Surface Defect Detection Algorithm for Aluminum Profiles Based on Deep Learning

Qin Dong
School of Information Engineering, Yancheng Institute of Technology, Yancheng, 224000, China
E-mail: dong882021@163.com

The surface quality of aluminum profiles directly affects the performance and safety of the final product. Efficient and accurate surface defect detection has become particularly important for ensuring product quality. In order to solve the low efficiency and low accuracy of traditional detection methods, on the basis of the original You Only Look Once version 5 algorithm, this study conducts surface defect detection on aluminum profiles and optimizes it from three perspectives: anchor box mechanism, data augmentation, and coordinated attention. To solve the poor defect detection effect of small target, the loss function is adjusted and the final optimization algorithm is obtained. The results showed that the mean average precision, recall, precision and F1 values of different types of ablation experiments were 0.99, 0.90, 0.94 and 0.91, respectively. The detection accuracy of the traditional CenterNet method was 94.5%, which was relatively high, but the number of parameters was large and the calculation speed was too slow, corresponding to 14.2M and 93.2%. Simulation analysis showed that the highest detection accuracy, false detection rate, and missed detection rate of the research method for 10 types were 99.2%, 1.3%, and 1.4%, respectively. The successful application of this method can provide reference for the surface defect detection in other materials, which has broad promotion value.

Povzetek: Opisan je izboljšan algoritem globokega učenja za zaznavanje površinskih napak na aluminijastih profilih. Na podlagi algoritma You Only Look Once version 5 (Yolov5) je bil optimiziran sistem za zaznavanje napak z uporabo mehanizma sidrnih okvirov, bogatenja podatkov in koordinirane pozornosti.

1 Introduction

In industrial production, aluminum profiles play a crucial role in many fields such as construction, transportation, aerospace, etc. due to their lightweight, corrosion resistance, and high strength characteristics. The surface quality of aluminum profiles directly affects its performance and application effectiveness. It is crucial to ensure that the aluminum profile surface is flawless during the production process. However, traditional manual detection methods have many shortcomings in terms of efficiency, accuracy, and consistency [1]. With the development of industrial automation and intelligent manufacturing technology, automated Surface Defect Detection (SDD) has become the key to improving the quality and efficiency of aluminum profile production [2]. Deep Learning (DL) has shown excellent performance in image processing and computer vision, providing new solutions for aluminum profiles SDD. DL methods can learn defect features from a large amount of image data, automatically identify and classify defects. This improves detection accuracy and efficiency, which also provides the possibility for further automated production [3]. The research aims to explore and develop a SDD method for

aluminum profiles based on DL. DL methods are used for training and testing to achieve an efficient, accurate, and reliable defect detection system. It can automatically identify and classify various types of surface defects in actual production environments. The research contains four parts. The first part is a literature review, summarizing the current research achievements in defect detection based on DL. The second part is the research methodology. An SDD method based on improved Yolov5 is constructed. The third part is the result analysis, mainly conducting simulation analysis on the research method. The fourth part is the conclusion, summarizing the research results and shortcomings. The innovation of the research mainly lies in the following two aspects. The first is to optimize the You Only Look Once Version 5 (Yolov5) through anchor box mechanism, data augmentation, and coordinated attention to improve the SDD accuracy for aluminum profiles. The second is to design an Enhanced Feature Pyramid Network (EFPN) to better preserve the semantic information of small targets.

2 Related works

DL has been widely applied in SDD for different objects. In recent years, there have been many innovations and

advancements. Chen et al. developed a low contrast SDD method based on DL to solve the low efficiency in ceramic curved parts. The fuzzy repair network was used to reduce the blurriness on surfaces. The multi-scale detail contrast enhancement was applied to highlight the feature information of defect areas. The research results indicated that the accuracy in identifying cracks and bulges in ceramic curved parts was 94.23% and 96.86% [4]. Lv and Song proposed a learning method that integrated a small number of lenses with AM to address the unsatisfactory performance of traditional DL methods in detecting surface defects on bars. Convolutional neural networks (CNNs) and relational networks were used to extract image features, calculate image similarity, and predict image categories, distinguishing pseudo defects from real defects according to the background information. The average detection accuracy was 97.36%, which was 7.81% higher than the traditional method [5]. Wei et al. developed a DL defect detection method based on fast region CNN to address the unsatisfactory performance of traditional computer vision methods in steel SDD. The interested region pool was weighted to eliminate region misalignment caused by quantization. Then it was combined with deformable convolution to adapt to different shapes. Simulation experiments showed that the accuracy of this method in detecting surface defects on steel was 97.3%, effectively improving the detection accuracy [6]. To improve the automatic detection accuracy, Block et al. designed a metal part surface imprinting defect detection and classification system based on Retina Net. The average detection accuracy was 76.3%. The detection accuracy and recall for severe defects were 90.3% and 92.4%, respectively, which were better than current detection methods [7]. Many scholars have also used DL methods such as CNN to detect surface defects in materials such as ceramic tiles.

Significant research results have been achieved. Wan et al. proposed a detection method based on improved YOLOv5s to address the difficulty of detecting surface defects on ceramic tiles. The Attention Mechanism (AM) module and small-scale detection layer were added to build a lightweight ceramic tile defect detection system. This method could compensate for the small texture features and insufficient information of ceramic tile defects, improving the ceramic tile SDD accuracy [8]. Akram et al. constructed an image detection method based on optical CNN to achieve automatic detection of photovoltaic module defects in electroluminescent images. The experimental results showed that this method had lower working environment requirements and faster image detection speed, which facilitated automatic detection of different defects [9]. Rahman et al. proposed an evaluation method based on semantic segmentation DL to address the low efficiency in surface corrosion detection of facilities. The red, green, and blue features were used to optimize the classifier. The research results indicated that this method could detect and evaluate infrastructure corrosion, improving the surface corrosion detection efficiency [10]. Hu et al. developed an unsupervised automatic defect detection method for fabrics based on deep convolutional generative adversarial networks to address surface defects in textured materials. The encoder component was used to reconstruct the query image. The residual image was created to highlight the defect area. The research results indicated that this method could effectively and automatically detect defects in fabrics. The overall effect was better than other conventional methods [11]. Based on the summary of the above literature, the specific contents are shown in Table 1.

Table 1: Summary of literature results

Author	Years	Research Contents	Key indicator
Lv and Song.[5]	2019	In this paper, a learning method that integrated a small number of lenses with attention mechanism is proposed.	The average detection accuracy was 97.36%.
Hu et al. [11]	2019	An unsupervised fabric defect detection method based on deep convolution generation adversarial network is proposed. Encoders are used to query images and the defect areas are highlighted by creating residual graphs.	/
Wei et al. [6]	2020	A deep learning defect detection method based on fast regional convolutional neural network is designed.	The detection accuracy was 97.3%.
Akram et al. [9]	2020	This paper proposes a method based on isolation deep learning and transfer deep learning.	The average accuracy reached 98.67%.
Block et al. [7]	2021	A defect detection and classification system based on Retina Net is designed.	The average detection accuracy was 76.3%, and the detection accuracy and recall value of severe defects were 90.3% and 92.4%, respectively.
Rahman et al. [10]	2021	A detection and evaluation method based on semantic segmentation deep learning is proposed, which uses	The surface corrosion detection efficiency was improved.

		red, green and blue features to optimize the classifier.	
Wan et al. [8]	2022	This paper proposes a detection method based on improved YOLOv5s, adding a focus mechanism module and a small-scale detection layer to build a lightweight tile defect detection system.	/
Chen et al. [4]	2023	A low contrast defect detection method based on deep learning is proposed. The fuzzy repair network is used to reduce the fuzzy degree of the surface, and the multi-scale detail contrast enhancement algorithm is used to highlight the feature information of the defect area.	The accuracy of crack and bulge defect identification for ceramic curved parts was 94.23% and 96.86%, respectively.

