World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3103

CONTAINER DAMAGE DETECTION BASED ON DEEP RESIDUAL NETWORKS AND REAL-TIME OBJECT DETECTION ALGORITHMS

XingYu Zhao
School of Computer Science and Technology, Shandong University of Technology, Zibo 255049, Shandong, China.

Abstract: Container damage detection is a critical component for enhancing automation and transport safety in modern smart ports. To address practical challenges such as complex background interference in port environments, a wide range of damage scales, and uneven data distribution, this study constructs a damage detection framework based on the deep residual network ResNet50 and the lightweight detection algorithm YOLOv8. The study first employs a Super-Resolution Generative Adversarial Network (SRGAN) to enhance image resolution, thereby improving the distinguishability of fine cracks and dents. By introducing a cross-entropy loss function and a cosine-annealed learning rate strategy, the model achieves robust learning for three major damage types: dents, perforations, and corrosion. Experimental results demonstrate that the ResNet50 model excels at feature extraction in complex backgrounds, achieving a classification accuracy of 76.99% on the validation set, with particularly outstanding performance in identifying dent-type damage with distinct features. Meanwhile, the YOLOv8 model, which incorporates an attention mechanism, exhibits significant advantages in inference speed and multi-object localization, with an average accuracy of approximately 0.85, effectively meeting real-time monitoring requirements. This study confirms the efficiency and practical value of deep learning technology in handling container image recognition tasks characterized by high dynamics and diverse operating conditions.
Keywords: Container damage detection; Deep learning; Model performance evaluation

1 INTRODUCTION

Against the backdrop of rapidly expanding global trade, containers serve as the indispensable core carriers of modern logistics and transportation networks. However, during frequent loading, unloading, and stacking operations at busy terminals, they are highly susceptible to mechanical impacts and chemical corrosion, inevitably leading to various surface damages such as dents, perforations, and rust. These structural defects pose a serious threat to overall cargo integrity and transportation safety. Currently, traditional manual inspection methods can no longer meet the large-scale, high-efficiency operational demands of automated smart ports due to their inherent low efficiency, high labor costs, and extreme susceptibility to complex environmental interference. Consequently, shifting from manual labor to intelligent, automated container damage detection has emerged as a critical component for enhancing port automation and ensuring secure transport operations globally[1-3].

Previous studies have extensively explored the application of convolutional neural networks in general-purpose object recognition and industrial surface defect detection. For instance, various deep learning architectures have been successfully utilized to address inspection tasks in manufacturing and railway systems. However, most existing research remains insufficient when it comes to effectively balancing detection accuracy and deployment costs within specific, highly dynamic port scenarios. Intelligent detection in this context still faces significant technical barriers that hinder practical application. These include severe background noise caused by port machinery, unpredictable complex weather conditions, an extremely wide range of damaged target sizes—from tiny cracks to large-area rust—and the high visual similarity of features across different damage types. Consequently, achieving robust recognition under such diverse and challenging operating conditions continues to be a formidable obstacle.

To directly address the aforementioned challenges, this study systematically proposes a composite evaluation framework based on deep residual learning and anchor-free detection mechanisms. Specifically, we construct an integrated recognition model combining the ResNet50 network and an optimized YOLOv8 algorithm. Furthermore, Super-Resolution Generative Adversarial Network (SRGAN) technology is strategically introduced to achieve feature enhancement in low-quality imaging environments, significantly improving the distinguishability of fine cracks. By establishing a preprocessed dataset of approximately 3,300 container images, we conduct comprehensive performance benchmarking[4-5]. A deep ResNet50 model is constructed as a classification baseline, while an optimized YOLOv8 model incorporating an efficient channel attention mechanism is developed. Finally, through quantitative metrics such as precision, recall, and confusion matrices, we thoroughly analyze the model's robustness and generalization limits, aiming to provide a highly reliable technical solution for real-world container damage image recognition.

2 ESTABLISHMENT AND SOLUTION OF CONTAINER DAMAGE IMAGE RECOGNITION MODEL

2.1 Problem Background

Container damage detection plays a crucial role in modern logistics. As the main carrier of global trade, containers are often subject to various external forces during long-term transportation, loading and unloading, resulting in surface damages such as cracks, dents, holes and rust. These damages not only affect the structural integrity of containers but also may lead to safety accidents or cargo losses in the transportation process. Therefore, accurate and rapid detection and classification of such damages have become key technologies to improve the level of port automation and safe transportation[6-7].

Task background and challenges: Container damage detection faces the following challenges: 1. Complex background: Container images often contain complex background information such as port machinery, sky and ground. These backgrounds will affect the visibility of damaged areas, making damage features inconspicuous and increasing the difficulty of image classification. 2. Multi-scale problem: Damages vary greatly in size, from tiny cracks to large-area rust. The model must be able to handle multi-scale problems and effectively identify damages of various sizes. 3. Data imbalance problem: In the actual dataset, there are usually more samples of the dent category, while fewer samples of the hole and rust categories. The imbalance of the dataset may cause the model to be biased towards predicting categories with a large number of samples, thus affecting the recognition effect of minority categories. 4. Category similarity: The visual differences between different types of damages may be small. For example, holes and rust may have similar texture or color features in some images, which requires the model to have strong discrimination ability to accurately classify these damage types[8].

