IEEE Open Journal of the
Industrial Electronics Society

Received 26 May 2022; revised 14 July 2022; accepted 21 July 2022. Date of publication 25 July 2022;
date of current version 10 August 2022. The review of this paper was arranged by Associate Editor Guanghui Wen.

Digital Object Identifier 10.1109/OJIES.2022.3193572

Surface Defect Detection in Sanitary Ceramics Based on Lightweight Object Detection Network

JINGFAN HANG [1], HAO SUN [2], XINGHU YU [3], JUAN J. RODRÍGUEZ-ANDINA [4] (Senior Member, IEEE), AND XIANQIANG YANG [1] (Member, IEEE)

[1]Research Institute of Intelligent Control and Systems, Harbin Institution of Technology, Harbin 150001, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China
[3]Ningbo Institute of Intelligent Equipment Technology Company Ltd., Ningbo 315200, China
[4]Department of Electronic Technology, University of Vigo, 36310 Vigo, Spain

CORRESPONDING AUTHORS: XIANQIANG YANG; JUAN J. RODRÍGUEZ-ANDINA (e-mail: xianqiangyang@hit.edu.cn; Juan.J.R@ieee.org)

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1308400 and in part by the Joint Funds of the National Natural Science Foundation of China under Grant U20A20188.

ABSTRACT Sanitary ceramic products, such as toilet and wash basin, are widely used in our daily life. Sanitary ceramics are expected to have some excellent physical properties, such as corrosion resistance, easy cleaning, and low water absorption. However, surface defects in sanitary ceramics are inevitable due to complex production processes and changing production environment. Therefore, surface defect detection must be performed in the manufacturing process of sanitary ceramics. There are many types of surface defects in sanitary ceramics, and different types of defects have large differences in characteristics and scales. Traditional detection methods with artificially designed features and classifiers are difficult to apply in this context. In addition, there are few studies on surface defect detection methods of sanitary ceramics based on deep neural networks. In this article, a lightweight real-time defect detection network based on the lightweight backbone MobileNetV3 is presented. The proposed network achieves multi-scale detection of surface defects in sanitary ceramics with a multi-layer feature pyramid. Combining region proposal network and anchor-free method, real-time defect detection is achieved. Finally, a detection head with channel attention structure and a low-level mixed feature classification strategy is used to perform defect classification with higher accuracy. Experimental results show that the proposed approach achieves at least 22.9% detection speed improvement and 35.0% average precision improvement while reducing memory consumption by at least 8.4% compared with the classic one-stage SSD, YOLO V3 and two-stage Faster R-CNN methods.

INDEX TERMS Sanitary ceramic, surface defect detection, deep learning.

I. INTRODUCTION

Sanitary ceramic products are indispensable in modern households. Though the market scale is huge, the fierce competition among production companies prompted them to innovate their production technology in order to produce higher quality products with higher efficiency. The production processes of sanitary ceramic products are very complex and surface defects may occur in any link of production. Products with surface defects entering the market cause economic losses and reduce the brand value of companies. Therefore, surface defect detection is essential in sanitary ceramic production processes.

Most of the sanitary ceramic manufacturers researched by the authors still detect surface defects in sanitary ceramics manually. The main reasons for this are as follows: firstly, the surface structure of sanitary ceramics is complex, some defects are difficult to detect in images, and the features of the same type of defects in different positions are not homogeneous; secondly, there are many kinds of surface defects of sanitary ceramics, and the characteristics of different defects

are quite different, which makes it difficult to design a unified feature description system to describe these defects using traditional image processing methods, or to detect multiple defects with a single detector; thirdly the surface defects of sanitary ceramics have the characteristics of non-structural, multi-scale, scale anisotropy and few samples, which make it more difficult to detect. With the development of deep learning technology, artificial neural networks have been widely used in many kinds of areas. Especially, convolutional neural networks (CNNs) have significant advantages in image processing tasks due to their parameter sharing structure and strong fitting ability, which makes it possible to detect surface defects of sanitary ceramics.

The main contributions of this work are as follows:

1) A lightweight sanitary ceramic surface defect detection network is proposed. Experimental results show that the network has faster detection speed and higher detection accuracy than the classic one-stage SSD and two-stage Faster R-CNN methods;

2) The proposed network has a small number of parameters, and achieves better detection results without the need for carefully tune a large number of hyperparameters, which makes it suitable for terminal devices with limited computing resources, and can be quickly applied to actual production environments;

3) The proposed network adopts a low-level mixed features reclassification strategy, which can significantly improve the detection accuracy without modifying network structure or increasing network parameters.

The remainder of the article is structured as follows. Previously reported related works are analyzed in Section II. The proposed method is presented in Section III and validated by the results of the experiments discussed in Section IV. Finally, Section V concludes the article.

II. RELATED WORKS

At present, there are many surface defect detection methods based on traditional image processing and deep learning methods. Tang et al. [1] proposed a method for ceramic valve spool defect detection based on template matching. The method achieves 8% missed detection rate in test data by performing object matching in a multi-level image pyramid. Zhang et al. [2] proposed a surface defect detection method for complex texture ceramic tiles based on traditional image processing methods, using canny edge detection and defect saliency detection, and support vector machines for defect classification. Chen et al. [3] proposed a ceramic decals defect detection method also based on traditional image processing methods. It eliminates imaging glare through homomorphic filtering, and the careful design of a decals out-of-bounds detection pipeline. In recent years, the development of deep learning technology led to its application in surface defect detection with good results in terms of classification accuracy. Lin et al. [4] constructed a multi-scale cascaded CNN based on MobileNetV2, with a reduced number of parameters thanks to the use of a lightweight backbone. Xiao et al. [5] proposed a method for surface defect detection based on Mask R-CNN and image pyramid. It fuses the image and feature pyramids, which improves the detection effect of multi-scale objects. Tabernik et al. [6] proposed a surface defect detection method based on image segmentation. In this case, the network performs coarse-grained segmentation of low-resolution feature maps after multiple pooling operations on the input image. Huang et al. [7] proposed a surface defect detection system for ceramic cell phone backplane based on YOLOv3-tiny. The detection program is running on an ARM embedded platform, achieving 89.9% recognition accuracy and 2 fps detection speed. Huang et al. [8] proposed a ceramic substrate defect detection method based on an unbalanced data set. It uses unsupervised K-means clustering to oversample and downsample unbalanced data to achieve the balance of training samples. The ResNet-based deep neural network is used for feature extraction and defect detection. Li et al. [9] proposed a two-stage defect detection method, which uses shared features in the defect discovery and defect classification stages to achieve efficient defect detection. Stephen et al. [10] proposed a simple CNN model for the detection of ceramic tiles surface defects. In the detection experiment, the method achieved a classification accuracy of 99.43%, but for high-resolution input images, resulting in a long detection time, 16 s.