Based on the above research, DL models have relatively efficient, accurate, and reliable results in object SDD. In addition, the most advanced technology with the best performance in the above literature is based on Isolation Deep Learning and Transfer Deep Learning (BIDL-TDL). However, there are challenges in its ability to generalize, and it can perform better in one field or task. Therefore, in order to solve the limitations of generalization ability in the above detection methods, this study utilizes the high precision and consistency of DL to detect small or difficult to detect defects, and optimizes it from five aspects: anchor box mechanism, data enhancement, coordinated attention, loss function and structure. In turn, it will promote the development of industrial automation and intelligent manufacturing.

3   SDD of aluminum profiles based on improved Yolov5

The Yolov5 algorithm is used for SDD of aluminum profiles. To improve the defect detection accuracy, Yolov5 is improved, mainly from three aspects: anchor box mechanism, data augmentation method, and AM, to obtain the You Only Look Once version 5-Coordinated Attention (Yolov5-CA). To more accurately identify these small flaws, the loss function is adjusted to reduce sensitivity to the location of minor flaws, resulting in the You Only Look Once version 5-Coordinated Attention-Enhanced Feature Pyramid Network (Yolov5-CA-EFPN).

3.1 SDD of aluminum profiles based on Yolov5-CA

Aluminum profiles are extensively applied in industrial production. Its surface quality directly affects product quality and safety. With the continuous progress of DL technology, SDD algorithms for aluminum profiles have also ushered in broader application prospects. Single-stage object detection algorithm refers to combining target localization and classification into one task in object detection technology based on DL, which simplifies the entire detection process and improves detection efficiency [12]. The You Only Look Once (Yolo) series, as a representative of single-stage object detection algorithms, has vital significance in the development of object detection network models. Since the launch of Yolo, after multiple version iterations, each version has brought better solutions for object detection tasks. These improvements not only improve detection accuracy and efficiency, but also enhance the ability to process small targets and dense scenes [13]. In Yolov5, a comprehensive upgrade is made to the Yolo series, mainly involving input and output terminals. These upgrades have made Yolov5 one of the popular choices in the current object detection, as shown in Figure 1.
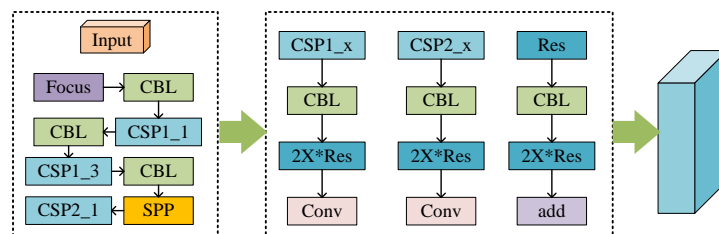

Figure 1: Yolov5 model structure

To improve the defect detection accuracy, Yolov5 is improved, mainly from three aspects: anchor box mechanism, data augmentation method, and AM. The SDD requires selecting appropriate anchor boxes to obtain higher quality models. The original anchor boxes in Yolov5 are preset based on the Common Objects in Context (COCO) dataset with contextual information. However, there are significant differences between the aluminum profile surface defect dataset and the COCO dataset. Therefore, the preset anchor box is modified to fit

the current dataset [14]. Before determining the anchor box size, the aspect ratio of the dataset target is considered. The aluminum profile surface defect dataset used contains multiple types of defects. The average aspect ratio of 10 defects is calculated. The results show that the average aspect ratio of the dataset is about 9. In response to the original clustering method in Yolov5,

1-Intersection over Union (1-IoU) replaces Euclidean distance to present the distance from the anchor box to the cluster center. This change helps to obtain the anchor box size of the aluminum profile surface defect dataset. The anchor box size based on (1-IoU) distance is displayed in Figure 2.
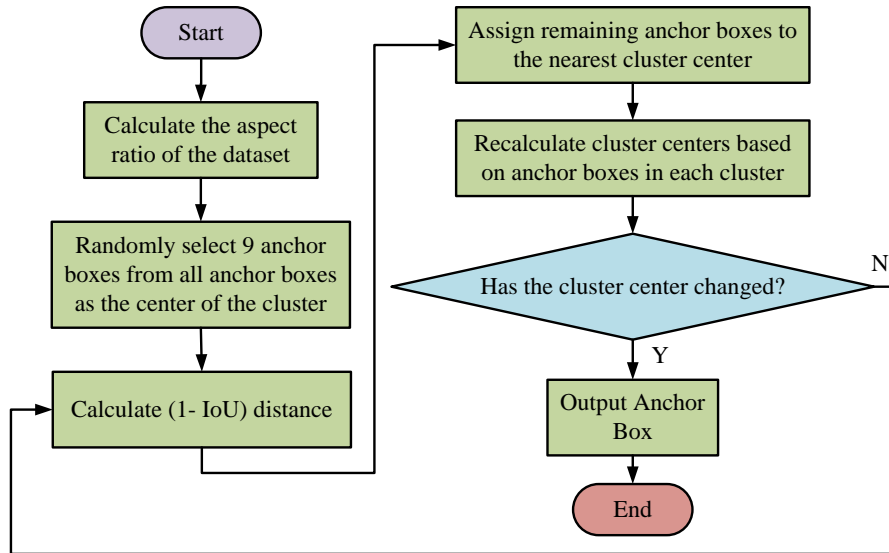
Figure 2: Calculation process of anchor box size based on (1-IoU) distance

From Figure 2, the aspect ratio of the dataset is first calculated. Nine anchor boxes are randomly selected as cluster centers. The (1-IoU) distance from other anchor boxes to the cluster center is calculated. Then, based on the (1-IoU) distance, anchor boxes are classified until the cluster center stabilizes. The final anchor box can be obtained. Yolov5 uses the original 4-segment Mosaic (Mosaic-4) as the data augmentation method. By concatenating four defect images together, it can enrich the background information of object detection and improve the training speed. However, during the training process, the model performs feature extraction calculations on gray areas, resulting in a slower training speed. Therefore, the study uses an improved 9-segment data Mosaic (Mosaic-9). Nine different defective images are concatenated and cropped, effectively reducing gray areas, avoiding useless feature extraction calculations, and improving training speed. Mosaic-9 also enhances the effectiveness of small object detection, as the concatenated images increase the defect types, thereby enhancing small object detection.