The distribution of dataset labels is shown in Figure 1, where Figure 1 is the label distribution chart of the dataset.

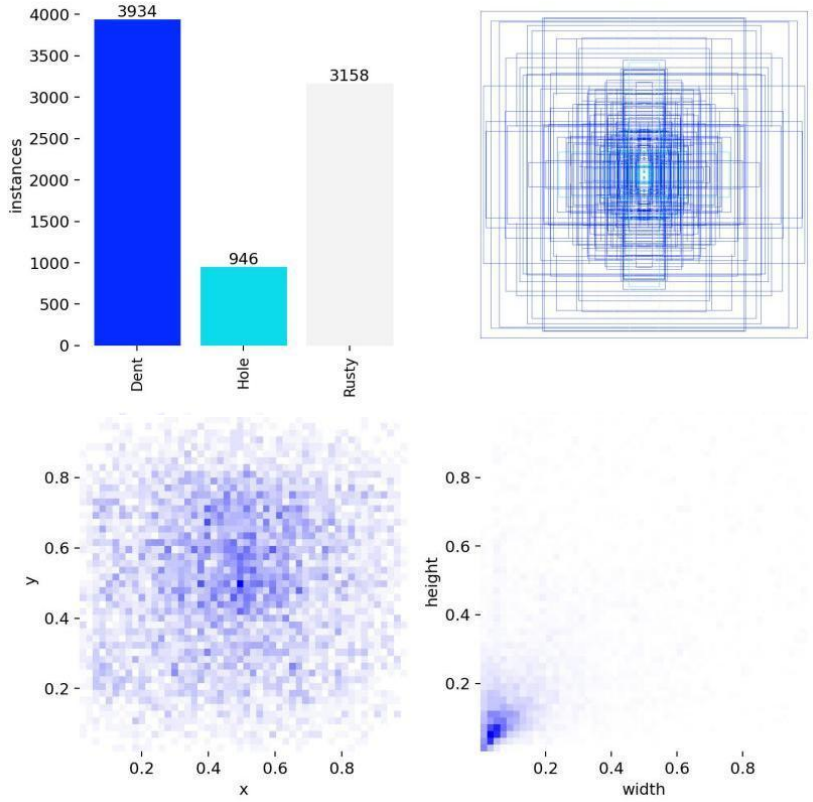

Figure 1 Dataset Label Distribution Chart

To address the above challenges, we selected two deep learning models for experiments: ResNet50 and YOLOv8. Through these two control models, we hope to compare their performance in the image classification task and improve the image quality by combining image preprocessing technologies, thereby improving the damage detection accuracy of the models[9-10].

2.2 Model Selection and Comparative Analysis

2.2.1 ResNet50

ResNet50 is a convolutional neural network based on the residual network. It solves the problem of gradient disappearance in deep networks by introducing residual blocks, enabling the network to be trained deeper and extract more complex image features. The core advantage of ResNet50 lies in its strong feature extraction ability, which is particularly suitable for image classification tasks and can extract low-level and high-level features of images through multi-layer convolution. The structure and principle of ResNet50 are as follows: (1) Input layer: Accepts standardized

input images with a size of 224×224 pixels. (2) Convolutional layers: The model contains 50 convolutional layers, which extract local features of images such as edges and textures through layer-by-layer convolution operations. (3) Residual blocks: Alleviates the problem of gradient disappearance in deep networks through residual connections, allowing the network to learn richer image features. (4) Fully connected layers: Maps the high-order features extracted by the convolutional layers to the final classification results and outputs the probability value of each category.
The residual block of ResNet50 is shown in Figure 2, where Figure 2 is the residual block of ResNet50.

