Surface Defect Detection Algorithm for Strip Steel Based on Improved YOLOv7 Model

Zhu Wang, Weisheng Liu

***Abstract*—This research proposes a refined deep learning framework aimed at boosting the precision and efficacy of detecting surface imperfections in strip steel. This method integrates enhancement and simplification techniques inspired by the You Only Look Once version 7 (YOLOv7) detection method, resulting in significant enhancements in the model's accuracy, speed, and flexibility. The substitution of ELAN with Bottleneck Transformer 3 (BoT3) leads to improved accuracy and mean Average Precision (mAP) values, while also introducing a more lightweight network architecture. The incorporation of the Involution mechanism enhances the model's feature extraction capabilities, thereby improving its ability to recognize small targets through the utilization of local perceptual fields. The ASPP_CA architecture leverages a multi-scale feature fusion technique along with an attention mechanism to reduce model parameters and enhance inference speed. Furthermore, it extends the model's receptive field, allowing it to capture additional visual information. The enhanced algorithm, denoted as YOLOv7-IBA, demonstrates empirical results that underscore its superiority over the three current state-of-the-art detection techniques in identifying surface flaws on strip steel. The accuracy has been improved to 82.9%, representing a significant increase of 7.2% compared to the previous performance. Furthermore, the mean mAP value has experienced a 3.2% increase, reaching a total of 79.9%. Moreover, there has been a remarkable 8.8% improvement in efficiency. The adoption of this approach holds the potential to enhance both the precision and productivity of strip surface flaw detection, while also providing valuable methodological support for the advancement of other related disciplines.**

***Index Terms*—ASPP_CA, BoT3, Involution, Object Detection, YOLOv7**

I. Introduction

With the rapid development of industrial automation, strip steel is widely utilized in industry, construction, transportation, chemical industry, and other fields. However, surface defects such as cracks, bubbles, scratches, rust, and oxidation can gradually worsen, posing a significant safety risk. Some defects, such as tiny pits and micro-cracks, are challenging to detect and may go unnoticed. These defects have the potential to worsen over time, leading to safety accidents. These defects also impact the corrosion and wear resistance of the final product. Therefore, there has been a continuous pursuit of advanced detection technologies for surface defects in steel strips.

The manual method of defect detection can be both inaccurate and inefficient; thus, it is essential to develop algorithms that can automatically detect surface defects on production lines in real-time [1].

Studies on the detection of surface defects in strip steel can be categorized into traditional methods and deep learning-based methods. The traditional method has drawbacks, such as low efficiency, high error rates, and demanding skill requirements for inspectors [2]. In contrast, deep learning-based target detection methods can rapidly and accurately identify surface defects without requiring advanced skills. He et al. [3] proposed a novel defect detection system based on deep learning to identify steel plate defects. The system can capture specific categories and detailed information to identify steel defects by integrating multiple levels of features. Tasi et al [4] proposed a fast regularity metric for defect detection on non-textured and uniformly textured surfaces. These two methods were used to detect defects by only a single discriminative feature. It avoided the complexity of using classifiers in high-dimensional feature spaces. On the other hand, the method did not require learning from a set of defective and non-defective training samples. Guo et al. [5] proposed a transformer-based approach for detecting defects on steel surfaces. The researchers used a transformer in conjunction with global contextual information to enhance functionality and improve the detection of faulty targets. A method for calculating the surface corrosion area of steel bridges was proposed by Son and his colleagues [6]. Tong et al. [7] proposed a defect detection model using an optimal Gabor filter. The model can be significantly less computationally intensive and effectively address fabric detection problems by utilizing optimal Gabor filters. Compared to the method above, the "You Only Look Once" (YOLO) series represents the latest technology in single-stage object detection. Compared to two-stage detectors such as the region-based CNN (R-CNN) family, single-stage detectors integrate region and object classification in a simple architecture to achieve faster inference speeds [8]. Such algorithms have achieved significant success in the past decade [9]. In 2015, Redmon et al. [10] proposed YOLOv1, a pioneering single-stage object detection algorithm. The follow-up to YOLOv3 [11] introduced the residual module and FPN [12] architecture. While it improved the speed and accuracy of target detection,

Manuscript received June 8, 2023; revised January 12, 2024.

This work was supported by the Special Fund for Scientific Research Construction of University of Science and Technology Liaoning.

Zhu Wang is a postgraduate student at the College of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China. (e-mail: 13050011069@163.com).

Weisheng Liu is a professor in the College of Computer Science and Software Engineering at the University of Science and Technology Liaoning, Anshan, CO 114051, China (corresponding author to provide fax: 0412-5929809; e-mail: succman@163.com).

it was relatively ineffective at detecting small targets. YOLOv5 [13] has optimized and enhanced the network structure, feature pyramid, and training strategy to improve detection speed and accuracy further. However, a more precise defect classification capability is needed to detect surface defects in strip steel.

YOLOv7, an evolution of the object detection algorithm previously referred to as YOLOv5, introduces a series of substantial advancements. Compared to YOLOv5, YOLOv7 features an improved model structure, anchor frame modeling, data augmentation, and model optimization. This leads to enhanced detection accuracy and faster detection speeds for the target detection task. Furthermore, due to its compatibility with various defect types, the YOLOv7 model stands out in terms of deployment convenience. However, the adopted anchor-based method needs to improve in detecting smaller targets. In contrast, the global detection method must be more effective for detecting defects distributed locally or at the edges of the strip. Therefore, there is an urgent need for an effective method to improve the detection capability of small targets and efficiently identify multiple categories of targets when detecting surface defects on strip steel. To address this issue, we propose YOLOv7-IBA, an enhanced approach based on deep learning networks, to improve the accuracy and efficiency of detecting surface defects in strip steel. Among them, the BoT3 module is used instead of the ELAN module. Using a lightweight network structure, this substitution improves the model's accuracy rate and mAP value. The involution mechanism enhances the model's feature extraction and improves its ability to recognize small targets by expanding the local perceptual field. Additionally, ASPP_CA is designed to reduce the model parameters and improve the speed of inference by integrating multi-scale feature fusion and an attention mechanism. Meanwhile, the sensory field of the model has been expanded, allowing it to capture more image details. The experimental results demonstrate that this method outperforms three existing state-of-the-art detection methods in identifying defects on strip steel surfaces. Specifically, it achieves an accuracy of 82.9% (an increase of 7.2%) and a mAP value of 79.9% (an increase of 3.2%). Furthermore, the efficiency has improved by 8.8%. The data shows that YOLOv7-IBA is a practical and feasible method for detecting steel-strip surface defects. It can also serve as a valuable reference for advancing other related fields.