AM has important implications in DL models. This network model can focus on certain features of the input image and assign higher weights to them. Meanwhile, for irrelevant feature information, lower weights are assigned to ignore it. Coordinated Attention (CA) introduces positional information into the AM, integrating channel attention and spatial attention. CA decomposes channel attention into two parallel one-dimensional feature maps

(FMs), namely, X and Y, and then encodes them into two attention FMs. This effectively extracts long-range related features from the input FM in a spatial direction. CA can represent any intermediate vector $X = [x_1, x_2, ... x_c]$ as a vector $Y = [y_1, y_2, ... y_c]$ with important feature information. In the CA, the first step is information embedding. This step processes the input feature map with a size of $H \times W \times C$ through Global Average Pooling (GAP) operation. The feature map is segmented in the spatial dimension to form a horizontal X feature map and a vertical Y feature map. For the X feature map in the horizontal direction, after applying a pooling operation with size $H \times 1$, the final FM size obtained is $H \times 1 \times C$. Similarly, the Y feature map in the vertical direction becomes $1 \times W \times C$ after undergoing $1 \times W$ pooling operation. The calculation method is shown in equation (1).

$$\begin{cases} Z_c^h(h) = \dfrac{1}{W} \sum_{i=0}^{W} x_c(h,i) \\ Z_c^w(w) = \dfrac{1}{H} \sum_{j=0}^{H} x_c(j,w) \end{cases} \tag{1}$$

In equation (1), $Z_c^h(h)$ and $Z_c^w(w)$ represent the GAP values in the horizontal and vertical directions. x_c refers to the input feature vector. W represents the weight generated for each channel. σ is the Sigmoid function

[15]. Then, the attention is generated. X and Y FMs are concatenated in the spatial direction. The 1×1-convolution operation is applied to reduce the dimensionality. An intermediate feature map f with a

size of $\dfrac{C}{r}\times1\times(W+H)$ is obtained through the BN

layer. The expression is shown in equation (2).

$$f = \sigma\left(F_1\left(\left[z^h, z^w\right]\right)\right) \tag{2}$$

In equation (2), F_1 represents the calculation result of the BN layer. Afterwards, f is decomposed into FMs f_h and f_w. σ obtains attention weights g^h and g^w in both horizontal and vertical directions. The dimensions are $C\times1\times W$ and $C\times H\times1$, respectively, as shown in equation (3).

$$\begin{cases} g^h = \sigma\left(F_h\left(f^h\right)\right) \\ g^w = \sigma\left(F_w\left(f^w\right)\right) \end{cases} \tag{3}$$

In equation (3), F_h and F_w respectively represent transforming f^h and f^w into transformation functions with the same dimension as the input. The next step is to modify the feature map of CA. The dimensions of g^h and g^w are converted to $C\times H\times W$. Combined with residuals, the input features are connected to obtain the final attention feature, as expressed in equation (4).

$$y_c = x_c \times g^h \times g^w \tag{4}$$

In equation (4), y_c is the final attention feature. x_c, g^h and g^w represent the elements corresponding to C, H, and W. Thus, the CA is introduced into the Yolov5 model, resulting in the Yolov5-CA model, as shown in Figure 3.
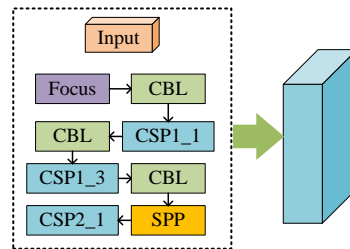

Figure 3: Yolov5-CA model structure

After introducing CA into the Yolov5 model, the Yolov5 to extract important features is enhanced, which is more conducive to SDD.

3.2 SDD of aluminum profiles based on Yolov5-CA-EFPN

In response to the small defect detection, especially in situations where the single sample type and background contrast are not obvious, the Yolo algorithm continues to focus on improving accuracy. For the aluminum surface defect dataset, small defects including dirty spots and scratches are more common, as shown in Figure 4.

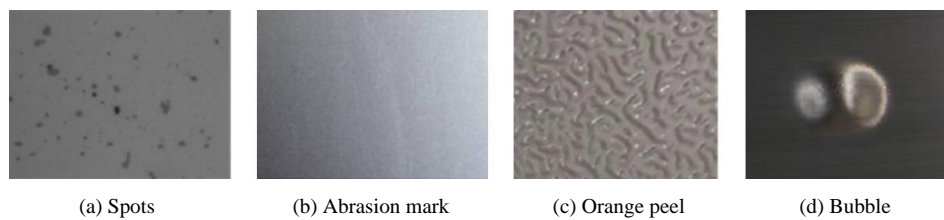(a) Spots   (b) Abrasion mark   (c) Orange peel   (d) Bubble

Figure 4: Common surface defects of aluminum materials

Figures 4 (a), 4 (b), 4 (c), and 4 (d) present dirty spots, abrasion marks, orange peel, and bubbles in the surface defects of aluminum materials. In addition, it also includes non-conductive, corner leakage, leakage, spray, paint bubbles, pitting, and discoloration. To more accurately identify these small flaws, the Yolov5-CA model is further improved, including adjusting the loss function to reduce sensitivity to the location of minor flaws and improve feature capture ability. After

experiencing these improvements, the detection performance for small defects is improved. The loss function of Yolov5 mainly includes rectangular box loss L_{CIoU}, target confidence loss L_{obj}, and category loss L_{cls}. Yolov5 divides the feature map into grids. Each grid outputs a vector that includes position, target probability, and category prediction [16]. Based on these three parts, the loss is calculated, and then added together to obtain the total loss function, as expressed in equation (5).

$$L_{loss}\left(t_p,t_g\right)=\sum_{k=0}^{K}\left[\alpha_k^{balance}\alpha_{box}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{kij}^{obj}L_{CIoU}+\alpha_{obj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{kij}^{obj}L_{obj}+\alpha_{box}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{kij}^{obj}L_{cls}\right] \tag{5}$$

In equation (5), t_p and t_g represent the vectors corresponding to the predicted box and the true box, respectively. K, S^2 and B respectively represent the output feature map, grid, and anchor box on each grid. $\alpha_k^{balance}$ is used to describe the weight of balancing the output FMs at each scale. α is the weight of the corresponding loss function. The rectangular box loss is represented by the Complete Intersection over Union (CIoU), which measures the error between the prediction box and the calibration box [17]. The target confidence loss is calculated by binary cross entropy using the target confidence p_o of the prediction box and the corresponding IoU of the target box. The expression is shown in equation (6).

$$L_{obj}\left(p_o,p_{IoU}\right)=BCE_{obj}^{sig}\left(p_o,p_{IoU};w_{obj}\right) \tag{6}$$

In equation (6), $BCE(\cdot)$ represents binary cross entropy. p represents the center distance between boxes A and B. The category loss is similar to the confidence loss, calculated from the category score of the prediction box and the true category label of the target box, as expressed in equation (7).

$$L_{cls}\left(c_p,c_g\right)=BCE_{cls}^{sig}\left(c_p,c_g;w_{cls}\right) \tag{7}$$

In equation (7), c is the diagonal length of the minimum enclosing rectangle. Based on equation (7), it can make the position difference of small target defects more sensitive, which affects the model accuracy in predicting target categories. The dataset of surface defects on aluminum profiles contains many small target defects. Most of them are not standard rectangles, causing the bounding box to carry some background information. Meanwhile, the target information is usually concentrated at the bounding box center, while the background information is located around the periphery [18-19]. To describe the weights of different pixels, a two-dimensional distribution is used to model the bounding box, with the center pixel having the highest weight and gradually decreasing to the boundary. (cx,cy) stands for the center coordinate of the horizontal bounding box. w and h stand for the width and height of the bounding box. The inscribed ellipse of the bounding box is shown in equation (8).

$$\frac{\left(x-cx\right)^2}{\left(\frac{w}{2}\right)^2}+\frac{\left(y-cy\right)^2}{\left(\frac{h}{2}\right)^2}=1 \tag{8}$$

For a two-dimensional distribution, equation (9) displays the probability density function.