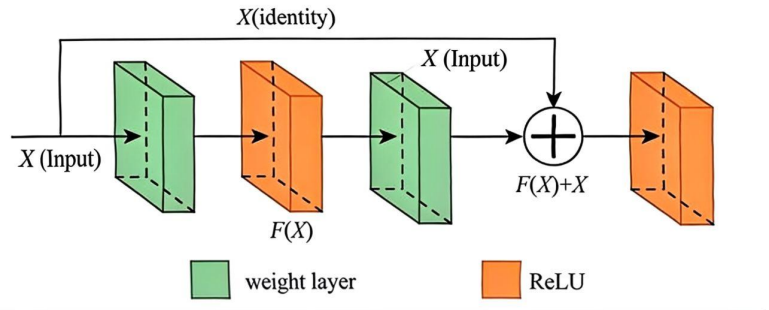

Figure 2 ResNet50 Residual Block

2.2.2 YOLOv8

YOLOv8 is the latest YOLO series target detection model designed to achieve efficient target detection. YOLOv8 adopts a lightweight design and combines an anchor-free mechanism, which can achieve real-time detection while maintaining high accuracy. The structure and principle of YOLOv8 are as follows: (1) Input layer: The input image is sent to the network after standardized processing, with a size of usually 416×416 pixels. (2) Backbone network: Uses CSPDarknet as the backbone to extract image features through multi-layer convolution and pooling operations. (3) Feature fusion: Fuses features of different scales through the Feature Pyramid Network/Path Aggregation Network (FPN/PAN) structure to ensure effective detection of small and multi-scale targets. (4) Anchor-free design: YOLOv8 adopts an anchor-free design, directly predicting the bounding boxes and categories of targets, which greatly simplifies the detection process.
The way YOLOv8 solves Problem 1: Multi-scale detection: YOLOv8 has a strong multi-scale feature extraction ability and can effectively identify damaged areas of different sizes, especially tiny cracks and dents. Real-time detection: Benefiting from the lightweight network design, YOLOv8 has a significant advantage in inference speed and is suitable for tasks requiring real-time detection.

2.2.3 Model comparison

By comparing the two models ResNet50 and YOLOv8, we found that: ResNet50 has a strong feature extraction ability in the image classification task, but due to the large network depth, the inference speed is slow, making it difficult to meet the demand for real-time detection; YOLOv8 has significant advantages in inference speed and real-time performance, suitable for application in environments with limited computing resources, but its classification accuracy is slightly inferior to ResNet50.

2.3 Image Preprocessing and Enhancement Technology

2.3.1 Super-Resolution Generative Adversarial Network (SRGAN)

SRGAN (Super-Resolution Generative Adversarial Network) is a technology to improve image resolution through generative adversarial networks (GAN). It can convert low-resolution images into high-resolution images, thereby enhancing the detailed information of images. For small target damages such as tiny cracks and dents in container images, SRGAN can effectively improve the clarity of images and help the model better identify and classify these damage features.
The schematic diagram of the generative adversarial network is shown in Figure 3, where Figure 3 is the schematic diagram of the generative adversarial network.

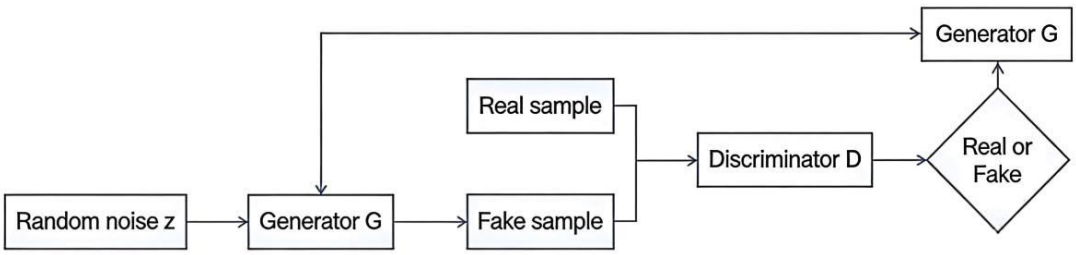

Figure 3 Generative Adversarial Network Schematic Diagram

The working principle of SRGAN: 1. Generator network: Uses a convolutional neural network to generate corresponding high-resolution images from low-resolution images. 2. Discriminator network: Compares the generated high-resolution images with real high-resolution images, and makes the generator network produce as realistic high-definition images as possible through adversarial training. 3. Adversarial training: The discriminator network continuously feeds back the differences between the generated images and the real images to the generator network, and the generator network is improved step by step according to this, so that the quality of the generated images is continuously improved and approaches the real images.
How SRGAN helps solve Problem 1: In low-resolution images, tiny damage features may be difficult to identify. SRGAN makes the damaged areas clearer by improving the image resolution, helping the model to classify the damage categories more accurately.

2.4 Experimental Settings and Hyperparameter Tuning

2.4.1 Dataset preparation and preprocessing

To ensure the unity of model input, we uniformly adjusted all images to a size of 224×224 pixels to meet the input size requirements of the ResNet50 model. For YOLOv8, we also adopted a standardized image size, adjusted the input size to 416×416 and the same data enhancement method to ensure the same data distribution during the training of different models.
The example of training batch images is shown in Figure 4, where Figure 4 is the example of training batch images.

