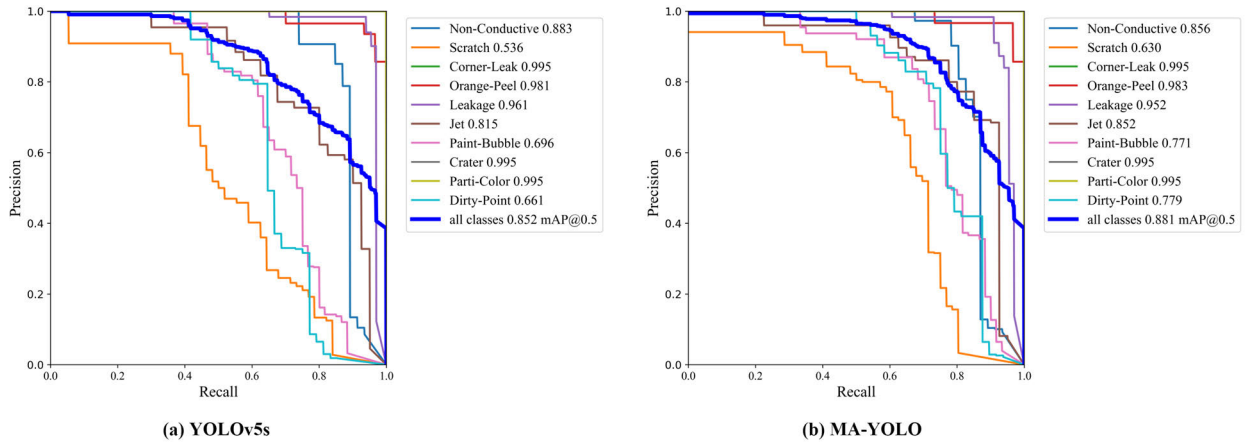


FIGURE 8. Comparison of PR curves for YOLOv5s and MA-YOLO.

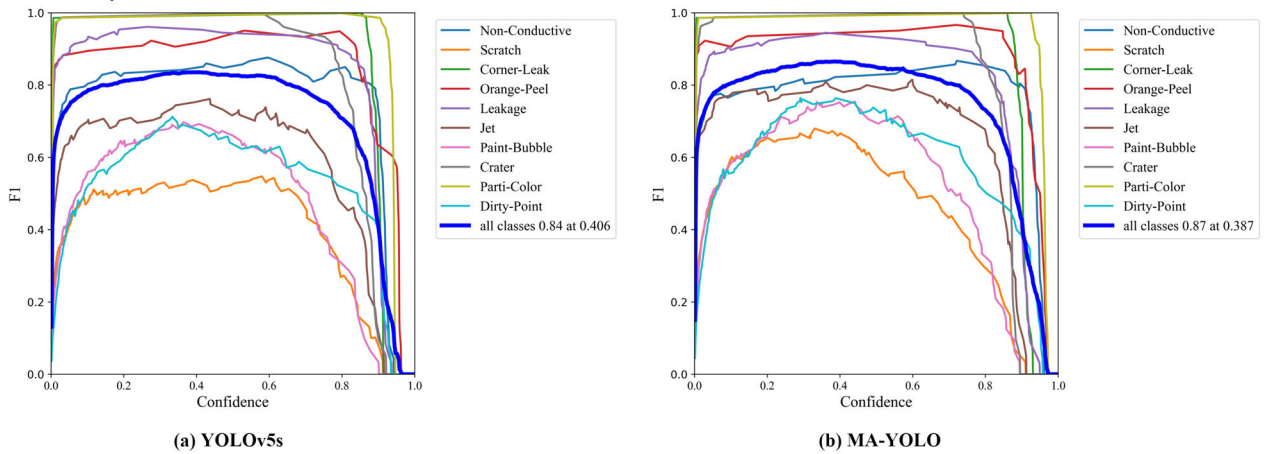


FIGURE 9. Comparison of F1 score curves for YOLOv5s and MA-YOLO.

TABLE 3. Results of ablation experiments for each module on the al\_dataset test set.

Method	K-Means++	Refine FPN	MA Attention	mAP <sub>50</sub> (%)	Model Size (M)	FPS
A				85.2	14.2	38
B	√			85.5	14.2	38
C	√	√		86.9	14.5	34
D	√	√	√	88.1	14.6	34

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$AP = \int_0^1 P(R)d(R) \quad (5)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

$$F1 \text{ score} = \frac{2 \times P \times R}{P + R} \quad (7)$$

In our experiments, we not only used the evaluation metrics mentioned above but also conducted a more comprehensive performance analysis of MA-YOLO. In this paper, we use model size and FPS to comprehensively evaluate the change in model lightweight after MA-YOLO significantly improved mAP<sub>50</sub>.

## V. RESULTS AND DISCUSSION

### A. OVERALL PERFORMANCE COMPARISON ANALYSIS

In the experimental environment provided in this paper, we have visually analyzed the performance of YOLOv5s and MA-YOLO on AL\_Dataset. Figure 8 shows the PR curves of YOLOv5s and MA-YOLO. It can be observed that MA-YOLO improves by 2.9% compared with YOLOv5s on mAP<sub>50</sub>, which is a significant improvement. Figure 9 shows the F1 score curves of YOLOv5s and MA-YOLO. On the whole, the curves of all categories are improved, and the F1 score is generally increased by 3%. By comparing the PR curve and F1 score of YOLOv5s and MA-YOLO, it can be concluded that MA-YOLO is significantly better than YOLOv5s in the overall detection performance of the network.