II. Materials and Methods

A. *ASPP_CA*

During strip steel inspection, the ASPP_CA technique is effectively employed to enhance inspection precision. Since the widths of the strips vary, it is necessary to detect strips of different sizes. By employing the ASPP_CA technique, detection accuracy can be improved by utilizing multi-channel feature maps at various scales. This enables the network to detect strips of different sizes and positions. By integrating ASPP [14] and CA [15], model performance in complex scenes can be improved for more effective target detection. Specifically, ASPP is utilized for multi-scale feature extraction, extraction to detect objects at various

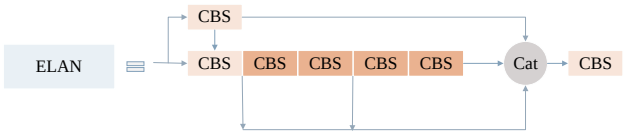

Fig.1. ELAN Structure Diagram

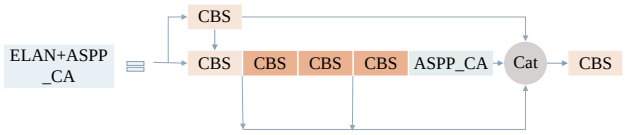

Fig.2. ELAN Add ASPP_ CA structure diagram

locations and sizes, while CA employs an attention mechanism to minimize redundant information and reduce misclassification. This can improve the accuracy of target detection and effectively reduce the computational burden of the model.

Furthermore, the ASPP_CA method not only enhances the robustness and generalization of the network but also improves the correlation among the feature map channels. Our research focuses on detecting surface defects on strips using the ASPP_CA module. By sampling the input feature maps and applying average pooling to the results, we can enhance the ability to detect small and dense defects in complex backgrounds. Meanwhile, in the second branch of the ELAN layer, the last 3×3 convolution module is replaced by ASPP_CA to improve the detection accuracy. The ELAN layer structure consists of two branches. The first branch is 1×1 CBS, which adjusts the number of channels. The second branch consists of multiple CBSs, which first undergo a 1×1 convolution, followed by four 3×3 convolutions for feature extraction. Finally, the outputs of both branches are combined using a concatenation operation, as illustrated in Figure 4 below. Given that the strip defects dataset contains numerous small and dense defects that can easily be overlooked, we replaced the last 3×3 convolution module in the second branch with ASPP_CA. ASPP_CA can conduct multi-scale processing on the input features and achieve channel-attention weighting, effectively reducing redundancy and noise. This improves the performance of target detection or classification, especially in identifying small and dense defects against dense backgrounds. This is shown in Figure 5 below.

ASPP_CA is a channel attention module constructed in this paper. The ASPP module and the channel attention mechanism are combined. During the fusion process, the ASPP module generates multiple feature maps. Meanwhile, the channel attention mechanism helps the model better learn key information from these feature maps and assign weights to obtain more accurate information. The ASPP module contains multiple parallel branches, each with a different dilation rate, to capture a broader range of features while preserving pixel-level information. The CAModule, on the other hand, utilizes the output of the ASPP branch to prioritize channel-specific information, thereby further enhancing network performance. The ASPP_CA module, which includes ASPP and CAModule, effectively extracts multi-scale features and applies channel-specific information weighting. This enables the network to adapt more effectively to various input data. Figure 6 illustrates the structure of the ASPP_CA module, which comprises multiple ASPP branches