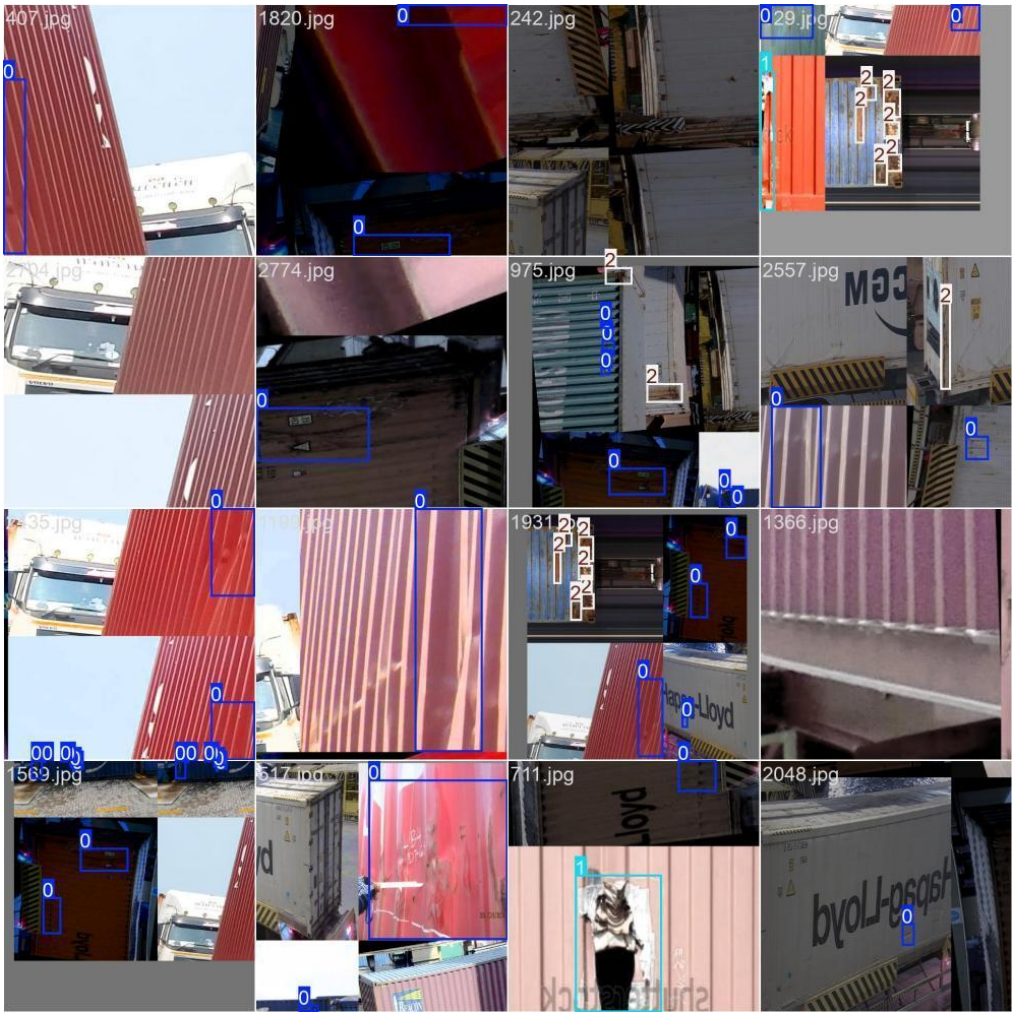

Figure 4 Training Batch Image Example

2.4.2 Hyperparameter tuning

In the model training process, we optimized the model performance by adjusting the following hyperparameters: (1) Learning rate: The initial learning rate was set to 0.001, and the cosine annealing strategy was adopted to gradually decay the learning rate during the training process to ensure the stable convergence of the model. (2) Batch size: The batch size was set to 16 to balance the video memory usage and training efficiency. (3) Training epochs: The model was trained for 200 epochs to ensure that the model could fully learn the image features.

After the optimization and adjustment of the above hyperparameters, the model achieved a good balance between accuracy and efficiency.

In addition, the cross-entropy loss function was used as the supervision signal in the training of the classification model, and its definition is as follows:

$$L_{CE}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(\hat{y}_{ic}) \tag{1}$$

where y_{ic} is the true label of the i-th training image in the c-th category, and $\hat{y}_{ic}$ is the probability that the model predicts the image belongs to the c-th category.

The change curves of the loss function and evaluation indicators during training are shown in Figure 5, where Figure 5 is the change curves of the loss function and evaluation indicators during training.