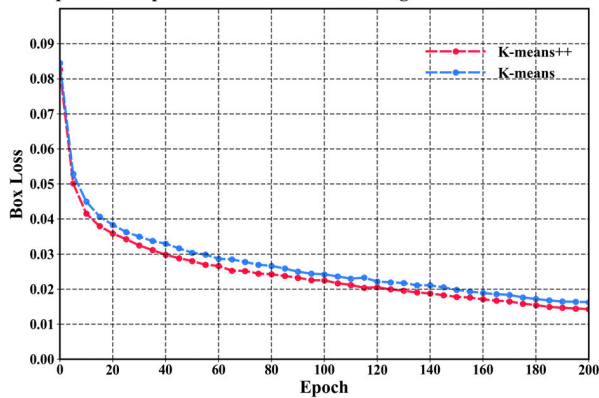
**TABLE 4.** Compare the effects of different attention mechanisms on the performance of surface defect detection for aluminum profiles.

Method	P (%)	R (%)	mAP <sub>50</sub> (%)	Model Size (M)
YOLOv5s	88.4	79.2	85.2	14.2
YOLOv5s + SE	88.2	80.3	85.5	14.4
YOLOv5s + CBAM	90.6	79.5	85.9	14.5
YOLOv5s + CA	91.1	79.9	86.1	14.5
YOLOv5s + MA	90.9	81.5	86.3	14.5

**TABLE 5.** Performance of mainstream object detection algorithms on the lv\_data test set.

Method	P (%)	R (%)	F1 score (%)	mAP <sub>50</sub> (%)	Model Size (M)	FPS
SSD300	<b>95.8</b>	51.1	67.2	84.1	95.2	34
Faster R-CNN	54.1	<b>87.9</b>	65.8	80.1	108.5	7
YOLOv3	90.1	78.7	82.6	84.6	120.7	14
YOLOv4	92.8	59.9	67.1	70.0	244.6	10
YOLOv5s	88.4	79.2	82.7	85.2	<b>14.2</b>	<b>38</b>
YOLOX_s	87.5	82.9	84.7	86.7	16.3	31
YOLOv6s	76.9	72.5	73.9	78.6	40.6	32
YOLOv7	87.8	55.4	63.9	71.6	74.9	27
DETR	71.1	83.7	77.8	80.5	497.3	8
MA-YOLO	90.8	81.1	<b>84.9</b>	<b>88.1</b>	14.6	34

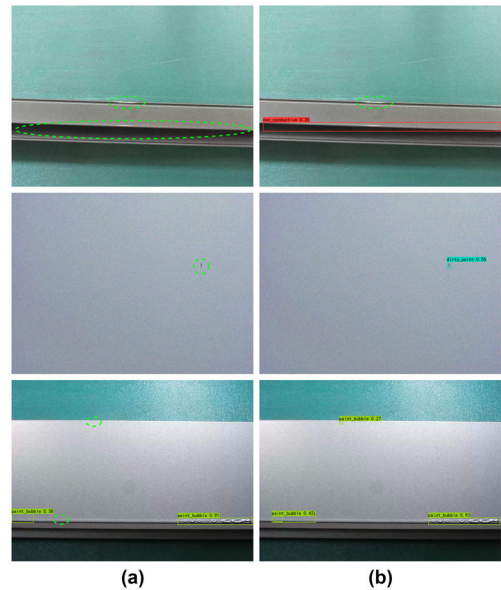
Compare the impact on box loss between using K-Means and K-Means++