The surface defect detection of sanitary ceramics can be regarded as an image-based object detection task. For example, the classic two-stage object detection network Faster R-CNN [11] can be used with this purpose. It first extracts potential object regions through a region proposal network (RPN), then the detection head is used to classify and rectify the position of the objects. The network has good detection accuracy, but because of the large number of parameters of its fully connected layer, it implies the use of a significant amount of computing resources. In addition, since Faster R-CNN is an anchor-based method, it consumes a lot of time when performing non-maximum suppression (NMS), which makes it difficult to apply to tasks with high real-time requirements. Finally, since its detection head makes predictions in the last feature layer of the network, it is not good at small objects detection. Feature Pyramid Networks (FPN) [12] provide a good idea for solving multi-scale target detection problems, but do not solve the real-time requirements of anchor-based methods. On the other hand, since object detection needs to be performed at multi-scale feature maps, the anchor-based method takes more time to complete the detection in images, which makes it difficult to be applied in industry. The Single Shot MultiBox Detector (SSD) [13] performs target detection at multi-scale feature maps and is capable of carrying out multi-scale target detection tasks. The SSD300 network achieves 59 fps detection speed, much faster than Faster R-CNN. However, although the SSD network has better detection results for small objects, its overall detection accuracy is not as good as that of Faster R-CNN. In addition, it is difficult to tune the hyperparameters of the prior boxes, which makes SSD hard to use in actual industrial production. For the classic one-stage detection network YOLOv3 [14], the

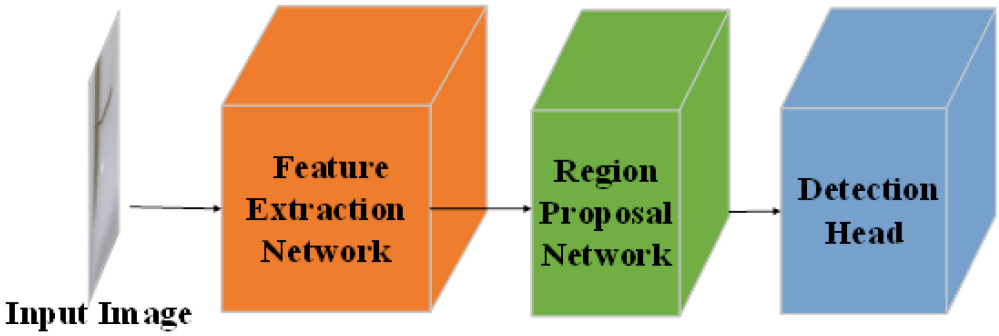


FIGURE 1. The overall structure of the network.

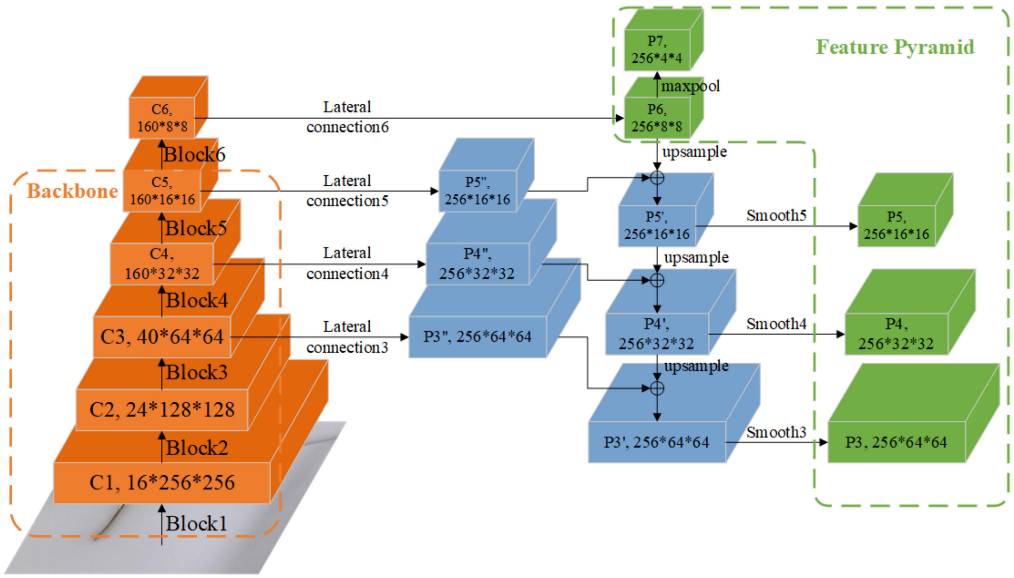


FIGURE 2. Feature extraction network.

network has a good balance between speed and accuracy. Compared with the Faster R-CNN and SSD networks, detection speed is faster but detection accuracy is still not as good as that of Faster R-CNN. It is worth mentioning that the fully connected layer of YOLOv3 contains a large number of weight parameters, which is a significant drawback for terminal devices with limited computing resources. The lightweight detection network MobileNetV3 [15] effectively reduces the amount of network parameters through depthwise separable convolution, which is very suitable for deployment in mobile terminals with limited computing resources. CenterNet [16] is an anchor-free target detection method that locates objects by finding their center and estimating the discretization error of object center and object size. Since there is no NMS step, the inference speed of this network is much higher than that of anchor-based methods. Yan et al. [17] proposed an infrared and visible image fusion algorithm which obtained fusion results suitable for human visual perception. Zhang et al. [18] proposed a novel approach to addressing the fusion of multi-focus images in either registered or mis-registered cases which provides better visual perception and higher objective evaluation.

Surface defect detection in sanitary ceramics belongs to image detection task. Considering the detection speed, computing resource consumption, hyperparameter adjustment, etc., the existing methods are difficult to apply to the surface defect detection of sanitary ceramics.

Inspired by the above networks, a lightweight CNN suitable for surface defect detection in sanitary ceramics is proposed in this article. It provides a solution that can be used in practical industrial production and narrows the gap between theory and application.