$$f\left(X|\mu,\Sigma\right)=\frac{\exp\left(-\frac{1}{2}\left(X-\mu\right)^T\Sigma^{-1}\left(X-\mu\right)\right)}{2\pi|\Sigma|^{\frac{1}{2}}} \tag{9}$$

In equation (9), X is the coordinate (x,y) of the Gaussian distribution. μ and Σ are the mean vector and covariance matrix. When $\left(X-\mu\right)^T\Sigma^{-1}\left(X-\mu\right)=1$, the probability density function can be expressed as an inscribed ellipse. Meanwhile, the horizontal bounding box can be modeled as a two-dimensional Gaussian distribution (2D-distribution) that satisfies $N\left(\mu,\Sigma\right)$. For the similarity calculation between bounding boxes, it can be regarded as the distance between corresponding Gaussian distributions. Assuming there are Gaussian distributions $\mu_1=N_1\left(m_1,\Sigma_1\right)$ and $\mu_2=N_2\left(m_2,\Sigma_2\right)$, the distribution distance between the two is shown in equation (10).

$$W_2^2\left(N_1,N_2\right)=\left\|\left(\left[cx_1,cy_1,\frac{w_1}{2},\frac{h_1}{2}\right]^T,\left[cx_2,cy_2,\frac{w_2}{2},\frac{h_2}{2}\right]^T\right)\right\|_2^2 \tag{10}$$

In equation (10), $W_2^2\left(N_1,N_2\right)_2^2$ represents the distance metric, which is transformed into exponential form to measure the similarity between Gaussian distributions. That is, the value of $W_2^2\left(N_1,N_2\right)_2^2$ is assigned to $[0,1]$, which is called NWD distance. The expression is shown in equation (11).

$$NWD\left(N_1,N_2\right)=\exp\left(-\frac{\sqrt{W_2^2\left(N_1,N_2\right)}}{C}\right) \tag{11}$$

In equation (11), C represents a constant related to the dataset. When detecting small defects, the accuracy of small object detection often decreases. This is because in deep neural networks, small features are often lost after multiple convolutions and pooling operations. The feature pyramid network constructs the pyramid structure of the feature map through a shrinking process and an expanding process. Feature fusion is performed at different levels to connect deep and shallow information [20]. However, a standard feature pyramid network may still result in small target information loss when handling multiple up sampling and down sampling processes. The EFPN improves this problem by introducing inter layer skip connections and multi-scale connections, thereby better preserving the semantic information of small targets. For the k-level scale features, the feature map received by the l-layer is calculated, as shown in equation (12).

$$P_k^l = Conv\left(Concat\left(P_k^0, ..., P_k^{l-1}\right)\right) \qquad (12)$$

In equation (12), *Concat* represents the concatenation operation of FMs in the $l-1$ layer. *Conv* is a 3×3 -convolution operation. Due to the increased network burden, parameter expansion, and gradient vanishing caused by the network model connection method, EFPN adopts a solution called $\log_2 n$ skip layer connection method to address these challenges. In all levels k , the l -th layer receives FMs from up to

$\log_2^l +1$ advanced layers, as shown in equation (13).

$$P_k^l = Conv\left(Concat\left(P_k^{l-2^n}, ..., P_k^{l-2^0}\right)\right) \qquad (13)$$

Based on equation (13), the complexity of the $\log_2 n$ skip layer connection method can be reduced, which is more conducive to the deeper development of the network. EFPN uses Queen-fusion to achieve cross scale connections, while fusing features from the same and adjacent levels. The fusion process is shown in Figure 5.
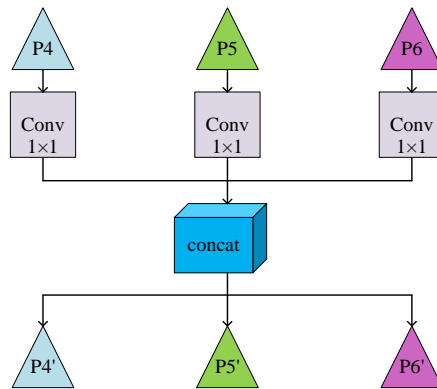

Figure 5: The fusion process of EFPN

From Figure 5, the down sampling results of P4 layer, the up-sampling results of P6 layer, and the features of P5 layer and P4 layer after 1×1 convolution are fused into P5 layer. Up sampling is achieved using the bi-linear interpolation, while down sampling is achieved using max pooling. Based on the principle of EFPN, two skip layer connection methods and a cross scale Queen-fusion connection method are used to improve the small target defect detection accuracy. The PAN structure of the Neck module in Yolov5-CA is replaced with EFPN to form the Yolov5-CA-EFPN. Figure 6 shows the connection between Yolov5-CA backbone module and EFPN and the improvement process of YOLOV5-CA-EFPN model.
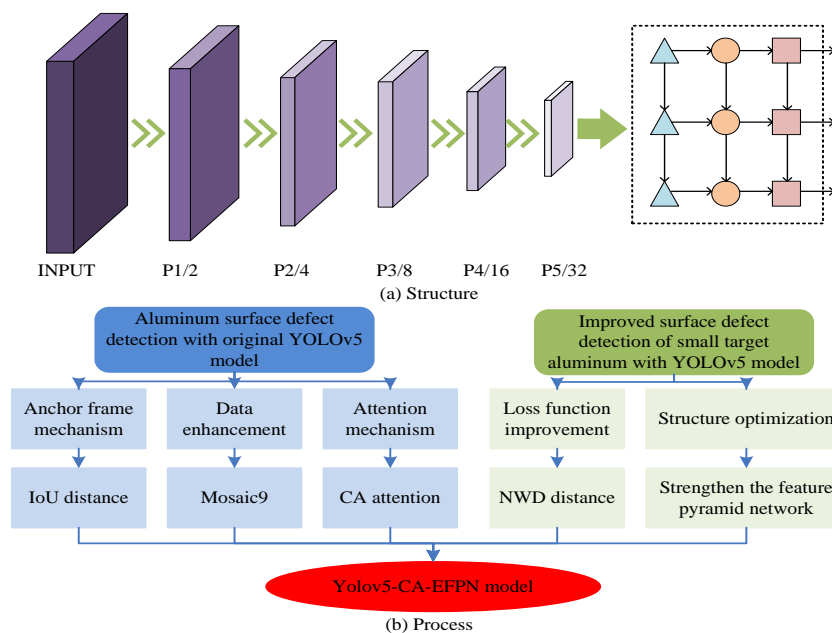

Figure 6: Structure and flow of Yolov5-CA-EFPN model

From Figure 6, the Yolov5-CA-EFPN model adopts an alternative EFPN structure in the Neck module. This change may have an interesting impact on the performance and feature learning ability of the model, further affecting the detection accuracy and overall performance.

4 Analysis of SDD for Aluminum Profiles Based on Yolov5-CA EFPN

To verify the effectiveness of Yolov5-CA-EFPN, it is compared with other methods to validate its superiority. Furthermore, it is incorporated into the simulation analysis to prove the actual detection effect.

4.1 Performance analysis of the improved algorithm

The SDD analysis for aluminum profiles based on Yolov5-CA-EFPN first conducts performance analysis. The relevant parameters are set to ensure accuracy and effectiveness. Table 2 displays the specific parameter settings.

