Current Science
CS 5 (5), 4409-4420 (2025)
Received 10 June 2025 | **Revised 17 July** 2025 | **Accepted 10 August** 2025 | **Online Available 14 September** 2025
https://doi.org/10.52845/CS/2025-5-5-8

Current Science

Current Science

OPEN ACCESS

ISSN(O) 2795-8639

Original Article

Towards Improving YOLOv8-Based Defect Detection Methods for Aluminum Profiles

Ziming Qu[1,2], Xufeng Zhang[1,2], Fafa Zhang[1,2], Hui Wang[1,2*]

[1]School of Artificial Intelligence, Anhui University, Jiulong Road, Hefei, 230601, Anhui, China

[2]Anhui Provincial Key Laboratory of Security Artificial Intelligence, Anhui University, Jiulong Road, Hefei, 230601, Anhui, China

[*]Corresponding Author: Hui Wang

Abstract:

In the field of aluminum surface defect detection, current detection models com- monly suffer from insufficient detection accuracy, weak real-time performance, and excessive parameter size. To address these issues, we propose a novel method based on the YOLOv8 object detection model. Firstly, an EMA (Efficient Multi- scale Attention) module is incorporated to enhance feature extraction efficiency and accuracy. Secondly, the conventional convolution is replaced with Repa- rameterized Convolution, which employs a multi-branch structure during the training phase to strengthen feature extraction capabilities while adopting a single-branch structure during the validation phase to accelerate detection speed through parameter re-parameterization. Finally, a composite regression loss func- tion is implemented that combines the CIoU loss and the WIoUv3 loss to enable dynamic anchor box screening, which effectively resolves inaccurate label assign- ment caused by variations in sensitivity of the IoU in defect types of different sizes. Experimental results demonstrate that compared with the baseline model, our method achieves a 4.6% increase in mean average precision (mAP), attains a processing speed of 94.0 frames per second (FPS), and maintains a compact model size of merely 3.12 MB. These advancements validate that the proposed algorithm exhibits superior recognition accuracy, rapid detection speed, and minimal memory footprint, demonstrating significant potential for industrial deployment.

Keywords: Surface Defect Detection, Deep Learning, YOLOv8, attention mechanism.

1. Introduction

Aluminum profiles, as structural materials widely used in construction, automotive, aerospace, and other fields, have a crucial position in manufacturing, and their qual- ity directly affects the stability and safety of the entire structure [1–4]. However, in the production process due to the influence of various factors, such as process flow, aluminum profiles often have various defects, resulting in early failure or performance degradation of aluminum profiles [5, 6]. The causes of defects in aluminum profiles mainly include oxide film defects due to insufficient surface protection during the man- ufacturing process, and surface imperfections due to die wear or improper extrusion process. These defects are usually small and tend to appear in multiple places on the surface. In order to ensure the surface quality of aluminum profiles, surface defect detection has become an important process in the manufacturing of aluminum pro- files, to find and deal with these defects in a timely manner. The traditional detection method is manual sampling, but due to the artificial heavy reliance on experience and labor-intensive, prone to misdetection, omission. Currently, the industry has entered the era of digitalization and intelligence, and visual inspection technology is gradu- ally replacing the traditional manual inspection method due to its high recognition accuracy and reliable detection performance [7–9]. Traditional visual inspection tech- nology has the problems of high cost, complex operation, and high environmental requirements, and it is difficult to realize the visualization process inspection. With the development of industrialization, machine vision inspection is widely used in sur- face defect detection due to the advantages of accurate and contactless measurement [10–13]. Early machine vision inspection extracts defect features through the image processing of defect photos, although it can detect specific defects more accurately, but in the situation of aluminum profile defects of many kinds and different scales, there are problems such as poor generalization, complex feature extraction, and low detection accuracy, which are limited in actual production and are difficult to meet demand [14, 15].