III. METHOD

A. NETWORK ARCHITECTURE

The proposed system shown in Fig. 1 consists of three parts, namely feature extraction network, RPN, and detection head. Fig. 2 shows the feature extraction network. It uses MobileNetV3_Large [15] as backbone, and constructs a

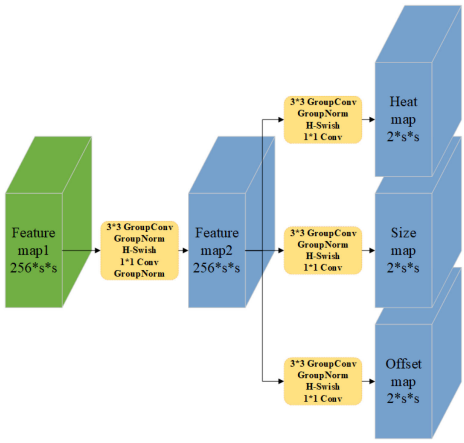


FIGURE 3. RPN network.

TABLE 1 Backbone Composition

Block	Operator	Input size	Output size
Block1	conv2d,3*3 bneck,3*3	$3*512^2$	$16*256^2$
Block2	bneck,3*3 bneck,3*3	$16*256^2$	$24*128^2$
Block3	bneck,5*5 bneck,5*5 bneck,5*5	$24*128^2$	$40*64^2$
Block4	bneck,3*3 bneck,3*3 bneck,3*3 bneck,3*3 bneck,3*3 bneck,3*3 bneck,5*5	$40*64^2$	$160*32^2$
Block5	bneck,5*5 bneck,5*5	$160*32^2$	$160*16^2$
Block6	bneck,3*3 bneck,3*3	$160*16^2$	$160*8^2$

five-layer feature pyramid, P3 to P7 in Fig. 2. According to the output feature map size of each layer in the backbone, it can be divided into five blocks. The feature maps generated by each block can be represented as C1 to C5, respectively. Different from [12], due to the large size of C2, which does not meet the lightweight goal, it does not participate in the construction of the feature pyramid. Therefore, block6 is added to the original backbone. C3–C5 are transformed into three temporary feature maps P3"-P5" through the lateral connection 3–5. C6 is transformed into feature map P6 through the lateral connection 6. And P6 is transformed into feature map P7, which has a larger receptive field through max pool. Afterwards, the feature map P6 is subjected to continuous bilinear interpolation from top to bottom to obtain feature maps with larger size, which are added to the three temporary feature maps to obtain feature maps P5,' P4,' P3,' respectively. These are transformed into the final feature maps P5-P3 through the smooth operation. For an RGB image of size 512*512, the output sizes of all feature maps are 256*64*64, 256*32*32, 256*16*16, 256*8*8 and 256*4*4, respectively.

RPN is shown in Fig. 3. The input to this network is the five-level feature maps output from the feature extraction network. Input feature map1 is transformed into feature map2, which has the same size as feature map1 through a depthwise separable convolution. Feature map2 is sent to the three branches to complete object center point detection, object size regression, and center position discretization error regression tasks, respectively. The central position where there may be a foreground can be determined through the generated heatmap. The specific position of the potential foreground targets in the current feature map can be determined according to the regression results of the size and offset maps at the corresponding positions. These potential foreground targets are used for subsequent network training or inference.

The detection head is shown in Fig. 4. The original detection head of Faster R-CNN [11] is too heavy for a lightweight detection network and it is hard to train with a small dataset. In this case, a detection head with a channel attention module has been designed. The input of the detection head is the sub-feature map with fixed size obtained from feature maps P3-P7 after ROI extraction. Input feature map1 is transformed into feature map2 and feature map3 through the continuous SE module [19] and expand depthwise separable convolution [15]. Afterwards, the final classification result is obtained through two fully connected layers. It is worth mentioning that the developed detection head only performs the computation of object confidence without further position rectification of the bounding boxes.

B. IMPLEMENTATION DETAILS

1) FEATURE EXTRACTION NETWORK

The backbone is divided into 6 blocks according to the feature map size of each layer. The composition of each block is shown in Table 1. All lateral connections are implemented through 1*1 pointwise convolution. Through lateral connections, the resulting number of channels in the feature map is 256, which is convenient for subsequent processing with RPN. The purpose of the smooth operation is to reduce the aliasing effect of upsampling [12], which is essentially a 3*3 convolution operation. In addition, due to the small amount of data used for network training and the limited GPU memory size, network training can only be performed with a small batch size, so all batch normalization layers are replaced by group normalization [20], thereby reducing the adverse effects caused by small batch size.

2) RPN

It uses the heatmap with two channels to predict the center point of the potential object. The judgment principle of the object center is that the score of the object center pixel in the foreground channel is greater than 0.5 and not less than the score of its 8 neighbor pixels. In particular, after performing a softmax operation on the heatmap generated by RPN and obtaining all pixel scores of the foreground channel, the 3*3 maxpooling operator is used to obtain the pooled foreground channel. Finally, the pixel scores of the foreground channel and the pooled foreground channel at the same position are

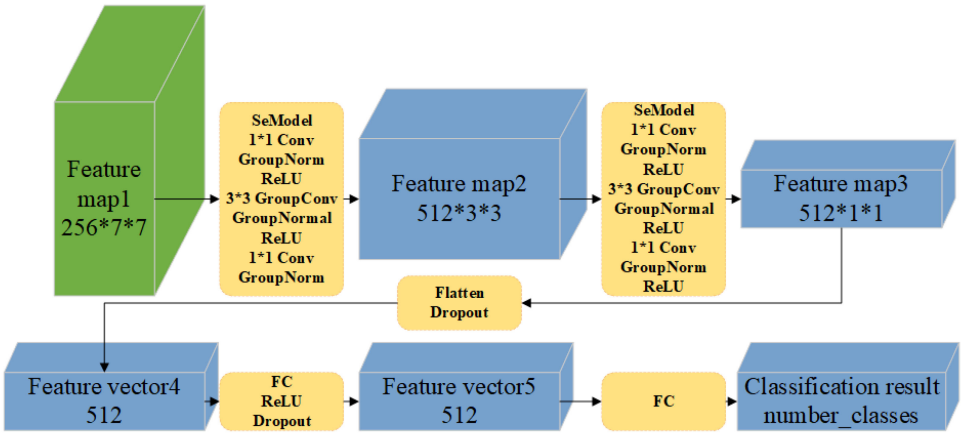


FIGURE 4. Detection head.

