

Received January 26, 2022, accepted February 9, 2022, date of publication February 17, 2022, date of current version March 22, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3152552

Improved YOLOV4-CSP Algorithm for Detection of Bamboo Surface Sliver Defects With Extreme Aspect Ratio

YIJING GUO¹, YIXIN ZENG², FENGQIANG GAO³, YI QIU¹, XUQIANG ZHOU², LINWEI ZHONG², AND CHOUJUN ZHAN¹, (Member, IEEE)

¹School of Information Science and Technology, Tan Kah Kee College, Xiamen University, Zhangzhou 363105, China

²School of Informatics, Xiamen University, Xiamen 361005, China

³School of Aerospace Engineering, Xiamen University, Xiamen 361005, China

Corresponding author: Fengqiang Gao (fengqianggao@foxmail.com)

This work was supported in part by the National Science Foundation of China Project under Grant 61703355, in part by the Natural Science Foundation Project of Zhangzhou 2019 under Grant ZZ2019J34, and in part by the Program for Young Excellent Talents in University of Fujian Province under Grant 201847.

ABSTRACT Bamboo surface defect detection provides quality assurance for bamboo product manufacture in industrial scenarios, an integral part of the overall manufacturing process. Currently, bamboo defect inspection predominantly relies on manual operation, but manual inspection is very time-consuming as well as labor-intensive, and the quality of inspection is not guaranteed. A few visual inspection systems based on traditional image processing have been deployed in some factories in recent years. However, traditional machine vision algorithms extract features in tedious steps and have poor performance along with poor adaptability in the face of complex defects. Accordingly, many scholars are committed to seeking deep learning methods to accomplish surface defect detection. However, existing deep learning object detectors struggle with specific industrial defects when directly applied to industrial defect detection, such as sliver defects, especially for ones with extreme aspect ratios. To this end, this paper proposes an improved algorithm based on the advanced object detector YOLOV4-CSP, which introduces asymmetric convolution and attention mechanism. The introduction of asymmetric convolution enhances the feature extraction in the horizontal direction of the bamboo strip surface, improving the performance in detecting sliver defects. In addition, convolutional block attention module (CBAM), a hybrid attention module, which combines channel attention with spatial attention, is utilized to promote the representation ability of the model by increasing the weights of crucial channels and regions. The proposed model achieves outstanding performance in the general categories and excels in the hard-to-detect categories. Some enterprise's bamboo strip dataset experiments verify that the model can reach 96.74% mAP for the typical six surface defects. Meanwhile, we also observe significant improvements when extending our model to aluminum datasets with similar characteristics.

INDEX TERMS Bamboo defect, defect inspection, object detection, asymmetric convolution, attention mechanism.

I. INTRODUCTION

China has the wealthiest bamboo resources globally, and its bamboo forest area, stock volume, and bamboo timber production all ranks first in the world [1]. In the production process of traditional bamboo products such as bamboo

The associate editor coordinating the review of this manuscript and approving it for publication was Sudipta Roy.

chopsticks, bamboo furniture, bamboo flooring, bamboo storage boxes, bamboo shoots [2], bamboo acts as an essential raw material. Emerging applications such as bamboo charcoal, bamboo vinegar, and bamboo fiber [3] have promoted the further development and utilization of bamboo resources. As the contradiction between wood supply and demand has become more intense, the mitigation plan of replacing wood with bamboo has made bamboo strip vital raw material in

industrial production. It can be seen that the bamboo industry has broad prospects for development and economic efficiency. However, defects are inevitable during the storage and production of bamboo products due to human operation, environment, or other factors. The surface defects of bamboo strip refer to the external defects of the product, which are different from the typical product surface. Putting the defective products into subsequent production will cause product stagnation or functional impairment and even cause safety accidents. Consequently, surface defect detection is a significant step to achieve quality control and improve production efficiency. The surface defect detection of bamboo strip belongs to the field of industrial surface defect detection. Industrial defect inspection refers to object detection in industrial application scenarios where the object categories are pre-specified defect types. The defect categories mainly include the typical six types of bamboo surface defects in this article.

The human eyes can usually distinguish surface defects under suitable light conditions. In most bamboo production workshops, the detection of bamboo strips' surface defects is implemented by manual visual inspection. Manual inspection requires experienced engineers, and the inspection standards are not the same, which is prone to subjective inter-examiner variations. Besides, workers are prone to fatigue and cannot work as long as machines. In general, manual inspection consumes a mass of workforce, time and cost. The examination can't promise quality. Many scholars are aware of this problem and propose some machine vision algorithms based on traditional image processing to complete defect detection, improving detection efficiency. Yet, the feature extraction steps of these methods are cumbersome and cannot sufficiently identify the complex and variable defects. In recent years, domestic and foreign enterprises have developed machine vision products for defect detection, such as Omron, DALSA, Cognex and HALCON, etc. The cost of this visual equipment is always too expensive to bear for some small and medium-sized enterprises. The advent of Industry 4.0 promotes intelligent industrial manufacturing and intelligent defect detection. Latterly, deep learning-based object detection algorithms have brought new blood into industrial defect inspection and made remarkable breakthroughs in defect inspection [4] of steel, aluminium, textiles, etc. However, the object detection algorithms based on deep learning have not been widely used in the surface inspection of the bamboo strip. Moreover, the existing algorithms based on deep learning have not explicitly designed modules for hard-to-detect categories, such as bamboo sliver defects, which implies that the detection performance of sliver bamboo defects needs to be improved.

Traditional surface defect detection methods based on image processing cannot adapt well to industrial inspection's complex and volatile environment. To raise the automation level of bamboo strips' surface defect detection and promote the development of related industries, it is of great significance to explore bamboo strips' surface defect detection algorithms based on deep learning. This article chooses

YOLOV4-CSP, a one-stage detector with excellent speed and accuracy as the baseline. After analyzing the characteristics of the surface defects of the bamboo strip, we find that most of the defects which are difficult to be accurately identified are strip-shaped. The characteristic inspires us to design an asymmetric convolution module and introduce an attention mechanism. Based on this, we propose an improved version of YOLOV4-CSP.

The improved YOLOV4-CSP not only follows the advantages of automatic learning characteristics of YOLOV4-CSP but also has better adaptability, especially for sliver defects. The addition of two targeted techniques (i.e., asymmetric convolution and attention mechanism) strengthens the feature extraction of the horizontal direction of the bamboo surface and diverts the attention to the salient features, contributing the model to coping with hard-to-detect samples. The added modules belong to the plug-and-play module, which does not significantly change the network structure and introduces no additional hyperparameter. Thus, the improved network training process is the same as the original one, enabling the deep learning object detector to detect bamboo strips' surface defects in industrial scenarios with the slightest modifications. The main contributions of our work are summarized as follows:

(i) We design an asymmetric convolution module, which is suitable for sliver defect detection. The module boosts the model's performance via enhancing feature extraction on the horizontal direction of the bamboo strip surface.