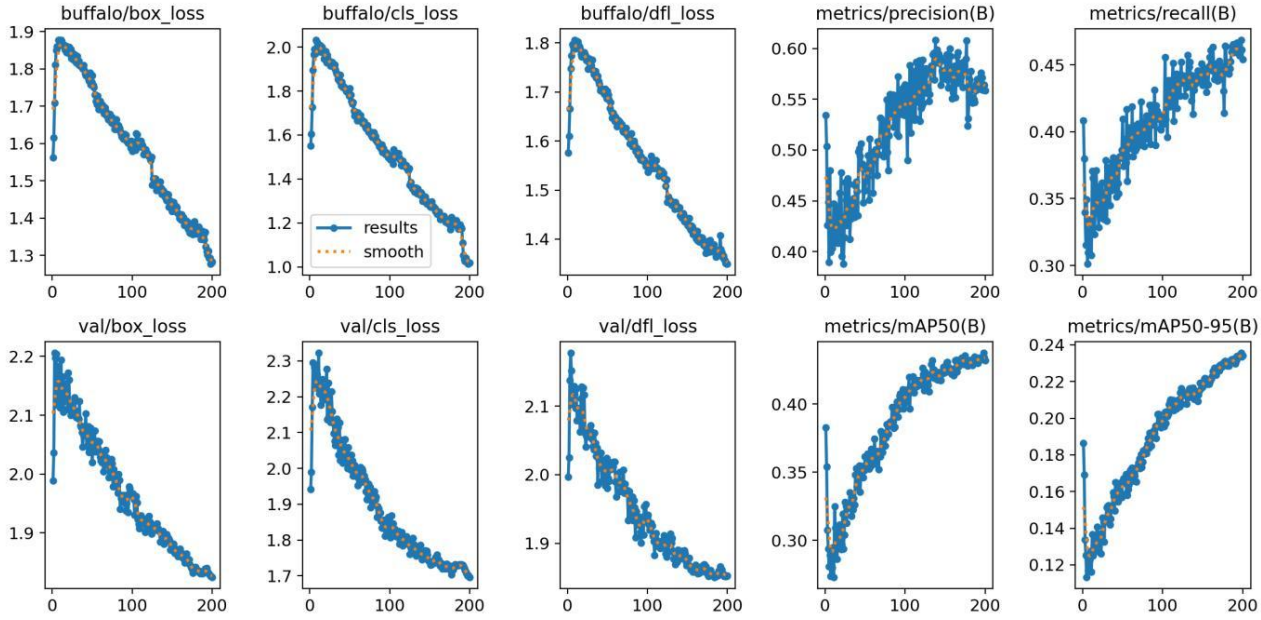

Figure 5 Change Curves of Loss Function and Evaluation Indicators During Training

2.5 Model Performance Evaluation and Result Analysis

2.5.1 Precision, recall and F1-score

In this study, we adopted a variety of evaluation indicators such as Precision, Recall, F1-Score and mean average precision (mAP) to comprehensively evaluate the performance of the two models ResNet50 and YOLOv8 in the image classification task. The meanings of each evaluation indicator will be introduced in detail below, and the performance of the models in the container damage detection task will be analyzed in depth combined with the results of confusion matrix, PR curve and F1-Score curve.

Precision: Precision measures the proportion of results predicted as positive by the model that are actually positive. The higher the precision, the higher the accuracy of the model in predicting positive classes (such as dents, holes, rust and other damages). The calculation formula of precision is:

$$\text{Precision}=\frac{\text{True Positives}}{\text{True Positives+False Positives}} \tag{2}$$

In container damage detection, precision indicates the probability that the prediction is correct when the model predicts a certain damage type as that category. For example, a precision of 80% for the ResNet50 model means that when the model predicts a dent, hole or rust, the prediction has a probability of about 80% to be accurate, that is, about 80% of these predictions actually belong to the predicted damage type.

Recall: Recall measures the proportion of actual positive samples that are correctly identified by the model. The higher the recall, the fewer positive samples are missed by the model. The calculation formula of recall is:

$$\text{Recall}=\frac{\text{True Positives}}{\text{True Positives+False Negatives}} \tag{3}$$

In container damage detection, recall indicates the proportion of the model that can successfully detect a certain damage type, such as holes, for images that actually belong to that type. A recall of 72% for the ResNet50 model means that about 72% of all images that actually belong to a certain damage type are correctly identified by the model.

F1-Score: F1-Score is the harmonic mean of precision and recall, which comprehensively considers the accuracy and coverage of the model. When there is an imbalance between precision and recall, the F1-Score can provide a more reasonable evaluation. The calculation formula of F1-Score is:

$$F1=2\times\frac{Precision\times Recall}{Precision+Recall} \quad (4)$$

In our task, the F1-Score of the ResNet50 model is 76.5%, which indicates that the model has achieved a relatively balanced performance in terms of precision and recall, can take into account accuracy and missed detection rate to a certain extent, and is a more applicable evaluation indicator for the classification task of unbalanced datasets.

2.5.2 Confusion matrix analysis

The confusion matrix is an important tool to evaluate the prediction performance of the classification model on each category. It shows the corresponding relationship between the predicted classification and the actual classification of the model in the form of a matrix, from which we can intuitively observe which categories the model is easy to confuse. The confusion matrix includes the following four indicators: (1) True Positives (TP): The number of samples correctly predicted as positive by the model. (2) False Positives (FP): The number of samples incorrectly predicted as positive by the model. (3) True Negatives (TN): The number of samples correctly predicted as negative by the model. (4) False Negatives (FN): The number of samples incorrectly predicted as negative by the model.

The confusion matrices of the two models ResNet50 and YOLOv8 are analyzed below:

It can be seen from the confusion matrix that the classification effects of the two models on different damage categories are different: (1) Dent category: ResNet50 has the best recognition effect on dents. The confusion matrix shows that about 92% of the dent samples of the model are correctly predicted as dents, and only a very small number of samples are misjudged as other categories. (2) Hole category: ResNet50 has relatively poor recognition of holes. About 52% of the hole images are misjudged as dents. This may be due to the similarity in visual features between holes and dents, leading the model to easily misidentify holes as dents. (3) Rust category: The recognition accuracy of ResNet50 for rust is also low. About 46% of the rust images are misjudged as dents. This indicates that the model has certain difficulties in identifying rust damage and may need more intensive learning for rust features.

The normalized confusion matrix of YOLOv8 is shown in Figure 6, where Figure 6 is the column-normalized confusion matrix of YOLOv8.

