FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X

Advancements in Steel Surface Defect Detection: An Enhanced YOLOv5 Algorithm with EfficientNet Integration

Fei Ren[1], ZiAngZhang[2], Jiajie Fei [3], HongSheng Li[4] and Bonifacio T. Doma Jr.[5*]
[1,5]School of Information Technology, Mapua University, Manila, Philippines; renfeivision@outlook.com[1] , btdoma@mapua.edu.ph[5]
[2,3,4]School of Automation, Nanjing Institute of Technology, NanJing, China

***Correspondence:** Bonifacio T. Doma Jr.; btdoma@mapua.edu.ph

ABSTRACT- Steel surface defect detection is of utmost importance for ensuring product quality, cost reduction, enhanced safety, and heightened customer satisfaction. To address the limitations of traditional steel surface defect detection algorithms, which often yielded singular detection results and suffered from high miss detection rates, we proposed an enhanced Yolov5 steel surface defect detection algorithm. In this approach, this paper employed the EfficientNet network as a replacement for the Yolov5 backbone network. Subsequently, we trained and tested this modified network on a steel surface defect dataset to mitigate the challenges associated with high miss detection rates and underperforming evaluation metrics. Our experimental findings underscored the superiority of the improved algorithm, particularly when compared to Yolov5. This enhanced algorithm exhibited substantial improvements across several key performance metrics, including Precision, Recall, mAP@0.5, parameter count, and pt file size. Noteworthy achievements included a 6.39% increase in Precision for Yolov5-EfficientNetB4, a remarkable 7.75% improvement in Recall for Yolov5-EfficientNetB0, and a 5.57% boost in mAP@0.5 for Yolov5-EfficientNetB6. Additionally, the pt file size for Yolov5-EfficientNetB0 saw a substantial 39.65% reduction, although it was important to note that the inference time for the improved algorithm increased. Among the models, Yolov5-EfficientNetB6 struck the best balance in terms of performance.

Keywords: Steel surface defect, Improved Yolov5, EfficientNet, Detection algorithm.

ARTICLE INFORMATION

Author(s): Fei Ren, ZiAngZhang, Jiajie Fei , HongSheng Li and Bonifacio T. Doma Jr.;

Received: 08/12/2023; **Accepted**: 08/02/2024; **Published**: 28/03/2024;
e-ISSN: 2347-470X;
Paper Id: IJEER 0812-03;
Citation: 10.37391/IJEER.120137
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-12/ijeer-120137.html

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

The iron and steel industry plays a crucial role in China's economic development and is intertwined with various sectors of the manufacturing industry. Steel, as a fundamental material, impacts industries ranging from mining and energy to construction and home appliances. However, steel often exhibits surface defects like scratches and cracks during production and use, undermining its quality and performance. Effectively monitoring these defects is essential to maintain steel quality. Current methods, including manual, infrared, and eddy current detection, have limitations [1-2].

Recent advancements in machine vision, artificial neural networks, and deep learning revolutionized image processing, increasing efficiency and accuracy. Scholars explored deep learning for steel surface defect detection. FU G et al. [3] proposed a CNN model combined with multiple receptive fields to obtain the deep semantic features of the target and achieve rapid classification and detection of steel surface defects. However, the accuracy rates dropped on images with severe camera noise, non-uniform illumination, and motion blur. Xing Jianfu [4] built a strip steel surface defect classification model based on AlexNet, expanded the data of various strip steel defects, and produced a strip steel surface defect dataset. Comparative experiments verified that this model improved the strip steel surface defect classification ability compared to traditional methods, but further refinement was needed in the derivation of system functions and the classification of defect types. Akhyar et al. [5] added the FPN structure to RetinaNet and proposed a defect detection network based on an improved RetinaNet. However, due to the complexity of the overall model and the large amount of calculation, it could not meet the requirements of real-time detection. Weimer et al. [6] evaluated the effectiveness of 5–11-layer CNN networks in defect detection, but the method they proposed extracted small patches of images and classified them separately, which was inefficient. The Yolo network, a One-Stage detection algorithm, gained popularity due to its speed and accuracy in steel defect detection. Xu et al. [7] replaced the original feature extraction network with a lightweight MobileNet network based on Yolov3. They used the Inception structure to improve the real-time performance of detection and improved the ability to extract small target features. However, better detection results could not be obtained on mobile devices with poor performance, and there was still room for improvement in terms of lightweighting.

To tackle the multi-category issue in steel surface defects, this paper leveraged the EfficientNet network to replace the Yolov5 backbone. The feature maps of three scales of steel surface defects were extracted through the EfficientNet network, and then fused by the feature pyramid structure in the Yolov5

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X

network to improve the overlapping target detection ability. We conducted comparative experiments involving EfficientNetB0-B7 and Yolov5 to assess fusion effects.