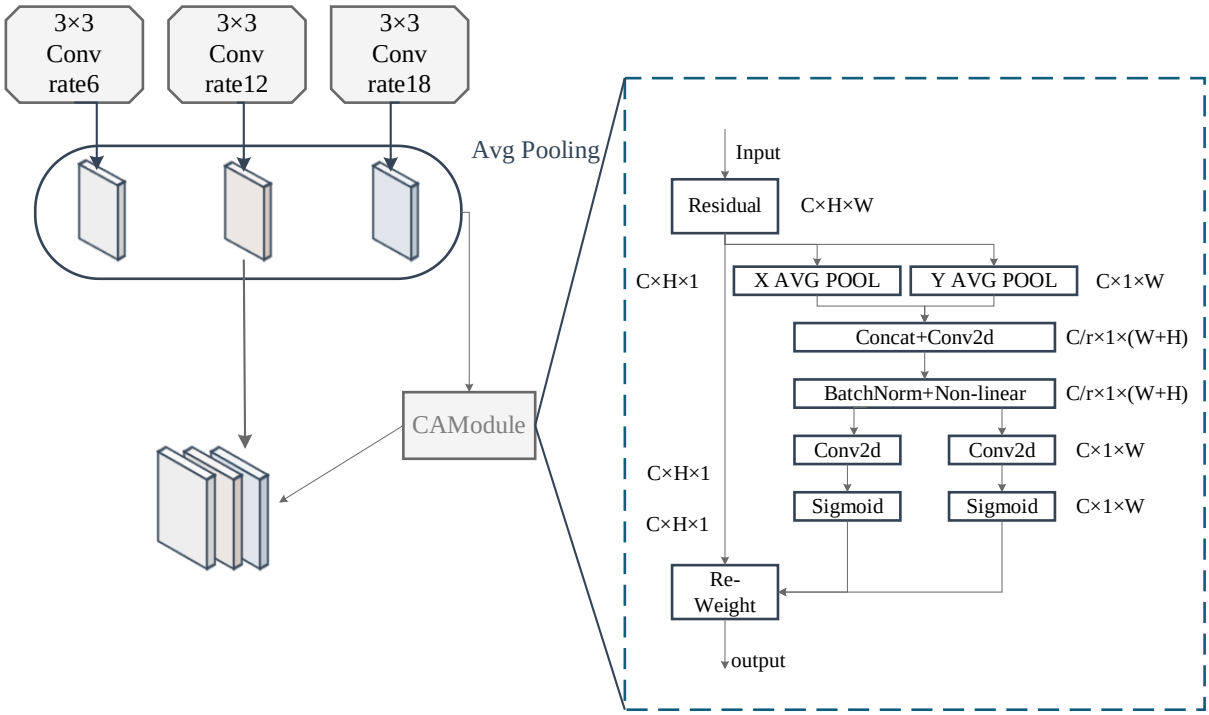

Fig.3. ASPP_ Structure diagram of CA module

and the CAModule. The ASPP branches utilize multiple convolutional kernels with varying expansion rates to extract multi-scale features, which are then combined. The CAModule module utilizes the stack of outputs from the ASPP branches to weight the information of each channel, thus obtaining the channel-specific information weighting to enhance the network's performance further. To sum up, the ASPP_CA module serves as an effective feature extraction tool with diverse applications across numerous tasks. In conclusion, the ASPP_CA module is a powerful feature extraction module with various applications in various tasks. Through multi-scale feature extraction and channel-specific information weighting, it can effectively enhance the performance of neural networks, especially when dealing with datasets of diverse and complex nature.

B. BoT3

To tackle the challenges of significant dimensional variations and complex surface textures in the strip steel defects dataset, we propose optimization techniques for BoT3. These techniques aim to enhance feature extraction and more comprehensively understand semantic data more efficiently. This successfully tackled the challenge of distinguishing between targets and backgrounds within the model while mitigating concerns related to false positives and missed detections. BoT3 streamlines both the encoder and decoder components of the model. The model's maintainability and scalability are enhanced by incorporating adaptive techniques and modular functions. The BoT3 module we employ predominantly comprises Convolutional blocks and Bottleneck Transformers (BoT), as depicted in Figure 4.

BoT stands for Bottleneck Transformer [16], a neural network module employed for image classification and object detection tasks. It represents an improved and optimized version of the Transformer network architecture.

The BoT comprises three essential components: Expansion, Multi-Head Self-Attention (MHSA), and Contraction, as depicted in Figure 5. Multi-Head Self-Attention is a pivotal component of the BoT model, characterized by an enhanced channel attention mechanism derived from the self-attention mechanism. The primary formula can be expressed as follows:

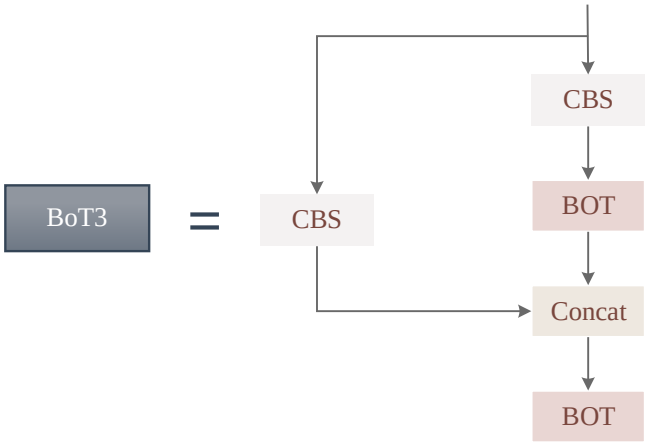

Fig.4. Structural diagram of BoT3

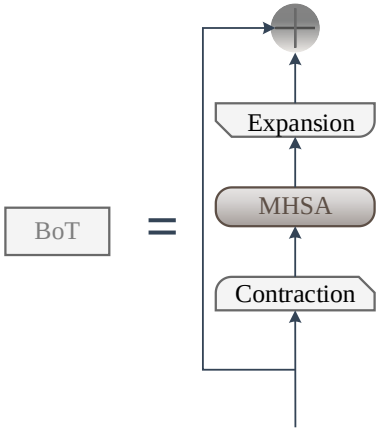

Fig.5. BoT structure diagram

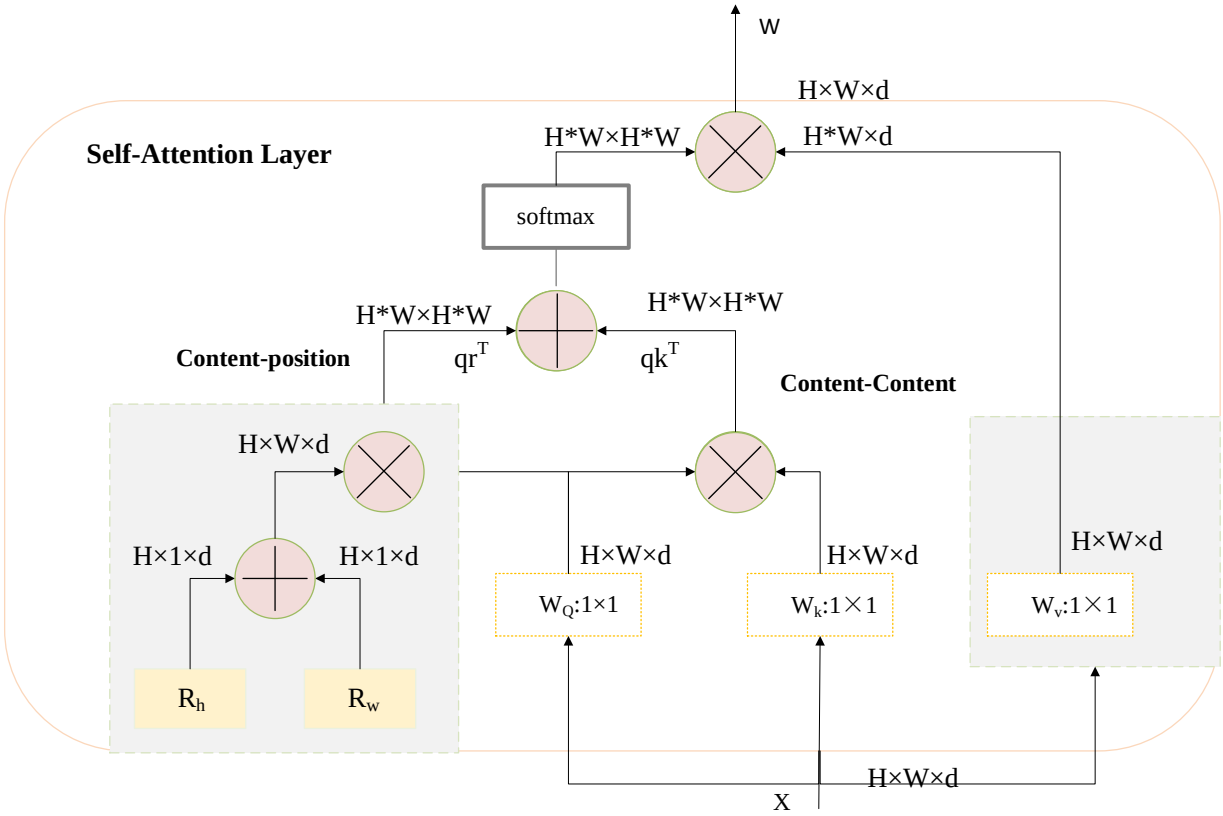

Fig.6. Structure of MHSA

$$Attention(Q,\ K,\ V) = softmax(\frac{QK^T}{\sqrt{d}})v \quad (1)$$

$$Eij = \frac{\exp(X i, j)}{\sqrt{d_k}} \quad (2)$$

For input feature maps $X \in ¡^{H\times W\times d}$, wherein H, W and d represents the respective dimensions of the individual tokens in the input feature matrix, including their height and width, The number of tokens is $H \times W$. The attention logic of the MHSA layer is $qr^T + qk^T$, wherein q, k and r represents the query, key, and location encoding, respectively. $\oplus$ and $\otimes$ represents the summation of elements and the multiplication of matrices, respectively. By computing the similarity between q and every token, the attention weight is ultimately multiplied and added to the input value to yield the output result. Notably, the gray boxes, which encompass position encoding and value projection, are the only three elements that are absent in non-local layers [17]-[18]. In the BoT model, position encoding and value projection are exclusively used within the MHSA layer.

C. *Involution*

Due to the variations in size and shape among strip surface defects, the utilization of traditional fixed convolutional kernels encounters challenges in efficiently extracting crucial spatial features from these defects. Conversely, the Involution network employs trainable convolutional kernels that can adapt their size and shape to accommodate defects of varying sizes and shapes. This significantly enhances the accuracy and robustness of defect detection. When detecting surface defects on strip steel, the involution excels at extracting spatial features from the affected areas. Moreover, it can dynamically adjust the size and shape of convolutional kernels, resulting in more precise object positioning and detection. Despite the rapid progress of deep neural networks, conventional convolutional operations, with their ability to densely pool local features from a given input, continue to play a crucial role in deep learning. However, traditional convolutional operations encounter challenges such as heightened computational complexity and excess redundant channels. These challenges can be effectively addressed with