Accompanied by the rapid development of artificial intelligence technology, deep learning has been increasingly applied in the field of industrial defect detection. Liu [16] proposed an improved dual-CNN model fusion framework and an efficient target detection model based on hierarchical attention and contextual information for accu- rate classification and identification of aluminum profile defects and surface defect detection. Wu Xiaocheng [17] proposed an improved algorithm model KCC-YOLOv5 for surface defect detection of aluminum profiles to solve the problems of low accu- racy, slow detection speed, and insufficient detection capability of small target defects in traditional aluminum profile surface defect detection. By optimizing the anchor frame, introducing a global attention module and a lightweight upsampling operator, the model significantly improves the detection accuracy of small target defects while maintaining a high detection speed, and effectively improves the leakage detection of small target defects on the surface of aluminum profiles. Xue Yang [18] significantly improved the algorithm's ability to extract information about defects while maintain- ing a high detection speed by introducing a dual-attention mechanism, improving the activation function, and optimizing the convolutional module.Neuhause [19] proposed a convolutional neural network (CNN)-based method for the classification and detec- tion of surface defects. A simple video camera was used to record extruded profiles during production and a neural network architecture was used to differentiate between flawless surfaces and surfaces containing a variety of common defects, while the neu- ral network was successfully applied to recognize defects in video frames. Chun Wang [20] proposed a lightweight strip steel surface defect detection model YOLOv8-VSC, based on the YOLOv8n target detection framework, using the lightweight VanillaNet [21] network as the backbone feature extraction network, reducing the complexity of the model by reducing the unnecessary branching structure, and providing a highly efficient and practical solution for strip steel surface defect detection . Xiang Kuan [22] proposed an improved Faster RCNN deep learning network for the detection of 10 defects on aluminum surface. Firstly, the region of interest calibration is used instead of the region of interest pooling to reduce the distortion of the feature map and improve the model's localization accuracy for small target defects. Kun Xie [23] proposed an improved YOLOv5s model for aluminum surface defect detection, which fuses the ShufflenNetV2-Block algorithm into the backbone network backbone to reduce the computational complexity of the model, and adds the SE attention mech- anism to focus the attention on the defect-related regions to better distinguish the differences between the categories and improve the classification performance and detection efficiency. It also reduces the size and memory occupation of the model, which is easy to deploy on mobile. Gang Deng [24] proposed a method of using IoU- K-means++ algorithm instead of K-means algorithm for clustering anchor frames, which can obtain anchor frames that better fit the surface defects of aluminum alloy profiles, and improve the quality of the anchor frames for small targets and the detec- tion ability of small targets. Based on the limitations of the above research, this paper proposes a high-precision aluminum surface defect detection model. The model takes YOLOv8 as the baseline and introduces the EMA attention mechanism mod- ule to reduce redundant calculations, weaken the influence of background factors, and strengthen the feature extraction capability of aluminum profile defects. Mean- while, the latest RepConv (Reparameterized Convolution) is integrated into YOLOv8. With the reparameterization technique, RepConv is able to simplify the multi-branch structure into single-branch structure when inference is performed, which significantly reduces the computation and memory consumption, and improves the inference speed. In addition, the loss function is improved by applying a combination of WIoU and CIoU instead of the original regression loss function, which enhances the small target detection capability and improves the detection accuracy.

Therefore, this paper highlights the following contributions.

- Proposed a novel aluminum surface defect detection model based on YOLOv8, integrating the latest RepConv into YOLOv8. Through the reparameterization technique, RepConv simplifies the multi-branch structure into a single-branch struc- ture during inference, significantly reducing computational overhead and memory consumption while enhancing detection efficiency.
- Incorporated the EMA mechanism into YOLOv8's C2f module, improving feature extraction and defect localization capabilities. The channel-wise attention enhancement demonstrates superior performance on fine-grained defect patterns.