Table 2: Parameter settings

Number	Project	Case
(1)	Deep learning framework	1.8.0
(2)	Cuda	11.3
(3)	Python	3.8
(4)	Graphics card	RTX3090
(5)	System	Ubuntu
(6)	Batch_size	64
(7)	CPU	i9-10920X
(8)	Defect image	640×640

Based on the relevant parameter settings in Table 2, ablation experiments are first conducted. The Mean Average Precision (mAP), recall, precision, and F1 value (F1) are compared. Figure 7 displays the results.
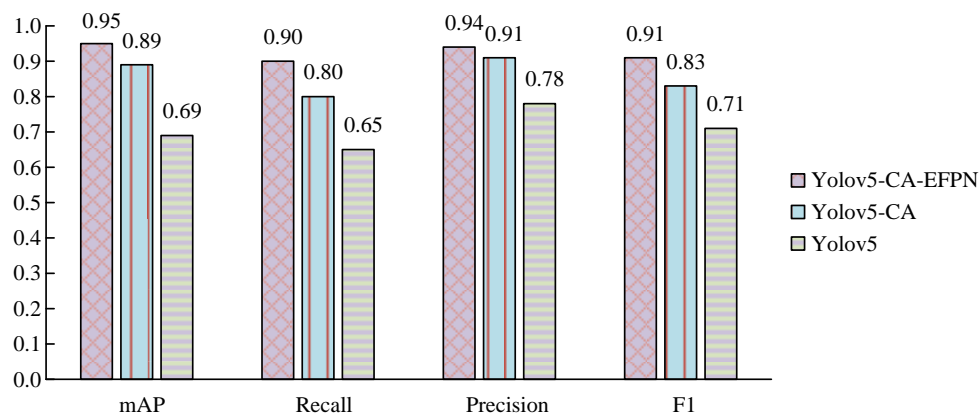

Figure 7: Experimental results of ablation using three algorithms

From Figure 7, the mAP values of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.69, 0.89, and 0.99, respectively. The recall values of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.65, 0.80, and 0.90, respectively. The precision of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.78, 0.91, and 0.94, respectively. The F1 values of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.71, 0.83, and 0.91, respectively. The improved Yolov5-CA-EFPN has the best performance. To enhance the persuasiveness, the loss and mAP change during the training process are compared. Figure 8 displays the results.
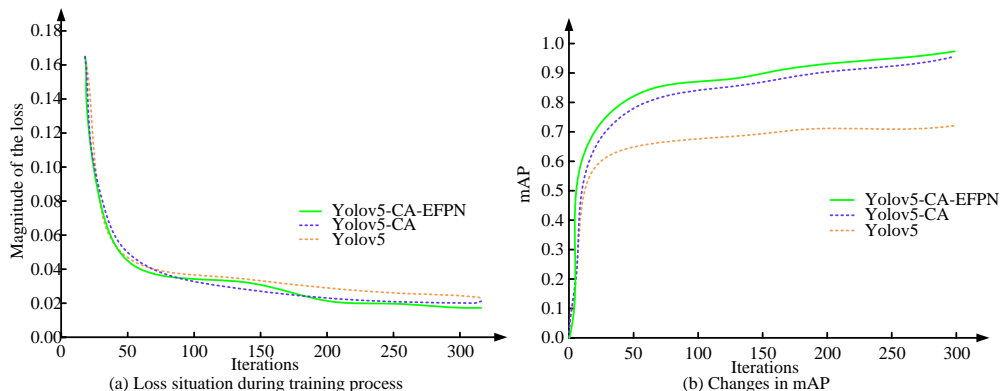
Figure 8: Comparison of training results for different algorithms

Figures 8 (a) and 8 (b) respectively represent the loss change and mAP changes. From Figure 8, the loss values of the three algorithms almost overlapped in the first 160 iterations. As the iterations increase, the loss of Yolov5-CA-EFPN is the lowest and tends to stabilize faster. When iterating 300 times, the loss values of Yolov5-CA-EFPN, Yolov5-CA, and Yolov5 algorithms were 0.0173, 0.0204, and 0.0288. The mAP values of the three algorithms in descending order were Yolov5-CA-EFPN, Yolov5-CA, and Yolov5. The mAP value of Yolov5-CA-EFPN stabilized faster. When the iteration was 300 times, the mAP values of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.0986,

0.0965, and 0.0721, respectively. The above results indicate that using NWD distance instead of IoU-based YOLO5-CA loss function in the Yolov5-CA-EFPN model can effectively reduce the sensitivity of small target position deviation. The training process of Yolov5-CA-EFPN is faster and more stable, with smaller loss values and the best overall performance, making it more suitable for SDD on aluminum profiles.

Considering the training performance of Yolov5-CA-EFPN, the CPU usage during its operation is analyzed to comprehensively compare the algorithm performance. Figure 9 displays the CPU usage.
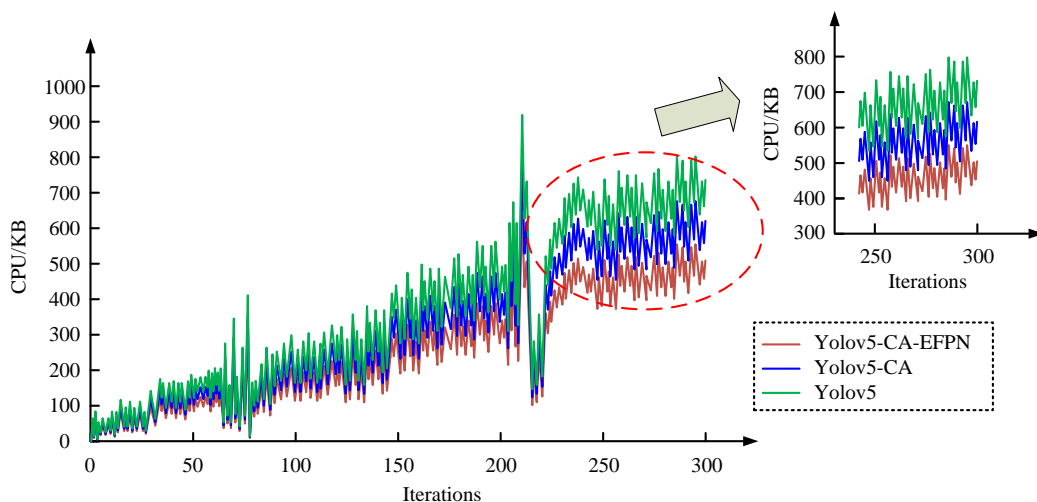

Figure 9: CPU usage during the training process of three algorithms

From Figure 9, the CPU usage of Yolov5-CA-EFPN was relatively close to Yolov5-CA and Yolov5. The difference was not significant. When Yolov5-CA-EFPN had the lowest loss value and the highest mAP value, it

did not increase more CPU usage. It indicates that Yolov5-CA-EFPN has the best comprehensive performance, which is more suitable for detecting surface defects on aluminum profiles.

To scientifically validate the performance of the proposed method, three mainstream deep learning methods are selected for comparative experiments, namely, CNN, CenterNet (CNT), and Single Shot Multi-Box Detector (SSMD). The performance results of SDD methods for different aluminum profiles based on deep learning are compared, as shown in Table 3.