**FIGURE 10.** Compare the box loss curves of K-Means and K-Means++.

**B. ABLATION EXPERIMENTS**

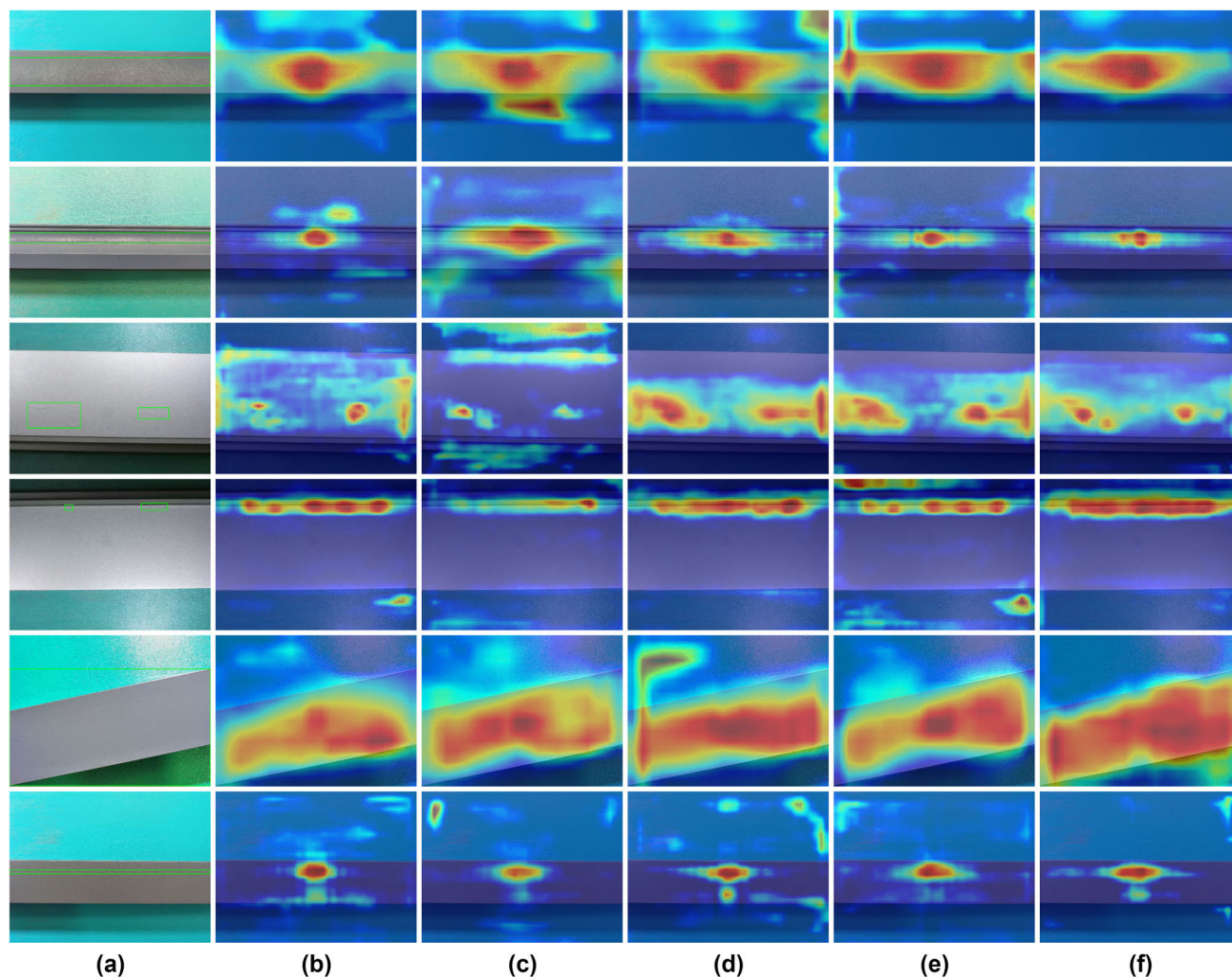
In the previous section, this paper demonstrated the overall performance advantage of MA-YOLO through visual analysis of the PR curve and F1 score curve. To further verify the effectiveness of the various improvement strategies proposed in this paper, we conducted ablation experiments on the AL\_Dataset test set. The evaluation metrics were mAP<sub>50</sub>, model size, and FPS, and the specific experimental results are shown in Table 3. Among them, Method A represents the baseline YOLOv5s, with a mAP<sub>50</sub> of 85.2%, a model size of 14.2 M, and an FPS of 38. Method B improved on YOLOv5s by using the K-Means++ clustering algorithm, achieving a 0.3% improve in mAP<sub>50</sub> without increasing computational complexity. This indicates that optimizing the initial anchor box is beneficial for improving the performance of anchor-based object detection algorithms.



**FIGURE 11.** Some comparison examples of detection results before and after the improvement of FPN, the green elliptical dashed line in the figure marks the area of missed detection defects (False Negative). (a) before improvement, there are a total of 5 missed defects; (b) after improvement, one more larger aspect ratio target and 3 more smaller targets were recalled.

Method C improved the multi-scale FFN based on Method B. On the one hand, due to the addition of the concat operation, the extra computation was introduced, resulting in a 0.3M increase in model size and a 3 FPS decrease. On the other hand, the multi-scale feature fusion significantly improved mAP<sub>50</sub> by 1.4%. Method D added the MA attention module on top of Method C, forming the MA-YOLO proposed in this paper. It achieved a 1.2% increase in mAP<sub>50</sub> while





**FIGURE 12.** Heatmaps analysis of different attention mechanisms. (a) manually marked GT in the images; (b) YOLOv5s; (c) YOLOv5s+SE; (d) YOLOv5s+CBAM; (e) YOLOv5s+CA; (f) YOLOv5s+MA.

maintaining the same FPS, effectively improving detection accuracy for surface defects on APs. This also proves that the MA attention module proposed in this paper can effectively improve the detection of surface defects on APs. Finally, compared with the baseline YOLOv5s, MA-YOLO increased  $mAP_{50}$  by 2.9% while only increasing the model size by 0.4M and decreasing FPS by 4, still meeting the real-time requirements of industrial detection. Next, we conducted individual experimental analyses on each improvement strategy in MA-YOLO, including the K-Means++ clustering algorithm, the improved multi-scale FFN, and the MA attention module. The detailed experimental results are as follows.