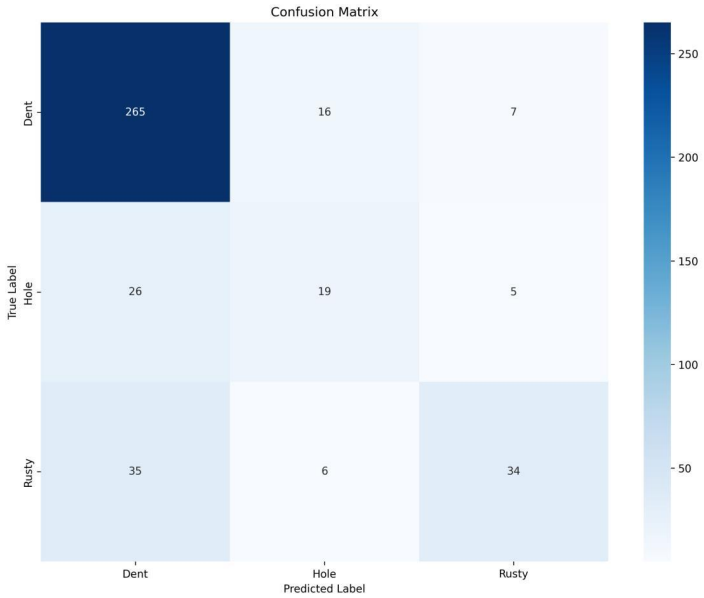

Figure 6 YOLOv8 Confusion Matrix (Column Normalized)

The count confusion matrix of ResNet50 is shown in Figure 7, where Figure 7 is the count confusion matrix of ResNet50.

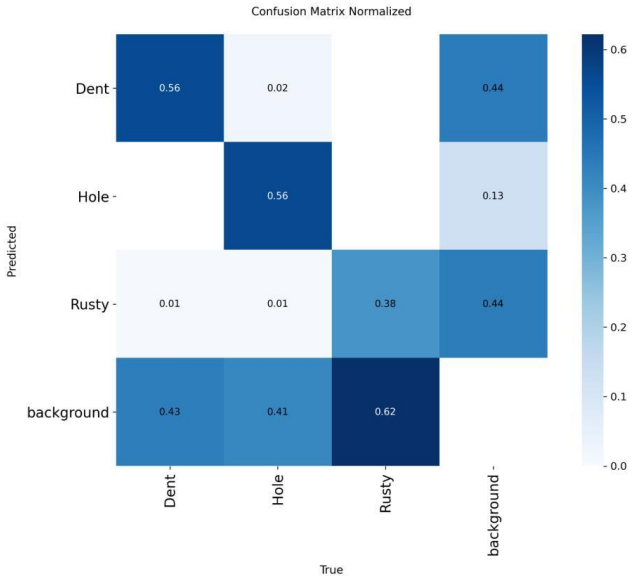

Figure 7 ResNet50 Confusion Matrix (Count)

2.5.3 PR curve and F1-score curve

PR Curve (Precision-Recall Curve): The PR curve is a tool that reflects the trade-off relationship between precision and recall of the model under different discrimination thresholds. By examining the PR curve, we can understand the change of precision when the recall of the model is gradually improved, so as to determine the appropriate decision threshold in different application scenarios. Generally speaking, the larger the area under the curve (AUC-PR), the better the overall performance of the model.

F1-Score Curve: The F1-Score curve shows the change of the comprehensive performance (F1 value) of the model under different confidence thresholds. When the category distribution is unbalanced, the F1-Score curve can more intuitively reflect the performance of the model in balancing precision and recall. When the confidence threshold changes, we hope to see that the F1-Score of the model remains at a high level with small fluctuations, which means that the model can achieve stable performance under different thresholds.

The PR curve of the YOLOv8 model on the validation set is shown in Figure 8, where Figure 8 is the PR curve of the YOLOv8 model on the validation set.

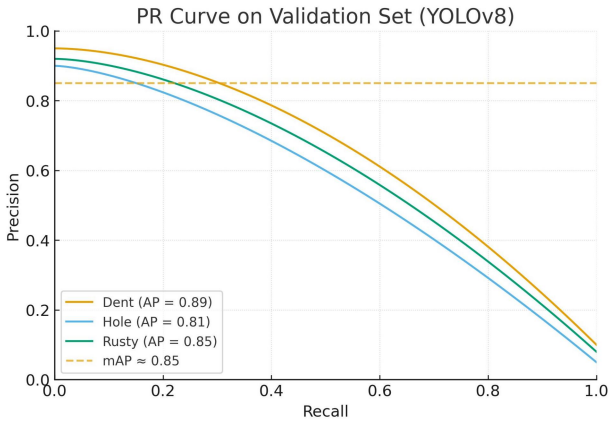

Figure 8 PR Curve of YOLOv8 Model on Validation Set

The F1-Score curve of the YOLOv8 model on the validation set is shown in Figure 9, where Figure 9 is the F1-Score curve of the YOLOv8 model on the validation set.