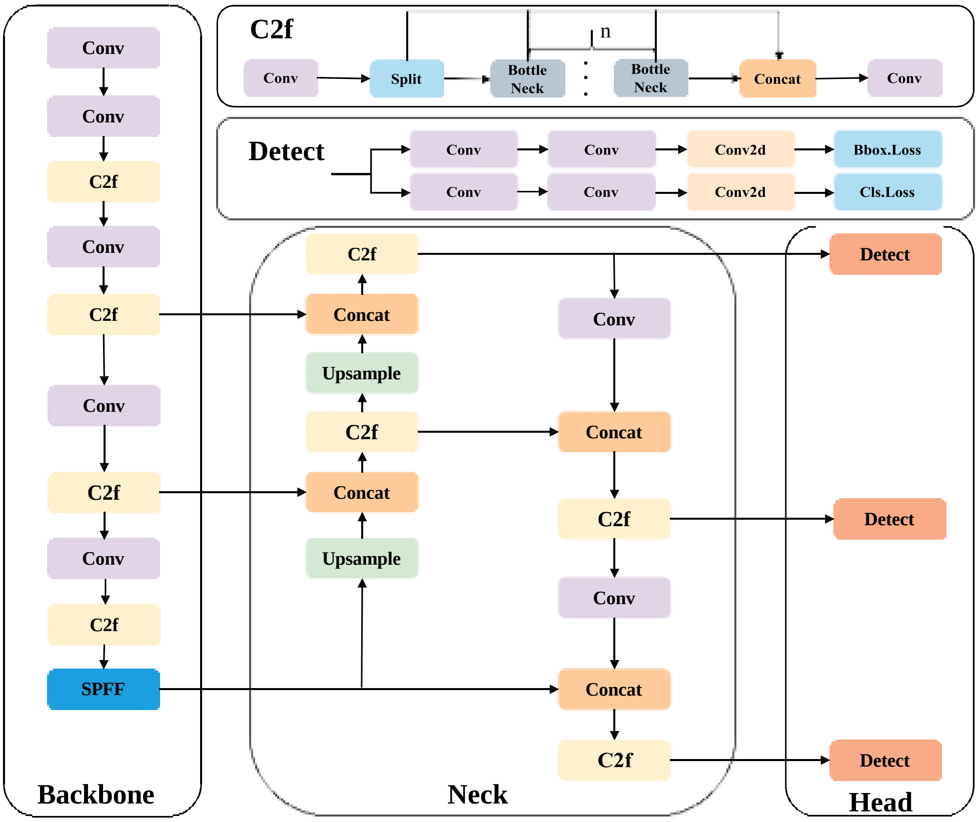


Fig. 1: YOLOv8 network architecture.

- Optimized the regression loss function by hybridizing CIoU and WIoU, effectively increasing small target detection accuracy. The combined loss function achieves better bounding box regression stability across varying defect sizes.

2. Methodologies

2.1 YOLOv8 Algorithm

YOLOv8 is a new algorithm in the YOLO family and is an improvement on YOLOv5 by Ultralytics. It is faster, more accurate, and easier to train and tune than previous versions. As shown in Figure 1, the architecture of YOLOv8 consists of three main parts: the backbone network, the neck, and the detection head. The backbone network is responsible for feature extraction; the neck fuses these features; and the detection head performs the final detection. In the backbone network, YOLOv8 introduces the SPPF module to improve detection accuracy and replaces the C3 module in YOLOv5 with the lighter C2f module. It also adjusts the convolution kernel of the first con- volutional layer from 6x6 to 3x3. The neck removes the 1x1 downsampling layer in

YOLOv5. The detection head is changed from coupled to decoupled and from Anchor- Based to Anchor-Free. In terms of loss function, YOLOv8 adopts a combination of DFL Loss and CIoU Loss, which incorporates the idea of regression loss.

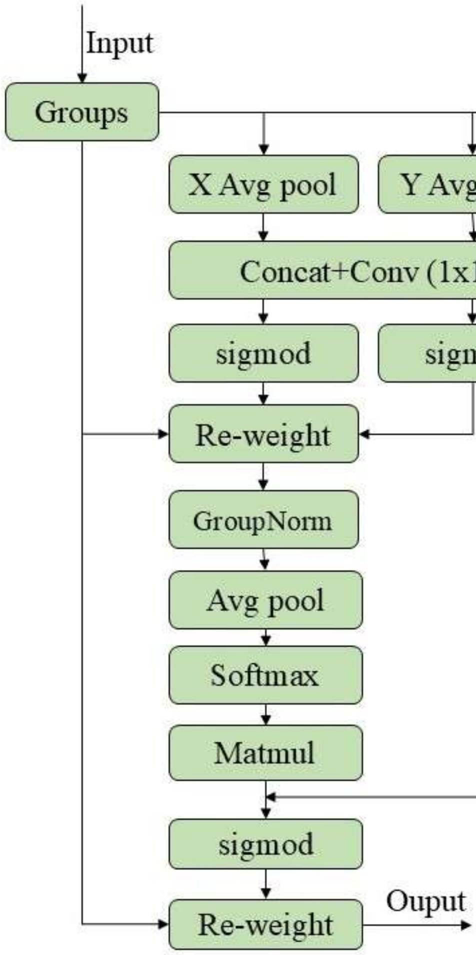


Fig. 2: EMA Attention Mechanism.

2.2 EMA Attention Mechanism

Attention mechanism [25] is a widely used data processing method in the field of machine learning. By assigning different weights to different regions of the input image, it enables the model to focus more on the key information and also helps to avoid the overfitting problem and enhance the stability and adaptability of the model.

EMA Efficient Attention Mechanism [26] possesses the functional module of multi- scale feature fusion, which enables the model to acquire and integrate large-scale and small-scale feature information at the same time, and thus significantly improves its detection ability for different. This significantly improves the detection capability of the model for different sized targets.

In terms of channel dimension, the EMA performs a grouping process on the input feature map X ∈ RC×H×W , dividing it into sub-feature groups of G. Each sub-feature group focuses on specific semantic information, thus helping the model to learn the information of different channels more effectively. The grouping method can be expressed as X = [X0, X1, · · · , XG−1], where for each Xi ∈ RC/(G×H×W). During the feature extraction process, three parallel subnetwork structures are adopted: two 1×1 convolution branches and one 3×3 convolution branch. For the 1×1 convolution branches, each channel feature is encoded along two spatial directions, respectively, and a one-dimensional global average pooling operation is performed; while in the 3×3 convolution branch, only a 3×3 convolution kernel is used to obtain the multi-scale feature representation.