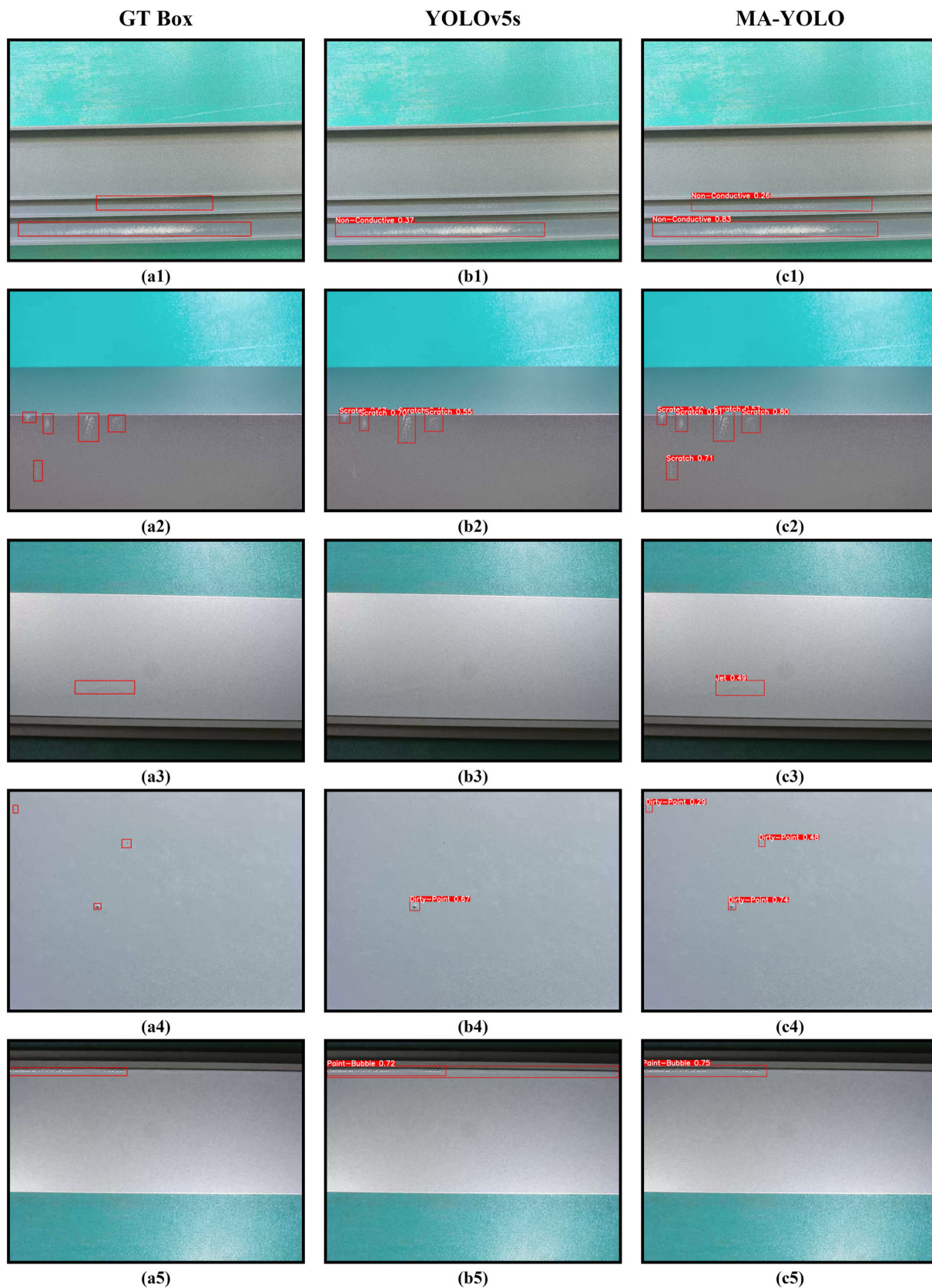
#### 1) EFFECTIVENESS OF K-MEANS++:

The initial setting of anchor boxes is crucial for anchor-based object detection algorithms. This paper introduces the K-Means++ clustering algorithm to alleviate the problem of local optima in the K-Means algorithm, effectively improving the detection accuracy and convergence speed of the

network. In Figure 10, we compare the effects of the KMeans algorithm and the K-Means++ algorithm on the box loss during model training. The results show that the KMeans++ clustering algorithm effectively accelerates the convergence speed of the network and further improves the detection accuracy.

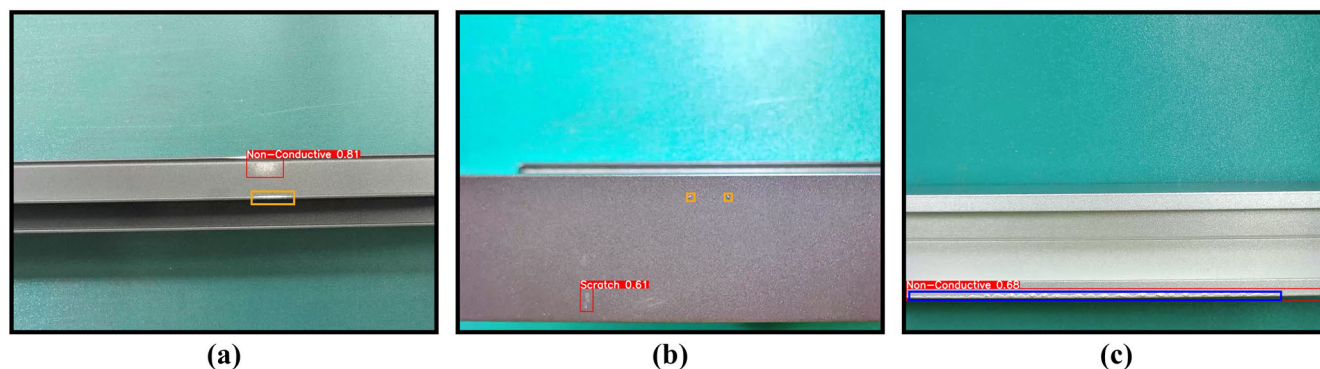
#### 2) EFFECTIVENESS OF IMPROVING FPN

In order to better address the problem of large-scale variations in surface defects of APs, this paper has improved the FPN to enhance the fusion of feature information at different scales. As shown in Figure 11, by comparing the detection results of three groups (a is the detection result before FPN improvement, and b is the detection result after FPN improvement), we found that the improved FPN can effectively alleviate the problem of object scale variation, and also improve the detection performance of small objects, thus significantly reducing the phenomenon of missed detection.