2. DEEP LEARNING YOLOV5 ALGORITHM

The network structure of Yolov5 followed the design principles of the Yolo [8-9] series, which consists of four key components: input, Backbone, Neck, and prediction result output. To expedite model convergence, Mosaic data augmentation is applied at the input stage. In the backbone network, Yolov5 draw inspiration from CSPNet, similar to Yolov4 [10-11], but introduces two CSP structures and incorporates the Focus module. Specifically, in the Yolov5 architecture, CSP1_X is utilized for the Backbone, while CSP2_X is implemented for the Neck. This configuration in the Neck phase facilitates multi-scale feature fusion through the amalgamation of top-down up-sampling from the FPN structure and bottom-up down-sampling from the PAN structure. Consequently, it seamlessly integrates both semantic and feature information from feature maps of varying sizes. To provide a visual representation of the network structure, please refer to *figure 1*.

For defining the boundary loss function at the output stage, the CIOU_loss is employed. After the predictions, a weighted NMS (Non-Maximum Suppression) is applied to filter prediction frames, ultimately outputting the coordinates of the frame with the highest confidence for successful target detection.

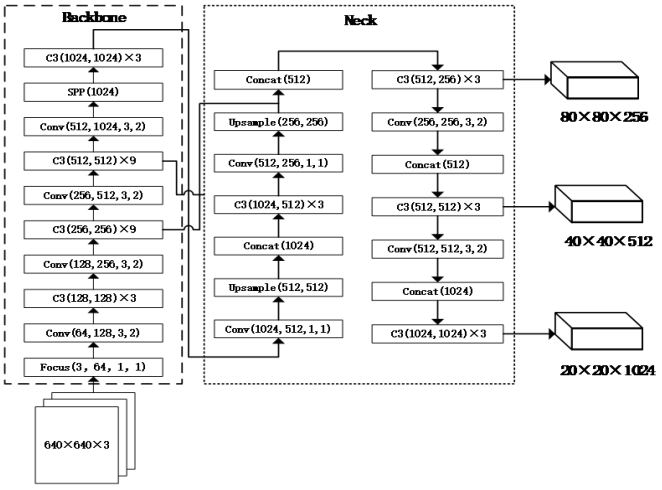


Figure 1. Yolov5 network structure

3. IMPROVED DEEP LEARNING ALGORITHMS

3.1. EfficientNet deep learning network

When dealing with limited data, deep learning models typically aim for higher accuracy by increasing the depth and width of the model or enhancing the image resolution. However, conventional convolutional neural networks require manual adjustments for these three dimensions, and improper tuning can lead to reducing model performance and efficiency. To address these challenges, TanM et al. introduced a groundbreaking network architecture called EfficientNet [12] in 2019. EfficientNet represented a significant innovation in network modelling as it established a direct relationship between the depth of the model and the number of layer structures, the width of the model and the number of convolution kernels in the convolution operation, and the image resolution and input image size.

In essence, high-resolution images encompass more information, demanding a larger receptive field and a deeper network structure to capture more comprehensive features. EfficientNet employs composite coefficients to consistently expand the width, depth, and resolution of the network. The specific expansion rules are detailed in *equation (1):*

$$\begin{aligned} &depth: d = a^{\phi} \\ &width: w = \beta^{\phi} \\ &resolution: r = \gamma^{\phi} \\ &a \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned} \tag{1}$$

In *equation (1)*, α 、 β 、 γ represent constants determined through neural architecture search, while ϕ signifies the composite coefficient. EfficientNet has achieved an impressive Top5 accuracy of 97.0% on the ImageNet [13] dataset. When compared to models achieving similar accuracy levels, EfficientNet reduces the number of parameters by 7/8 and shortens training time by 5/6. This demonstrates the network's remarkable efficiency.

The overall architecture of the EfficientNet base network was visually represented in *figure 2*. It boasted a relatively intricate structure, primarily composed of 16 MBConv (Mobile Inverted Bottleneck Convolution) modules. Within this module, an attention mechanism inspired by SENet (Squeeze and Excitation Network) was introduced, commonly referred to as the SE module. *Figure 3* provides a visual representation of the comprehensive structure of the MBConv module. Within the mobile inverted bottleneck convolution operation, the input feature matrix undergoes several critical steps:

1. Firstly, it goes through a point-by-point convolution with a 1x1 kernel, adjusting the output channel dimension based on the expansion ratio.
2. If the expansion ratio exceeds 1, batch normalization is applied, followed by the Swish activation function.
3. Subsequently, depth convolution is executed, followed by compression and excitation operations, as depicted in *figure 4*.
4. Finally, a 1x1 point-by-point convolution is employed to reduce dimensionality and restore the original channel dimensions.

