Steel Surface Defect Detection Method based on YOLOv11-MobileNetv4

Huanxin Zhou, Wenhao Li, Shuiyi Wei, Guangquan Men, Yuhe Wang, and Jiaqi Li*

School of Civil Engineering, University of Science and Technology Liaoning, Anshan 114051, China

*lijiaqi@ustl.edu.cn

Abstract

Steel occupies a central position in modern industry, but due to long-term use, cracks, corrosion and other defects often appear on its surface, threatening structural safety. Traditional manual inspection methods are inefficient and prone to misdetection and leakage, therefore, efficient and accurate inspection techniques are urgently needed. In this paper, we propose a lightweight steel surface defect detection model YOLOv11-MobilNetv4 based on the combination of YOLOv11 and MobileNetv4. Experimental results show that YOLOv11-MobileNetv4 shows a faster detection speed suitable for application in mobile devices, and the mAP reaches 0.714, which is comparable. This study provides an effective solution for steel defect detection and lays the foundation for subsequent lightweight applications of deep learning models.

Keywords

Steel Defect Detection; Deep Learning; YOLO Algorithm; MobileNet; Lightweight Model; Computer Vision.

1. Introduction

Steel, a material that occupies a central position in modern industry, is widely used in many industrial fields such as buildings, bridges, machinery manufacturing, transportation, energy facilities, and so on, due to its unique properties such as excellent strength, good plasticity, and toughness. However, with the passage of time, steel structures will inevitably experience performance degradation during long-term use. This degradation may be manifested in the form of cracks on the surface of the steel structure, corrosion, and spalling of the surface material and other forms of damage. These damages not only greatly affect the overall aesthetics and solidity of steel structure buildings, but also threaten people's lives and property safety. Therefore, timely and accurate detection of apparent defects on steel structures to ensure their structural integrity and safety has become an important task in the field of modern engineering.

In the past, this task mainly relied on manual inspection, however, a large number of practical experiences and facts show that the efficiency of manual inspection is extremely low, and it is difficult to meet the needs of large-scale, high-efficiency and high-precision inspection. Missed and misdiagnosis occurs from time to time. However, with the rapid development of computer information technology and the in-depth application of artificial intelligence technology, defect detection algorithms based on deep learning have emerged. This algorithm is able to automatically learn and identify various apparent defects in steel structures by training a large number of data samples, and it demonstrates multiple advantages such as high accuracy, intelligence, robustness, scalability and data-driven in detection. It can accurately recognize tiny defects and avoid miss detection and misdetection, while significantly improving the detection efficiency, and is applicable to various types and sizes of steel structures. Therefore, this technology will become the mainstream of apparent defect detection in the future, helping to improve the efficiency and accuracy of steel structure inspection, reduce costs, and provide a solid guarantee for project quality and safety.

In recent years, YOLO-based target detection algorithms have been developed to achieve faster and more accurate detection. Feng Yingbin [1] et al. proposed HPDE-YOLO based on the improvement of YOLOv8n. It makes the model detection faster, can extract more feature information and improve the ability of the model to process the features, and improves the accuracy of the detection of defects on small targets. Liu Wenzhao [2] et al. used a lightweight convolutional module MSConv and M-BiFPN network for fusion of deep and shallow feature information. Reducing the number of parameters and computational complexity of the model, a lightweight steel surface defect detection model YOLO-LSNet was proposed . Tao Youfeng [3] et al. proposed the RCD-YOLO algorithm. By designing a lightweight network R_HGNet as the backbone feature extraction network to improve the feature extraction efficiency, introducing the upsampling operator CARAFE and designing the dilation residual module C2f-DWR in the Neck network to improve the quality and richness of the output features, and finally using the Inner-CIOU loss function to improve the regression accuracy of the bounding box. However there is still no better lightweight model that can be integrated into mobile devices for steel defect detection. In this paper, we propose a lightweight steel surface defect detection method based on the combination of YOLOv11 and MobilNetv4 that can be implemented through mobile terminals.