(ii) We combine the attention mechanism (CBAM) containing spatial attention and channel attention with YOLOV4-CSP to select discriminative features via adaptively weighting feature maps, which further benefits sliver defect detection.

(iii) Targeted experiments verify the effectiveness of the proposed method, and our method achieves 96.74% mAP on six types of typical bamboo strip defects. Moreover, our model performs favorably when transferring to the aluminum profiles dataset with similar characteristics, implying that our method may solve other sliver defect detection.

The rest of the paper is organized as follows. Section 2 (Related Work) introduces the current research status of bamboo surface defect inspection and object detection based on deep learning. The network structure is shown in detail in Section 3 (Methodology), including the design of asymmetric convolution and the introduction of the attention mechanism. Section 4 (Experiments) is devoted to show ablation experiments and analyze module design-related issues. Summary of this paper and future work are presented in Section 5 (Conclusions).

II. RELATED WORK

With the massive applications of bamboo in the industry, bamboo surface defect detection has witnessed many developments in recent years to guarantee the quality of bamboo products. The research on bamboo surface defect detection is mainly based on traditional image processing, adopting artificial design features combined with

classifiers to achieve detection [5]. The feature extraction step is divided into the structural method, statistical method, filtering method, and modeling method. The structural method usually extracts features via edge detection [6], [7] and morphological operations [8]. The statistical method broadly analyzes the histogram, LBP [9], and grayscale covariance matrix GLCM [10] features of bamboo strip surfaces. While the filtering method is composed of spatial domain filtering [11] and frequency domain filtering, such as Fourier transform [12], Gabor transforms and Wavelet transforms [13]. The modeling method establishes a random field model, inversed scatter model or fractal body for feature extraction. Based on feature extraction, these defect detection algorithms then classify defects by threshold judgment method, SVM classifier, or BP network [8], [14], etc. Furthermore, some methods have been proposed for specific types of defects in bamboo strips. The defective edge detection system designed by Sun *et al.* focuses on the detection of edge defects in bamboo [15]. The system uses an optical fiber amplifier to detect the intensity of light leaking from the gap between the contact plate and the edge of the bamboo strip. It is determined whether there is a bamboo strip edge defect according to the preset amplifier threshold. The feature extraction of these above methods requires elaborate design. These methods are merely effective for certain classes but suffer from poor adaptability, insufficient generalization ability, and harsh imaging conditions. The object detectors based on Deep learning can provide a solution to these issues. To this end, the article proposes a bamboo surface's defect inspection algorithm, which exploits the advantages of the automatic learning of the object detectors based on deep learning.

Over the years, object detection algorithms based on deep learning have received extensive attention from researchers. Existing object detectors are mainly divided into one-stage detectors and two-stage detectors. The former first generates region proposals, then extracts features according to the region candidate proposals, and finally returns classification and positioning results, a coarse-to-fine process. Common two-stage detectors are RCNN [16], Fast RCNN [17], Faster RCNN [18], etc. While the latter achieves end-to-end detection, obtaining classification and locations only after a single CNN operation. Common one-stage detectors are: YOLO series [19]–[22], SSD [23], CenterNet [24], Retinanet [25], etc. Generally speaking, two-stage detectors have advantages in detection accuracy while inferior in speed to one-stage detectors; One-stage detectors have better detection speed than two-stage detectors but suffer from a decrease in accuracy. In April 2020, Alexey *et al.* proposed YOLOV4 to achieve the optimal balance of accuracy and speed in the COCO dataset. In November of the year, Wang *et al.* researched CSP-ized [26] YOLOV4 and model scaling and proposed Scaled-YOLOv4 [27], which can be applied to different computing devices to achieve optimal performance.

Object detectors based on deep learning have been successfully applied to several fields, such as video monitoring, autonomous driving, and industrial defect inspection.

In recent times, deep learning-based object detectors have been widely studied to provide solutions for industrial defect detection, such as fabric defect detection [28], steel defect detection [29], and wood defect detection. To the best of our knowledge, there are few studies on deep learning to achieve end-to-end detection for the bamboo defect. Gao *et al.* [30] propose improved CenterNet for bamboo surface defect detection. They design an auxiliary detection module based on training from scratch and fuse the main part of the pre-training model with a connection mode of attention mechanism to improve the detection performance of CenterNet in a small amount of bamboo surface defect data. However, this model does not study the difficulties in the surface defects of bamboo, such as sliver defects. Existing object detectors either use k-means or its improved method k-means++ [31] to re-cluster to calculate anchors that are more suitable for the dataset or optimize the convolution operation to expand the sampling range to cope with the diversity of scale transformations [32], [33] for sliver defects. However, few targeted modules have been proposed for sliver defects, which regularly stretch in the horizontal direction. On the other hand, the current state-of-art detectors do not work well when directly ported to industrial scenarios with bamboo defects. Most general object detectors are developed based on general object recognition and are not applicable for specific application scenarios such as industrial defect detection. As a result, we propose the improved YOLOv4-CSP model based on the advanced detector YOLOv4-CSP and design pertinent modules that facilitate the detection of sliver defects to achieve optimal accuracy.

III. METHODOLOGY

In this section, the structure of the defect detection network will be described in detail. Firstly, we will introduce the overall structure of the improved YOLOV4-CSP, followed by the part of the network with asymmetric convolution. Finally, the attention mechanism in the network will be presented.

A. NETWORK ARCHITECTURE

The network structure of our model generally follows the design of YOLOV4-CSP. Wang *et al.* do CSP-ization in the neck based on YOLOV4, who build the CSPSP and CSPPAN structures and develop YOLOV4-CSP [27], reducing the computational cost by about 40%. As shown in Figure 1, the network is divided into three parts, including feature extraction(backbone), feature enhancement(neck) and detection. CSPDarknet53 is selected as the backbone of YOLOv4, YOLOv4-CSP and our model due to its excellent computations, inference speed and accuracy. CSPDarknet53 is obtained by the fusion of Darknet53 and CSPNet, whose core idea is partitioning the feature map of the base layer into two parts and then merging them through the cross-stage hierarchy to reduce duplicate gradient information while promoting inference speed. In Darknet53, the output of the residual layer (named bottleneck in this paper) is obtained by adding the initial input and the results of the