This process concluded with connection inactivation and skip operations, exclusively applied to the last layer of MBConv within each stage. This approach introduced variability in the model's depth while simultaneously reducing the time required for model training.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X

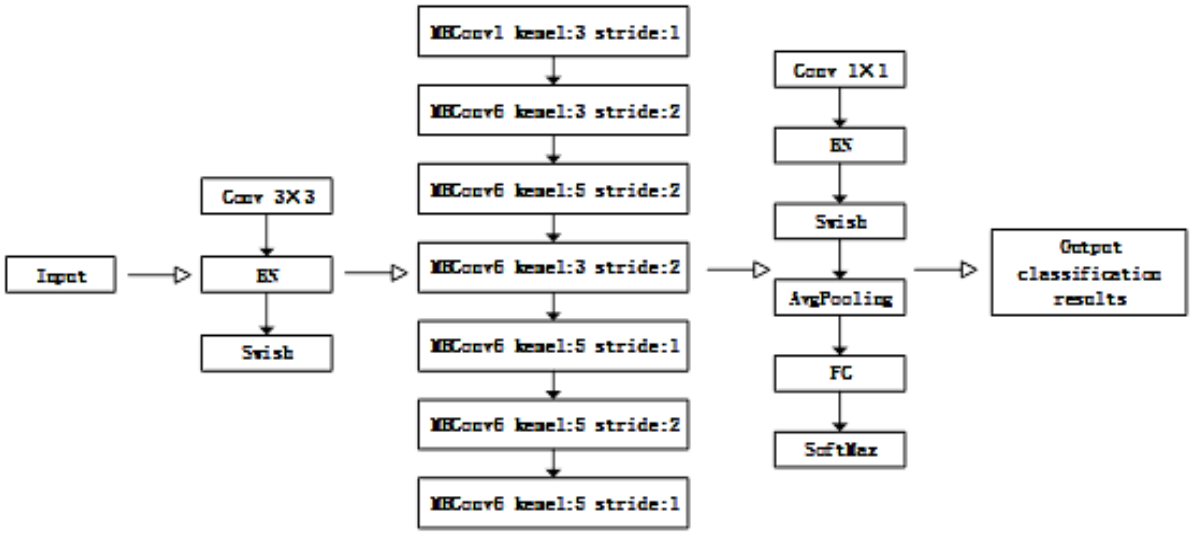


Figure 2. EfficientNet structure

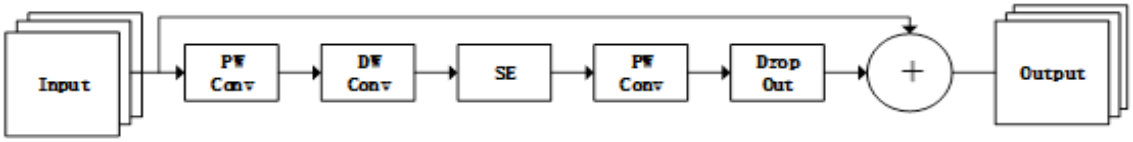


Figure 3. MBConv module

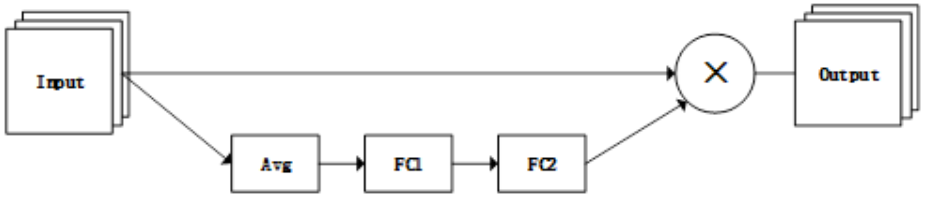


Figure 4. SE module

3.2 Deep learning network based on Yolov5-EfficientNet

In the past, the majority of research in steel surface defect detection relied on traditional image processing methods, resulting in limited generalization ability within most detection models. This paper addressed a detection task involving six distinct defect types, each varying in shape, color, and scale. Detecting complex and diverse steel surface defects directly with existing detection networks demanded a high level of network generalization.

EfficientNet models, spanning from B0 to B7, constitute a series of convolutional neural network models that utilize Compound Scaling. They create models of different sizes by simultaneously adjusting depth, width, and resolution. B0 represents the smallest model with relatively fewer parameters and computational demands, making it well-suited for resource-constrained environments. Slightly larger than B0, B1 introduces a width multiplier, augments the channel count, and enhances performance, catering to mobile devices and embedded systems. B2 to B7 further increases both width and depth while maintaining heightened resolution for improved accuracy and performance, albeit requiring more computational resources. This family of models offers flexible options to strike a balance between performance and computational cost, accommodating various computer vision tasks and resource limitations.

Experiments revealed that when the Yolov5 network was employed for steel surface defect detection, a substantial portion of true positive samples was erroneously classified as negative, indicating a significant issue with "missed detection". To tackle these challenges, the Yolov5 backbone network was replaced with the Efficient Net network. The feature maps extracted were subsequently passed through multiple convolutions and a feature fusion module up sampling to reach the output prediction module, as shown in *figure 5*.