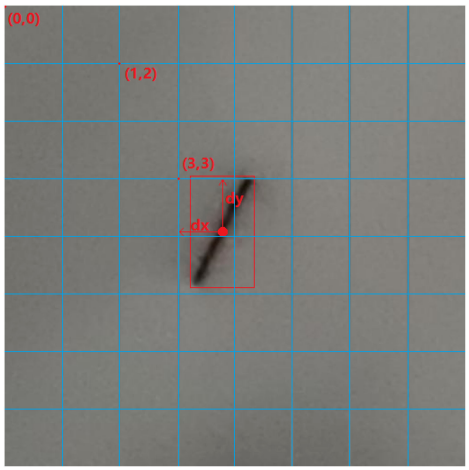


FIGURE 5. Example of coordinate discretization error.

compared. The pixels with the same score that are greater than 0.5 are the center of the potential objects. The sizemap and offsetmap regression branches of the RPN are both 2-channel maps, where channel 0 of sizemap corresponds to the object height h, channel 1 corresponds to the object width w, channel 0 of offsetmap corresponds to the discretization error dy, and channel 1 corresponds to the discretization error dx. An example of coordinate discretization error can be seen in Fig. 5. The center of the object is located in the pixel at position (3, 3). Through the coordinate discretization errors dy and dx, the center position of the object can be further refined as (3+dy, 3+dx). According to the position of the potential objects obtained by RPN, the corresponding ROIs in the current feature map, which can be used for training and inference in the subsequent detection head, are intercepted by ROI Align [21].

There are three loss calculations involved in the training of RPN [16]:

$$L_{RPN} = L_{heatmap} + \lambda_{sizemap} L_{sizemap} + \lambda_{offsetmap} L_{offsetmap} \quad (1)$$

where $L_{heatmap}$ is heatmap loss, $L_{sizemap}$ is sizemap loss, $\lambda_{sizemap}$ is sizemap loss weight, $L_{offsetmap}$ is offsetmap loss, and $\lambda_{offsetmap}$ is offsetmap loss weight, respectively.

Firstly, improved focal loss is used to calculate heatmap loss [16]:

$$L_{heatmap} = \frac{-1}{N} \sum \begin{cases} (1-\hat{y})^{\alpha} \log(\hat{y}), y = 1 \\ (1-y)^{\beta} \hat{y}^{\alpha} log(1-\hat{y}), y \neq 1 \end{cases} \quad (2)$$

where N is the number of positive objects in the feature map; y is the score for the current pixel to contain the object center, whose construction method is the same as in [16]; $\hat{y}$ is the confidence that the predicted pixel contains the object center; α is the difficulty and easy sample balance factor; and β is the negative sample weight factor. $(1-\hat{y})^{\alpha}$ and $\hat{y}^{\alpha}$ are the difficulty and easy sample balance coefficients. When positive or negative samples are easy to judge, $(1-\hat{y})^{\alpha} \to 0$ or $\hat{y}^{\alpha} \to 0$. $(1-y)^{\beta}$ is negative sample weight. When the negative sample center is far from the target center, $y \to 0$ and $(1-y)^{\beta} \to 0$.

Smooth L1 loss is used to calculate sizemap and offsetmap losses. Only the regression loss of positive samples is considered [22]:

$$L = \sum \begin{cases} 0.5(y-\hat{y})^2, |y-\hat{y}| < 1 \\ |y-\hat{y}| - 0.5, |y-\hat{y}| \geq 1 \end{cases} \quad (3)$$

where y represents the actual size and position parameters of the objects in the current feature map, and $\hat{y}$ is the predicted value of the corresponding parameter. When the difference between predicted value and target value is large ($|y-\hat{y}| \geq 1$), L1 loss ($|y-\hat{y}| - 0.5$) is used; otherwise ($|y-\hat{y}| < 1$), L2 loss ($0.5(y-\hat{y})^2$) is used.

3) DETECTION HEAD

The input of the detection head is a fixed-size feature map, sampled from the output feature maps P3-P7 of the feature extraction network. The sampling method is ROI Align [21], and the sampling frames are determined by the RPN network. Specifically, when training the detection head, the positive potential objects, whose intersections over unions (IoUs) are

TABLE 2 **Comparison of Detection Accuracy**

Method	Backbone	AP	AP^{50}	AP^{75}	AP^{small}	AP^{medium}	AP^{large}	AR^{1}	AR^{10}	AR^{100}	AR^{small}	AR^{medium}	AR^{large}
Faster R-CNN	VGG16	17.6	38.8	14.0	7.8	17.4	25.3	19.3	21.2	21.2	10.7	21.9	28.6
SSD	VGG16	24.2	43.8	24.4	**30.8**	28.0	19.9	27.9	29.6	29.6	**34.7**	35.3	23.2
YOLO V3	Darknet-53	23.1	43.1	22.5	23.6	24.6	22.8	25.6	27.3	27.3	26.9	28.8	25.8
Our method	MobileNetV3_Large	**32.7**	**60.6**	**32.6**	**27.3**	**34.7**	**36.6**	**37.3**	**39.4**	**39.4**	33.8	**41.7**	**42.1**

greater than the positive threshold, are assigned labels according to the defect category. Similarly, negative potential objects, whose IoUs are lower than the negative threshold, are assigned background labels of 0. The quantitative restrictions of positive and negative samples are determined by the hyperparameters pos_roi_num and neg_roi_num separately. When all positive and negative samples are determined, they are sent to the detection head for training. The classification loss is the only loss when training the detection head:

$$L_{Head} = L_{cls} \tag{4}$$

Cross entropy loss is used to compute the classification loss:

$$L_{cls} = \sum_{n=1}^{N} \sum_{i=0}^{C-1} y_i^n \log\left(\hat{y}_i^n\right) \tag{5}$$

where N is the number of samples; C is the number of defect categories (including the background category); $y^n = [y_0^n, y_1^n, \ldots, y_{C-1}^n]$ is the one-hot label representation of sample n (When sample n belongs to class i, $y_i^n = 1$; otherwise, $y_i^n = 0$); and $\hat{y}_i^n$ is the confidence that sample n belongs to category i.

In the inference stage, considering the conflict of feature requirements between localization and classification tasks [23], [24], the initial classification results of ROIs are not used as final results. Specifically, NMS processing is performed in all ROIs in the five feature maps of the same image according to their initial classification confidence, to determine the final ROIs. Final ROIs are used to perform feature resampling in the lower-level features (P3), which have more detailed information and high-level semantic information. After that, a reclassification on the final samples is performed. Reclassification results are the final detection results.