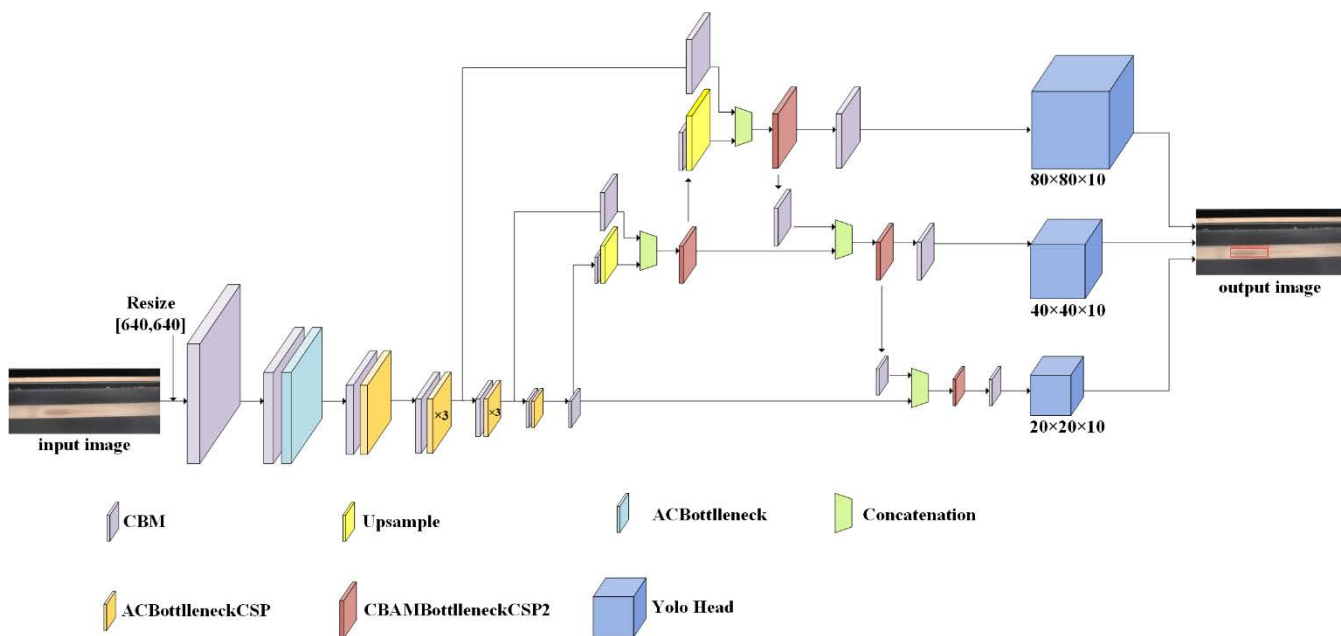


FIGURE 1. The network structure of improved YOLOV4-CSP.

residual block as shown in Figure 2(a). At each stage of CSPDarknet53(called BottleneckCSP in this paper), feature maps at the base layer are separated into two parts as shown in Figure 2(b). One part sequentially passes through a convolutional block, several residual blocks, and a convolution operation. While the other part firstly undergoes a convolution operation, then combines with the last part. Finally, the part goes through a transition layer(a convolution block) to get the ultimate output. The number of residual layers of each stage is 1,2,8,8,4 in Darknet53. Of note that in order to achieve the best trade-off between accuracy and speed, the first CSP stage (BottleneckCSP) is replaced with the original residual structure (Bottleneck) in our model. On the basis of the CSP-ization, our model introduces an asymmetric convolution to the 3×3 convolution of the residual block of each stage and constructs a new Bottleneck and BottleneckCSP, named ACBottleneck and ACBottleneckCSP, which are shown in Figure 2(c). The new residual layer is designed for promoting the influence of bamboo strip salient features in the horizontal direction. It is in line with the issue that bamboo surface defects are mostly sliver and the weights in the horizontal direction of feature maps are more vital than vertical ones.

The feature enhancement network adopts the idea of multi-scale fusion, applying CSPPAN to realize the effective fusion of bamboo strip spatial information and semantic information. Compared with FPN, PAN adds a bottom-up aggregation path, which can obtain richer locations information. The CSP-ization on the neck mainly reflects the design of the residual layer(BottleneckCSP2). The features from different feature pyramids are integrated as the input of this module. It is also divided into two parts. One passes through the residual block without shortcut connection, the other through

a convolution operation, and then combined with the former. Ultimately, the outcome is acquired through the transition layer(a convolution block). We introduce a hybrid attention mechanism and construct CBAMBottleneckCSP2 to replace the original module BottleneckCSP2 to calibrate the channel and space weights better. The attention mechanism and SPP duplicate functionally in increasing the receptive field, and we remove CSPSP for a more straightforward network design.

Chien-Yao Wang *et al.* also investigate model scaling and propose Scaled-YOLOv4 [27], which can be easily deployed on GPUs with different computing power. Researchers usually increase the depth or width of the network to enhance the feature representation capability (via controlling the number of BottleneckCSP in backbone and the number of BottleneckCSP2 in neck). Theoretically, deeper and wider networks tend to yield higher detection accuracy. However, this is not always the case, as there are also issues to consider, such as small datasets being prone to overfitting in large networks. In this paper, taking the small scale of the bamboo strip defect dataset, the solid color background of the images, and the high requirement of defect detection speed into account, we scale down the YOLOV4-CSP. The model depth scale factor is 0.33, so the number of residual blocks of each CSP stage changes from 1,2,8,8,4 to 1,1,3,3,1. The number of sub-modules (i.e., ACBottleneck, ACBottleneckCSP, and CBAMBottleneckCSP2) of the improved YOLOV4-CSP can be seen in Figure 1, where “ $\times 3$ ” means that there are three such modules stacked.

B. ASYMMETRIC CONVOLUTION BLOCK

Feature extraction is the first part of the whole model, and high-quality features are significant for the subsequent