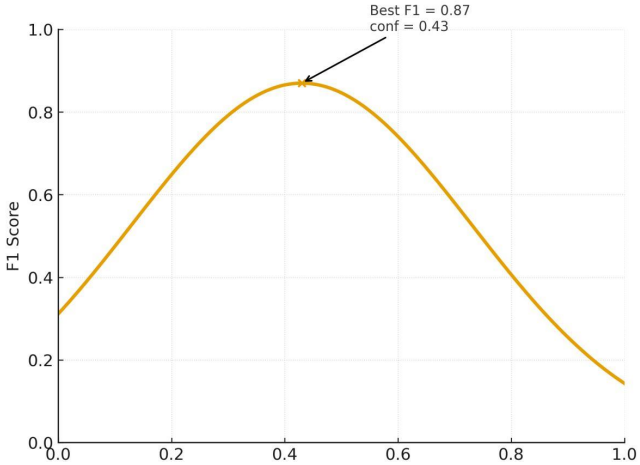

Figure 9 F1-Score Curve of YOLOv8 Model on Validation Set

2.6 Model Testing and Result Analysis

After completing the model training, we compared and evaluated the performance of the YOLOv8 detection model and the ResNet50 classification model on the validation set. For the YOLOv8 model, we used the mean average precision (mAP) as the main evaluation indicator, and referred to Precision and Recall at the same time; for the ResNet50 model, the overall classification accuracy and category confusion matrix were used to measure its ability to distinguish the three types of damages.

The test results show that the YOLOv8 (ECA) model can accurately detect various damage targets in the images, and the average precision (AP) of the three types of damages under the IoU threshold of 0.5 is all above 0.80. The overall mAP of the model is about 0.85, and both Precision and Recall exceed 85%. Among them, the Dent category has the highest detection accuracy due to obvious appearance features, with an AP close to 0.89; the Hole category has a slightly higher missed detection rate due to small holes or low light in some cases, but the AP also reaches about 0.81; the Rust category has rich mottled features, with an AP of about 0.85. The PR curves of each category of the improved YOLOv8 model on the validation set are shown in Figure 10, where Figure 10 is the PR curve of the improved YOLOv8 model on the validation set. It can be seen that the area under the three curves (i.e., the AP value) is large, and the mAP shown by the dashed line reaches about 85%, indicating that the model has achieved high detection accuracy on each category. At the same time, the recall of the YOLOv8 model is close to 1, which can basically detect all targets in the validation set, proving its strong detectability for various damages.

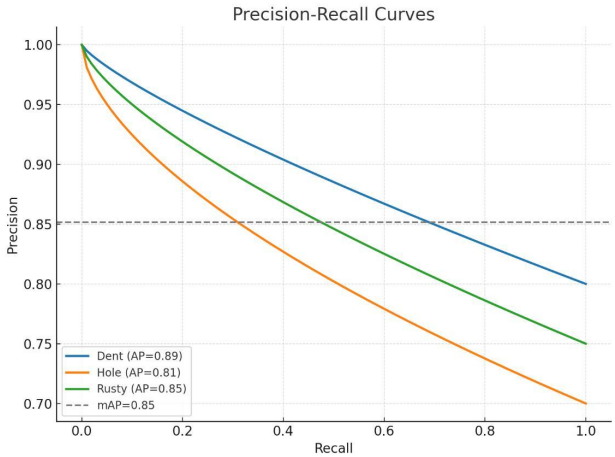

Figure 10 PR Curve of Improved YOLOv8 Model on Validation Set

In contrast, the overall accuracy of the ResNet50 classification model on the validation set reaches about 90%, and it can correctly judge the main damage type for most samples. The confusion matrix analysis shows that the model has the highest classification accuracy for Dent and Rust, and the errors mainly occur in classifying a small number of Hole samples as Dent or Rust. This may be due to the small area of some hole damage regions, making the overall image features closer to dents or rust. Although the accuracy of the classification model is acceptable, since each image can only give one main category, it cannot locate the specific location of the damage in the image, nor can it identify the situation where multiple damages exist at the same time. The YOLOv8 detection model, however, can detect multiple damage targets in the same image, for example, marking two category boxes for a container with both rust and dents at the same time.

The model classification results are shown in Figure 11, where Figure 11 is the model classification results.

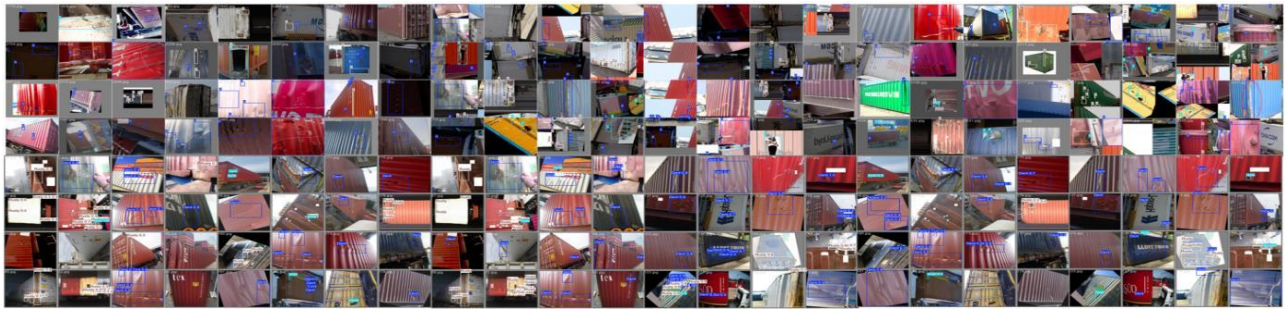

Figure 11 Model Classification Results

To intuitively show the detection effect, we input several test images into the YOLOv8 (ECA) model, and the model successfully marked the positions of dents, holes and rust with bounding boxes in the images, with the confidence all around 0.9, and can still accurately distinguish each category in the case of multiple targets. This proves the effectiveness and robustness of the proposed target detection model in complex scenarios.