These improvements enhanced the network's capacity to extract features from small, defective targets while significantly reducing missed detection rates and other detection metrics.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X

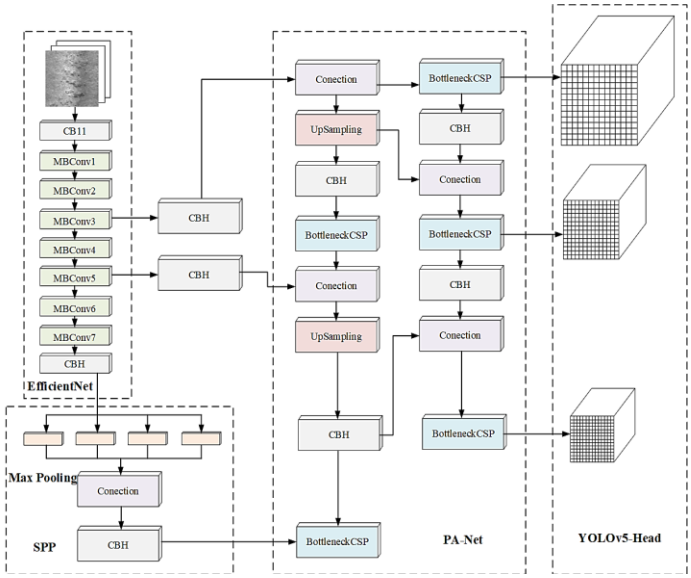
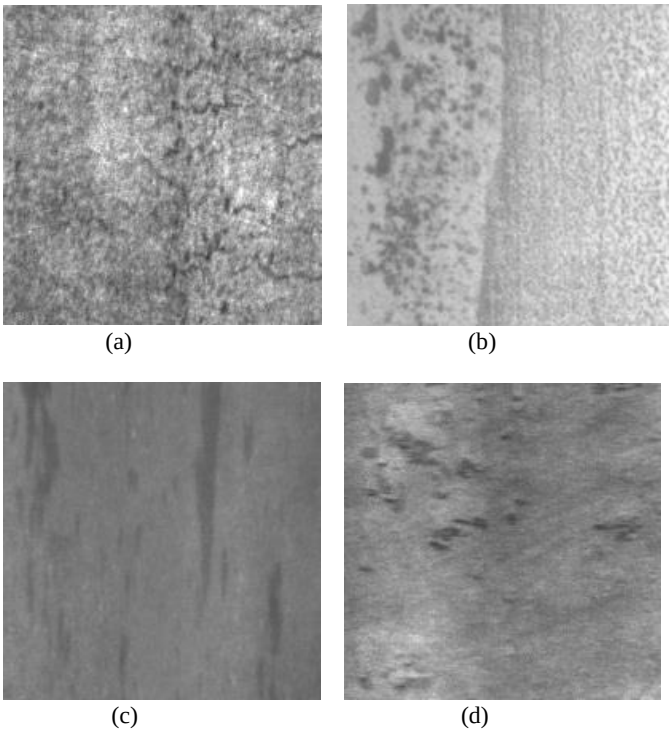


Figure 5. EfficientNet-YOLOv5 network structure diagram

4.Experimental verification and comparison

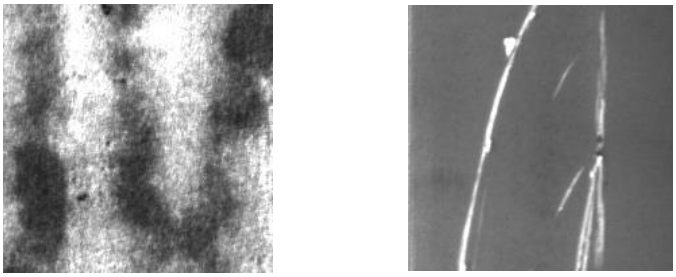
4.1 Experimental environment and data set

The experiment was conducted on a Windows 10 system, utilizing an Intel i7-11700 CPU operating at 2.50GHz. The GPU employed was an NVIDIA GeForce RTX 3080 Ti, with 32 GB of available RAM. The development environment consisted of PyCharm Community 2018.3.5 as the compiler and Python 3.8 as the interpreter.



(a) (b)

(c) (d)



(e) (f)

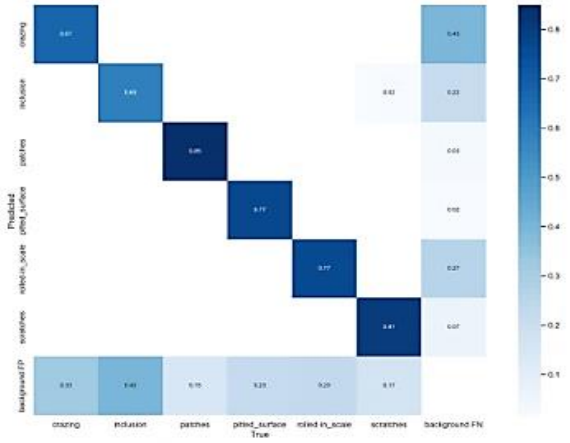
Figure 6. NEU-CLS data set (a) crazing; (b) pitted_surface; (c) inclusion; (d) rolled-in_scale; (e) patches; (f) scratches.

