

# Improved the Notched Character Defect Detection Algorithm of YOLOv8

Hao Jiang, Lu Liu \*

Tianjin University of Technology and Education, China

## ABSTRACT

Engraved character marking technology is a method of creating permanent identifiers on workpiece surfaces by directly removing material or causing plastic deformation through physical action. This technology is widely used in industrial fields for product traceability, quality control, and anti-counterfeiting certification. However, long-term operation of marking equipment leading to aging and fluctuations in process parameters can easily cause character defects such as broken strokes and distortions, directly affecting the readability and reliability of the identification. This paper addresses the issue of low detection accuracy for common character defects in the original YOLOv8 model by optimizing it. Experiments show that the optimized model significantly improves detection accuracy for various defect types compared to the original model, achieving satisfactory detection performance.

## KEYWORDS

Engraved Characters; Defect Detection; YOLOv8; Detection Accuracy

## 1. INTRODUCTION

Defect detection is a core task within the field of computer vision. It is an application branch of object detection tasks. Object detection technology finds extensive applications in industrial automation, medical imaging, and other domains. In recent years, with the rapid development of deep learning, the accuracy and efficiency of object detection algorithms have been substantially improved. Particularly with the support of large-scale datasets and powerful computational resources, numerous deep learning-based object detection algorithms have emerged. These algorithms are primarily divided into two categories: traditional object detection algorithms and deep learning object detection algorithms [1].

Before the rise of deep learning, object detection algorithms mainly relied on hand-crafted features and traditional machine learning models. Common traditional object detection algorithms include:

- Haar features [2] with Adaboost classifier [3]: This was one of the early object detection methods, widely used in tasks like face detection. It scans the image using a sliding window approach and classifies using a weighted combination of simple features. Its advantage lies in fast computation speed, but its performance is poor for complex scenes and multi-scale objects.
- DPM (Deformable Part Models) [4]: DPM is a part-based detection model that describes various parts of an object and their geometric relationships to achieve detection. DPM showed good performance on relatively complex detection tasks, but it relies on manual feature extraction, which made it difficult to handle large-scale, diverse detection tasks.

Traditional object detection methods, limited by the representational capacity of manual feature extraction and model complexity, exhibited significant limitations when dealing with complex scenes or multi-scale, multi-shape objects. With the advent of deep learning algorithms, traditional methods have gradually been replaced by neural network-based deep learning approaches.

The introduction of deep learning, particularly Convolutional Neural Networks (CNNs), has fundamentally changed the research direction of object detection algorithms. Deep learning-based algorithms automatically learn features from images and, through training on large-scale data, greatly enhance the accuracy and robustness of object detection. Deep learning-based object detection algorithms can be broadly classified into two types: two-stage detectors and one-stage detectors.

Representative models of two-stage detectors are the R-CNN [5] series. Their core concept is to divide object detection into two steps: first generating a series of candidate regions in the image, and then classifying these regions and performing bounding box regression. One-stage object detection algorithms treat the detection task directly as a regression problem, outputting both the class and bounding box coordinates in a single pass, significantly improving detection speed. Main one-stage detectors include the YOLO (You Only Look Once) [6] series and SSD (Single Shot Multibox Detector) [7]. SSD is a multi-scale, one-stage detection algorithm that detects objects of different sizes using feature maps at multiple resolutions. YOLO is a typical representative of one-stage detectors. Unlike two-stage algorithms, YOLO takes the entire image as input, performs object detection in one go using global features, and outputs the class and bounding box coordinates for all objects at the network's output layer, offering very high real-time performance.

## **2. DATASET AND CHARACTER DEFECTS**

### **2.1. Introduction to the Dataset**

Data is the core driver of performance for the YOLOv8 object detection model. The data's quality, diversity, and scale directly determine the model's detection accuracy and generalization capability. High-quality annotated data ensures the model learns accurate bounding boxes and class features. Data augmentation techniques (such as geometric transformations, color perturbations, and Mosaic augmentation) further expand data diversity and is useful for optimizing the model's sensitivity to small objects [8]. In practical applications, it is necessary to fine-tune model with domain-specific data and address challenges like annotation costs and privacy concerns, balancing efficiency and effectiveness through data cleaning, synthetic data, or semi-supervised learning. Ultimately, the synergistic optimization of data and model architecture (e.g., anchor box design, input image resolution adjustment) is key to achieving efficient and accurate detection.