2. Method

2.1 Computer Vision

Computer Vision is a technology designed to mimic the visual system of living things, enabling computers to “see” and understand images and video content. As an important branch of Artificial Intelligence (AI) and Machine Learning (ML), the goal of Computer Vision is to extract valuable information from still images or moving videos, analyze and understand it in depth, and based on this information, make appropriate decisions or perform specific tasks.

Currently, computer vision mostly relies on deep learning models such as convolutional neural networks (CNN). These models are trained and optimized by processing a large amount of well-labeled image data, thus gradually improving the ability to recognize and understand complex real images. With the abundance of training data and increased computational power, these models have achieved impressive results in many areas.CNNs are deep learning models that are particularly suitable for processing data with grid structure, such as images and videos [4,5].CNNs automatically detect spatial hierarchical features in the input data through mechanisms such as local receptive fields, weight sharing and pooling. Local receptive fields mean that each neuron responds to only a small region of the input; weight sharing ensures that features of the same type are consistently recognized throughout the input space; and pooling is used to reduce the spatial dimensionality of the data, thereby reducing computation and preventing overfitting. These features allow CNNs to efficiently learn hierarchical representations from raw data without the need for manual feature engineering.

Compared to traditional machine learning algorithms, CNNs are not only able to automatically extract features, but also learn more complex nonlinear relationships. Traditional machine learning methods usually require manual feature engineering and have limited model complexity to capture subtle patterns in high-dimensional data. In addition, CNNs tend to outperform traditional methods on large datasets, especially in computer vision tasks such as image classification, target detection, and semantic segmentation. Figure 1 shows the basic structure of a CNN with the VGG16 network as an example.

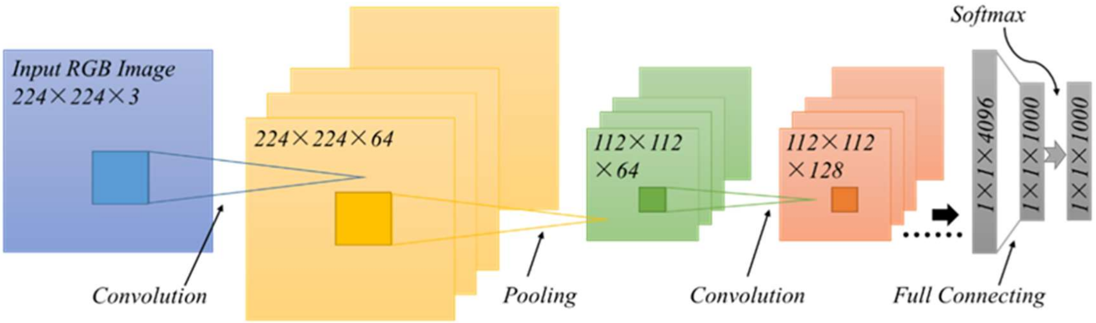

Figure 1. Basic structure of a CNN

2.2 Dataset

The NEU-DET steel surface defect detection dataset [6] was selected by Song Kechen team of Northeastern University, which contains 1800 pictures, including Crazing, Inclusion, Patches, Pitted Surface, Rolled-in Scale, Scratches, and six common types of steel damage. All images in the dataset are in standard image format, which is convenient for subsequent computer vision tasks, and are accompanied by a label file that describes in detail the type and location of defects in the image.

2.3 YOLOv11

YOLOv11 is an important breakthrough in the field of target detection by the Ultralytics team, which combines advanced accuracy, speed, and efficiency to provide users with a powerful tool to tackle a wide range of computer vision tasks [7]. The model supports a wide range of computer vision applications, including target detection, instance segmentation, image classification, pose estimation, and oriented target detection.

In order to improve performance on mobile devices, YOLOv11 has significantly tuned the depth and width parameters of the model. This improvement allows the model to operate efficiently even in resource-constrained environments, thus meeting the demand for real-time processing and high accuracy in modern applications.