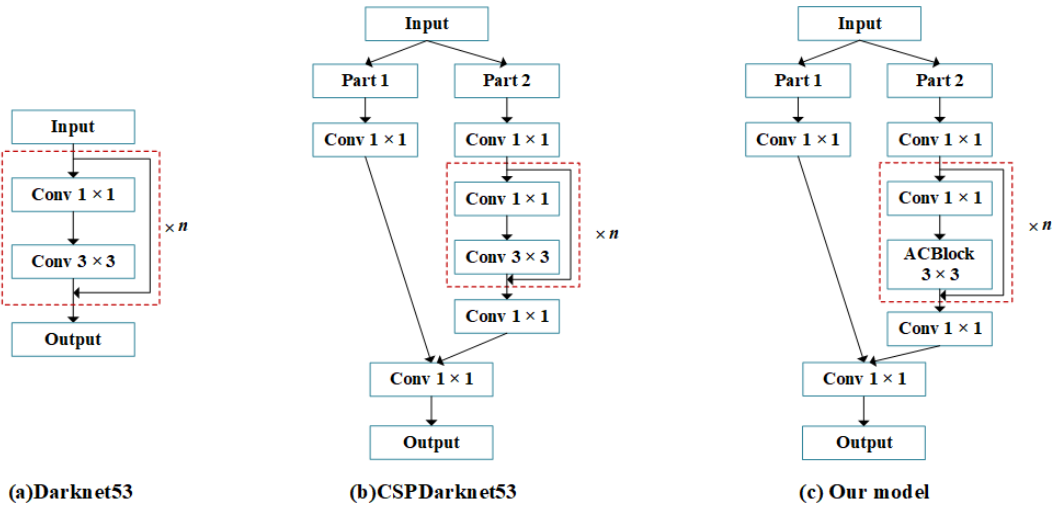


FIGURE 2. The structure of the residual layer in Darknet53, CSPDarknet53 and our model.

modules. The backbone network of the model, CSPDarknet53, is used to extract preliminary features of bamboo strips as this structure matches almost all optimal architecture features through analysis of network architecture search techniques [27]. In order to improve the learning ability of the network, we design an asymmetric convolution block and add the module to the residual block of the CSP stage. Asymmetric convolution was applied to reduce the sum of network parameters at the outset. The standard square convolution kernel ($d \times d$) is split into one-dimensional convolutions ($1 \times d$ and $d \times 1$), which can lessen the computational load of the network and raise the network training speed [34], [35]. In contrast, Ding *et al.* proposed ACBlock by integrating asymmetric convolution into square convolution from the perspective of convolution design and developed ACNet. The ACBlock [36] can enhance the characterization ability of the square convolution kernel by adding one-dimensional convolution in vertical and horizontal directions, thus enhancing robustness to rotational distortions and generalization ability to unseen data. Inspired by the concept of ACNet, we propose an asymmetric convolutional module more suitable for bamboo strip defect detection and combine it with the backbone.

The asymmetric convolution module in ACNet comprises three parallel layers with $d \times d$, $1 \times d$ and $d \times 1$ kernels, respectively, of which the outputs are summed up to enrich the feature space. This module enhances representability by reinforcing the magnitude of the skeleton of the convolution kernel (the positions of the crisscross of the convolution kernels), which plays a vital role in the model performance. Firstly, the feature map is padding to a suitable size. Then three feature maps of the same size are obtained via a square convolution kernel, a one-dimensional convolution kernel in the horizontal direction, and a one-dimensional convolution kernel in the vertical direction. After that, the operation results of these three branches are summed element by element. The final

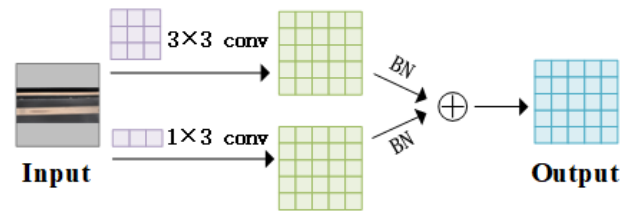


FIGURE 3. The structure of ACBlock.

fused results are regarded as the output of the asymmetric convolution module.

Instead of simply replacing the square convolution with the asymmetric convolution module of ACNet, we construct a new asymmetric convolution module for sliver defects in bamboo strip detection. We firstly analyze the effect of the asymmetric convolution block with either horizontal submodule or vertical submodule on the learning ability (verified in Section 4.2). Then, based on the extreme aspect ratio of sliver defects, the aspect ratio of half of the bamboo strip defects is more extensive than eight. Therefore, we remove the asymmetric convolution in the vertical dimension to reduce the interference of redundant information and heighten the influences of local power feature points in the horizontal dimension. The new asymmetric convolutional module contains square convolution and only one-dimensional convolutional branches added in the horizontal direction. We replace the 3×3 convolution block of the residual layer with an asymmetric convolution block and develop ACBottleneck as well as ACBottleneckCSP. The structure of ACBlock is presented in Figure 3. The feature maps are fed to the square convolution kernel and the horizontal direction 1D convolution kernel. Finally, the results of these two branches' operations after normalization are merged as the output of the asymmetric convolution module.

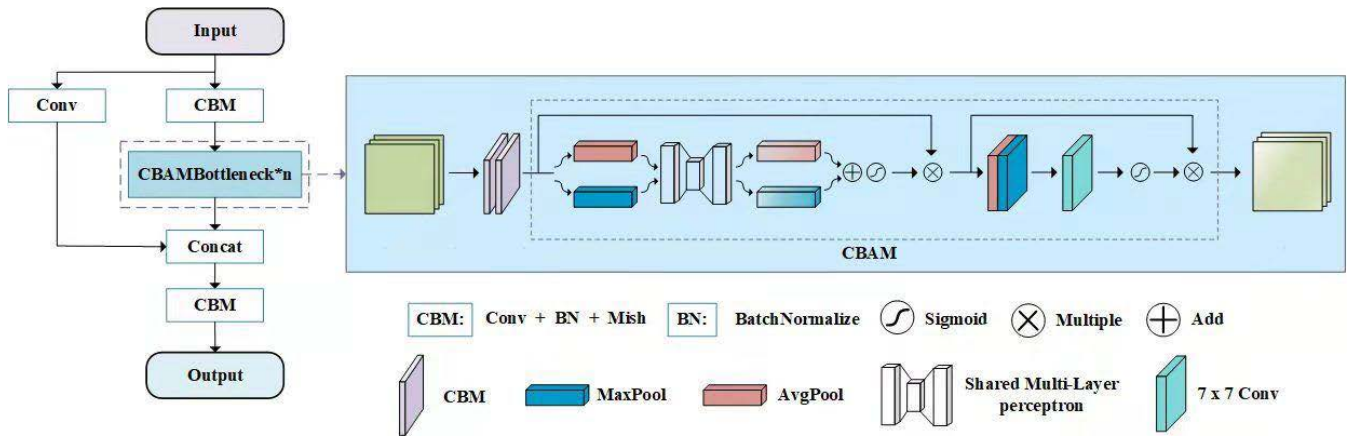


FIGURE 4. The network structure of CBAMBottleneckCSP2 and its sub-modules, CBAMBottleneckCSP2 is drawn on the left and CBAMBottleneck is on the right.