### **2.2. Introduction to Engraved Character Defects**

The dataset used in this study comes from a real industrial production field, consisting of images of numeric characters on metal workpiece surfaces by engraved marking equipment. The dataset contains 2341 images, classified into five categories: Broken Stroke defect, Missing Stroke defect, Extra Stroke defect, Distorted defect, and Correct characters, covering typical character defects caused by equipment aging and process fluctuations. To ensure reliable model evaluation, the dataset was randomly split into training set and test set by 7:3 ratio.

In order to solve the challenges of a scarcity of defect samples and insufficient sample diversity due to a single background, two advanced data enhancement technologies, Mosaic and Mixup, were introduced into the model training stage. Mosaic enhancement is conducive to improving the model's perception of global features by randomly scaling, cropping, and stitching four training images. Mixup enhancement constructs training samples through linear interpolation, thereby enhancing its generalization and ability to resist overfitting. The synergistic application of these two strategies

significantly expands the distribution space of training data and lays a solid foundation for improving the robustness and accuracy of defect detection models in complex industrial scenarios.

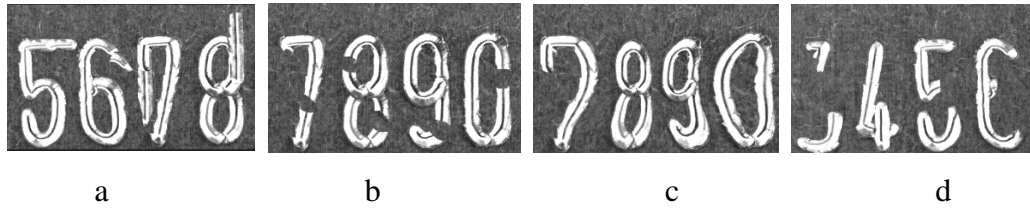


Figure 1. Example of a scratched character defect

### 3. THEORETICAL DESCRIPTION

#### 3.1. YOLOv8

YOLOv8 is a high-efficiency real-time target detection model launched by Ultralytics in 2023, which continues the core advantage of YOLO series of "Fast and Accurate", and realizes the balance of accuracy and speed through the improved CSPDarknet backbone network and Anchor-free detection head (53% mAP, 100+ FPS on COCO dataset). It supports multi-tasking extension of detection, segmentation, attitude estimation, etc. By combining Mosaic data augmentation with an adaptive feature pyramid, the lightweight version (e.g., YOLOv8n) enhances detection robustness for small targets and occluded scenes and is well-suited for edge computing devices [9]. The model provides comprehensive end-to-end development support, spanning from data annotation to TensorRT deployment, and is backed by an active open-source ecosystem. Its wide application in fields such as industrial quality inspection, autonomous driving, and security monitoring makes it an optimal choice for scenarios demanding real-time performance, flexibility, and a low technical threshold. It is particularly well-suited for rapid implementation in resource-constrained environments. The specific architecture is illustrated in the figure below.