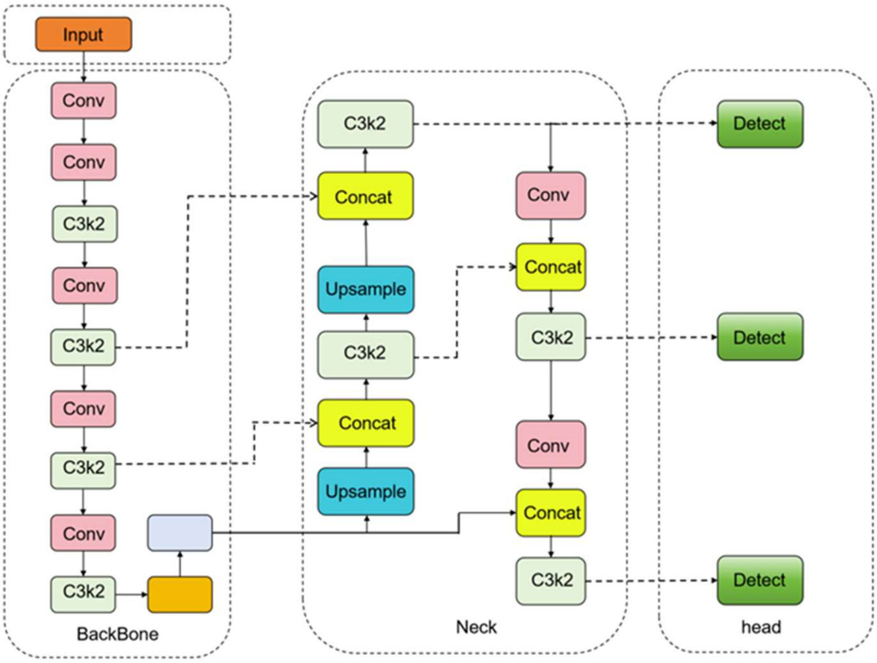

Figure 2. Network structure diagram of YOLOv11

Figure 2 shows the network structure of YOLOv11. The backbone network of YOLOv11 adopts C3k2 block, which integrates various state-of-the-art deep learning techniques, such as Depthwise Separable Convolutions, CNNs and deformed convolution, to further optimize the efficiency and effectiveness of feature extraction. The design of the C3k2 block improves computational efficiency, enabling YOLOv11 to extract features more quickly when processing images, and it has good adaptability under different computing resource environments. Whether on a high-performance GPU server or a more limited mobile device, it can maintain high performance and efficiency to meet the needs of different application scenarios.

2.4 MobilNetv4

As the focus of this research on lightweighting, MobileNetV4 (MNv4) is the latest generation of MobileNet convolutional neural networks designed for mobile devices, aiming to provide a versatile and efficient architecture that enables real-time interactive experiences while avoiding the uploading of private data to the public Internet [8].The core concept of the MobileNet family is Depthwise Separable Convolution, a technique that decomposes standard convolution into two more simplified processes: first, a deep convolution that processes each input channel independently, and second, a deep convolution that processes the outputs independently. This approach significantly reduces the number of parameters and the computational effort of the network.

Building on this foundation, MobileNetV4 introduces the Universal Inverted Bottleneck (UIB) module (Figure 3), which combines Inverted Bottleneck (IB), ConvNext, Feed Forward Network (FFN) and a novel Extra Depthwise variant. These innovations make the model structurally more flexible and efficient.

In addition, MobileNetV4 proposes a mobile version of Mobile MQA optimized for mobile gas pedals, which provides up to 39% inference acceleration compared to traditional MHSA. To enhance the generalization ability of the model, a novel knowledge distillation technique is introduced in the study, which achieves 87% ImageNet-1K classification accuracy through dataset mixing and balanced increase of intra-class data. But on CPU, MobileNetV4 runs much faster than its predecessor model, proving its extreme efficiency and usefulness on mobile devices. YOLOv11-MobilNetv4, which combines YOLO with this algorithm, also performs quite well in the detection experiments on steel surfaces.