**FIGURE 13.** Comparison of detection results for YOLOv5s and MA-YOLO on the AL\_Dataset. (a1-a5) ground truth boxes; (b1-b5) detection results for YOLOv5s; (c1-c5) detection results for MA-YOLO.





**FIGURE 14.** Failure case analysis. (a) and (b) False Negative cases; (c) Defect is detected but misclassified.

### 3) EFFECTIVENESS OF MA ATTENTION MECHANISM

The purpose of introducing the MA attention mechanism is to enhance the model's attention to defect features and suppress useless background and noise interference. In Table 4, we compared the effects of several different attention mechanisms on model performance, among which the MA attention mechanism improved P, R, and mAP<sub>50</sub> by 2.5%, 2.3%, and 1.1%, respectively. Figure 12 shows the heatmap visualization analysis results of different attention mechanisms, indicating that the MA attention mechanism can effectively alleviate the interference of complex backgrounds and focus on collecting feature information of different defects. This result further proves the effectiveness of the MA attention mechanism.

### C. COMPARISON WITH OTHER OBJECT DETECTION ALGORITHMS

In the previous section, we compared the performance differences between YOLOv5s and MA-YOLO. To comprehensively analyze the performance advantages of MA-YOLO, we compared its performance with that of mainstream object detection algorithms on the AL\_Dataset test set. The experimental results are shown in Table 5, where MA-YOLO achieved the best performance in the F1 score and mAP<sub>50</sub> evaluation metrics. Specifically, MA-YOLO outperformed the second-best algorithm YOLOX\_s by 1.4% and 0.2% in mAP<sub>50</sub> and F1 score, respectively, and outperformed the baseline algorithm YOLOv5s by 2.9% and 2.2%. Compared with other object detection algorithms, Faster R-CNN achieved the best results in the recall, but its precision was low, resulting in poor overall detection performance of the model. Moreover, because the model detection process is divided into two steps, its detection speed is the slowest, with an FPS of only 7. Unlike Faster R-CNN, SSD300 performs very well in precision, but its recall is poor, resulting in average overall detection performance of the model, with a mAP<sub>50</sub> of 84.1%. At the same time, due to the simple structure of the SSD300 model, its detection speed is fast, with an FPS of 34. YOLOv3 also achieved good results in detection performance, with a mAP<sub>50</sub> of 84.6%, but its

model is larger and its detection speed is slow. YOLOv4 and YOLO7 have similar performance in detection metrics, both with high precision and low recall, resulting in poor detection performance of the model. However, the model size of YOLOv7 is only 30.6% of that of YOLOv4, which has an advantage in terms of speed, with an FPS of 27. YOLOv6s is relatively balanced in terms of precision and recall, but its overall detection performance is mediocre, with a mAP<sub>50</sub> of 78.6%. As a Transformer-based end-to-end object detection network, DETR has good detection performance for large objects, so it also achieved good results in the defect detection of APs. However, the model size is 497.3 M, and the FPS is only 8. Through a comprehensive analysis of the overall performance metrics of various object algorithms, our MA-YOLO achieved the best detection performance while maintaining a leading detection speed, with an FPS of 34, second only to YOLOv5s, which meets the real-time requirements of industrial detection.

To verify the effectiveness of MA-YOLO more intuitively, the detection results of YOLOv5s and MA-YOLO are visualized and analyzed on the AL\_Dataset test set in this paper. The detection results of the model are shown in Figure 13. In this paper, five groups of APs surface defect images are compared, containing 15 images. From the detection results in the first to fourth rows of Figure 13, it is clear that YOLOv5s suffers from a general defect miss detection phenomenon, especially when the defect size is small and the network has difficulty effectively identifying the defect. In contrast, MA-YOLO enhances the model's focus on minor defects on the surface of APs by introducing the MA attention module, thus significantly optimizing the defect miss detection phenomenon. Taking the example of smaller-sized scratch defects, experimental test results show that MA-YOLO can detect scratch defects with a minimum size of  $42 \times 42$  in the original image (size:  $2560 \times 1920$ , defects smaller than 42 pixels cannot be recognized by the detection algorithm at the same confidence level). In addition, the detection results in the fifth row of Figure 13 show that MA-YOLO also improves defect misdetection compared to YOLOv5s. The analysis suggests that by optimizing the

multi-scale FFN, the in-depth information is more effectively fused with the shallow information, thus avoiding the false detection of the background as a defect. In conclusion, this paper has demonstrated through extensive experiments and visualized analysis that MA-YOLO performs better than the current mainstream object detection algorithms in the scenario of APs surface defect detection.