C. ATTENTION MECHANISM

The feature enhancement network further refines the features and raises representation power based on the backbone. We utilize CSPPAN, an excellent parameter aggregation path method, as the infrastructure of the neck. Besides, we apply the attention mechanism in the neck rather than the backbone, as richer semantic information of high-level features can induce the network to learn distinctive features properly. Attention plays a critical role in human perception [37]. Humans construct their cognition through a sequence of partial glimpses, naturally converting their focus on salient regions in complicated scenes. Motivated by human visual mechanisms, attention mechanisms have been extensively studied and broadly applied to computer vision tasks, such as image classification, object detection, semantic segmentation, object tracking.

Squeeze-and-excitation networks (SENet) [38] is the pioneer of channel attention, which generates attention mask across the channel domain and use it to select essential channels [39]. This module adaptively adjusts the weights of each channel by constructing the channel-wise relationships, thereby extracting key features. However, the attention captures spatial information via global average pooling, ignoring the local information within each channel. Woo *et al.* proposed convolutional block attention module(CBAM) [40], which introduced spatial attention complementary to the original channel attention. This module is a classic hybrid attention mechanism(channel attention & spatial attention) telling the network “what and where to focus on”. The intuition behind the mechanism is predicting channel and spatial attention masks separately and using it to select important features. The weights of channel attention and spatial attention in CBAM are calculated independently. Based on this, Misra *et al.* proposed convolutional triplet attention [41] to capture cross-dimension interaction. The mechanism models attention for the channel dimension C and the spatial dimension W , the channel dimension C and the spatial dimension H , and spatial dimensions H and W through three branches,

respectively. The refined feature maps are obtained by aggregating the three branches by simple averaging. The mechanism is an efficient and lightweight module that achieves improvements on large-scale datasets ImageNet, MS COCO datasets. We compare the above three attention mechanisms and select CBAM as the attention module in our model, which can boost the performance to the greatest extent(see detail in section 4.3).

In CBAM, given an input feature map, it generates the channel attention mask(1D tensor, $S_C \in R^C$) and the spatial attention mask(2D tensor, $S_S \in R^{H \times W}$) in turn. The channel sub-module adopts two pooling types, average pooling and max pooling, to gather global object features and distinct ones. Both feature descriptors are then sent to the multi-layer perceptron, and the results of the two branches are merged by element-wise addition. The sigmoid function normalizes the final output. The spatial sub-module likewise uses max pooling and average pooling and then concatenates them along the channel axis. Sequentially, the intermediate feature maps pass through a 7×7 convolution operation and are excited by the sigmoid function. This paper argues that the high-level rich semantic information is more conducive to constructing attention modules (verified in section 4.3). Therefore, we insert CBAM into the feature-enhanced network and build CBAMBottleneckCSP2 to replace the corresponding module BottleneckCSP2 in YOLOV4-CSP. The structure of CBAM-BottleneckCSP2 and its sub-modules (CBAMBottleneck and CBAM) are shown in Figure 4. The feature maps are also separated into two branches. One branch is subjected to convolution block, then channel attention and spatial attention are performed in turn at the CBAMBottleneck. The other one passes through a convolutional operation. The concatenated feature maps of the two branches after convolution operation are regarded as the output of CBAMBottleneckCSP2.

IV. EXPERIMENTS AND DISCUSSIONS

The main application scenarios of our research focus on defect inspection of the bamboo strip when it has already been

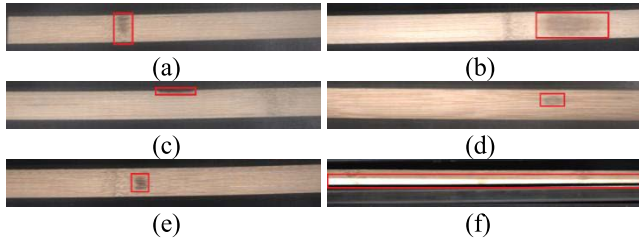


FIGURE 5. Six defects of bamboo strip: (a) black node; (b) concave bamboo yellow; (c) crack edge; (d) mildew; (e) scar; (f) tile.

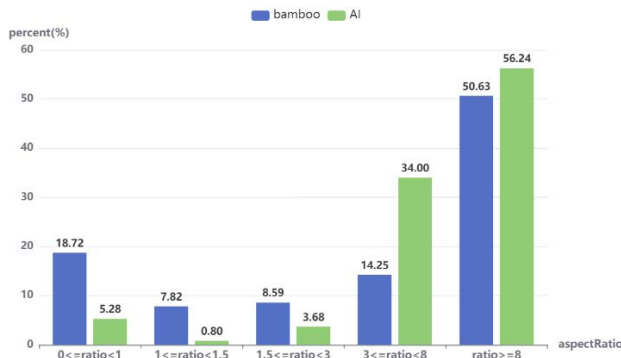


FIGURE 6. Comparison of aspect ratio between bamboo defects and aluminum defects.

formed and located in manufacturing workshops. We evaluate the improved YOLOV4-CSP in an enterprise's bamboo surface defect dataset. The dataset contains 1069 images, including six types of defects, namely black node, concave bamboo yellow, crack edge, mildew, scar and tile. The black node defect appears as a dark rectangular area, which generally covers the width of the bamboo strip. In most cases, the concave bamboo yellow defect is located in the middle of the bamboo strip, which is an oval and non-striped area of bamboo yellow. The crack edge defect is a small-area crack that is approximately straight and located at the edge of the bamboo strip. The mildew defect has a smaller gray value than the average area, with various shades. The scar defect usually consists of multiple thin, short, and dark stripes. The tile defect looks sliver in shape and lighter color (white or yellowish) than common areas. The sample of each type of defect is shown in Figure 5. Six defects are evenly distributed: about 178 pieces of each type are collected. The image size is floating up and down on the scale of 1024×450 . Each image is composed of the front and side of the bamboo strip and may contain multiple defects. The statistics on aspect ratios of bamboo defects are shown in Figure 6. We observe that more than 60% of the bamboo defect aspect ratio exceeds 3, and nearly 50% of the bamboo defect aspect ratio exceeds 8. This extreme aspect ratio issue differs from general objects, presenting challenges for strip workpiece detection.

In this paper, we propose the improved YOLOv4-CSP to deal with the above issue. To thoroughly evaluate the effectiveness of our final model, we conduct targeted ablation

experiments. Considering the relatively small scale of the dataset and the significant differences in brightness and contrast among samples, We first perform data augmentation via linear enhancement. Sequentially, we describe the ablation experiments of the asymmetric convolutional module design in detail. Then the investigations concerning the attention mechanism are shown. Next, we verify that the final design of improved YOLOv4-CSP outperforms other baselines without bells and whistles. Also, we extend the model to the aluminum dataset with sliver defects and examine the effect of sub-modules on the model.