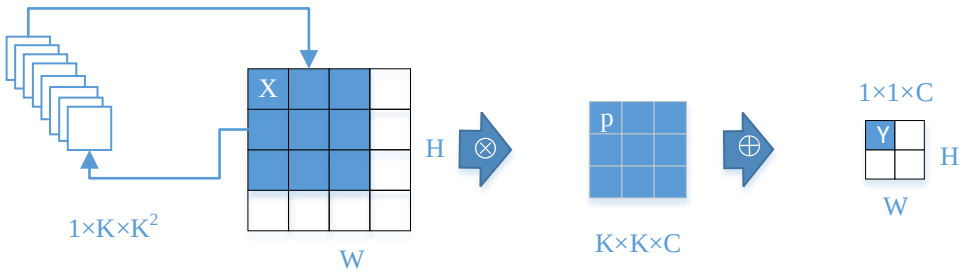

Fig.7. Schematic Diagram of Involution

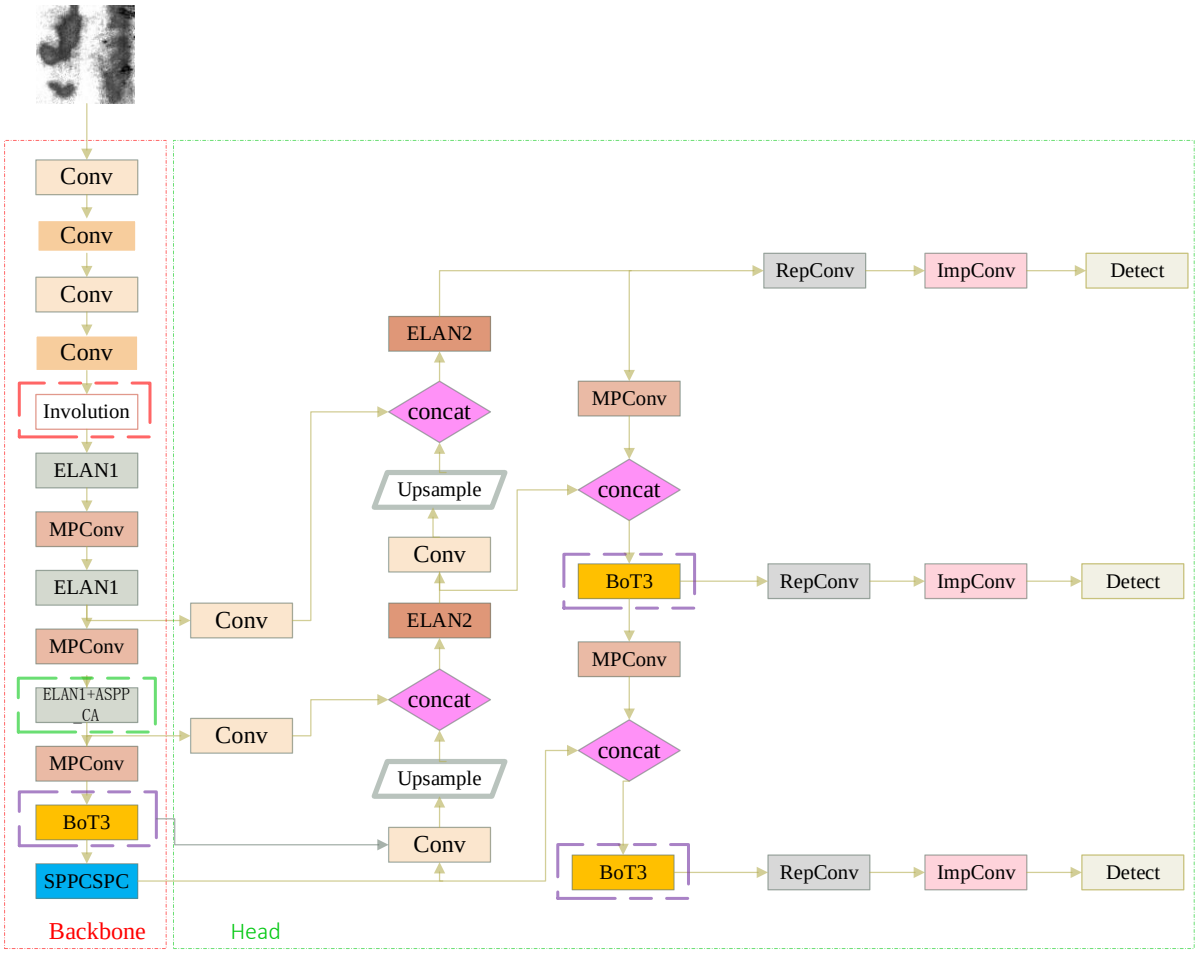

Fig.8. Improved YOLOv7-IBA structure diagram

Involution. Involution's parameters can be learned through training and can dynamically perform operations at specific pixel locations. This approach substantially diminishes redundancy and enhances the efficiency and generalization capabilities of the model, both about channel and feature spaces of the input. Hence, we have opted to utilize Involution to enhance the model's accuracy in detecting surface irregularities on strip steel.

The Involution core is specifically designed for the pixels Xi, j located at the corresponding coordinates (i, j), However, this information is shared uniformly across all channels. G calculates the count of groups that share the same Involution core for each group. When validating input for multiplication and addition operations with Involution, the output feature map of the Involution is expressed by the following formula:

$$Y_{i,j,k} = \sum_{(u,v)\in V_k} P_{i,j,u+[k/2],v+[K/2],[KG/C]} X_{i+u,j+v,k} \quad (3)$$

The format for the kernel is as follows:

$$P_{i,j} = \varnothing(X_{\varphi_{i,j}}) = W_1\sigma(W_0 X_{i,j}) \quad (4)$$

In 2021, Li Duo and his colleagues introduced a novel neural network operator named Involution [19]. Compared to convolution, this approach is lighter in weight and more efficient. It applies to various visual task models, enhancing their effectiveness and efficiency, as illustrated in Figure 7. Our model, incorporating Involution, significantly reduces network parameters and computational complexity. Our model, incorporating Involution, significantly reduces network parameters and computational complexity.

D. Model Reconstruction

To enhance the YOLOv7 network's performance in detecting small or densely packed objects, we introduce an improved YOLOv7-IBA algorithm and optimize the network structure. In contrast to the original YOLOv7 algorithm, the enhanced YOLOv7-IBA algorithm improves the network's resolution and feature expression capabilities, leading to better target localization. Specific optimizations are highlighted using differently colored rectangular boxes in the framework of the network structure, as shown in Figure 8.

The involution convolution, added within the red box of the network structure, enhances the model's accuracy and efficiency by improving feature expression and detection precision. Furthermore, the inclusion of the BoT3 module, surrounded in the purple box, provides an upgrade over the previous green-labeled ELAN module, capturing a wider spectrum of contextual information. This enhancement improves the precision of the model's object detection. Within the green box, we have integrated the ASPP_CA module into the original ELAN module to obtain multi-scale target information and expand the perception range. This configuration demonstrates robust performance in detecting dense or small targets. With the optimizations and enhancements implemented, YOLOv7-IBA has demonstrated a significant improvement in its ability to detect surface

defects on strip steel compared to the original YOLOv7 algorithm, resulting in improved target detection accuracy.

III. Experiments

A. DataSets

Surface defect detection is a critical aspect of steel production, with the primary objective being to identify and correct surface defects. This process plays a crucial role in improving steel products' overall quality and safety. The NEU-DET dataset utilized in this experiment was sourced from the Surface Defect Database provided by Northeastern University [20]. This dataset comprises six common surface abnormalities typically observed on hot-rolled steel strips, including rolled-in scale (Rs), patches (Pa), crazing (Cr), pitted surface (Ps), inclusions (In), and scratches (Sc). This dataset accurately represents the surface imperfections commonly encountered in most hot-rolled steel strips. The dataset consists of 1800 images and includes comprehensive annotations that provide essential details such as the type, location, size, and quantity of defects. Accurate annotation data is crucial for effectively training and testing the algorithm. The dataset includes 300 samples for each type of defect, with multiple defects present in each image sample. All images have dimensions of 200 by 200 pixels, and the data is split into training and testing sets in an 8:2 ratio.

B. Experimental Configuration

The experiment used a GEFORCE RTX 3090 graphics card, and PyTorch 1.13.1 was utilized as the deep learning framework with Cuda 11.7 for enhanced performance. The experiments ran for a total of 300 iterations.