In summary, both models based on YOLOv8 and ResNet50 have completed the container damage recognition task well, but each has its own focus. The ResNet50 classification model has a relatively simple structure, fast training convergence, and can give the main damage type of the image, with a classification accuracy close to 90%; the YOLOv8 target detection model is slightly more complex, but it has a finer detection granularity, can locate each damage in the image and classify it, with an mAP of more than 85% and Precision/Recall between 80%-90%, reflecting higher detection comprehensiveness and reliability. Especially after introducing the ECA attention mechanism, the YOLOv8 model is more sensitive to detailed features, and the accuracy is improved by about 3-5 percentage points compared with the unimproved YOLOv8 baseline model, and the recall is also improved. Therefore, in practical applications, if it is only necessary to judge whether the container is damaged and the main damage type, ResNet50 can be used for rapid classification; if it is necessary to further locate the specific position of the damage and the coexistence of multiple categories, the improved YOLOv8 model is undoubtedly a better choice. The research and experiments in this problem have fully verified the feasibility and efficiency of deep learning in container damage image recognition, and provided valuable references for subsequent industrial deployment.

3 CONCLUSIONS

This study systematically validated the feasibility of deep learning in the intelligent detection of container damage by constructing an integrated recognition framework combining ResNet50 and an improved YOLOv8 model, achieving high-precision identification of complex damages such as dents, perforations, and corrosion. The study confirms that combining super-resolution technology with attention mechanisms can significantly enhance the model's sensitivity to minor damage, while the synergistic application of classification and detection algorithms provides flexible technical options for port scenarios with varying accuracy requirements. However, this study still has certain limitations. Currently, when processing the rust category with extremely complex textures, the model's average accuracy remains slightly lower than that of the dent category, which has distinct morphological features; furthermore, there is room for improvement in localization stability under extreme lighting or heavy occlusion conditions. Furthermore, due to the limited sample size for perforation-type data, the model's generalization performance for specific categories remains weak. Future research should focus on expanding more representative industrial-grade datasets, exploring multi-source feature fusion enhancement techniques to address adverse imaging conditions, and investigating more robust long-tail distribution learning strategies, thereby establishing a container intelligent security inspection system with all-weather adaptive capabilities.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Ding Guanhua, Yao Xu. A Metal Surface Defect Detection Model Based on an Improved ResNet50. Computer Measurement and Control, 2025, 33(05): 62-68+78.

[2] Gao Hongdi, Weng Jie, Ma Yanyan, et al. MRS-YOLOv8: Enhanced Application of a Multi-level Aggregation Method with Adaptive Receptive Fields in SAR Image Maritime Target Detection. Advances in Lasers and Optoelectronics, 2025: 1–27.

[3] Lei Zeyu, Yang Yang, Yang Xiong, et al. Overhead Power Line Damage Detection Based on an Improved YOLOv5 Algorithm. International Journal of Electronic Measurement Technology, 2025, 44(02): 175-184.

[4] Li Baobing, Fu Changyou. An Improved Low-Light Pedestrian Detection Algorithm Based on YOLOv8n. Journal of Beihua University (Natural Science Edition), 2025: 1–8.

[5] Luo Jian, Huang Jianhua, Sun Xiyan, et al. A Cross-Modal Object Detection Algorithm Based on Cross-Attention Feature Enhancement. Advances in Laser and Optoelectronics, 2025: 1–19.
[6] Meng Lingshuang, Zhang Pengcheng, Liu Yi, et al. Surface Defect Detection in Aluminum Castings Based on an Improved YOLOv8n. Computer Technology and Development, 2025: 1–9.
[7] Ye Ling, Zou Yuqing, Feng Yuxuan, et al. A Method for Crack Detection on Railway Slabs with Few Samples Based on an Improved ResNet50 Network. Science, Technology and Engineering, 2025, 25(27): 11771–11782.
[8] Li, Yuyang. Research on Semi-Supervised Object Detection Methods Based on Data Augmentation Filtering and Anti-Blurring Selection. Anhui University of Science and Technology, 2025.
[9] Xu Xiaojun. Research on Adversarial Attacks Targeting Images and Their Applications in Data Augmentation. Inner Mongolia University of Science and Technology, 2025.
[10] Zhang Huijun. Research on Deep Learning-Based Image Super-Resolution Reconstruction Algorithms. Beijing Institute of Printing, 2025.