In terms of spatial dimensions, EMA innovatively proposes a cross-space informa-tion fusion method, which realizes feature fusion by integrating features in different spatial dimensional directions and effectively handles both short- and long-range dependencies, a feature that exhibits a significant advantage in the task of dense small target detection. Specifically, EMA introduces two tensors from the outputs of the 1×1 convolutional branch and the 3×3 convolutional branch, and subsequently encodes the global spatial information in the output of the 1×1 convolutional branch using a 2D global average pooling operation, which is shown in formula 1.

$$z_c = \frac{1}{H \times W}\sum_{j=1}^{H}\sum_{i=1}^{W} x_c(i,j) \quad (1)$$

2.3 Reparameterization Module

The multibranch structure [27] can effectively extract multiscale features and improve model accuracy, but its complexity also brings problems such as difficult implementa- tion, slow inference, low memory utilization, and is not conducive to the application of channel pruning techniques. In order to solve these problems, this paper intro- duces a reparameterization module, which employs a multi-branch parallel structure in the model training phase to enhance the feature extraction capability, and fuses the multi-branch structure into a single-branch structure in the validation phase, thus improving the detection efficiency. Specifically, this paper replaces the ordinary Conv module with the RepConv module [28] in the Neck network part to improve the capa- bility to extract and merge features of the Neck network. The structure of RepConv is shown in Fig. 3, which adopts a convolutional reparameterization method using a parallel strategy of 3×3 Conv, 1×1 Conv, and BatchNorm at the training stage, while in the validation phase the BN layer and Conv are fused into one 3×3 Conv by the reparameterization module to achieve fast detection. The Conv formula is shown in Eq. 2, the BN layer formula is shown in Eq. 3, and the fusion process is shown in Eq. 4.

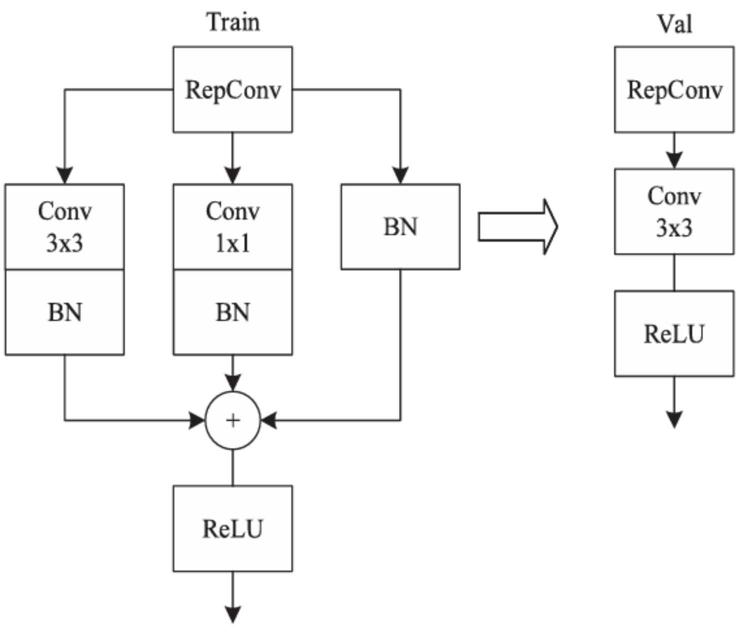


Fig. 3: Repconv Structure.

$$\begin{cases} \text{Conv}(x) = W(x) + b \\ \text{avg}(x) = \frac{1}{n}\sum_i^n x_i \end{cases} \tag{2}$$

$$\begin{cases} \sigma^2(x) = \frac{1}{n}\sum_{i=1}^n (x_i - \text{avg}(x))^2 \\ BN(x) = \alpha\frac{x-\text{avg}(x)}{\sqrt{\sigma^2(x)}} + \beta \end{cases} \tag{3}$$

$$f_{use} = \alpha\frac{x - \text{avg}(x)}{\sqrt{\sigma^2(x)}}(W(x) + b) + \beta \tag{4}$$

In the formula, denotes the input, denotes the convolution (Conv) weight, denotes the bias, denotes the mean of the input, denotes the variance of the input, and denotes the weight and bias of the batch normalization (BN), respectively, and denotes the fusion operation of convolution and batch normalization. In the transformation process of the model, the weight size of each branch needs to be converted to 3×3, where 1×1 convolution is used to realize the size conversion by filling operation. Finally, the weight values and bias of each branch are summed separately as the weight values and bias of the single-branch convolution.