C. HYPERPARAMETER TUNING

1) HYPERPARAMETERS OF FEATURE EXTRACTION NETWORK

As mentioned above, due to the small amount of data used for network training and the small batch size during training, it is difficult for the original batch normalization layer of the feature extraction network to accurately estimate the mean and variance of the objects, resulting in bad test results. Therefore, all batch normalization layers in the model are replaced with group normalization (GN) layers. For the grouping parameters in the GN layer, the experimental results shown in Table 3 in [20] are considered, that is, when the number of channels in each group is 16, the error rate is the smallest. Therefore, when setting the number of groups of each GN layer, while ensuring that the GN layer will not degenerate into Layer Normalization [25], to the extent possible each group of the

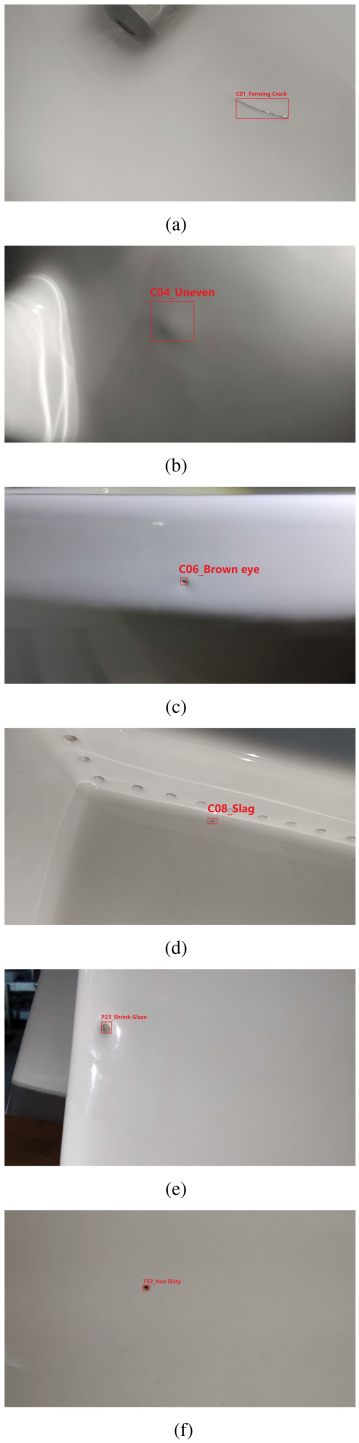


FIGURE 6. **Orginal image of surface defects. (a) C01_Forming Crack. (b) C04_Uneven. (c) C06_Brown eye. (d) C08_Slag. (e) P23_Shrink Glaze. (f) T52_Iron Dirty.**

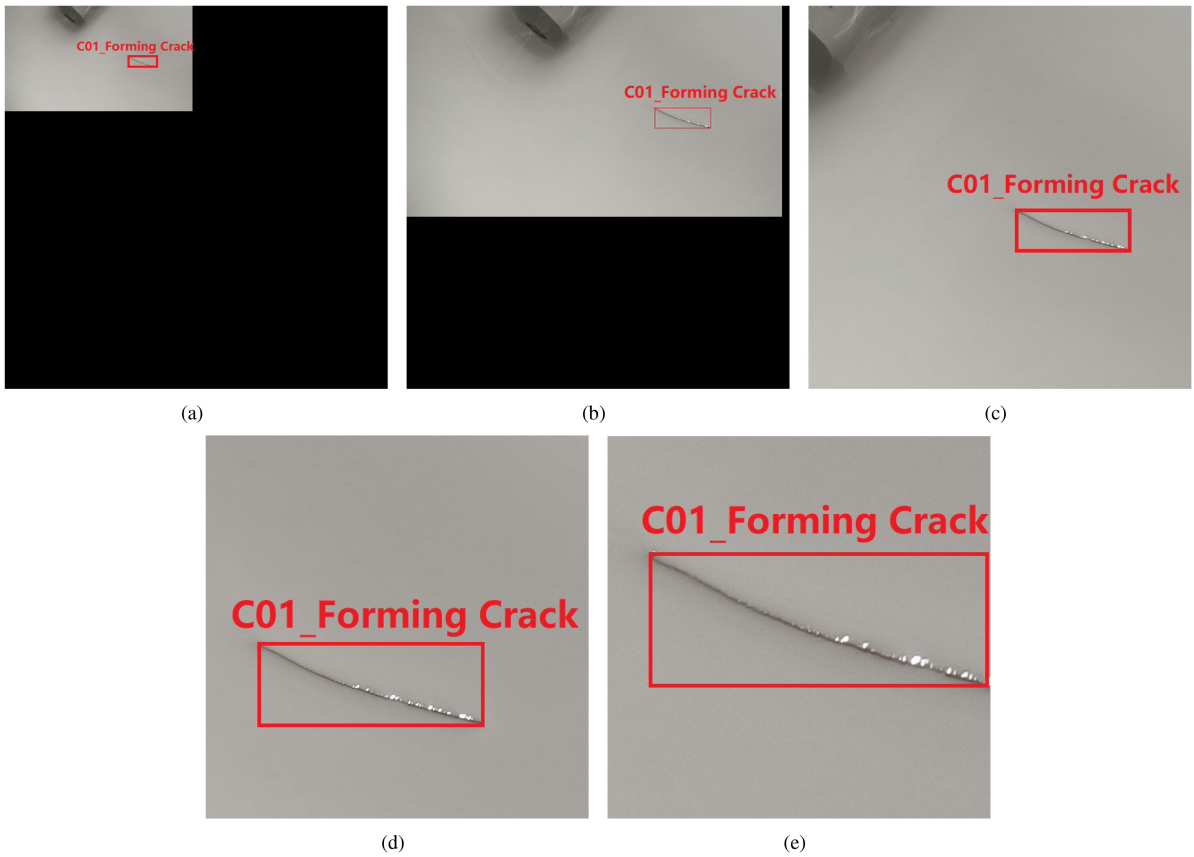


FIGURE 7. Cropped up images. (a) Image for P3. (b) Image for P4. (c) Image for P5. (d) Image for P6. (e) Image for P7.

GN layer is made to contain 16 channels. When the channel number of the input feature map is not divisible by 16, each group is intended to contain 8 channels, and so on.

2) HYPERPARAMETERS OF RPN

Firstly, for the improved focal loss, the heatmap construction method in [16] is considered, and $\alpha = 2$, $\beta = 4$. Then, for the sizemap loss weights $\lambda_{sizemap}$ and the offsetmap loss weigths $\lambda_{offsetmap}$, the setting method in CenterNet is considered, hence $\lambda_{sizemap} = 0.1$ and $\lambda_{offsetmap} = 1$.

3) HYPERPARAMETERS OF DETECTION HEAD

The input feature map size of the detection head is set to 256*7*7 according to [21]. The positive and negative sample thresholds are set to 0.5 and 0.45, respectively, which means ROIs with foreground confidence higher than 0.5 are regarded as positive samples, those lower than 0.45 are regarded as negative samples, and all others do not participate in training. The quantitative restriction hyperparameters of positive and negative samples are set to 16 and 48, respectively, again according to [21]. Since the proposed method is anchor-free, RPN extracts fewer ROIs, and the above two hyperparameters usually do not work in images with few defective objects.