#### D. FAILURE CASE ANALYSIS

Although MA-YOLO has shown the best detection performance in many experiments, MA-YOLO still has the phenomenon of missed detection and false detection in some cases due to various complex situations in industrial defect detection scenarios. The specific details are shown in Figure 14. For the failed defect detection cases, this paper profoundly analyzes and tries to explore the reasons for missed detection and false detection of the model. In Figure 14, a and b represent cases of model miss detection, and the missed defects are marked using yellow boxes in this paper. In case a, the analysis suggests that the missed defect in the image is more similar to the reflection of light from the surface of the APs, which causes MA-YOLO to overlook the defects. In case b, even the human eye could not clearly identify the minor defects on the aluminum profile's surface, making it difficult for MA-YOLO to make an accurate detection. In Figure 14, case c shows a misclassified case, and the defects is marked using blue box. It can be seen that MA-YOLO misclassified the PB defect as the NC defect, which is probably due to the inconsistent characteristics of the continuous and single PB defects and the fact that the size of the PB defect and the NC defect in case c is relatively similar, thus leading to the misdetection of the model. Through the reasonable analysis of failure cases, we realize the shortcomings of MA-YOLO. In future work, we will optimize it to reduce the missed detection and false detection of the model.

#### VI. CONCLUSION

The detection of defects on the surface of APs has an important influence on the quality of APs products. In this paper, we propose MA-YOLO, an attention-guided algorithm for detecting defects on APs surfaces, to address issues such as a fuzzy definition of the defect area, large variation in defect scale, imbalance in defect aspect ratio, and high similarity between inter-class defects. Firstly, the original anchor boxes are optimized using the K-Means++ algorithm instead of the K-Means algorithm to make the initial anchor box settings fitter to the defects dataset, thereby alleviating the problem of an imbalance in the defect aspect ratio. Secondly, by improving the multi-scale FFN, not only the detection performance of high aspect ratio imbalance defects is improved, but also the adaptability to variation scale features is enhanced. Finally, a novel MA attention module is proposed to improve the detection accuracy of the network, especially for small object defects, by enhancing the model's focus on defect feature information and suppressing unwanted background and noise interference. Experiments on the surface defect dataset in APs

show that the MA-YOLO has better detection accuracy than current mainstream object detection algorithms. Compared to the baseline YOLOv5s, MA-YOLO improves mAP<sub>50</sub> and F1 score by 2.9% and 2.2%, respectively. Overall, MA-YOLO achieved significant improvement in mAP<sub>50</sub> while maintaining consistency with YOLOv5s in model lightweight, meeting the real-time requirements of industrial detection. In future work, we will improve and enhance the defect categories that are not detected well, thus improving the overall detection accuracy of the model.

#### ACKNOWLEDGMENT

The authors would like to express their gratitude for the research conditions provided by academician He Jifeng's studio, the Research Center for Internet of Things and Big Data Technology, Xijing University; the Xi'an Key Laboratory of High Precision Industrial Intelligent Vision Measurement Technology, Xijing University; and the Shaanxi Key Laboratory of Integrated and Intelligent Navigation.

#### REFERENCES

- [1] W. Zhou, Z. Shao, J. Yu, and J. Lin, "Advances and trends in forming curved extrusion profiles," *Materials*, vol. 14, no. 7, p. 1603, Mar. 2021.
- [2] F. Czerwinski, "Current trends in automotive lightweighting strategies and materials," *Materials*, vol. 14, no. 21, p. 6631, Jan. 2021.
- [3] M. Aamir, K. Giasin, M. Tolouei-Rad, and A. Vafadar, "A review: Drilling performance and hole quality of aluminium alloys for aerospace applications," *J. Mater. Res. Technol.*, vol. 9, no. 6, pp. 12484–12500, Nov. 2020.
- [4] T. Welo, G. Ringen, and J. Ma, "An overview and evaluation of alternative forming processes for complex aluminium products," *Proc. Manuf.*, vol. 48, pp. 82–89, Jan. 2020.
- [5] R. Chen, D. Cai, X. Hu, Z. Zhan, and S. Wang, "Defect detection method of aluminum profile surface using deep self-attention mechanism under hybrid noise conditions," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
- [6] R. Usamentiaga, D. G. Lema, O. D. Pedrayes, and D. F. Garcia, "Automated surface defect detection in metals: A comparative review of object detection and semantic segmentation using deep learning," *IEEE Trans. Ind. Appl.*, vol. 58, no. 3, pp. 4203–4213, May 2022.
- [7] Y. Gao, X. Li, X. V. Wang, L. Wang, and L. Gao, "A review on recent advances in vision-based defect recognition towards industrial intelligence," *J. Manuf. Syst.*, vol. 62, pp. 753–766, Jan. 2022.
- [8] Y. Li, Y. Xiang, H. Guo, P. Liu, and C. Liu, "Swin transformer combined with convolution neural network for surface defect detection," *Machines*, vol. 10, no. 11, p. 1083, Nov. 2022.
- [9] X. Fang, Q. Luo, B. Zhou, C. Li, and L. Tian, "Research progress of automated visual surface defect detection for industrial metal planar materials," *Sensors*, vol. 20, no. 18, p. 5136, Sep. 2020.
- [10] S. Guo, G. Ren, and B. Zhang, "Subsurface defect evaluation of selective-laser-melted Inconel 738LC alloy using eddy current testing for additive/subtractive hybrid manufacturing," *Chin. J. Mech. Eng.*, vol. 34, no. 1, p. 111, Nov. 2021.
- [11] E. D'Accardi, D. Palumbo, R. Tamborrino, and U. Galietti, "A quantitative comparison among different algorithms for defects detection on aluminum with the pulsed thermography technique," *Metals*, vol. 8, no. 10, p. 859, Oct. 2018.
- [12] W. Lou, C. Shen, Z. Zhu, Z. Liu, F. Shentu, W. Xu, T. Lang, Y. Zhang, Z. Jing, and W. Peng, "Internal defect detection in ferromagnetic material equipment based on low-frequency electromagnetic technique in 20# steel plate," *IEEE Sensors J.*, vol. 18, no. 16, pp. 6540–6546, Aug. 2018.
- [13] O. Kayode and E. T. Akinlabi, "An overview on joining of aluminium and magnesium alloys using friction stir welding (FSW) for automotive lightweight applications," *Mater. Res. Exp.*, vol. 6, no. 11, Nov. 2019, Art. no. 112005.