The experimental dataset used in this study was sourced from the NEU-CLS dataset, featuring a selection of six distinct defect types. Each defect category comprised 300 images, resulting in a total of 1,800 steel surface defect images, as depicted in *figure 6*.

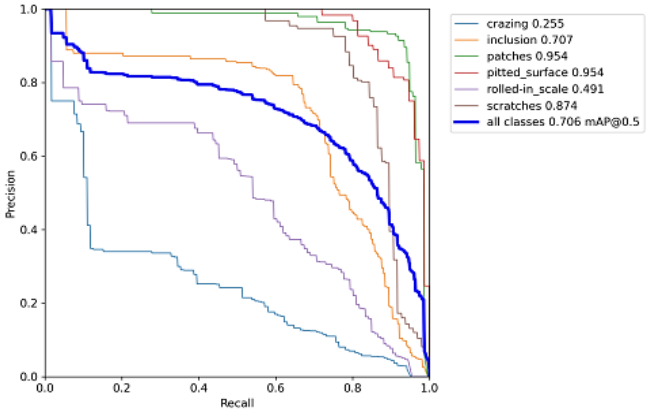
4.2 Experimental results

4.2.1. Enhanced Yolov5-EfficientNetB4 Model

After 200 training epochs to acquire the optimal weights, the performance on the test dataset yielded an mAP@0.5 of 0.706 and an F1 score of 0.670. These outcomes were visually represented in *figure 7*.

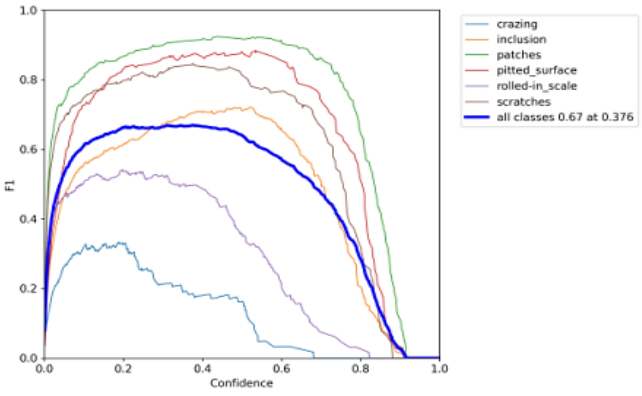


(a)

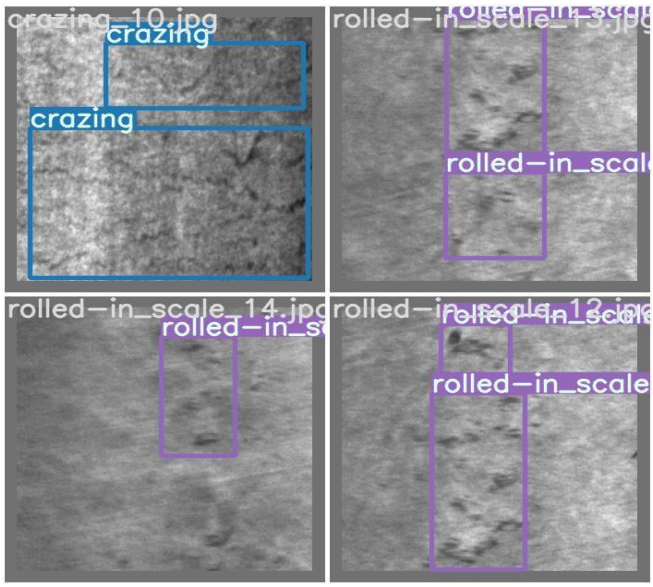


(b)

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X



(c)

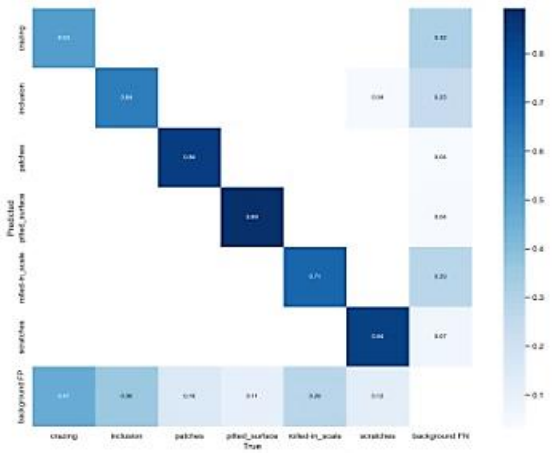


(d)

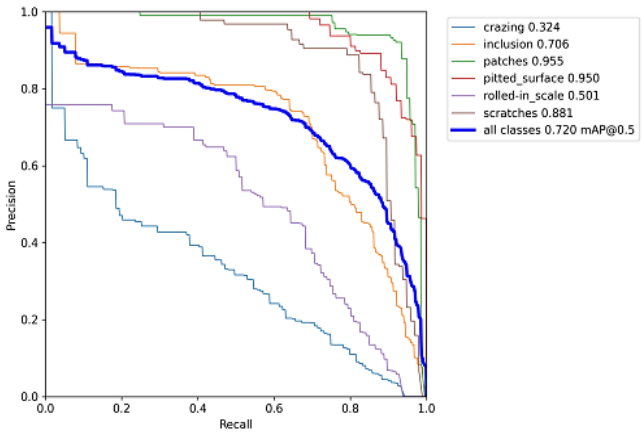
Figure7. Enhanced Yolov5-EfficientNetB4 result chart (a) confusion matrix; (b) PR curve; (c) F1 diagram; (d) Verification test chart.