The experiments are conducted on a computer with a single NVIDIA GeForce GTX 2080 GPU having 8GB memory. The calculation software environment is set with python 3.8.5, CUDA 10.1 and cuDNN 7.6.3. We implement all evaluated models in the PyTorch framework. The origin bamboo dataset is split into train set, validation set, and test set with 684, 171 and 214 images. The initial setting of the learning rate is 0.01, and the adjustment strategy of the cosine annealing learning rate is adopted. The momentum is set to 0.937, and weight decay is set to $5e-4$. We firstly train the network for 150 epochs with the above parameters. Then we set the learning rate to 0.001, momentum to 0.9, and fine-tune the network for 30 epochs. In the experiments, mean average precision (mAP@0.5) is calculated to evaluate the models.

A. IMAGE LINEAR ENHANCEMENT AND AUGMENTATION

The brightness of some images in the original dataset is far below the ideal level, so improving the image quality via image enhancement techniques [42] is imperative. This paper exploits the image grayscale linear transformation to optimize the dataset, a simple but efficient method. The grayscale transform function is shown in Equation 1. $G(x,y)$ represents the pixel value after the image grayscale transformation and $f(x,y)$ represents the pixel value before the transformation. The parameter α affects the image's contrast, and the parameter β affects the image's brightness. In this experiment, α takes 1.5 and β takes 10 to enhance image contrast and brightness.

$$G(x, y) = \alpha * f(x, y) + \beta(x, y) + \beta \quad (1)$$

Table 1 shows image enhancement and data augmentation experimental results. It brings improvements for most classes, showing the efficacy of image enhancement. Note that scar achieved a considerable improvement, lifting up by 13.33%. We attribute this phenomenon to the fact that scar is due to the traces left on the surface of bamboo by human factors, which are very similar to the bamboo strip texture. The results imply that enhancing the contrast and brightness of the image promotes the model to identify such background-like defects. We merge the original and enhanced images as the augmented dataset for sufficient training and validation in the following experiments. The augmented dataset has 1367 images for train, 343 images for validation, and 428 images for test.

TABLE 1. Accuracies of before and after image linear enhancement and augmentation.

Classes	YOLOV4-CSP	YOLOV4-CSP + aug	Accuracy ↑
mAP(%)	90.32	94.59	4.27
Scar(%)	85.37	98.70	13.33
Crack edge(%)	78.48	86.35	7.87
Concave bamboo yellow(%)	96.64	99.50	2.86
Black node(%)	99.50	99.24	-
Mildew(%)	82.46	85.54	3.08
Tile(%)	99.5	98.18	-

TABLE 2. Ablation study of different dimensions in asymmetric convolution.

Model	Square conv	Horizontal conv	Vertical conv	Parameters(M)	mAP(%)
YOLOV4-CSP + aug	✓			9.13	94.59
YOLOV4-CSP + aug	✓	✓		9.91	95.39
YOLOV4-CSP + aug	✓		✓	9.91	95.69
YOLOV4-CSP + aug	✓	✓	✓	9.91	95.23

TABLE 3. Comparison of different placement of asymmetric convolution block.

Model	Backbone	Neck	Parameters(M)	mAP(%)
YOLOV4-CSP + aug			9.13	94.59
YOLOV4-CSP + aug	✓		9.91	95.69
YOLOV4-CSP + aug		✓	9.75	95.47
YOLOV4-CSP + aug	✓	✓	10.52	95.66

B. ACBLOCK DESIGN

In this subsection, we further explore the practical method of integrating ACBlock. Table 2 summarizes the experimental results of different ACBlock morphologies for bamboo strip defect detection. Of note is that ACBlock in Table 2 is inserted into the residual structure of the backbone(Bottleneck and BottleneckCSP). We observe that models with different types of ACBlock achieve improvements simultaneously. However, adding asymmetric convolution in the horizontal direction yields the best results with considerably low computational cost. We believe this is closely related to the object instance characteristics: extreme aspect ratio, indicating that enhancing the features in the horizontal direction matters for improving strip workpiece detection. As mentioned in ACNet [36], adding one-dimensional asymmetric convolution may result in a stronger or weaker kernel skeleton owing to randomly initialized horizontal and vertical kernels. The authors of ACNet have empirically observed that adding horizontal as well as vertical asymmetric convolution is effective in ImageNet. In bamboo strip detection, we find that adding only asymmetric convolution in the horizontal axis is more helpful than adding both axes. We assume that adding asymmetric convolution in the vertical direction after adding asymmetric convolution in the horizontal direction may weaken the weights of features in the horizontal axis, which leads to the loss of some helpful information.

We continue to study how the model will behave if we add the asymmetric convolution onto other positions rather than the residual structure of the backbone. These experiments introduce asymmetric convolution into the standard 3 × 3 convolution in the residual block. We compare the

effects of the three placements: the asymmetric convolution in backbone, neck, or both backbone and neck. As shown in Table 3, the residual module with asymmetric convolution block placed at backbone has comparable accuracy but fewer parameters than that with both backbone and neck. Moreover, the asymmetric convolution residual structure at the backbone performs better than when placed at the neck, implying that the low-level semantic information and more accurate location information are more beneficial for the asymmetric convolution block. As a brief conclusion, we add asymmetric convolution in the horizontal direction based on the square convolution and deploy it on the backbone of the model in the following experiments.

C. ATTENTION MECHANISM DESIGN

To delve into the design of the attention mechanism module based on YOLOV4-CSP after adding asymmetric convolution block, we conduct a series of ablation experiments, the corresponding results presented in Table 4 and Table 5. We first reserve CSPSP and compare the performance of models with different attention mechanisms, namely SE, CBAM and triplet attention. From Table 4, we observe that integrating either SE or CBAM or triplet attention into YOLOv4-CSP can all achieve performance gains, demonstrating the effectiveness of dynamically adjusting the weights in the attention mechanism on the learning ability of the model. When introducing CBAM, the model performs slightly better than the other two attention mechanisms.

Given that CSPSP and the attention mechanism may have some functional duplication in increasing the recep-

TABLE 4. Comparison of different attention mechanisms when reserving CSPSP (baseline means YOLOV4-CSP adding asymmetric block after data augmentation).

Description	Parameters(M)	mAP(%)	mAP(%) ↑
baseline	9.91	95.69	-
+ SE	9.92	96.01	0.32
+ CBAM	9.92	96.22	0.53
+ Triplet attention	9.91	96.11	0.42

TABLE 5. Comparison of models with attention methods when removing CSPSP (baseline means YOLOV4-CSP adding asymmetric block after data augmentation).