Table 3: Comparison of performance results of surface defect detection methods for different aluminum profiles based on deep learning

Evaluating indicator	CNN	CNT	SSMD	Yolov5-CA-EFPN
mAP/%	95.8	94.5	91.9	95.3
Recall/%	88.6	89.6	88.1	93.2
Precision/%	97.4	74.7	92.7	93.6
FPS	25.6	48.7	128.1	262.4
Parameter quantity/M	41.5	14.2	24.4	7.3

From Table 3, the proposed method achieved the best comprehensive performance in terms of detection precision and detection speed, with corresponding mAP and recall of 95.3% and 93.2%, respectively. The simultaneous detection speed and parameter quantity were 262.4FPS and 7.3M, respectively. The CNT method had the best detection precision due to the maximum input image resolution, but the corresponding parameter quantity was large and the calculation speed was too slow, at 14.2M and 18.7FPS, respectively, which was not suitable for practical scene applications. The SSMD and the designed method had certain differences in precision and speed. Finally, in order to test the computational complexity of the designed method, the research is evaluated from the two indicators: time complexity and space complexity. The former is determined by the calculation amount of the method, and the latter is determined by the solving process. In addition, the study introduces the most advanced method for comparative experiment, namely the BIDL-TDL method. The results are obtained, as shown in Figure 10.

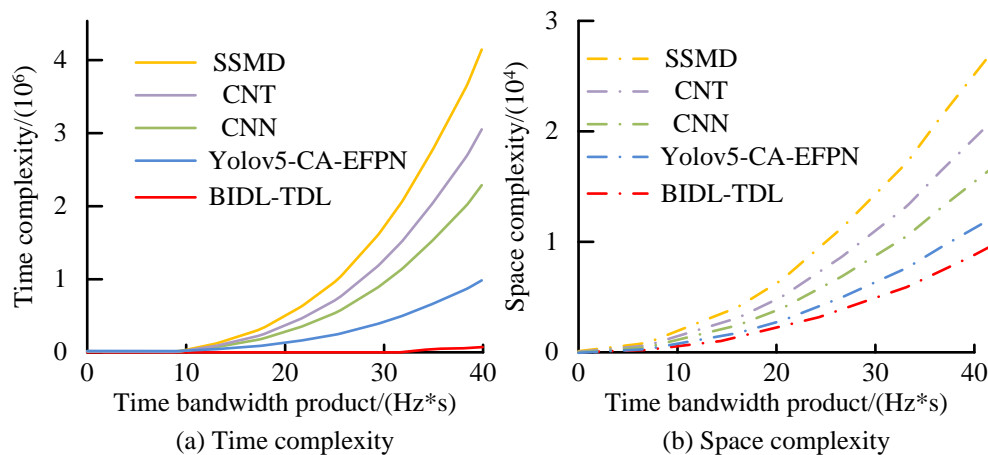(a) Time complexity   (b) Space complexity

Figure 10: Comparison of computational complexity results for different methods

Figure 10 (a) and Figure 10 (b) respectively show the time complexity and space complexity results for different methods. From Figure 10, with the continuous increase of time bandwidth product, the time complexity of the research method was smaller than that of other methods. In the later period, the time complexity curve of the research method tended to stabilize. However, the time complexity of the other methods increased continuously with the increase of the time bandwidth product. It indicates that the research method can effectively reduce the computation amount to a great extent. In addition, the spatial complexity of research method grew very slowly, which meant that the data stored by the research method was significantly reduced and had greater advantages. However, the spatial complexity of other methods continued to increase. When the time bandwidth product of BIDL-TDL method was micro 40Hz*s, the spatial complexity reached $1.5*10^4$. The above results show that the calculation amount and storage space required by the research method are obviously reduced, and the computational complexity is optimized. In summary, the proposed method has good performance and feasibility, which is more suitable for deployment in practical application scenarios.

4.2 Simulation analysis

In Windows 10, a Yolov5-CA-EFPN real-time detection system is designed in the PyQt5 environment. During system testing, an industrial camera AF12 provided by a certain technology company is used, which is fixed with a triangular stabilizing bracket. The system settings, as the core part of the deployment detection platform, include the following functions, such as importing model weight files, obtaining files to be detected, connecting industrial cameras, obtaining videos, and setting common parameters. The image/video detection visualization module compares the objects before and after detection, analyzes the results and records the type and quantity of defects currently detected. The detection accuracy bar is used to start and pause the detection. In the Windows 10 environment, the camera is connected to the laptop using USB 2.0. The following laptops, aluminum profiles, cameras, and tripods are deployed to obtain a complete testing platform. Table 4 displays the specific parameter settings.

Table 4: Specific parameter settings in the simulation experiment

Number	Camera parameters	Parameter value	Unit
(1)	Connecting line	USB 2-meter all copper tape magnetic ring anti-interference shielding wire	m
(2)	Sensor	Samsung S5K2L8SX03 CMOS sensor	/
(3)	Sensor Size	1/2.8"	/
(4)	Pixel size	1.28*1.28	um
(5)	Data format	MJPG/YUY	/
(6)	Focal length	3.97	mm
(7)	Dynamic frame rate	1920*1080/30FPS	/
(8)	F/NO	1.7	/

The OB1640L aluminum profile provided by a certain company is used for testing, with a thickness of 4.6mm and a weight of 1.25kg per meter. The aluminum profile itself has no defects. However, for the convenience of testing, some defects are artificially set, including 10 types: non-conductive, scratches, corner leakage, orange peel, leakage, spray, paint bubbles, pitting, discoloration, and dirty spots, each with 20 defects. The research method is used to detect them. The results are shown in Figure 11.
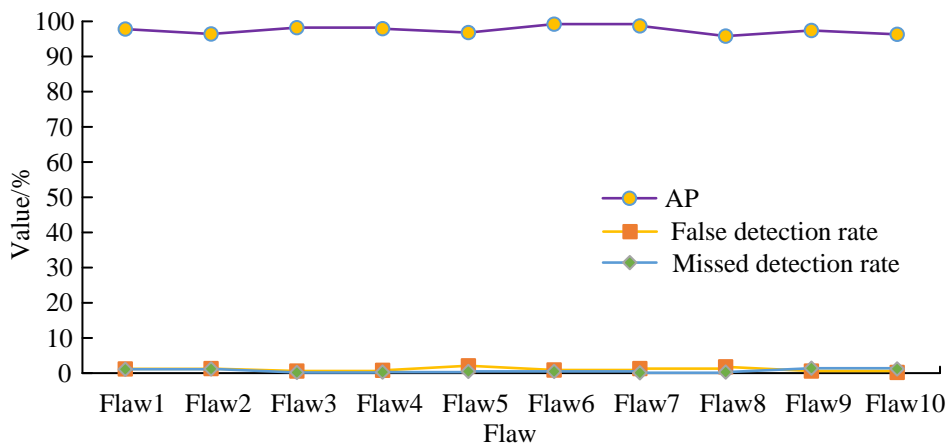

Figure 11: Test results of Yolov5-CA-EFPN in different defect types

From Figure 11, the Yolov5-CA-EFPN had high detection precision for 10 types. The highest and lowest values were 99.2% and 96.3%. The false detection rate and missed detection rate were both low. The highest and lowest false detection rates were 1.3% and 0.2%. The highest and lowest missed detection rates were 1.4% and 0.2%, respectively. The Yolov5-CA-EFPN has a good effect, which is conducive to detecting surface defects of aluminum profiles and ensuring the integrity of defect detection.

In addition, the Yolov5-CA-EFPN is used to test the confidence intervals of four typical defects and two other defects. The results are shown in Figure 12.