4.2.2. Enhanced Yolov5-EfficientNetB6 Model

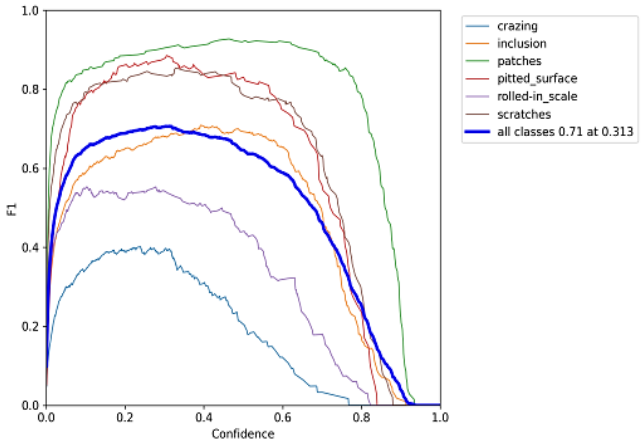
Following 200 training epochs to obtain the optimal weights, the performance on the test dataset yielded an mAP@0.5 of 0.720 and an F1 score of 0.71. These results were visually presented in *figure 8*.



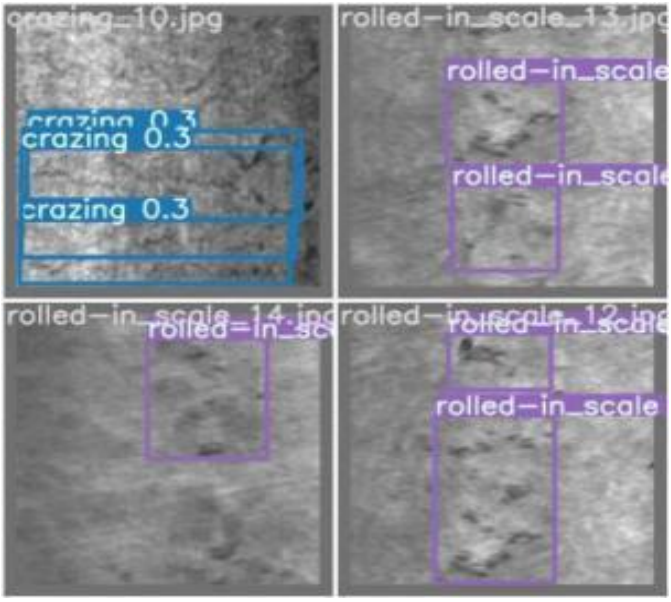
(a)



(b)



(c)



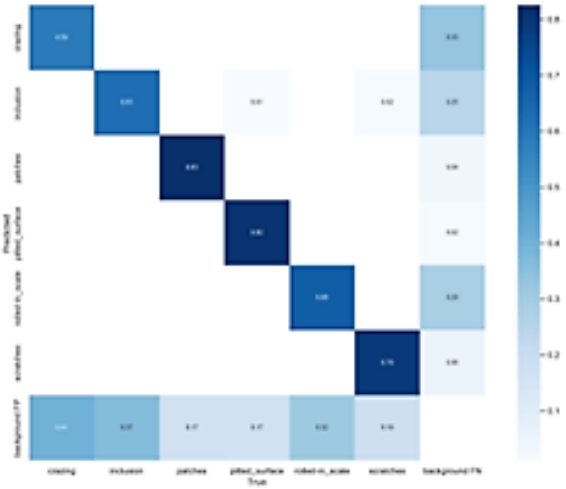
(d)

Figure 8. Enhanced Yolov5-EfficientNetB6 result chart (a) confusion matrix; (b) PR curve; (c) F1 diagram; (d) Verification test chart

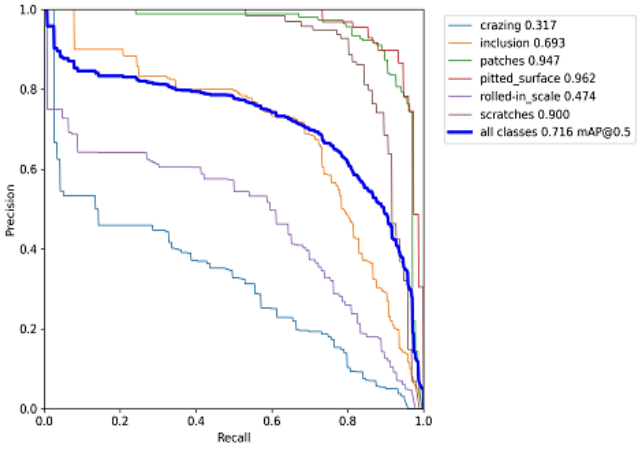
FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X