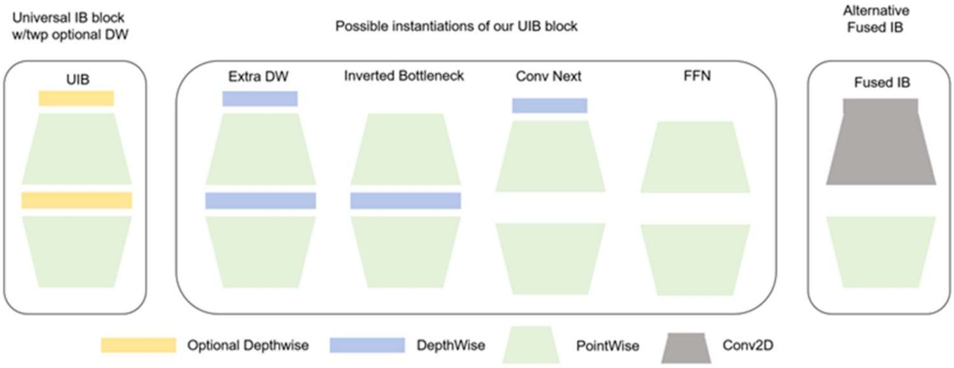

Figure 3. UIB Module

3. Result

Three models were trained for this experiment, all on a single Nvidia GeForce RTX 4060 Laptop GPU, YOLOv11, YOLOv11s and YOLOv11-MobilNetv4. The table below shows the performance of each model for different defect types (e.g., Crazing, Inclusion, Patches, Pitted Surface, Rolled-in Scale, Scratches) and mAP (mean of all types of defect accuracy).

The data in the table show that the three models have high recognition accuracies for patches and scratches, which are all over 0.9, but they are not good enough for crazing, which does not even reach the 0.5 level. This suggests that there are still challenges in detecting certain defect types.

In terms of overall performance, comparing the mAPs, we can see that the performance of YOLOv11-MobilNetv4 is relatively weak at 0.714. However, its model size is significantly lower than that of the other two models, which reflects its superiority in light weight. In addition, since YOLOv11-MobilNetv4 is equipped with fewer network layers, it has the fastest detection speed, which makes it a good potential for application in small devices or edge computing environments that require real-time processing.

Table 1. Performance of different models

	YOLOv11n	YOLOv11s	YOLOv11-MobilNetv4
crazing	0.421	0.467	0.352
inclusion	0.841	0.845	0.786
patches	0.928	0.940	0.915
pitted_surface	0.798	0.812	0.727
rolled-in_scale	0.627	0.626	0.600
mAP	0.762	0.773	0.714