Description	Parameters(M)	mAP(%)	mAP(%) ↑
baseline	9.91	95.69	-
+ SE	9.19	95.42	-
+ CBAM	9.20	96.74	1.05
+ Triplet attention	9.18	96.37	0.68

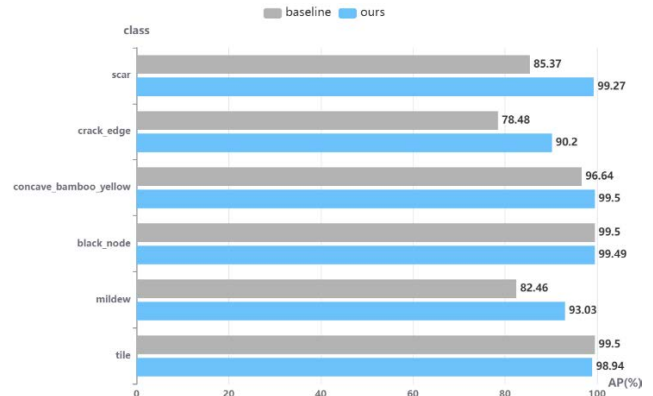
tive field in terms of the spatial dimension, we attempt to remove CSPSP and see the performance change. As shown in Table 5, the model with CBAM performs favorably against the other two attention modules. The intuition behind the phenomenon is that CSPSP raises representational capability in the spatial dimension via convolution operations with different kernel sizes, which can compensate for the spatial attention for SE. Therefore, removing CSPSP may lead to the loss of some crucial location information for the model with SE. We argue that the model with CBAM is superior to the one with triplet attention is two-fold. First, triplet attention builds cross-dimension interaction compared with CBAM yet may cause the loss of some informative features. We go back to the construction of cross-dimension interaction in the module, which includes rotation operation, while some bamboo features do not have rotation invariance. Second, the final output is obtained by averaging the three branches, which does not seem to work for defect detection of strip workpieces. As mentioned above, we believe that horizontal direction(w dimension) weights are more important in the scenario. The averaging operation may result in the loss of discriminative features along the w axis. The model combined with CBAM achieves better performance after removing CSPSP, which we attribute to reducing redundant information and enhancing salient features. In short, we adopt CBAM as the attention module for our model.

D. COMPARISON WITH OTHER BASELINES IN BAMBOO DEFECT DATASET

Throughout the ablation studies, we add the horizontal asymmetric convolution to the residual structure of the backbone and introduce CBAM to the neck meanwhile removing CSPSP as the final design. To thoroughly verify the effectiveness of our model, we compared the improved YOLOV4-CSP with some advanced one-stage detectors, which meet the requirements of real-time detection in industrial scenarios. As can be observed in Table 6, our model achieves optimal

TABLE 6. Comparison with state-of-the-art one-stage detectors in bamboo dataset.

Method	Parameters(M)	GFLOPS	mAP(%)
SSD	24.28	61.38	83.71
CenterNet	32.66	70.06	88.40
YOLOV4	63.96	59.78	90.83
YOLOV4-CSP	9.13	20.70	90.32
Ours	9.20	21.70	96.74

**FIGURE 7.** The comparison results of six defects between our model and the original YOLOv4-CSP.**TABLE 7.** Ablation study of the improved yolov4-CSP in aluminum dataset (we introduce CBAM while removing based on analysis above).

Description	mAP(%)	mAP(%) ↑
YOLOV4-CSP	92.99	-
+ ACBlock	93.61	0.62
+CBAM	95.54	2.55

accuracy, which is 13.03%, 8.34%, 5.91%, 6.42% higher than SSD, CenterNet, YOLOV4, and YOLOv4-CSP, respectively. Our model retains the lightweight characteristics of YOLOv4-CSP and achieves performance improvements at a low additional computational cost. The comparison results of six defects between our model and the original YOLOv4-CSP are shown in detail in Figure 7. Our model achieves substantial improvements on hard-to-recognize classes(mildew, crack edge, scar) and maintains comparable performance on easy-to-recognize classes(tile, black node). These results demonstrate that our model is adequate for bamboo strip defect detection.

E. EXTENSION TO THE ALUMINUM PROFILE DATASET

We also extend the improved model to the aluminum dataset. The dataset comes from the Tianchi Aluminum Profile Competition organized by the People's Government of Guangdong Province, China and Alibaba Group. The aluminum dataset contains five kinds of defects, namely non-conductivity, bottom leakage, scratches, pits, and variegated colors. From Figure 6, we observe that the aspect ratio of more than half of the defects in the dataset is greater than 8,

TABLE 8. Comparison with state-of-the-art one-stage detectors in aluminum dataset.

Method	Parameters(M)	Model Size(M)	mAP(%)
SSD	24.28	97.1	91.74
CenterNet	32.66	130.9	92.46
YOLOV4	63.96	256.2	93.40
YOLOV4-CSP	9.13	18.6	92.99
Ours	9.20	18.7	95.54

similar to the bamboo strip defect dataset. We divide the dataset with 867 images for training and 154 for validation and follow the same experimental protocol as specified at the beginning of Section 4. Table 7 and Table 8 show corresponding results on the aluminum dataset. The results show that our model also yields better accuracy than YOLOv4-CSP and outperforms other baselines with fewer parameters and smaller model size. The results show that our model infers more advanced features for sliver defect detection, implying that our model may provide a reference for other similar scenes.

V. CONCLUSION

Bamboo surface defect detection is of great significance to the ordinary operation of bamboo products in manufacturing workshops. The timely discovery of these defects can provide an early warning to workers. For intelligent defect detection, we propose improved YOLOV4-CSP, an object detector based on deep learning. We design a new asymmetric convolution block to enhance the ability to extract features in the horizontal direction owing to the sliver defect with an extreme aspect ratio. The attention mechanism, CBAM, is further introduced to learn salient features of the bamboo strip. Extensive experiments are conducted to verify the efficacy and efficiency of the improved YOLOV4-CSP algorithm. Of note, our model may provide a solution for sliver defect detection with similar extreme aspect ratios.