IV. EXPERIMENTS

A. DATASET

In order to verify the effectiveness of the proposed method, images collected by ourselves on an actual sanitary ceramics production site were used to train and test the network. Specifically, the six types of defects shown in Fig. 6 were used. Before network training and testing, images were adjusted and randomly cropped. Each defect category contains 998 images, which are divided into training, evaluation, and test sets in a ratio of 8:1:1 randomly. The whole test was carried out 5 times, and all samples in the test set have no intersection. The results in Table 2. are the average of 5 tests.

For the details of how images are resized and randomly cropped, let us take Fig. 6(a) as an example. In order to efficiently train the feature extraction network, we need to provide training objects with multiple sizes. Considering that small objects are usually detected in feature layers with large output size, the training objects for P3 layer should be the smallest, and the smaller the output size of the feature layer, the larger the training object required. For an object with a square bounding box, when the bounding box size is 28*28, its pixel size in the P3 layer is 3*3, which matches the size of the RPN convolution kernel, and we call the object with 28*28 pixel size as ideal object for P3 feature layer. But since the bounding box of the actual

target is not necessarily a square, we take the diagonal length as the benchmark for object scaling. Based on the above reasons, the ideal object sizes corresponding to the P3-P7 feature layers are 28*28, 56*56, 112*112, 224*224 and 448*448, respectively, and the ideal object diagonal lengths are 39, 79, 158, 316 and 633. Furthermore, to increase the diversity of sample sizes, we add random Gaussian noise ($g(\mu, \sigma)$) to the above ideal sizes. For the defect object shown in Fig. 5a, its diagonal length is 505 pixels, so the five scales of the image are $(39/505) * (1 + g(\mu, \sigma))$, $(79/505) * (1 + g(\mu, \sigma))$, $(158/505) * (1 + g(\mu, \sigma))$, $(316/505) * (1 + g(\mu, \sigma))$ and $(633/505) * (1 + g(\mu, \sigma))$, where $\mu = 0$, $\sigma = 0.05$. Then the image is resized with above scales. Finally, sub-images are randomly cropped up with a resolution of 512*512 (as shown in Fig. 7) from the scaled images, respectively.

In addition, during network training all training data are further enhanced by random rotation and color jittering. The angle of random rotation can be 90°, 180°, 270°or 0°. And color jittering consists of a random sequence of changes in brightness, contrast, and saturation.

B. MAJOR EXPERIMENTS

1) INDICATORS

The performance of the methods is quantified by coco evaluation method [26]. In coco evaluation system, AP is the mean average precision of all categories across 10 IoU thresholds, and AP is the main indicator that determines the challenge winner. AP^{50} is the mean average precision of all categories when IoU is 0.50. The definition of AP^{75} is similar to that of AP^{50}. AP^{small} is AP for small objects whose area is smaller than 32^2 pixels. The area of medium objects is between 32^2 and 96^2 pixels. And the area of large objects is larger than 96^2 pixels. AR is the maximum recall given a fixed number of detections per image, averaged over all categories and 10 IoU thresholds. So, AR^1 means the maximum recall when no more than one object is detected per image. AR^{small}, AR^{medium} and AR^{large} are the AP for small, medium, and large objects separately.

The computation resource consumption is measured by memory usage and floating-point operations. And the computation resource consumption is considered in the inference step.

2) DETECTION ACCURACY

The proposed method is compared with classic one-stage (SSD, YOLO V3) and two-stage (Faster R-CNN) object detection networks. The corresponding detection results are shown in Table 2. The training loss is shown in Fig. 8 and partial detection results are shown in Fig. 9.

It can be seen that in terms of AP the proposed method outperforms Faster R-CNN by 85.1%, SSD by 35.0% and YOLO V3 by 41.3%. It performs very close to SSD for AP^{small} (11.5% lower) and AR^{small} (2.4% lower). For all other metrics, the proposed method outperforms the classic ones by at least 15.4%.

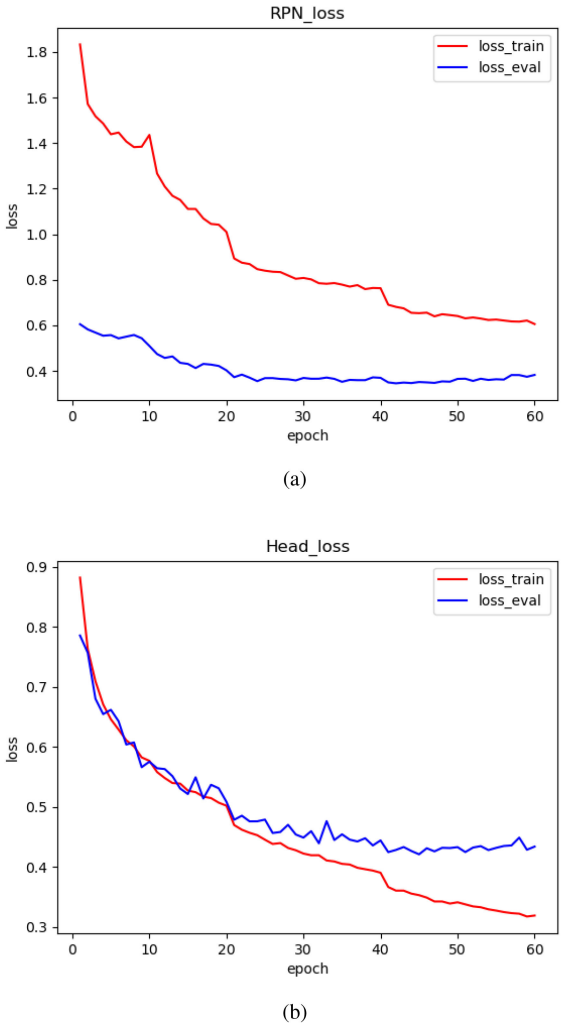


FIGURE 8. Training loss. (a) RPN loss. (b) Head loss.

3) DETECTION SPEED AND RESOURCE CONSUMPTION

The detection speed and resource consumption of the proposed method have been experimentally evaluated in a hardware platform consisting of an Intel i7-11700 CPU and an NVIDIA GeForce RTX 2060 SUPER GPU with 8 GB of memory. The software used was python3.7, pytorch1.9.0, and torchvision0.10.0 over Windows 10.