C. Performance Evaluation

For a comprehensive and objective performance assessment of strip steel surface defect detection, we use two metrics: Precision and mAP. The formulas for calculating these metrics are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$AP = \int_0^1 P(R)dR \tag{6}$$

$$mAP = \sum_{i=1}^{c} AP_i \tag{7}$$

In these formulas, TP represents the number of actual positive defects, FP represents the number of false positive defects, P(R) represents the precision-recall curve, I represents the defect category in this experiment, and c represents the number of defect categories considered in this experiment. The elements above are used to calculate the performance evaluation of defect detection.

D. Ablation Experiment

The IBA model incorporates YOLOv7, BoT3, ASPP_CA, and Involution, all utilizing convolutional modules. Ablation experiments were conducted to evaluate each module's effectiveness in improving the model's performance. In these experiments, specific modules were intentionally removed to evaluate the impact of module removal on the experimental outcomes. Table 1 displays the findings of the ablation study, providing insights into how these enhanced techniques affect the experimental outcomes.

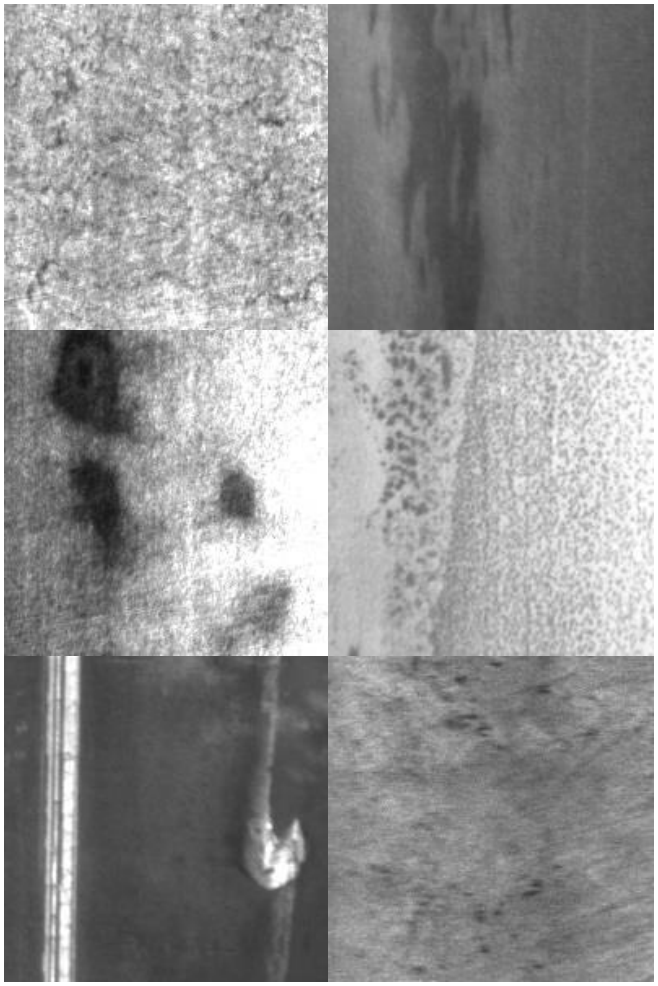

Fig.9. Dataset image

Table I shows the utilization of the YOLOv7 model, which integrates BoT3, ASPP_CA, and Involution techniques. Comparative experiments were conducted to evaluate a range of indicators, namely Rs, Pa, Cr, Ps, In, Sc, P, and mAP. The aim was to assess and compare their performance. We conducted comparative experiments using various indicators to determine these techniques' impact on identifying different categories of surface defects. We conducted comparative experiments using various indicators. The indicators include Rs, Pa, Cr, Ps, In, Sc, P, and mAP. The experimental results undeniably illustrate that the YOLOv7-IBA model has significantly improved in various metrics compared to other models using the same dataset. In summary, the YOLOv7-IBA model demonstrates exceptional detection capabilities and improved sensitivity in identifying surface defects on strip steel.

E. Comparative Experiment

This study conducted experiments to assess the impact of integrating an attention mechanism into the ASPP module of the strip steel surface defect detection model. The study evaluated the effect of three distinct attention mechanisms, namely GAM [21], ECA [22], and CA, on the performance of

TABLE I
ABLATION STUDY

Scheme	Rs	Pa	Cr	Ps	In	Sc	P	mAP
YOLOv7	0.750	0.788	0.569	0.895	0.686	0.852	0.757	0.767
YOLOv7-ASPP_CA	0.696	0.841	0.546	0.867	0.788	0.947	0.781	0.788
YOLOv7-BoT3	0.618	0.83	0.547	0.854	0.804	0.917	0.762	0.777
YOLOv7-Involution	0.659	0.839	0.566	0.897	0.827	0.879	0.777	0.794
YOLOv7-IBA	**0.780**	**0.88**	**0.647**	**0.934**	0.799	**0.935**	**0.829**	**0.799**

TABLE II
COMPARISON RESULTS WITH OTHER ATTENTION MECHANISMS

Model	Input size	Batch size	Epoch	Precision	mAP
YOLOv7-GAM	640*640	16	300	74.80	77.10
YOLOv7-ECA	640*640	16	300	75.70	77.20
YOLOv7-CA	640*640	16	300	74.60	77.70
YOLOv7-ASPP_CA	640*640	16	300	78.10	78.80

the model. As shown in Table II, including CA attention significantly improves the model's overall performance, particularly in accurately detecting small targets. Therefore, this article adopts the ASPP_CA module, which integrates a CA attention mechanism within the ASPP module. The experimental comparison showed that the attention mechanisms GAM and ECA had some impact, but their improvement compared to the CA attention mechanism was relatively modest. The CA attention mechanism itself is a type of channel attention mechanism. I am distinguishing it from ECA in its computational ASPP module. Creating the ASPP_CA module effectively handles multi-scale features and enhances the model's performance.