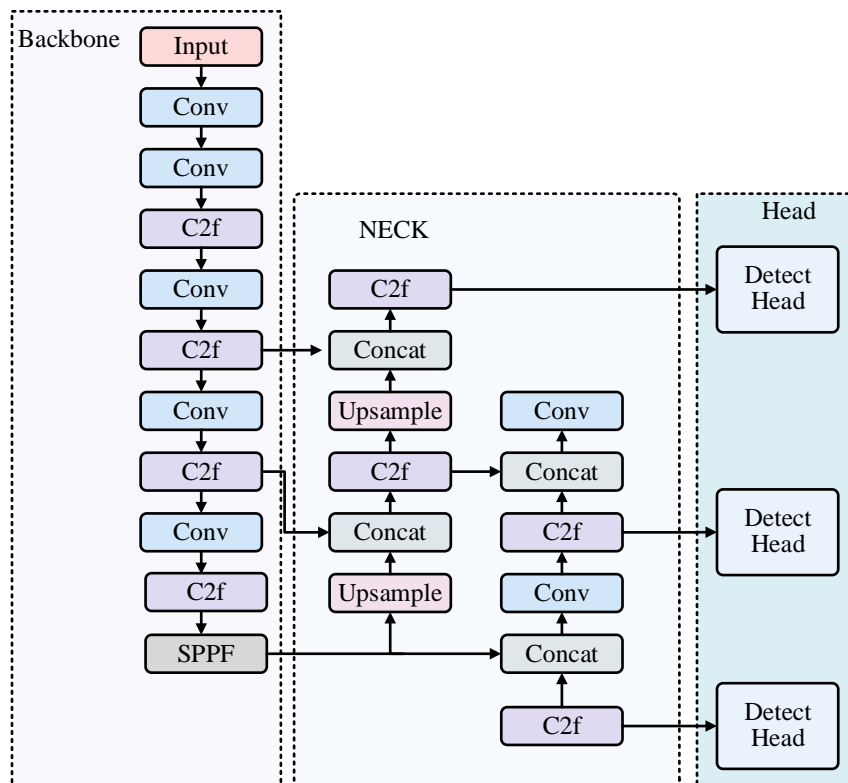
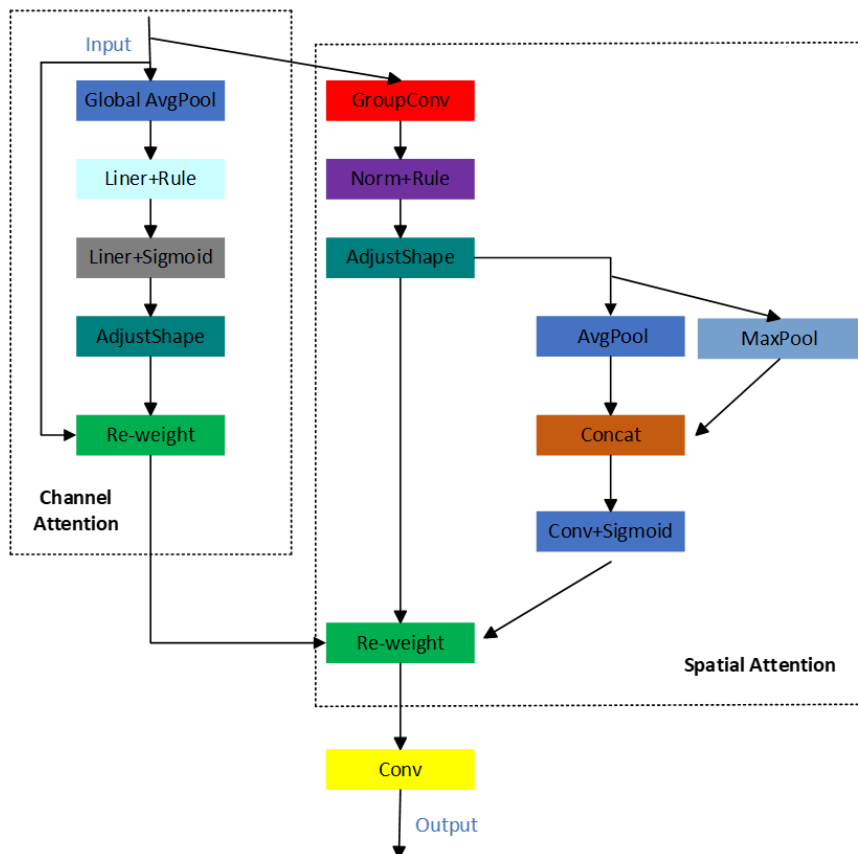