- [14] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, X. Zhao, A. Khaliq, M. Faheem, and A. Ahmad, "CNN based automatic detection of photovoltaic cell defects in electroluminescence images," *Energy*, vol. 189, Dec. 2019, Art. no. 116319.
- [15] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, "Machine learning for industrial applications: A comprehensive literature review," *Expert Syst. Appl.*, vol. 175, Aug. 2021, Art. no. 114820.
- [16] K. Ide, K. Nomura, H. Hosono, and T. Kamiya, "Electronic defects in amorphous oxide semiconductors: A review," *Phys. Status Solidi A*, vol. 216, no. 5, Mar. 2019, Art. no. 1800372.
- [17] Q. Luo, X. Fang, L. Liu, C. Yang, and Y. Sun, "Automated visual defect detection for flat steel surface: A survey," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 626–644, Mar. 2020.
- [18] H. Wang, J. Zhang, Y. Tian, H. Chen, H. Sun, and K. Liu, "A simple guidance template-based defect detection method for strip steel surfaces," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2798–2809, May 2019.
- [19] T. Shi, J.-Y. Kong, X.-D. Wang, Z. Liu, and G. Zheng, "Improved Sobel algorithm for defect detection of rail surfaces with enhanced efficiency and accuracy," *J. Central South Univ.*, vol. 23, no. 11, pp. 2867–2875, Nov. 2016.
- [20] L. Jayaweera, H. Priyankara, and P. Dadigamuwa, "Identification of defects in aluminum profiles using vision-based system," *GSI*, vol. 8, no. 7, pp. 1304–1311, Jul. 2020.
- [21] A. Chondronasios, I. Popov, and I. Jordanov, "Feature selection for surface defect classification of extruded aluminum profiles," *Int. J. Adv. Manuf. Technol.*, vol. 83, nos. 1–4, pp. 33–41, Mar. 2016.
- [22] L. Jiang, B. Yuan, W. Ma, and Y. Wang, "JujubeNet: A high-precision lightweight jujube surface defect classification network with an attention mechanism," *Frontiers Plant Sci.*, vol. 13, Jan. 2023, Art. no. 1108437.
- [23] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [25] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.
- [26] W. Liu, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [27] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [28] X. Kou, S. Liu, K. Cheng, and Y. Qian, "Development of a YOLO-V3-based model for detecting defects on steel strip surface," *Measurement*, vol. 182, Sep. 2021, Art. no. 109454.
- [29] B. Tang, L. Chen, W. Sun, and Z. Lin, "Review of surface defect detection of steel products based on machine vision," *IET Image Process.*, vol. 17, no. 2, pp. 303–322, Feb. 2023.
- [30] J. Hong, H. Zhang, and X. Xu, "Thermal fault prognosis of lithium-ion batteries in real-world electric vehicles using self-attention mechanism networks," *Appl. Thermal Eng.*, vol. 226, May 2023, Art. no. 120304.
- [31] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.
- [32] F. M. Neuhauser, G. Bachmann, and P. Hora, "Surface defect classification and detection on extruded aluminum profiles using convolutional neural networks," *Int. J. Mater. Forming*, vol. 13, no. 4, pp. 591–603, Jul. 2020.
- [33] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, "PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7448–7458, Dec. 2020.
- [34] R. Wei and Y. Bi, "Research on recognition technology of aluminum profile surface defects based on deep learning," *Materials*, vol. 12, no. 10, p. 1681, May 2019.
- [35] C. Duan and T. Zhang, "Two-stream convolutional neural network based on gradient image for aluminum profile surface defects classification and recognition," *IEEE Access*, vol. 8, pp. 172152–172165, 2020.
- [36] Z. Ma, Y. Li, M. Huang, Q. Huang, J. Cheng, and S. Tang, "A lightweight detector based on attention mechanism for aluminum strip surface defect detection," *Comput. Ind.*, vol. 136, Apr. 2022, Art. no. 103585.
- [37] Z. Li, B. Li, H. Ni, F. Ren, S. Lv, and X. Kang, "An effective surface defect classification method based on RepVGG with CBAM attention mechanism (RepVGG-CBAM) for aluminum profiles," *Metals*, vol. 12, no. 11, p. 1809, Oct. 2022.
- [38] T. Wang, J. Su, C. Xu, and Y. Zhang, "An intelligent method for detecting surface defects in aluminium profiles based on the improved YOLOv5 algorithm," *Electronics*, vol. 11, no. 15, p. 2304, Jul. 2022.
- [39] *Tianchi Data Sets*. Accessed: Jan. 9, 2023. [Online]. Available: <https://tianchi.aliyun.com/dataset/>
- [40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [41] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
- [42] T.-Y. Lin, "Microsoft COCO: Common objects in context," in *Computer Vision ECCV 2014 (Lecture Notes in Computer Science)*, vol. 8693, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740–755.
- [43] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esen, A. A. S. Awwal, and V. K. Asari, "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*.
- [44] J. Shi, J. Yang, and Y. Zhang, "Research on steel surface defect detection based on YOLOv5 with attention mechanism," *Electronics*, vol. 11, no. 22, p. 3735, Nov. 2022.
- [45] C. Wang, H. M. Liao, Y. Wu, P. Chen, J. Hsieh, and I. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1571–1580.
- [46] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.
- [47] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. ACM-SIAM Symp. Discrete Algorithms*, New Orleans, LA, USA, Jan. 2007, pp. 1027–1035.
- [48] Y. Zhang, B. Yuan, Y. Wang, F. Wang, and J. Guo, "A real-time oil content analysis method of cuttings based on deep learning," *IEEE Access*, vol. 10, pp. 132083–132094, 2022.
- [49] L. Huang and W. Huang, "RD-YOLO: An effective and efficient object detector for roadside perception system," *Sensors*, vol. 22, no. 21, p. 8097, Oct. 2022.
- [50] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
- [51] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.
- [52] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [53] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision ECCV 2018 (Lecture Notes in Computer Science)*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 3–19.
- [54] K. Khalil, O. Eldash, A. Kumar, and M. Bayoumi, "Designing novel AAD pooling in hardware for a convolutional neural network accelerator," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 30, no. 3, pp. 303–314, Mar. 2022.