4.2.3. Enhanced Yolov5-EfficientNetB7 Model

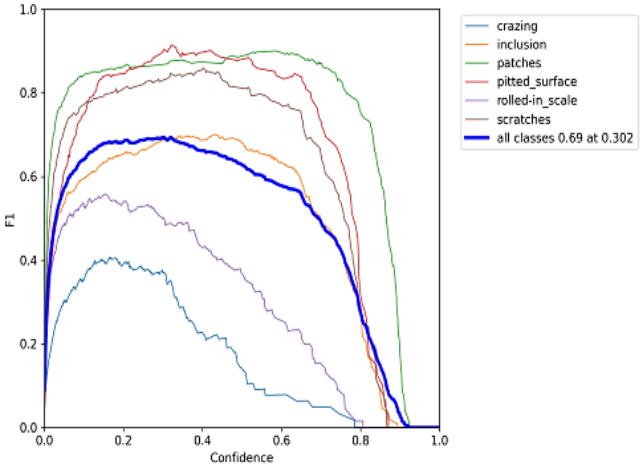
After 200 training epochs to acquire the optimal weights, the performance on the test dataset yielded an mAP@0.5 of 0.716 and an F1 score of 0.690. These results were visually presented in *figure 9*.



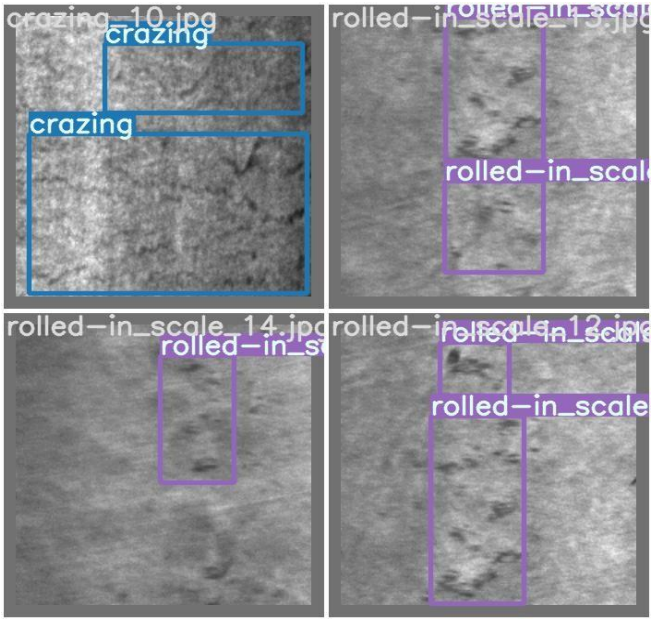
(a)



(b)



(c)



(d)

Figure 9. Yolov5-EfficientNetB7 result chart (a) confusion matrix; (b) PR curve; (c) F1 diagram; (d) Verification test chart.

5. DISCUSSION

The comparison of detection results between Yolov5-EfficientNetB0-B7 and Yolov5 was presented in *table 1*. The test results clearly indicated that the EfficientNet-Yolov5 network demonstrated improvements across various metrics when compared to Yolov5. EfficientNetB0-B7 exhibited distinct enhancements over Yolov5.

The comparison in *table 1* revealed several noteworthy observations. When compared to Yolov5:

Yolov5-EfficientNetB4 exhibited a 6.39% increase in Precision.

Yolov5-EfficientNetB0 achieved a significant improvement in Recall, reaching 7.75%.

Yolov5-EfficientNetB6 demonstrated a 5.57% increase in mAP@0.5.

Yolov5-EfficientNetB0 witnessed an 39.65% reduction in pt file size.The reduced parameter count and smaller model file size had significant advantages in resource-constrained deployment scenarios. This not only reduced storage requirements but also enhanced network transmission efficiency, making it more suitable for environments with limited resources, such as embedded systems and mobile devices. Consequently, it provided a more economical and efficient solution for practical applications.