REFERENCES

- [1] Z. Y. Li, T. Trinh, N. Li, J. Wu, Y. Feng, and Y. Hao, "Bamboo resources and utilization in major Asian countries," *World Bamboo Rattan Commun.*, vol. 18, no. 98, pp. 7–13, 2020.
- [2] J. Q. Wu, "Research on the competitiveness of China's bamboo industry based on international trade," Chinese Acad. Forestry, Beijing, China, Tech. Rep., 2010.
- [3] H. Y. Qiu, J. F. Xu, and Z. He, "Bamboo as an emerging source of raw material for household and building products," *Bioresources*, vol. 14, no. 2, pp. 2465–2467, 2019.
- [4] S. Qi, J. Yang, and Z. Zhong, "A review on industrial surface defect detection based on deep learning technology," in *Proc. 3rd Int. Conf. Mach. Learn. Intell.*, 2020, pp. 24–30.
- [5] B. Tang, J. Kong, and S. Wu, "Review of surface defect detection based on machine vision," *J. Image Graph.*, vol. 22, no. 12, pp. 1640–1663, 2017.
- [6] L. Chen and S. Song, "Defect detection of mahjong bamboo block based on image processing," *J. Guangxi Univ.*, vol. 42, pp. 1069–1077, Feb. 2017.
- [7] Q. XianSheng, S. Xin, L. Qiong, and H. Feng, "Online detection and sorting system of bamboo strip based on visual servo," in *Proc. IEEE Int. Conf. Ind. Technol.*, Feb. 2009, pp. 1–5.
- [8] B. Q. Huang, "Research on surface defect detection and color classification of long bamboo strips based on machine vision," Guangxi Normal Univ., Guilin, China, Tech. Rep., 2018.
- [9] H. Kuang, Y. Ding, R. Li, and X. Liu, "Defect detection of bamboo strips based on LBP and GLCM features by using SVM classifier," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 3341–3345.
- [10] Y. N. Li, X. J. Lin, J. P. Chen, and S. X. Chen, "Defect recognition of carbonized bamboo strips based on gray level co-occurrence moment and SVM," *J. Jimei Univ.*, vol. 22, pp. 49–54, Oct. 2017.
- [11] L. Yeni and Y. Shaowei, "Defect inspection system of carbonized bamboo cane based on LabView and machine vision," in *Proc. Int. Conf. Inf. Commun. Eng. (ICICE)*, Nov. 2017, pp. 314–317.
- [12] Z. Q. Yan, "Texture feature analysis and application of chordal section image of Moso Bamboo," Fujian Agricult. Forestry Univ., Fujian, China, Tech. Rep., 2014.
- [13] F. He, X. S. Qin, Q. Liu, and X. Song, "Bamboo strip defect detection algorithm based on wavelet multi-scale decomposition," *Mechatronics*, vol. 1, pp. 46–49, Oct. 2010.
- [14] Q. F. Lu, "Research on bamboo strips' defect recognition technology based on machine vision," Fujian Agricult. Forestry Univ., Fujian, China, Tech. Rep., 2019.
- [15] T. Sun, X. W. Mou, S. X. Song, and J. H. Li, "Defective edge detection system design for long bamboo batten," in *Proc. Int. Conf. Electron. Ind. Autom.*, 2017, pp. 246–250.
- [16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 580–587.
- [17] R. Girshick, "Fast R-CNN," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 91–99.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788.
- [20] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 7263–7271.
- [21] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Springer, 2016, pp. 21–37.
- [24] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6569–6578.
- [25] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.
- [26] C. Y. Wang, H. Y. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of cnn," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 390–391.
- [27] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 13029–13038.
- [28] J. Li, "Research on cloth defect detection algorithm based on deep learning," University of Electronic Science and Technology, Sichuan, China, Tech. Rep., 2020.
- [29] R. Hao, B. Lu, Y. Cheng, X. Li, and B. Huang, "A steel surface defect inspection approach towards smart industrial monitoring," *J. Intell. Manuf.*, vol. 32, no. 7, pp. 1833–1843, 2020.
- [30] Q. Gao, B. Huang, and W. Liu, "Bamboo strip surface defect detection method based on improved CenterNet," *Comput. Appl.*, vol. 41, no. 7, pp. 1–8, 2020.
- [31] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of carefuleseeding," in *Proc. Symp. Discrete Algorithms*, 2007, pp. 1–14.
- [32] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 764–773.
- [33] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable convnets V2: More-deformable, better results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jan. 2019, pp. 9308–9316.
- [34] E. L. Denton, W. Zaremba, J. Bruna, Y. LeCun, and R. Fergus, "Exploiting-linear structure within convolutional networks for efficient evaluation," in *Proc. Adv. Neural Inf. Proc. Syst.*, 2014, pp. 1269–1277.

- [35] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Speeding up convolutional neural networks with low rank expansions," *Computing*, vol. 4, no. 4, p. 8, 2014.
- [36] X. Ding, Y. Guo, G. Ding, and J. Han, "ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Jan. 2019, pp. 1911–1920.
- [37] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Rev. Neurosci.*, vol. 3, no. 3, pp. 201–215, 2002.
- [38] H. Jie, S. Li, S. Gang, and S. Albanie, "Squeeze-and-excitation networks," *Tech. Rep.*, 2018.
- [39] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," 2021, *arXiv:2111.07624*.
- [40] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
- [41] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. Hou, "Rotate to attend: Convolutional triplet attention module," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021.
- [42] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement methods," *IEEE Access*, vol. 8, pp. 87884–87917, 2020.



YI QIU received the B.S. degree from Zhejiang University, in 2007, and the M.S. degree in 2010. He has worked with the Motorola Hangzhou Research and Development Center, Meiya Pico, and AutoNavi, with four years of software development experience. Since 2014, he has been working with the Tan Kah Kee College, Xiamen University. He has engaged in developing industrial robot systems, computer vision, and the IoT for a long time.



XUQIANG ZHOU received the B.S. degree from Hangzhou Dianzi University, Hangzhou, China, in 2019. He is currently pursuing the M.S. degree in computer technology with the School of Information, Xiamen University, Xiamen, China.



YIJING GUO received the B.S. and M.S. degrees from Xiamen University, Xiamen, China, in 2002 and 2005, respectively. Since Spring 2015, he has been an Associate Professor with the School of Information Science and Technology, Tan Kah Kee College, Xiamen University, Zhangzhou, China. His current research interests include machine vision and intelligent algorithm.



YIXIN ZENG received the B.S. degree from Fujian Normal University, Fuzhou, China, in 2019. She is currently pursuing the M.S. degree in computer technology with the School of Information, Xiamen University, Xiamen, China.



LINWEI ZHONG received the B.S. degree from South China Agricultural University, Guangzhou, China, in 2019. He is currently pursuing the M.S. degree in computer technology with the School of Information, Xiamen University, Xiamen, China.



FENGQIANG GAO received the B.S. and M.S. degrees in detection technology and automatic equipment from Xiamen University, China, in 2007 and 2010, respectively, where he is currently pursuing the Ph.D. degree in systems engineering with the Department of Automation. His research interests include deep learning and computer vision detection.



CHOUJUN ZHAN (Member, IEEE) received the B.S. degree in automatic control engineering from Sun Yat-sen University, Guangzhou, China, in 2007, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, in 2012. After graduation, he worked as a Postdoctoral Fellow with The Hong Kong Polytechnic University. His research interests include complex networks, time series modeling and prediction, epidemic spreading, information diffusion, and machine learning.