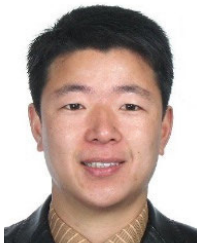


**LINGJIE JIANG** received the B.S. degree in computer science and technology from Xijing University, Xi'an, China, in 2021, where he is currently pursuing the M.S. degree in computer technology. His research interests include deep learning and object detection.



research interests include pattern recognition, machine learning, computer vision, and deep learning.

**BAOXI YUAN** received the B.S. degree in fire control engineering and the M.S. degree in signal and information processing from Air Force Engineering University, Xi'an, China, in 2000 and 2006, respectively, and the Ph.D. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2013. He is currently an Associate Professor with the School of Electronic Information, Xijing University, Xi'an. His current



Xi'an. His current research interests include intelligent measurement and control, and educational research.

**YUQIAN WANG** received the B.S. degree in aircraft and engine engineering and the M.S. degree in aerospace propulsion theory and engineering from Air Force Engineering University, Xi'an, China, in 2002 and 2006, respectively, and the Ph.D. degree in aerospace propulsion theory and engineering from Northwestern Polytechnical University, Xi'an, in 2010. He is currently an Associate Professor and the Deputy Director of the Postgraduate Department, Xijing University,



China. She received the Ph.D. degree in electrical engineering from Xi'an Jiaotong University, in 2013. She was with the Department of Information Engineering, Xijing University, where she is currently an Assistant Professor. She is the author of two books and more than 20 articles. She is responsible for or participated in many national projects. Her research interests include intelligent information acquisition and processing and the Internet of Things.

**YAHONG MA** was born in Fuping, Shaanxi, China. She received the Ph.D. degree in electrical engineering from Xi'an Jiaotong University, in 2013. She was with the Department of Information Engineering, Xijing University, where she is currently an Assistant Professor. She is the author of two books and more than 20 articles. She is responsible for or participated in many national projects. Her research interests include intelligent information acquisition and processing and the Internet of Things.



where she is currently pursuing the master's degree in electronic information. Her research interests include networks and information security.

**JIawei DU** received the B.E. degree from Xijing University, Xi'an, Shaanxi, China, in 2021, where she is currently pursuing the master's degree in electronic information. Her research interests include networks and information security.



and optimization theory.

**FENG WANG** received the B.S. degree in communications engineering from Air Force Engineering University (AFEU), Xi'an, China, in 2001, and the Ph.D. degree in information and communications engineering from Xidian University Xi'an, in 2016. From 2004 to 2018, he was a Lecturer with AFEU. Since 2018, he has been an Assistant Professor with the School of Electronic Information, Xijing University. His research interests include compressed sensing, wireless localization, and optimization theory.



and their implementation in GNU radio.

**JIANXIN GUO** received the B.S. degree in communications engineering from the Telecommunications Engineering Institute of the Air Force, Xi'an, China, in 1997, the M.S. degree in information and communications engineering from Air Force Engineering University, Xi'an, in 2000, and the Ph.D. degree in information and communications engineering from PLA Information Engineering University, Zhengzhou, China, in 2004. He is currently a Professor with the School of Electronic Information, Xijing University, Xi'an. His current research interests include cognitive radio technologies, MIMO-OFDM wireless communications, and their implementation in GNU radio.

...