Figure 2. Structure of YOLOV8

## 3.2. Optimized Model

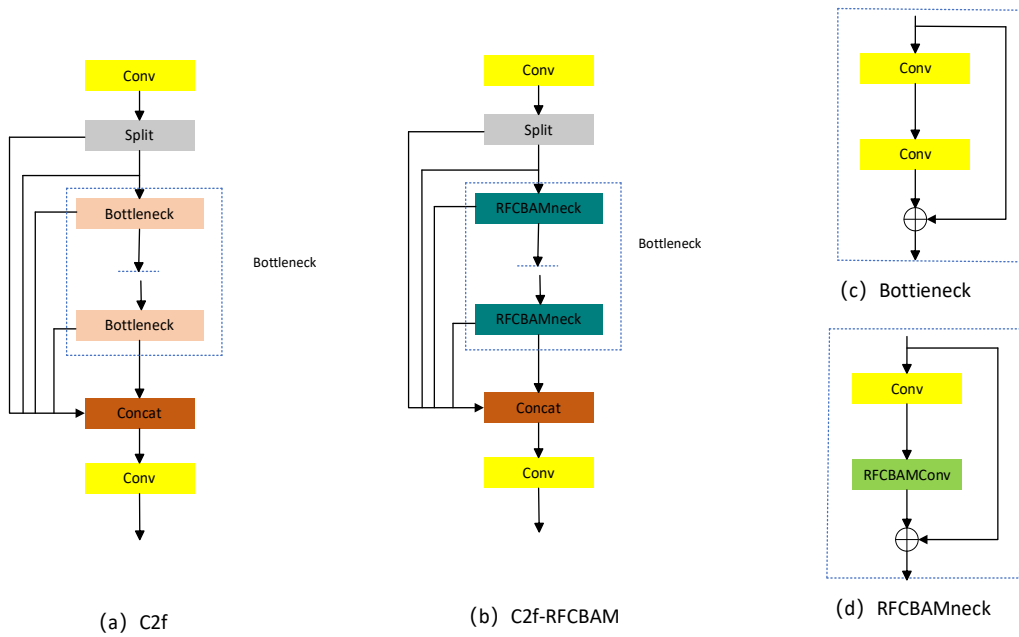
### 3.2.1. RFCBAMConv

RFCBAMConv [10] is an efficient feature extraction module designed to significantly improve the model's adaptability to complex scenes through combining multi-branch sensory field expansion with channel-space attention mechanism (CBAM). Its core captures multi-granularity context information and solves the problem of target scale change by concatenating dilated convolution or multi-scale convolution kernels with different expansion rates. Subsequently, it combines channel attention (dynamically enhancing key feature channels) and spatial attention (focusing on the target region) to realize adaptive screening and enhancement of features. This module takes into account multi-scale perception and dynamic feature optimization while maintaining lightweight computation, and is especially suitable for tasks such as small target detection, occluded scene detection and image segmentation. It can comprehensively balance the accuracy and efficiency better than similar modules (e.g., CBAM, RFB-Net), and is widely used in industrial quality inspection, remote sensing analysis and other fields that require robust feature expression [11]. The specific structure is shown in Figure 3.