Currently, widely used network models for object detection include YOLOv8, YOLOv7, YOLOX, and SSD. We have used these models and recent research to train them to identify strip steel surface defects. Next, we tested the effectiveness of our detection methods on a test dataset. After thoroughly comparing the models using precision metrics, it became evident that the YOLOv7-IBA model showcased significant improvements in detection efficiency. Through experimental comparison, it is evident that the YOLOv7-IBA model demonstrates significantly improved accuracy compared to traditional models. Compared to the latest YOLOv8 model, the YOLOv7-IBA model demonstrates an increase in accuracy of 8.1 percentage points. Additionally, compared to YOLOX and SSD [23], the YOLOv7-IBA model shows significant improvements in accuracy, with an increase of 10.1 and 8.3 percentage points, respectively. Furthermore, compared to recent publications such as YOLOv7-BES [24] and YOLOv5-v3 [25], our approach outperforms others in detecting surface defects in strip steel datasets. The YOLOv7-IBA model has demonstrated exceptional proficiency in identifying surface I imperfections on strip steel.

TABLE III
COMPARISON OF STATE-OF-THE-ART MODELS

Model	Input size	Batchsize	Epoch	Precision
YOLOv8	640*640	16	300	74.80
YOLOv7	640*640	16	300	75.70
SSD	640*640	16	300	74.60
YOLOX	640*640	16	300	72.80
YOLOv5-v3	640*640	16	300	75.40
YOLOv7-BES	640*640	16	300	79.20
YOLOv7-IBA	640*640	16	300	82.90

The enhanced YOLOv7-IBA model, introduced in this paper, is specifically designed to detect surface defects in strip steel. It has been trained using a dataset that is relevant to this domain. Following each training session, it is crucial to use a test set to evaluate the model's detection performance.

A diverse range of images was chosen from the test dataset to assess the generalization capability of the improved model. The detection results for the enhanced YOLOv7-IBA algorithm and the original YOLOv7 algorithm in various scenarios are presented in Figures 10 and 11, respectively. In contrast to the improved YOLOv7-IBA algorithm, the original YOLOv7 algorithm exhibits challenges regarding missed detections and false positives.

In summary, the improved YOLOv7-IBA network significantly reduces missed targets and false positives and has accurate detection capabilities for dense, blurry, and small-scale targets. Moreover, the upgraded YOLOv7-IBA architecture consistently achieves exceptional detection results, even in low-contrast and complex backgrounds, especially when identifying strip defects. According to the experimental findings, the improved YOLOv7-IBA algorithm outperforms existing models and holds significant practical potential for enhancing target detection accuracy. It effectively meets the requirements for target detection in the context of surface defects on strip steel.

IV. CONCLUSION

This study introduces an enhanced framework known as YOLOv7-IBA, which builds upon the original YOLOv7 model. By integrating the BoT3 module and retaining a consistent feature size, the framework achieves improved detection precision. Furthermore, the creation of