3. Loss Function

In the YOLOv8 algorithm, the bounding box regression loss function is used as CIoU Loss [29], and its complete loss function is defined as:

$$Loss_{CIoU} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha\theta \tag{5}$$

$$\theta = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{6}$$

$$\alpha = \frac{\theta}{(1 - \text{IoU}) + \theta} \tag{7}$$

In the formula, $LossCIoU$ denotes the value of the loss function. IoU is the inter- section and concatenation ratio, which is used to calculate the ratio between the intersection area and the concatenation area of the two boundary boxes. p^2 denotes the Euclidean distance between two centroids, where b is the centroid of the predicted box and b^{gt} is the centroid of the true box. α is the weighting function to balance the loss of different parts. θ is used to measure the consistency of the aspect ratio. w^{gt} and h^{gt} are the width and height of the real frame and w and h are the width and height of the predicted frame, respectively. c is the diagonal distance of the smallest closure region that can contain both predicted and real boxes.

The loss function for the regression of the bounding box is critical to the perfor- mance of the model in the target detection task.The CIoU loss function, although it takes into account the aspect ratio of the predicted box to the real box, its complex- ity increases the difficulty of network optimization. In addition, it heavily penalizes low-quality samples, which affects the model generalization performance. Therefore, this paper adopts the loss function of CIoU followed by WIoUV3 [30], which aims to measure the quality of anchor frames by using the dynamic non-monotonic focus- ing mechanism and outlier degree, which can evaluate and optimize the detection accuracy more precisely, and improve the generalization ability and robustness of the model. The outlier degree β is defined as:

$$\beta = \frac{L_{IoU}^{*}}{L_{IoU}}[0, +\infty) \tag{8}$$

In the formula, L∗IoU denotes the tensor separated from the computed image, while LIoU is the IoU loss function value of the prediction frame. The higher the quality of the anchor frame, the smaller its outlier. This means that the smaller part of the gradient gain will be assigned to anchor frames with larger outliers, thus effectively avoiding low quality samples from generating too large harmful gradients. By utilizing the outlier degree β and constructing a non-monotonic focusing coefficient r, applying it to the LWIoUV 1 loss function yields the WIoUV3 loss function LWIoUV 3:

$$L_{WIoUV3} = rL_{WIoUV1} \tag{9}$$

$$r = \frac{\beta}{\delta\sigma^{\beta-\delta}} \tag{10}$$

In the equation, both δ and σ are hyperparameters. When the outlier β is set to δ, the value of the nonmonotonic focusing coefficient r is 1. When the outlier β reaches a certain constant T, the anchor frames are able to obtain the highest gradient gain.

In the middle and late stages of training, the WIoUV3 loss function assigns smaller gradient gains to low-quality anchor frames, thus reducing the harmful gradients of these anchor frames. At the same time, the WIoUV3 loss function is able to focus on normal-quality anchor frames and improve the overall performance of the model by optimizing the gradient allocation to these anchor frames.

4. Experiments and Results

4.1 Experimental Settings

In this paper, we adopt the PyTorch deep learning framework and utilize NVIDIA GeForce RTX4090 GPUs, CUDA12.1, and programmed using Python3.11. The initial learning rate is set to 0.01, the learning rate momentum is 0.937, the weight decay coefficient is 5 × 10-4 , the batch size is set to 16, and 300 epochs are rowed in the experimental environment set in this paper.

4.2 Experimental Dataset

This study uses aluminum profile defect data provided by the Alibaba Cloud Tianchi Competition. The data set comprises 10 defect categories, including nonconductivity, scratch mark, impact damage, orange peel effect, exposed substrate, horizontal defor-mation or pressed mark, powder-protrusion, pitting, coating crack, and contamination spot, with an image resolution of 2560×1920. To address the issue of class imbalance in the dataset, this experiment employs data augmentation techniques such as ran-dom flipping, brightness transformation, and rotation-scaling, ultimately generating a balanced set of 3,416 defect samples. To ensure effective model training, the dataset is divided into training, validation, and test sets in a 7:2:1 ratio, consisting of 2,386 images for training, 684 for validation, and 346 for testing.

4.3 Evaluation Metrics

In the field of object detection, evaluating model performance is crucial and detection accuracy is one of the core considerations. This paper adopts widely recognized eval-uation standards in computer vision, including AP (Average Precision), mAP (Mean Average Precision), and FPS (Frames Per Second), as metrics for assessing model performance. Here, AP represents detection precision, mAP@0.5 denotes average detection precision, and FPS indicates detection speed.

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$AP = \int_0^1 P(r)\, dr \tag{13}$$

$$MAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{14}$$

Where, TP represents the number of correct detections, FN represents the number of missed detections, FP represents the number of misdetections, P is the accuracy rate, R is the completeness rate and N is the number of categories for the target detection task. A higher value of AP indicates that this category is better detected in this network model, and a higher value of mAP@0.5 reflects the effectiveness of this network model for target detection.

GFLOPS and FPS are two core performance metrics that measure the computa-tional complexity and real-time inference speed of the model, respectively. GFLOPS represents the total number of floating-point operations required by a model for a single inference. GFLOPS represents the total number of floating-point operations required by the model for a single inference, and FPS represents the number of images that the model can process per second. GFLOPS represents the total number of floating-point operations required by the model for a single inference, and FPS rep-resents the number of frames per second that the model can process, which directly determines real-time performance.