Based on *table 1*, YOLOv5-EfficientNetB0-B7 had longer inference times compared to YOLOv5. This may have been due to YOLOv5-EfficientNet having more parameters, leading to increased model complexity and consequently affecting inference speed.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X

Table 1. Comparison of results from different networks

Type	Precision	Recall	mAP@.5	parameters	InferenceTime（ms）	pt file size
Yolov5	0.689	0.671	0.682	20873139	**3.9**	40.1
Yolov5-EfficientNetB0	0.64	**0.723**	0.709	**12484867**	5.6	**24.2**
Yolov5-EfficientNetB1	0.68	0.692	0.703	13808869	5.1	26.7
Yolov5-EfficientNetB2	0.61	0.722	0.693	16943371	5.6	32.8
Yolov5-EfficientNetB3	0.649	0.712	0.708	24071211	6.9	45.5
Yolov5-EfficientNetB4	**0.733**	0.641	0.706	35134739	8.0	67.8
Yolov5-EfficientNetB5	0.651	0.706	0.704	47805115	8.9	92.1
Yolov5-EfficientNetB6	0.721	0.699	**0.720**	71131827	10.7	136
Yolov5-EfficientNetB7	0.693	0.703	0.716	92216465	11.8	177

For scenarios with extremely high demands on inference time, various alleviation strategies were employed. Model pruning effectively reduced model complexity and improved inference speed by eliminating redundant parameters. Model quantization transformed parameters into integer representations, significantly reducing model size, accelerating inference, and having limited impact on accuracy. Hardware optimization involved the use of higher-performance hardware such as GPUs or specialized neural network accelerators, markedly enhancing inference speed. Distributed inference distributed tasks to multiple devices for parallel processing, improving overall inference speed. These strategies, while maintaining accuracy, significantly reduced inference time, making them suitable for applications with stringent real-time requirements.

In summary, this experiment involved 200 epochs of training to select the best model and showcased performance improvements in Precision, Recall, and mAP@0.5 on the steel surface defect dataset. Yolov5-EfficientNetB6 stood out with improved Precision, Recall, and mAP@0.5 performance, along with relatively minor increases in inference time and pt file size, representing the best balance in terms of performance.YOLOv5-EfficientNetB6 demonstrated outstanding performance in precision, recall, and mAP@.5 metrics for object detection tasks, surpassing YOLOv5 comprehensively in all aspects (other improved algorithms achieved superiority in only one or two metrics over YOLOv5). Simultaneously, it exhibited moderate characteristics in terms of parameter count, inference time, and model file size, presenting a well-balanced performance. The model's equilibrium across various performance metrics positioned it as the optimal choice during testing, suitable for real-time applications. Striking a favourable balance between complexity and performance, it provided superior cost-effectiveness.

6. CONCLUSION

This paper proposed a steel surface defect detection algorithm based on Yolov5, utilizing the Efficient Net network. By replacing the Yolov5 backbone network with the Efficient Net series network, the algorithm was enhanced for steel surface defect detection. The Northeast China University's steel surface defect dataset was used to train and test the improved network, aiming to reduce the missed detection rate of steel surface defects and improve other related detection indicators. The experimental results showed significant improvements in Precision, Recall, mAP@0.5, number of parameters, and pt file size when compared to Yolov5. Yolov5-EfficientNetB4 exhibited a 6.39% increase in Precision, Yolov5-EfficientNetB0 achieved a 7.75% increase in Recall, and Yolov5-EfficientNetB6 demonstrated a 5.57% increase in mAP@0.5. Furthermore, the pt file size of Yolov5-EfficientNetB0 decreased by 39.65%. These findings indicated that improving the network structure could effectively enhance the steel defect detection performance of Yolov5. Future research endeavours will focus on reducing inference time, refining the network architecture, and further enhancing detection speed.

Funding: This research was funded by Office of Directed Research for Innovation and Value Enhancement (DRIVE) of Mapua University.

REFERENCES

[1] Xu Huan, Yin Chenbo, Li Xiangdong, et al. Research on Infrared Thermal Image Detection of Weld Defects Based on Finite Element Method [J]. Hot Working Process, 2019,48 (17): 122-127+133. DOI: 10.14158/j.cnki.1001-3814.2019.17.033.

[2] Huang Fengying. Quantitative Evaluation Method for Eddy Current Detection of Rail Surface Cracks [J]. China Railway Science, 2017,38 (02): 28-33.

[3] Fu G, Sun P, Zhu W, et al. A deep learning based approach for fast and robust steel surface defects classification [J] Optics and Lasers in Engineering, 2019, 121:397-405.

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 269-276 | e-ISSN: 2347-470X

[4] Xing Jianfu Surface defect recognition and system development of hot-rolled strip steel based on convolutional neural network [D]. Northeastern University, 2017.

[5] Akhyar F, Lin C Y, Muchtar K, et al. High effective single stage steel surface defect detection [C]//2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) IEEE, 2019: 1-4.

[6] Weimer D, Scholz Reiter B, Shpitalni M. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection [J] CIRP Annals, 2016, 65 (1): 417-420.

[7] Xu Qiang, Zhu Hongjin, Fan Honghui, et al. Research on improved Yolov3 network for surface defect detection of steel plates [J]. Computer Engineering and Application, 2020,56 (16): 265-272.

[8] Quach* DL ,Quoc NK,Quynh NA, et al.Evaluating the Effect of YOLO Models in Different Size Object Inspection and Feature-Based Classification of Small Objects[J].Journal of Advances in Information Technology,2023,14(5).

[9] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[10] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[11] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.

[12] Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR, 2019: 6105-6114.

[13] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.