(a) Scratches

(b) Jet flow

(c) Lacquer bubble

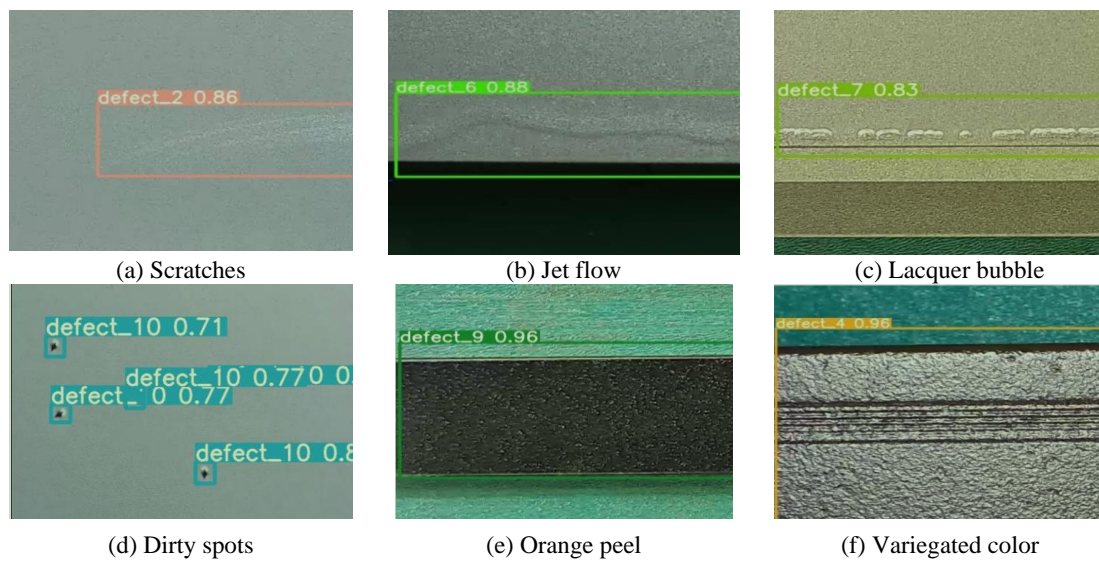(d) Dirty spots

(e) Orange peel

(f) Variegated color

Figure 12: Confidence results based on the Yolov5-CA-EFPN model for detecting four typical defects and two other defects

From Figure 12, the confidence intervals in the four typical detects of scratches, jet flow, lacquer bubble, and dirty spot were [80%, 90%], while the confidence intervals in the other two defects were all over 90%. This indicates that the proposed method is not only suitable for small target defect detection, but also has excellent application effects in overall dataset defect detection. Based on the above results, it can be concluded that the Yolov5-CA-EFPN model can effectively improve the detection performance of surface defects in aluminum profiles, which has strong robustness and generalization ability.

4.3 Discussion

As an important basic material industry, China's aluminum industry has developed rapidly in recent years, becoming a global aluminum production and consumption power. With the gradual recovery of the industry, aluminum downstream market is gradually active, and aluminum market demand will be driven by this substantial growth. According to the "2023-2029 China Aluminum market analysis and Investment Prospects Research Report", the cumulative value of China's aluminum production in 2023 reached 63.034 million tons, with a final total increase of 5.7% compared to the previous year. In addition, the sustained development of the Chinese economy and the improvement of household consumption levels will maintain a long-term growth trend in aluminum demand. Especially in new energy vehicles, rail transit, aerospace and other emerging fields, aluminum will play a greater role. However, in the actual production process, aluminum profiles are often affected by factors such as equipment and environment, inevitably resulting in

surface defects that affect the aesthetic appearance of aluminum profiles. In severe cases, it can also affect product quality and subsequent use. It is of great significance to design a real-time and accurate detection method for aluminum profile surface defects for industrial development and product quality improvement. Therefore, the Yolov5-CA-EFPN model is proposed to improve the detection effect of aluminum profile surface defects.

Firstly, in order to test the feasibility of using different modules for optimization, the ablation experiment is designed for evaluation. The results showed that the mAP values of Yolov5, YOLOV5-CA and YOLOV5-CA-EFPN were 0.69, 0.89 and 0.95, respectively. The recall values of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.65, 0.80, and 0.90, respectively. The precision values of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.78, 0.91, and 0.94, respectively. The F1 values of Yolov5, YOLOV5-CA and YOLOV5-CA-EFPN were 0.71, 0.83 and 0.91, respectively. In addition, the loss values of the above three methods were almost consistent in the first 160 iterations. With the increase of iterations, the loss value of Yolov5-CA-EFPN was the lowest and tended to stabilize faster. When iterated for 300 times, the loss values of YOLOV5-CA-EFPN, YOLOV5-CA and Yolov5 were 0.0173, 0.0204 and 0.0288, respectively. During the training process, the mAP values of the three algorithms were YOLOV5-CA-EFPN, YOLOV5-CA, and Yolov5 in order from high to low, and the mAP value of YOLOV5-CA-EFPN was stable faster. When the iteration was 300 times, the mAP values of Yolov5, YOLOV5-CA, and YOLOV5-CA-EFPN were 0.0986, 0.0965, and 0.0721, respectively. According to the ablation experiment results, it can be found that the

optimized module designed in the study can significantly improve the performance of the Yolov5-CA-EFPN model.

In order to further verify the performance of the research method, CPU usage experiments, detection precision, detection speed and computational complexity experiments are designed, and the following results are obtained. The CPU usage of YOLOV5-CA-EFPN was smaller than that of YOLOV5-CA and Yolov5, but the research method did not increase CPU usage at the minimum loss value and high mAP value. The mAP, recall, detection speed and parameter quantity were 95.3%, 93.2%, 14.2M and 18.7FPS, respectively. The CNT method had the best detection precision because the input image resolution was the largest, but the corresponding parameter quantity was large and the calculation speed was too slow, which was 14.2M and 18.7FPS respectively. It is not suitable for practical application scenarios. In addition, with the increase of the time bandwidth product, the time complexity of the research method was smaller than that of other methods, and the time complexity curve of the research method tended to be stable in the later period. However, the time complexity of the remaining methods increased continuously with the increase of the time bandwidth product. In addition, the growth rate of the spatial complexity of the research method was very slow, which meant that the amount of data stored by the research method was significantly reduced had greater advantages. However, the spatial complexity of other methods continued to increase. When the time bandwidth product of BIDL-TDL method was micro 40Hz*s, the spatial complexity reached $1.5*10^4$. The above results indicate that this research method can effectively reduce computational complexity to a large extent, and the amount of data to be stored is significantly reduced, which has greater advantages. Finally, the detection precision of 10 types of defects was simulated, and the highest and lowest values of the research method were 99.2% and 96.3%, respectively.

In summary, the research method can overcome the shortcomings of traditional detection techniques, which has far-reaching significance for the development of aluminum profile industry, and plays a huge role in new energy vehicles, aerospace and other emerging fields. However, there are still shortcomings in the research method. In the research, hyper-parameters are determined through genetic algorithms and used for evolutionary selection of hyper-parameters. The specific sensitivity analysis and mechanism of action of hyper-parametric image performance are still unclear. In future research, other advanced technologies can be used for sensitivity analysis.