4.4 Comparative Experimental Analysis

The main algorithms in the field of target detection, including Faster RCNN [31], which is a two-stage target detection algorithm, as well as single-stage target detec-tion algorithms such as SSD [32], the YOLO series, DETR [33], RT-DETR [34] and DINO [35] are selected for comparative experiments with the algorithms in this paper, and the Guangdong Industrial Wisdom Manufacturing Big Data Innovation Compe-tition's aluminum surface defects dataset as the evaluation standard, and the results of the comparison test are shown in Table 1. According to Table 1, it can be seen that the model in this paper achieves an optimal balance compared to other mainstream algorithms in the four key metrics of average accuracy, number of parameters, com-putation volume, and number of frames processed per second. Compared with the representative algorithm of the two phases, Faster RCNN, the model shows significant advantages in all the indexes, especially in the two indexes of the number of frames processed per second and the amount of computation, which makes the model more convenient to be deployed in the actual production environment. Compared with the baseline YOLOv8 model, the algorithm achieves a 4.6% improvement in average accu-racy and a 14 f/s improvement in frames per second, and the number of parameters is comparable to YOLOv8 and better than other models.

Table 1: Comparative Experiments.

Method	mAP@0.5/(%)	Params/M↓	GFLOPs/G↓	FPS(f/s)↑
Faster RCNN	79.1	137.08	37.1	15
SSD	75.2	26.25	62.5	44
YOLOv5	80.3	7.03	15.6	74
YOLOv7	79.8	6.02	13.1	77
YOLOv8	82.4	3.01	8.1	80
DETR	85.2	41	187	76
DINO	84.9	47	279	75

RT-DETR	85.9	20	60	81
Our method	**87.0**	**3.12**	**8.5**	**94.0**

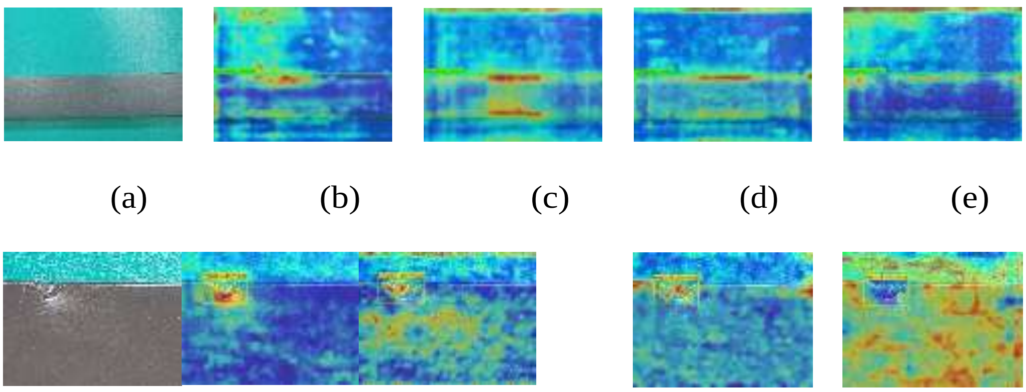
Table 2: Ablation Study.

Method	mAP@0.5/(%)	Params/M	GFLOPs/G	FPS(f/s)
YOLOv8	82.4	3.01	8.1	80.0
YOLOv8 + CWIoU	83.2	3.01	8.1	80.0
YOLOv8 + EMA	85.3	3.12	8.5	60.7
YOLOv8 + Rep	83.9	3.01	8.1	81.0
YOLOv8 + CWIoU + EMA	85.7	3.12	8.5	60.7
YOLOv8 + CWIoU + Rep	84.4	**3.01**	**8.1**	81.0
YOLOv8 + EMA + Rep	85.4	3.12	8.5	62.9
YOLOv8 + CWIoU + EMA + Rep	**87.0**	3.12	8.5	**94.0**