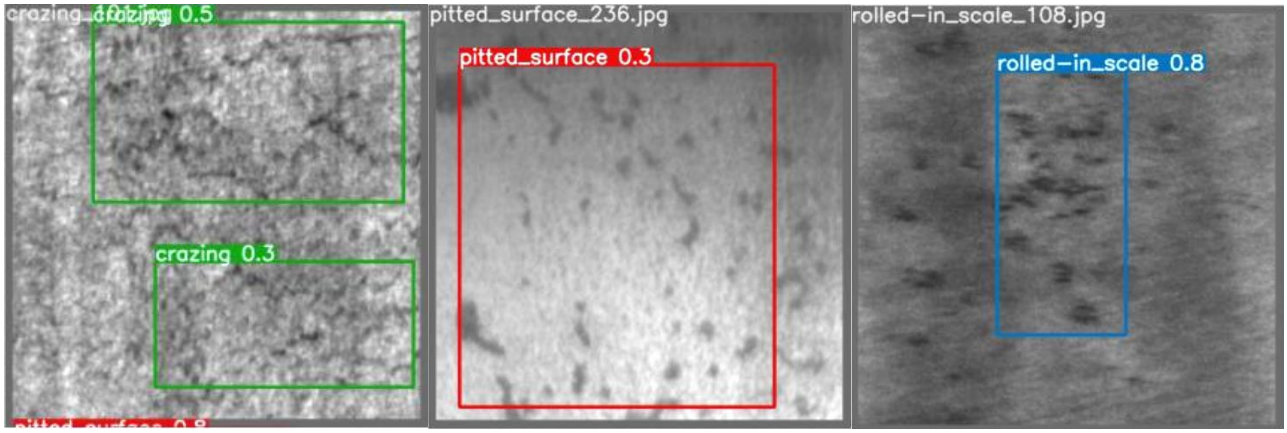

Fig.10. YOLOv7 detection effect diagram

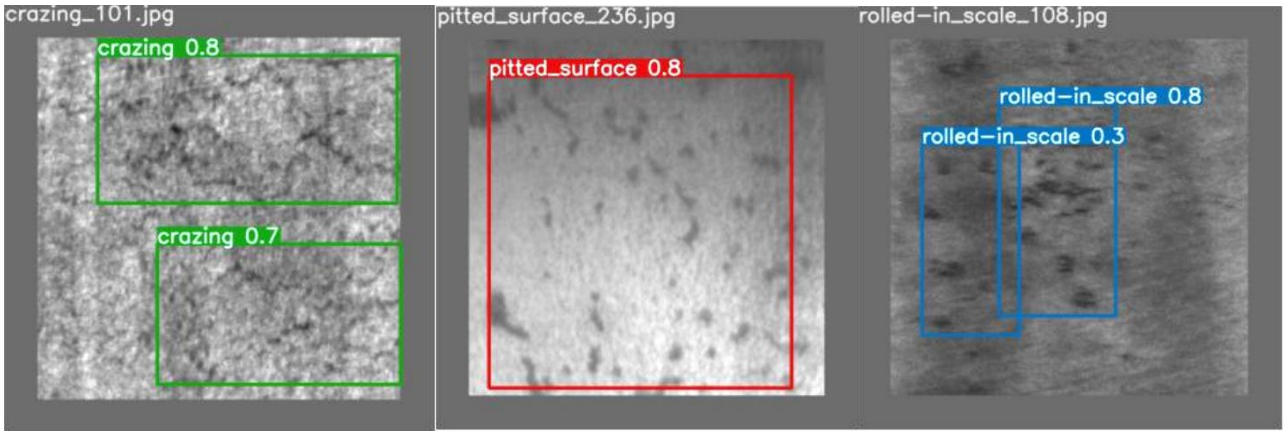

Fig.11. YOLOv7 IBA detection effect diagram

ASPP_CA contributes to a reduction in model size and an increase in inference speed for essential computer architecture modules. The adoption of involuted convolution technology has further improved the model's proficiency in detecting small targets, especially for the purpose of identifying surface defects on strip steel. Compared to the four conventional detectors, namely YOLOv8, YOLOv7, YOLOX, and SSD, our approach is uniquely tailored for surface defect detection in strip steel. It performs better by significantly improving accuracy and mAP values, ultimately leading to better and more accurate detection results.

Detecting surface defects on strip steel is challenging and complex, but our research presents innovative concepts and approaches to tackle this issue effectively. We are confident that as technology advances and we explore new possibilities, the efficiency and accuracy of surface defect detection in strip steel will continue to improve. This will undoubtedly add even greater value and enhance safety within the industrial detection field. We aim to provide valuable references and insights through our research while contributing to the development and progress of related fields.

REFERENCES

[1] S. Wang, X. J. Xia, L. Q. Ye, and B. Yang, "Automatic Detection and Classification of Steel Surface Defect Using Deep Convolutional Neural Networks," *Metals*, 11(3): 388, 2021.

[2] Z. P. Li, J. Zhang, T. Zhuang, and Q. Y. Wang, "Metal Surface Defect Detection," *2019 IEEE Pune Section International Conference (PuneCon)*, pp. 1-4 , Pune, India, 2019.

[3] Y. He, K. C. Song, Q. G. Meng, and Y. H. Yan, "An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features. [J]," *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(4): 1493-1504.

[4] D. M. Tsai, M. C. Chen, W. C. Li, and W. Y. Chiu, "A Fast Regularity Measure for Surface Defect Detection," *Machine Vision and applications* 23 (2012): 869-886.

[5] Z. X. Guo, C. S. Wang, G. Yang, Z. Y. Huang, and G. Li, "MSFT-YOLO: Improved YOLOv5 Based on Transformer for Detecting Defects of Steel Surface," *Sensors*22.9 (2022): 3467.

[6] H. Son, N. Hwang, C. M. Kim, and C.W. Kim, "Rapid and Automated Determination of Rusted Surface Areas of a Steel Bridge for Robotic Maintenance Systems," *Automation in Construction* 42 (2014): 13-24.

[7] L. Tong, W. K. Wong, and C. K. Kwong, "Differential Evolution-Based Optimal Gabor Filter Model for Fabric Inspection," Neurocomputing 2016, 173, 1386-1401.

[8] P. Soviany and R. T. Ionescu, "Optimizing the Trade-Off between Single-Stage and Two-Stage Deep Object Detectors Using Image Difficulty Prediction," *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Timisoara, Romania, pp. 209-214, 2018.

[9] J. Q. Fan, T. J. Huo, and X. Li, "A Review of One-Stage Detection Algorithms in Autonomous Driving," *In Proceedings of the 2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, Hangzhou, China, pp. 210-214, 18-20 December 2020.

[10] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-art for Real-time Object Detectors," *arXiv* 2022, arXiv:2207.02696.

[11] J. Redmon, and A. Farhadi, "YOLOv3: An Incremental Improvement [J] ," *arXiv preprint arXiv:*180402767, 2018.

[12] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936-944, Honolulu, HI, USA, 2017.

[13] G. Jocher, A. Chaurasia, A. Stoken, X. Tao, Lorna, and Laughing, et al. "Ultralytics/YOLOv5: V7. 0—YOLOv5 SOTA Realtime Instance Segmentation," *Zenodo* (2022).

[14] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and Alan L Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional

Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (2016): 834-848.

[15] Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[16] A. Srinivas, T. Y. Lin, N. Parmar, J. Shlens, and A. Vaswani, "Bottleneck Transformers for Visual Recognition," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[17] X. L. Wang, R. Girshick, A. Gupta, and K. He, "Non-Local Neural Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[18] C. H. Xie, Y. X. Wu, L.V. D. Maaten, A. L. Yuille, and K. He, "Feature Denoising for Improving Adversarial Robustness," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 501-509, Long Beach, CA, USA, 2019.

[19] D. Li, J. Hu, C. H. Wang, and X. T. Li, "Involution: Inverting the Inherence of Convolution for Visual Recognition," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12316-12325, Nashville, TN, USA, 2021.

[20] K. C. Song, and Y. H. Yan, "A noise Robust Method Based on Completed Local Binary Patterns for Hot-rolled Steel Strip Surface Defects," *Applied Surface Science* 285 (2013): 858-864.

[21] Y. C. Liu, Z. R. Shao, and N. Hoffmann, "Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions," *arXiv preprint arXiv:2112.05561* (2021).

[22] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[23] L. Wei, A. Dragomir, E. Dumitru, S. Christian, and R. Scott, "SSD: Single Shot MultiBox Detector," *European Conference on Computer Vision* (2015).

[24] Y. Wang, H.Y. Wang, and Z. H. Xin, "Efficient Detection Model of Steel Strip Surface Defects Based on YOLO-V7," *IEEE Access* 10 (2022): 133936-133944.

[25] D. Wu, W. K. Ma, M. H. Li, R.T. Li, and Y. Li, "Steel Surface Defect Detection Based on Improved YOLOv5," *Journal of Shaanxi University of Science and Technology* 41.2 (2023): 8.