Results are shown in Table 3. It can be seen that for image size 512*512 and batch size 1, the detection speed of the proposed method is 118.3%, 34.6% and 22.9% faster than that of Faster R-CNN, SSD and YOLO V3, respectively. Memory consumption is reduced by 46.2%, 67.2% and 8.4%, respectively, compared to Faster R-CNN, SSD and YOLO V3. Regarding the amount of network parameters, for the proposed network is only 4.6% of that of Faster R-CNN, 25.3% of that of SSD, and 10.1% of that of YOLO V3.

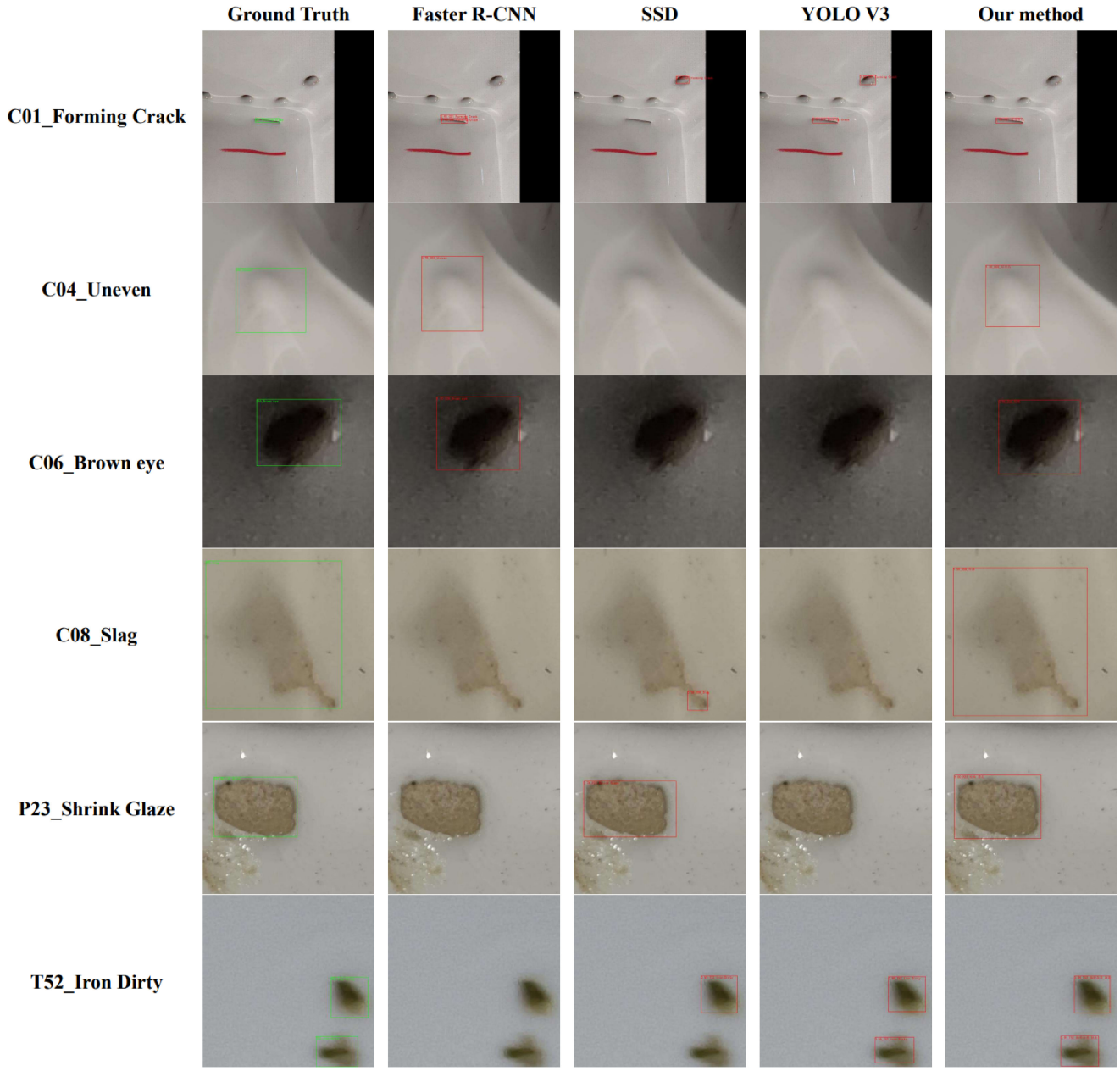


FIGURE 9. Detection results.

TABLE 3 Comparison of Computing Resource Consumption

Method	FPS	GPU memory consumption	Parameters	Weights	GFlops	Batch size	Input resolution
Faster R-CNN	15.9	2135MB	136.73M	521.58MB	83.0	1	512*512
SSD	25.8	3506MB	24.64M	94.01MB	88.0	1	512*512
YOLO V3	28.4	1255MB	61.53M	234.74MB	49.6	1	512*512
Our method	**34.8**	**1149MB**	**6.24M**	**23.80MB**	**5.1**	1	512*512

Finally, regarding the amount of floating-point operations, the proposed method only needs to perform 6.1% of that of Faster R-CNN, 5.7% of that of SSD and 10.2% of that of YOLO V3.

C. ABLATION EXPERIMENTS

In order to verify the beneficial effect of the low-level mixed feature reclassification strategy on classification accuracy, ablation experiments were conducted. From the corresponding

TABLE 4 Detection Head Ablation Experiments

Reclassification	AP	AP^{50}	AP^{75}	AP^{small}	AP^{medium}	AP^{large}	AR^{1}	AR^{10}	AR^{100}	AR^{small}	AR^{medium}	AR^{large}
	27.7	56.3	23.9	18.0	26.9	**36.8**	34.4	36.3	36.3	24.4	36.3	**43.6**
✓	**32.7**	**60.6**	**32.6**	**27.3**	**34.7**	36.6	**37.3**	**39.4**	**39.4**	**33.8**	**41.7**	42.1

experimental results shown in Table 4, it can be seen that using the reclassification strategy improves AP by 17.9%. It can be clearly concluded that the reclassification strategy improves detection accuracy.

V. CONCLUSION

A lightweight sanitary ceramics surface defect detection network has been proposed in this article. The network uses feature pyramid to achieve multi-scale objects detection, and the anchor-free method to define ROIs. Higher classification accuracy is achieved by adopting a low-level mixed feature reclassification strategy. The network does not use fancy tracks, and does not require elaborate anchor design or hyperparameters tuning, which makes it easy to use. The experiments show that compared to Faster R-CNN, it achieves 118.3% increase in detection speed and 85.1% increase in AP with 46.2% less GPU memory consumption and 93.9% less floating-point operations. Compared to classic SSD method, the proposed network achieves 34.6% increase in detection speed and 35.0% increase in AP with 67.2% less GPU memory consumption and 94.3% less floating-point operations. Compared to classic YOLO V3 method, the proposed network achieves 22.9% increase in detection speed and 41.3% increase in AP with 8.4% less GPU memory consumption and 89.8% less floating-point operations. These results demonstrate that the proposed network can achieve real-time defect detection with good accuracy and low computing resources consumption. Therefore, it is very suitable for industrial detection scenarios, as it achieves good detection results easily, fast, and at low cost.