5 Conclusion

With the continuous development and optimization of DL technology, SDD for aluminum profiles based on DL is expected to become one of the key technologies to improve industrial production quality and efficiency. To improve the detection accuracy and more accurately identify small defects, the anchor box mechanism, data augmentation method, and AM are improved relying on the Yolov5. The loss function is adjusted and the Neck structure is optimized to obtain the Yolov5-CA-EFPN detection model. The results show that compared with Yolov5, YOLOV5-CA-EFPN had the best performance, and its mAP, recall, precision and F1 value were increased by 26%, 25%, 16% and 20% respectively. Compared with Yolov5 and Yolov5-CA, the Yolov5-CA-EFPN had the lowest loss value and tended to stabilize faster. The results showed that Yolov5-CA-EFPN had a detection precision of over 96.3% for 10 types, with false detection and missed detection rates below 1.4%. Yolov5-CA-EFPN can accurately and efficiently detect surface defects on aluminum profiles, greatly improving the efficiency and accuracy of defect detection, and helping to improve product quality and production efficiency. In addition, this algorithm can reduce reliance on manual inspection and lower production cost. It is of great significance to the aluminum profile production industry, providing new technological solutions and ideas for the intelligent manufacturing. There are still some shortcomings in this study, such as the failure to identify defect types based on detection. Future research will start from this aspect to better improve the process flow and repair defects. Meanwhile, the functionality of the research method is scalable. In future research, testing visualization and dataset generation functions can be added to improve efficiency in the industrial field, and other real-time application scenarios can be actively explored.

References

[1] C. Wang, D. Luo, Y. Liu, B. Xu, and Y. Zhou, "Near-surface pedestrian detection method based on deep learning for UAVs in low illumination environments," Optical Engineering, vol. 61, no. 2, pp. 023103.1-023103.19, 2022. https://doi.org/10.1117/1.OE.61.2.023103.

[2] Y. Yang, Z. Liu, M. Huang, Q. Zhu, and X. Zhao, "Automatic detection of multi-type defects on potatoes using multispectral imaging combined with a deep learning model," Journal of Food Engineering, vol. 336, no. Jan, pp. 1-8, 2023. https://doi.org/10.1016/j.jfoodeng.2022.111213

[3] J. He J and J. Yang, "Network Security Situational Level Prediction Based on a Double-Feedback Elman Model," Informatica: An International Journal of Computing and Informatics, vol. 46, no. 1, pp. 87-93, 2022. https://doi.org/10.31449/inf.v46i1.3775

[4] W. Chen, B. Zou, C. Huang, J. Yang, L. Li, J. Liu, and X. Wang, "The defect detection of 3D-printed ceramic curved surface parts with low contrast

based on deep learning," Ceramics International, vol. 49, no. 2, pp. 2881-2893, 2023. https://doi.org/10.1016/j.ceramint.2022.09.272

[5] Q. Lv and Y. Song, "Few-shot learning combine attention mechanism-based defect detection in bar surface," ISIJ International, vol. 59, no. 6, pp. 1089-1097, 2019. https://doi.org/10.2355/isijinternational.isijint-2018-722

[6] R. Wei, Y. Song, and Y. Zhang, "Enhanced faster region convolutional neural networks for steel surface defect detection," ISIJ international, vol. 60, no. 3, pp. 539-545, 2020. https://doi.org/10.2355/isijinternational.isijint-2019-335

[7] S. B. Block, R. D. D. Silva, L. Dorini, and R. Minetto, "Inspection of imprint defects in stamped metal surfaces using deep learning and tracking," IEEE Transactions on Industrial Electronics, vol. 68, no. 5, pp. 4498-4507, 2021. https://doi.org/10.1109/TIE.2020.2984453

[8] G. Wan, H. Fang, D. Wang, J. Yan, and B. Xie, "Ceramic tile surface defect detection based on deep learning," Ceramics International, vol. 48, no. 8, pp. 11085-11093, 2022. https://doi.org/10.1109/ist48021.2019.9010098

[9] M. W. Akram, G. Li, Y. Jin, X. Chen, and A. Ahmad, "Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning," Solar Energy, vol. 198, no. Mar, pp. 175-186, 2020. https://doi.org/10.1016/j.solener.2020.01.055.

[10] A. Rahman, Z. Y. Wu, and R. Kalfarisi, "Semantic deep learning integrated with RGB feature-based rule optimization for facility surface corrosion detection and evaluation," Journal of computing in civil engineering, vol. 35, no. 6, pp. 4021018.1-4021018.15, 2021. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000982

[11] G. H. Hu, J. F. Huang, Q. H. Wang, J. R., Li, and Z. J. Xu, "Unsupervised fabric defect detection based on a deep convolutional generative adversarial network," Textile Research Journal, vol. 90, no. 3-4, pp. 247-270, 2019. https://doi.org/10.1177/0040517519862880

[12] H. Zhang, G. Qiao, S. Liu, Y. Lyu, L. Yao, and Z. Ge, "Attention-based vector quantization variational autoencoder for colour-patterned fabrics defect detection," Coloration Technology, vol. 139, no. 3, pp. 223-238, 2023. https://doi.org/10.1111/cote.12644

[13] X. X. Li, Q. Yang, Z. Lou, and W. J. Yan, "Deep learning based module defect analysis for large-scale photovoltaic farms," IEEE Transactions on Energy Conversion, vol. 34, no. 1, pp. 520-529, 2019. https://doi.org/10.1109/TEC.2018.2873358

[14] W. Tang, Q. Yang, X. Hu, and W. Yan, "Deep learning-based linear defects detection system for large-scale photovoltaic plants based on an edge-cloud computing infrastructure," Solar Energy, vol. 231, no. Jan, pp. 527-535, 2022. https://doi.org/10.1016/j.solener.2021.11.016

[15] W. Tang, Q. Yang, K. Xiong, and W. Yan, "Deep learning based automatic defect identification of photovoltaic module using electroluminescence images," Solar Energy, vol. 201, no. May, pp. 453-460, 2020. https://doi.org/10.1016/j.solener.2020.03.049

[16] X. Ma, N. Kittikunakorn, B. Sorman, H. Xi, A. Chen, M. Marsh, A. Mongeau, N. Piché, and D. Skomski, "Application of deep learning convolutional neural networks for internal tablet defect detection: High accuracy, throughput, and adaptability," Journal of Pharmaceutical Sciences, vol. 109, no. 4, pp. 1547-1557, 2020. https://doi.org/10.1016/j.xphs.2020.01.014

[17] A. Yan, P. Rupnowski, N. Guba, and A. Nag, "Towards deep computer vision for in-line defect detection in polymer electrolyte membrane fuel cell materials," International Journal of Hydrogen Energy, vol. 48, no. 50, pp. 18978-18995, 2023. https://doi.org/10.1016/j.ijhydene.2023.01.257

[18] M. Alipour and D. K.Harris, "Increasing the robustness of material-specific deep learning models for crack detection across different materials," Engineering Structures, vol. 206, no. Mar.1, pp. 110157.1-110157.14, 2020. https://doi.org/10.1016/j.engstruct.2019.110157

[19] M. W. Akram, G. Li, Y. Jin, X. Chen, and A. Ahmad, "CNN based automatic detection of photovoltaic cell defects in electroluminescence images," Energy, vol. 189, no. Dec.15 Pt.2, pp. 116319.1-116319.15, 2019. https://doi.org/10.1016/j.energy.2019.116319

[20] L. Mei, "Model construction of higher education quality assurance system based on fuzzy neural network," Informatica: An International Journal of Computing and Informatics, vol. 48, no. 10, pp. 153-164, 2024. https://doi.org/10.31449/inf.v48i10.5676