4.5 Ablation Experiments

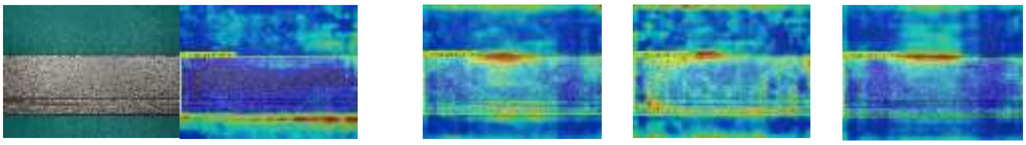
The results of the ablation experiment are shown in Table 2. The first row represents the experimental results of the original YOLOv8 model, and the second to the seventh rows represent the ablation experiment part. As can be seen from the data in Table 2, after introducing the EMA attention mechanism module into the c2f module of the YOLOv8 algorithm, the mAP@0.5 of the model is increased by 2.9%. This indicates that the EMA attention module enhances the network's ability to perceive multi-scale information. When traditional Conv is replaced by Repconv in the YOLOv8 algorithm network, mAP@0.5 increases by 1. 55%. This shows that the Repconv module can learn more deep features without increasing the scale of the model. After integrating WIoU into the boxxLoss CIoU function in the YOLOv8 algorithm, mAP@0.5 increases by 0. 8%. This shows that the CWIoU loss function has better performance on the surface defect dataset of aluminum profiles studied in this paper. After adding the CWIoU loss function, the EMA attention mechanism module and Repconv, compared to the original YOLOv8 algorithm, the mAP@0.5 increases by 4. 6%. In conclusion, the results of the ablation experiment fully demonstrate that, compared with the original YOLOv8n model, the improved model has certain advantages in the four key indicators of mAP@0.5, the number of parameters, GFLOPs, and frame rate.

Original	Model A	Model B	Model C	Model D



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)



(k) (l) (m) (n) (o)

Fig. 4: Heat map of ablation experiments.

Fig. 4 illustrates the experimental results of the thermograms for the three types of defects. Group A is the thermogram of the original YOLOv8 model, which has unsatisfactory detection results, especially for small target defects. group B intro- duces an improved loss function CWIoU based on group A, which leads to improved detection performance, but accuracy is still unsatisfactory. group C further adds the EMA attention mechanism module, which enhances the model's attention to small targets and inconspicuous defects.Group D, on the other hand, replaces the original conv module with Repconv on the basis of this C, which significantly improves the model's performance, in which the detection accuracies of stray color, angular leakage of the bottom, and orange peel reach 99.5%, 66.6%, and 97.3%, respectively.

5. Conclusion

In order to solve the problems of low recognition accuracy, slow detection speed, and difficult deployment in the aluminum profile defect detection process, this paper pro- posed an improved defect detection algorithm model based on YOLOv8. The model firstly employs the EMA attention module to enhance the feature extraction ability of the model, so that it can more accurately extract feature information in complex backgrounds. RepConv, instead of the traditional Conv, is implemented in the train- ing phase to enhance the feature extraction ability. The single-branching structure is used to accelerate the speed of detection. And finally, the loss function of CIoU fol- lowed by WIoUV3 is used, which aims to measure the quality of anchor frames by using the dynamic non-monotonic focusing mechanism and outlier degree, which can more accurately assess and optimize the detection accuracy, and improve the model's generalization ability and robustness. The experimental results show that the accu- racy of the improved model reaches 87. 0%, and the size of the model is only 3.12

M. Compared with the original model, the new model effectively has the advantages of high recognition accuracy, fast detection speed, and small model memory occupa- tion in aluminum profile defect detection, which has the potential to be deployed in industrial applications.

6. Acknowledgements

The authors thank the anonymous reviewers for their valuable comments and helpful suggestions.

Author Contributions Ziming Qu, Xufeng Zhang and Fafa Zhang: Methodology, Experiments, Writing original draft; Hui Wang: Conceptualization, Writing, Review and Editing.

Data Availability The data supporting the validation of this study is available. The complete data sets used for this study are available from the corresponding author on reasonable request.

7. Declaration

Funding and Competing of Interests The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. The authors have no competing interests to declare those are relevant to the content of this article.

Ethical and Informed Consent for Data Used This article does not contain any studies with human participants carried out by any of the authors.

References

1. Jing, H., Jia, G., Liu, X., et al.: Application of spotted hyena optimization algo- rithm in aluminum profile loading. Machinery Design & Manufacture, 1–6 (2025). [Online; accessed 2025-05-05]
2. Zhou, W., Shao, Z., Yu, J., Lin, J.: Advances and trends in forming curved extrusion profiles. Materials **14**(7), 1603