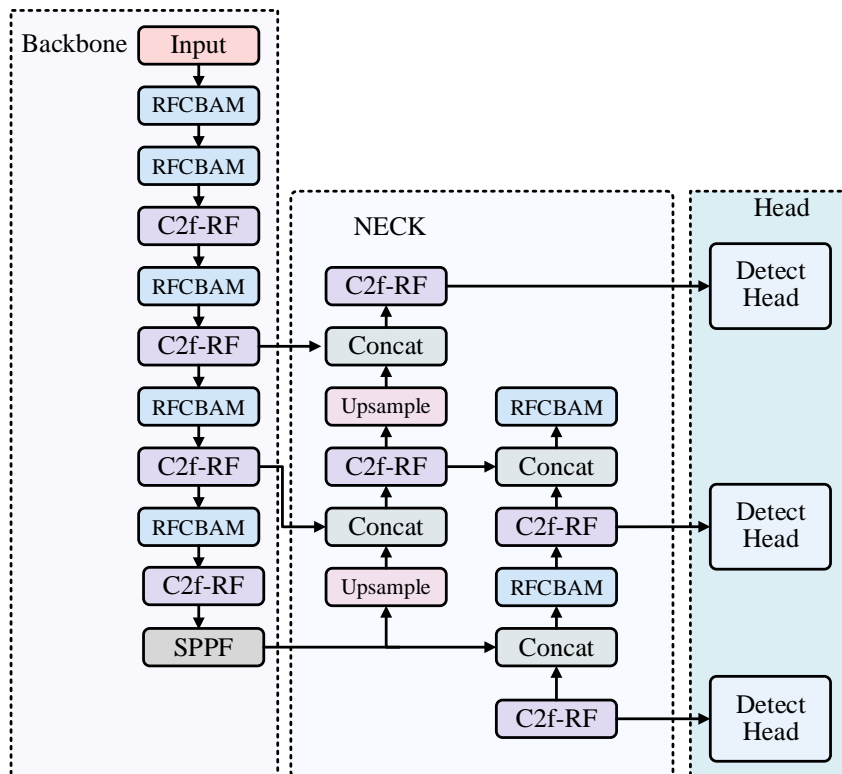
**Figure 3.** Schematic diagram of RFCBAMConv structure

By replacing the second convolutional layer in the Bottleneck module with an RFCBAM module (dubbed RFCBAMneck), the improved C2f module achieves enhanced feature extraction capabilities. The specific structure is shown in Fig. 4.



**Figure 4.** Schematic diagram of the C2f module, the RFCBAMneck module, and the submodules 3.2.2. Improved YOLOV8

Instead of convolution, RFCBAMConv is used and C2f-RFCBAM is designed to replace C2f in the backbone network. The network takes into account the importance of different features and enhances an ability of the network to extract features, as shown in Fig. 5.



**Figure 5.** Improved model structure

## 4. EXPERIMENTAL RESULTS AND EQUIPMENT INFORMATION

### 4.1. Equipment Information

The experiment adopts cloud-based GPU training, and the specific experimental parameters and related experimental equipment are shown in Table 1. The training configuration is based on the Ubuntu 18.04 system and CUDA 11.3 acceleration environment, using the PyTorch 1.10 framework with the Timm 0.9.8 image model library, and set up 200 training cycles (epochs) for the image tasks (such as classification or detection), with 224×224 resolution input and a batch size of 32. The model is optimized with an initial learning rate of 0.01, and the convergence stability is improved by closing Mosaic data enhancement (close\_mosaic=1) at the end of training. The overall configuration takes into account the framework version compatibility (MMDetection 2.0.0 + Python 3.8) and the GPU memory efficiency, and is suitable for the SOTA model training of medium-sized datasets, which needs to be coupled with at least. It is suitable for SOTA model training on medium-sized datasets, and requires a NVIDIA graphics card with at least 24GB of graphics memory to realize efficient computing.