REFERENCES

[1] L. Tang, Q. Li, W. Lu, R. He, F. Gong, and D. Zhang, "Research and implementation of ceramic valve spool surface defect detection system based on region and multilevel optimisation," *Nondestruct. Testing Eval.*, vol. 34, no. 4, pp. 401–412, 2019.

[2] H. Zhang, L. Peng, S. Yu, and W. Qu, "Detection of surface defects in ceramic tiles with complex texture," *IEEE Access*, vol. 9, pp. 92788–92797, 2021.

[3] X. Chen, Y. Zhang, L. Lin, J. Wang, and J. Ni, "Efficient anti-glare ceramic decals defect detection by incorporating homomorphic filtering," *Comput. Syst. Sci. Eng.*, vol. 36, no. 3, pp. 551–564, 2021.

[4] Z. Lin, H. Ye, B. Zhan, and X. Huang, "An efficient network for surface defect detection," *Appl. Sci.*, vol. 10, no. 17, 2020, Art. no. 6085.

[5] L. Xiao, B. Wu, and Y. Hu, "Surface defect detection using image pyramid," *IEEE Sensors J.*, vol. 20, no. 13, pp. 7181–7188, Jul. 2020.

[6] D. Tabernik, S. Šela, J. Skvarč, and D. Skočaj, "Segmentation-based deep-learning approach for surface-defect detection," *J. Intell. Manuf.*, vol. 31, no. 3, pp. 759–776, 2020.

[7] W. Huang, C. Zhang, X. Wu, J. Shen, and Y. Li, "The detection of defects in ceramic cell phone backplane with embedded system," *Measurement*, vol. 181, 2021, Art. no. 109598.

[8] Y.-P. Huang, C.-M. Su, H. Basanta, and Y.-L. Tsai, "Imbalance modelling for defect detection in ceramic substrate by using convolutional neural network," *Processes*, vol. 9, no. 9, 2021, Art. no. 1678.

[9] F. Li, F. Li, and Q. Xi, "Defectnet: Toward fast and effective defect detection," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 2507109.

[10] O. Stephen, U. J. Maduh, and M. Sain, "A machine learning method for detection of surface defects on ceramic tiles using convolutional neural networks," *Electronics*, vol. 11, no. 1, 2021, Art. no. 55.

[11] 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, doi: 10.1109/TPAMI.2016.2577031.

[12] T.-Y. 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.*, 2017, pp. 2117–2125.

[13] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

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

[15] A. Howard et al., "Searching for mobilenetv3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[16] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.

[17] H. Yan, J.-X. Zhang, and X. Zhang, "Injected infrared and visible image fusion via l_1 decomposition model and guided filtering," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 162–173, 2022.

[18] X. Zhang, H. He, and J.-X. Zhang, "Multi-focus image fusion based on fractional order differentiation and closed image matting," *ISA Trans.*, pp. S0019–0578, 2022, doi: 10.1016/j.isatra.2022.03.003.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[20] Y. Wu and K. He, "Group normalization," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 742–755, 2020.

[21] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

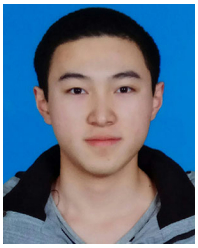
[22] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[23] Y. Wu et al., "Rethinking classification and localization for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10186–10195.

[24] G. Song, Y. Liu, and X. Wang, "Revisiting the sibling head in object detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11560–11569, doi: 10.1109/CVPR42600.2020.01158.

[25] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.

[26] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.



JINGFAN HANG received the B.E. degree in automation from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2019.

He is currently working toward the Ph.D. degree in control science and engineering with the Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin, China. His research interests include computer vision and deep learning.

HAO SUN received the B.E. degree in automation from the Shandong University of Science and Technology, Qingdao, China, in 2011, the M.S. degree in control theory and engineering, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2013 and 2020, respectively.

His research interests include image processing, computer vision, pattern recognition, machine learning, and visual servo.

XINGHU YU was born in Yantai, China, in 1988. He received the M.M. degree in osteopathic medicine from Jinzhou Medical University, Jinzhou, China, in 2016 and the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2020.

He is currently the Chief Executive Officer with the Ningbo Institute of Intelligent Equipment Technology Company, Ltd., Ningbo, China. He has authored more than ten technical papers of conference proceedings and refereed journals, including the IEEE TRANSACTIONS JOURNALS, and holds more than 20 invention patents. His research interests include switched systems, intelligent control, and biomedical image processing.

JUAN J. RODRÍGUEZ-ANDINA (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from the Technical University of Madrid, Madrid, Spain, in 1990 and the Ph.D. degree in electrical engineering from the University of Vigo, Vigo, Spain, in 1996.

He is currently a Professor with the Department of Electronic Technology, University of Vigo. From 2010 to 2011, he was on Sabbatical Leave as a Visiting Professor with the Advanced Diagnosis, Automation, and Control Laboratory, North Carolina State University, Raleigh, NC, USA. From 2015 to 2017, he was with the Harbin Institute of Technology, where he delivered summer courses. He has authored more than 170 journal and conference articles, and holds several Spanish, European, and U.S. patents. His research interests include the implementation of complex control and processing algorithms and intelligent sensors in embedded platforms.

Prof. Rodríguez-Andina has coauthored the article awarded the 2017 IEEE Industrial Electronics Magazine Best Paper Award. He was also the recipient of the 2020 Anthony Hornfeck Award from the IEEE Industrial Electronics Society. From 2016 to 2021, he was the Vice President for Conference Activities of the IEEE Industrial Electronics Society. He was the Editor-in-Chief of the IEEE Industrial Electronics Magazine (2013–2015) and an Associate Editor for the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS (2008–2018). He is an Associate Editor for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and the IEEE Open Journal of the Industrial Electronics Society.

XIANQIANG YANG (Member, IEEE) received the B.E. degree in automation and M.E. degree in control theory and control engineering from the Shandong University of Science and Technology, Qingdao, China, in 2008 and 2011, respectively, and the Ph.D degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2015.

He is currently a Professor with the School of Astronautics, Harbin Institute of Technology. His research interests include identification, soft sensor development, and image processing.