(2021)
3. Czerwinski, F.: Current trends in automotive lightweighting strategies and materials. Materials **14**(21), 6631 (2021)
4. Zhao, Y., Sun, G., Wang, H., et al.: Aluminum profile surface defect detection based on cda-yolov8. Laser & Optoelectronics Progress, 1–14 (2025). [Online; accessed 2025-05-05]
5. Chen, R., Cai, D., Hu, X., Zhan, Z., Wang, S.: Defect detection method of alu- minum profile surface using deep self-attention mechanism under hybrid noise conditions. IEEE Transactions on Instrumentation and Measurement **70**, 1–9 (2021)
6. Usamentiaga, R., Lema, D.G., Pedrayes, O.D., Garcia, D.F.: Automated surface defect detection in metals: A comparative review of object detection and semantic segmentation using deep learning. IEEE Transactions on Industry Applications **58**(3), 4203–4213 (20 22)
7. Chen, M., Wang, H., Cheng, J., *et al.*: Research progress on subsurface defect detection and suppression technology for fused silica optical components process- ing. Journal of Mechanical Engineering **57**(20), 1–19 (20 21)
8. Chen, Y., Ding, Y., Zhao, F., Zhang, E., Wu, Z., Shao, L.: Surface defect detection methods for industrial products: A review. Applied Sciences **11**(16), 7657 (2021)
9. Li, H., Zhou, M., Chen, C., et al.: Aluminum profile defect detection based on improved yolov5. Modular Machine Tool & Automatic Manufacturing Technique (2), 194–199 (20 25)
10. Li, J., Zheng, P., Li, Y., et al.: Research on aluminum profile surface defect detec- tion algorithm combining ghostbottleneck and attention mechanism. Modern Manufacturing Engineering (3), 115–123 (2025)
11. Dong, J., Sang, F., Guo, Q., et al.: A review of lightweight research on deep learning-based object detection algorithms. Journal of Frontiers of Computer Science & Technology, 1–34 (2025). [Online; accessed 2025-05-05]
12. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, pp. 580–587 (2014). https://doi.org/10.1109/CVPR.2014.81
13. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analy- sis and Machine Intelligence **39**(6), 1137–1149 (2017) https://doi.org/10.1109/TPAMI.2016. 2577031
14. Guo, Y.: Design of online product surface defect detection system based on machine vision. Master's thesis, Soochow University (2014)
15. Li, X.: Research progress and industrial applications of machine vision. Digital Communication World (11), 79–80146 (2021)
16. Liu, X., He, W., Zhang, Y., Yao, S., Cui, Z.: Effect of dual-convolutional neu- ral network model fusion for aluminum profile surface defects classification and recognition. Mathematical Biosciences and Engineering **19** (1), 997–1025 (2022)
17. Wu, X., Liang, D., Liang, D., Li, P.: Aluminum profile surface defect recognition method based on yolo deep learning model. Machine Design & Research **37**(2), 34–39 (2021)
18. Xue, Y., Ding, K., Li, Q., Yang, J., Li, J.: Piston rod surface defect detection based on improved yolov5s. Modern Manufacturing Engineering (11), 104–112 (2023)
19. Neuhauser, F.M., Bachmann, G., Hora, P.: Surface defect classification and detection on extruded aluminum profiles using convolutional neural networks. International Journal of Material Forming **13**(3) (2020)
20. Wang, C., Liu, H.: Yolov8-vsc: A lightweight strip steel surface defect detection algorithm. Journal of Frontiers of Computer Science & Technology **18**(1), 151–160 (2024)
21. Chen, H., Wang, Y., Guo, J., et al.: Vanillanet: The power of minimalism in deep learning. arXiv, 12972 (2023)
22. Xiang, K., Li, S., Luan, M., *et al.*: Aluminum surface defect detection method based on improved faster rcnn. Chinese Journal of Scientific Instrument **42**(1), 191–198(2021)
23. Xie, K., Fang, K., Chen, J., et al.: Aluminum surface defect detection method based on improved yolov5s. Manufacturing Technology & Machine Tool (1), 179– 184 (2024)
24. Deng, G., Zhao, Q., Qi, G., et al.: Research

on aluminum alloy profile surface defect detection method based on yolov5. Modern Manufacturing Engineering (11), 120–128 (2023)
25. Ren, H., Wang, X.: A review of attention mechanisms. Journal of Computer Applications **41**(S01), 6 (2021)
26. Ouyang, D., He, S., Zhan, J., et al.: Efficient multi-scale attention module with cross-spatial learning. arXiv, 2305–13563 (2023)
27. Wang, H., Wang, B., Ge, C.: Reparameterized yolov8 road damage detection algorithm. Computer Engineering and Applications **60**(5), 191–199 (2024)
28. Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J.: Repvgg: Making vgg- style convnets great again. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 13728–13737 (2021)
29. Zhang, Y.-F., Zhang, Z., Jia, Z., *et al.*: Focal and efficient iou loss for accu- rate bounding box regression. Neurocomputing (2022) https://doi.org/10.1016/j.neucom.2022.07.042
30. Tong, Z., et al.: Wise-iou: Bounding box regression loss with dynamic focusing mechanism. arXiv, 2301–10051 (2023)
31. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(6), 1137–1149 (2017)
32. Liu, W., Anguelov, D., Erhan, D., *et al.*: Ssd: Single shot multibox detector. In: Computer Vision – ECCV 2016. Springer, ??? (2016)
33. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020 (2020)
34. Zhao, Y., *et al.*: Detrs beat yolos on real-time object detection. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 16965–1697 (2024)
35. Zhang, H., et al.: Dino: Detr with improved denoising anchor boxes for end-to-end object detection. arXiv, 2203–03605 (2022).