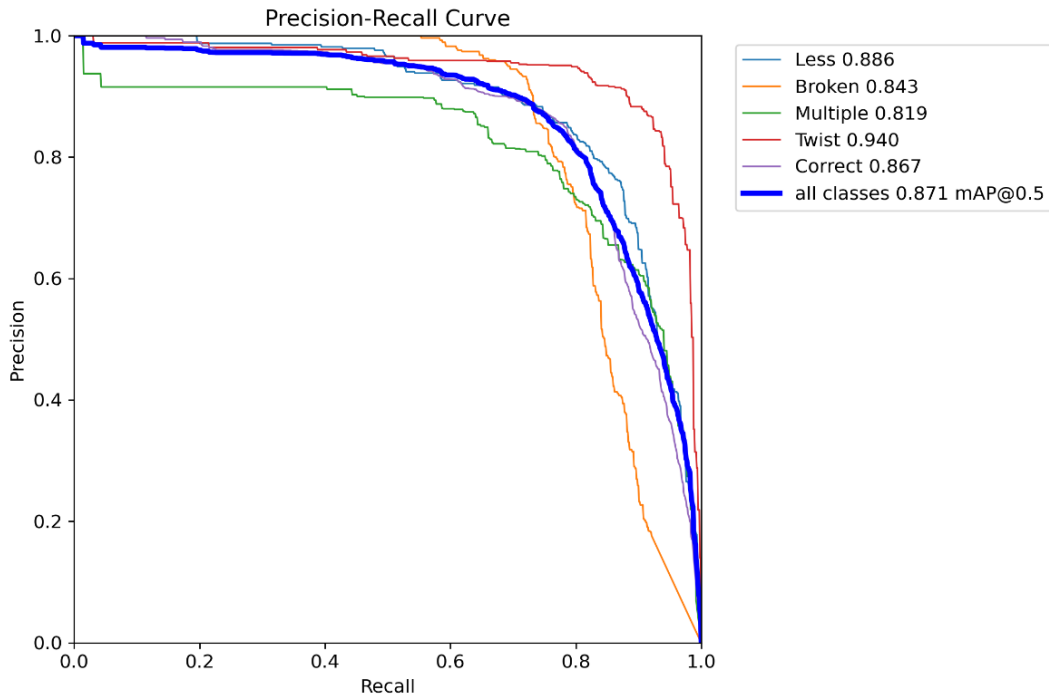
**Table 1.** Three Scheme comparing

Parameters	Set up
system	Ubuntu 18.04
timm	0.9.8
Python	3.8
pytorch	1.10
cuda	11.3
mmdet	2.0.0
epochs	200
close_mosaic	1

### 4.2. Analysis of Experimental Results

Figure 6 shows the performance of the object detection model in industrial defect detection: the detection results of the base YOLOv8 model on engraved characters. The Precision-Recall curve indicates good overall model performance, with a mean Average Precision (mAP@0.5) of 0.871, demonstrating high comprehensive accuracy in localization and recognition tasks.

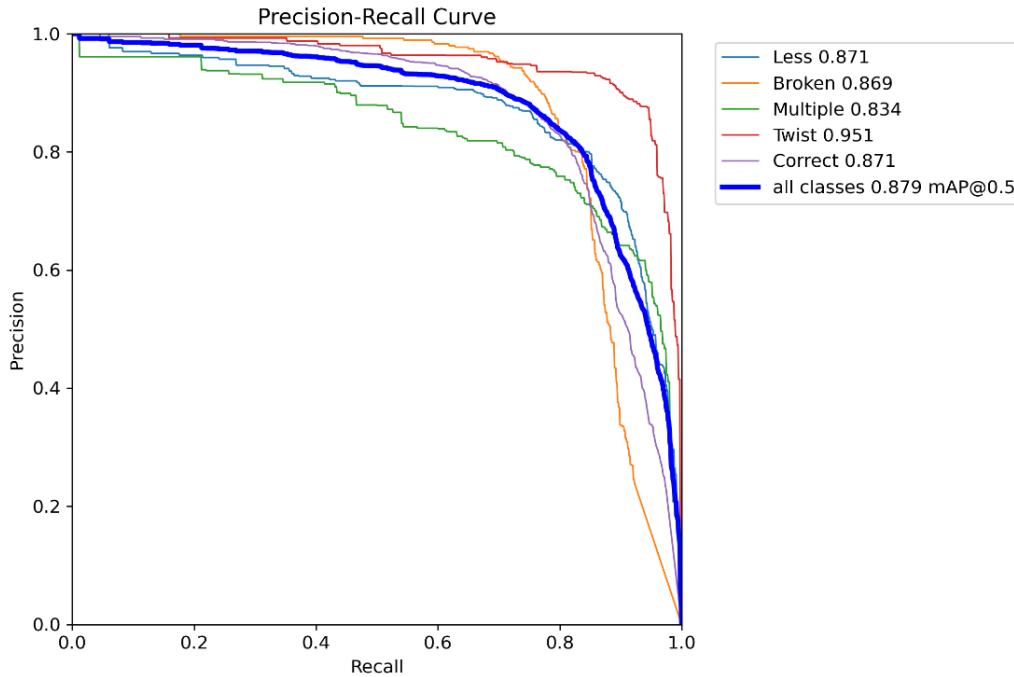
Analyzing the precision by category, the "Twist" category achieved the highest precision (0.940), indicating the model's strongest capability in recognizing distorted characters. The "Less" and "Correct" categories achieved 0.886 and 0.867 respectively, showing stable recognition performance. The "Broken" and "Multiple" categories were slightly lower, at 0.843 and 0.819 respectively, possibly due to the complexity of broken or overlapping characters affecting detection consistency.



**Figure 6.** Analysis of results before improvement

Analysis based on Figure 7 shows that the improved YOLOv8 model demonstrates comprehensive and significant performance enhancement in the engraved character detection task. The model's mAP increased from 0.871 to 0.879, indicating a stronger overall detection capability. The improvement is particularly evident in specific categories. The precision for the previously lower-performing "Broken" category increased substantially from 0.843 to 0.869, suggesting that the optimization enhanced the model's ability to learn features of incomplete characters, overcoming this specific challenge. The precision for the "Multiple" category also improved from 0.819 to 0.834, indicating better capability in recognizing overlapping characters. Notably, the precision for the "Twist" category reached an excellent 0.951, maintaining the model's superior recognition ability for complex character morphologies. Furthermore, the precision for the "Less" and "Correct" categories remained stable at high levels (both 0.871).

Considering the recall ranges, the model maintained high precision without sacrificing the coverage of positive samples, demonstrating robust generalization performance. In summary, the model improvements were effective, particularly in significantly enhancing the detection robustness for challenging cases like "Broken", laying a more solid foundation for applying the model to industrial inspection scenarios with extremely high reliability requirements.



**Figure 7.** Analysis of improved results

## 5. CONCLUSION

This paper focuses on the detection requirements for typical defects occurring in engraved characters during processing and use. Due to engraving process deviations, equipment wear, or external forces, characters are prone to various morphological defects such as missing strokes (Less), broken strokes (Broken), multiple engravings (Multiple), and distortions (Twist), which severely impact the reliability of character recognition and traceability systems. Addressing the issues of missed detections and false alarms with the original YOLOv8 model on morphologically variant characters, this research introduced a backbone optimization based on the RFCBAMConv module to enhance the model's ability to represent and fuse multi-scale defect features. This was combined with a channel-spatial attention mechanism to strengthen the focus on defect regions and suppress background interference. Experimental results show that the improved model achieved an mAP@0.5 of 87.9% on the self-built engraved character defect dataset, an increase of 0.8 percentage points over the baseline model. Notably, the detection precision for the "Broken" category significantly increased by 2.6%, and the precision for the "Twist" category reached 0.951, demonstrating excellent recognition robustness for deformed characters. This study provides an effective solution for detecting small, multi-type character defects in industrial identification scenarios, holding significant practical value for enhancing the accuracy and applicability of automated quality inspection systems.

## REFERENCES

- [1] Zou Z, Chen K, Shi Z, et al. Object detection in 20 years: A survey [J]. Proceedings of the IEEE, 2023.
- [2] Lienhart R, Maydt J. An extended set of haar-like features for rapid object detection [C]//Proceedings. international conference on image processing. IEEE, 2002, 1: I-I
- [3] Rätsch G, Onoda T, Müller K R. Soft margins for AdaBoost [J]. Machine learning, 2001, 42: 287-320.
- [4] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.
- [5] Girshick R. Fast R-CNN [C]. International Conference on Computer Vision, 2015:
- [6] Cortes C. Support-Vector Networks [J]. Machine Learning, 1995.

- [7] Girshick R, Iandola F, Darrell T, et al. Deformable part models are convolutional neural networks [C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2015: 437-446.
- [8] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-78
- [9] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector [C]. European Conference on Computer Vision. Springer International Publishing, 2016: 21-37.
- [10] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1-6.
- [11] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020: 1-6.