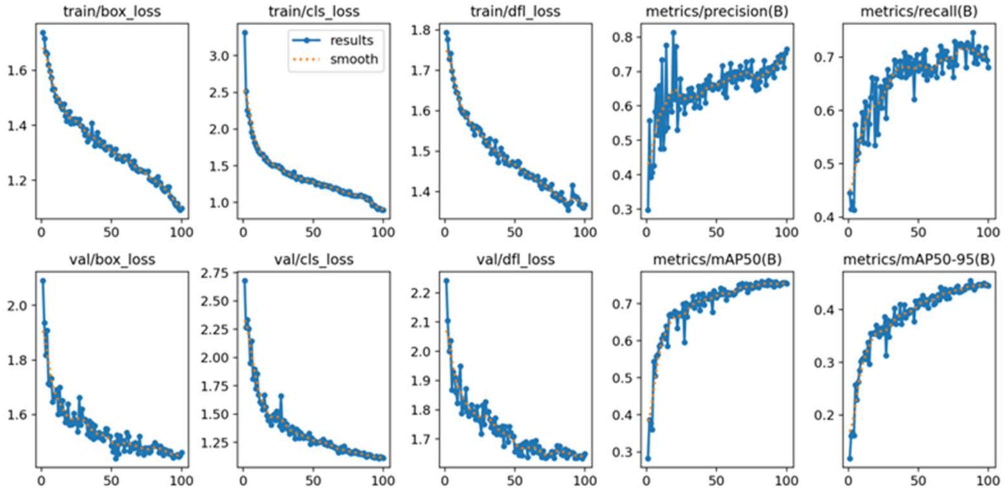

Figure 4. Loss values and mAP for YOLOv11n

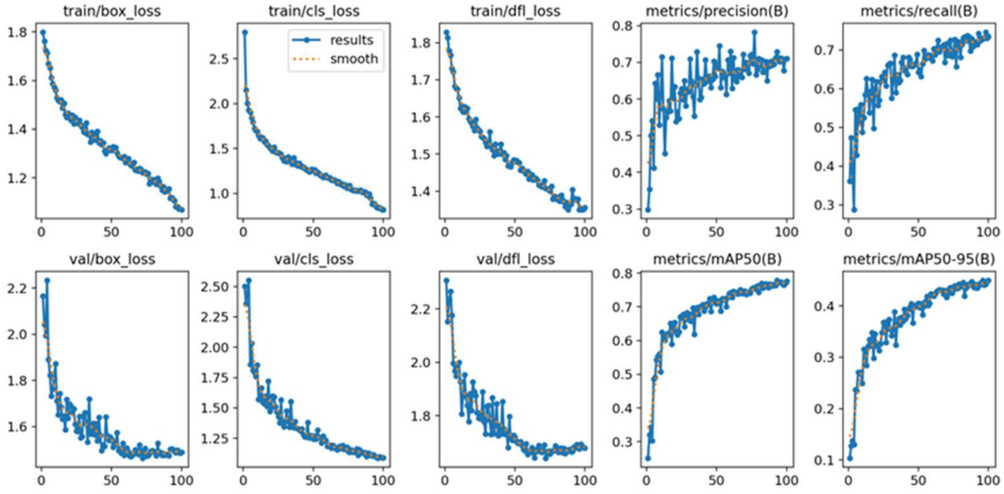

Figure 5. Loss values and mAP for YOLOv11s

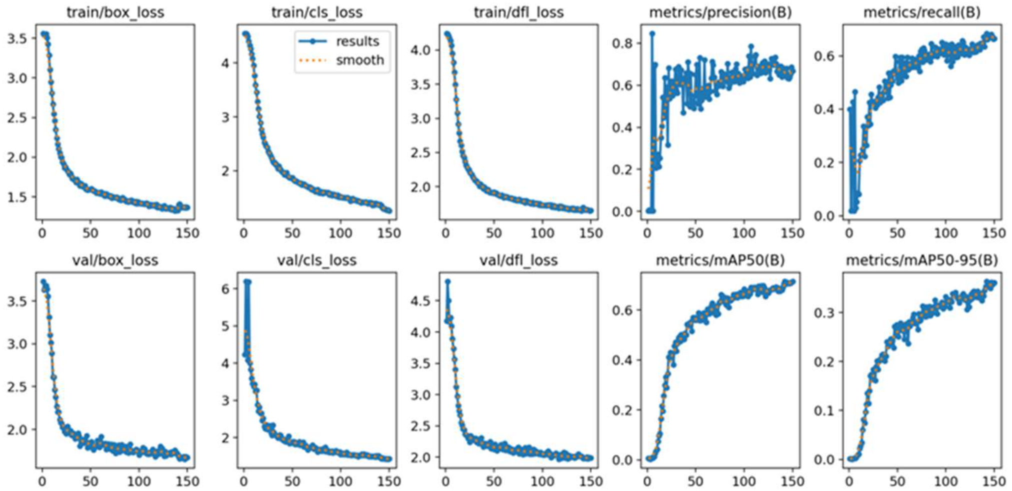

Figure 6. Loss values and mAP for YOLOv11-MobilNetv4

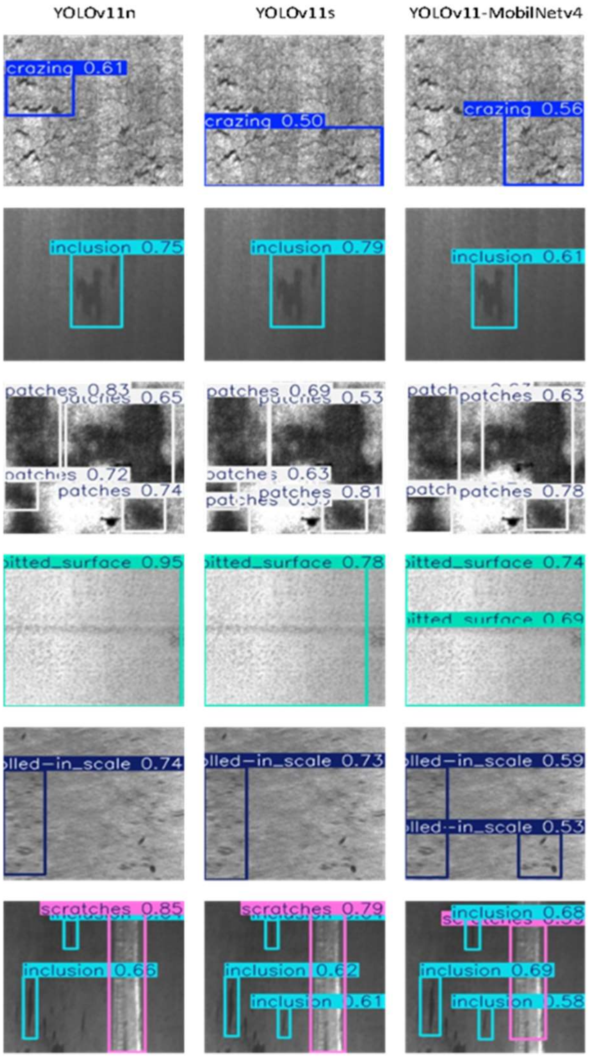

Figure 7. Detection results

We selected a number of images from the target detection dataset, and then completed the model training based on the pre-training model of YOLO v11. Figures 4 to 6 show the loss value curves and the mAP value curves of the three models, respectively. loss value is the loss of the model detection process, which is an important indicator used to optimize the performance of the model during the training process, and the smaller the loss value is, it usually implies that the model's performance is better. It can be observed that with the deep learning of the model, the loss value curves of the three models eventually decrease gradually and tend to be smooth and stable. At the same time, the mPA50 index of all three models eventually reaches more than 0.7, indicating that the three models can maintain a good level of overall performance. Figure 7 shows an example of the detection results, it can be seen that all three models can better recognize the cracks on the steel, and the model of YOLOv11-MobileNet v4 shows better competitiveness.

4. Conclusion

In this study, a lightweight model YOLOv11-MobilNetv4 is proposed for steel surface defect detection. experimental results show that the model has a comparable mAP of 0.714, which shows a significant advantage in terms of model size and computational complexity over YOLO v11s and YOLO v11n, making it particularly suitable for deployment in resource-limited mobile devices or edge computing environments. With its demonstrated efficiency and flexibility, it remains promising for a wide range of applications and shows great potential for future engineering inspections.

Acknowledgments

This work was funded by the following programs:

Innovation and Entrepreneurship Training Project of University of Science and Technology Liaoning (Research on Health Monitoring and Data Analysis of Campus Buildings Based on Multi-source Data);

The Basic Research Program for Universities of the Educational Department of Liaoning Province (Grant Number JYTQN2023241);

And the Outstanding Young Scientist Program of the University of Science and Technology Liaoning (Grant Number 2023YQ03).

References

[1] Feng Yingbin, Liu Wenze. Steel Surface Defect Detection Algorithm Based on HPDE-YOLO [J]. Journal of Shenyang Ligong University, 2025, 44(1): 31-38. (In Chinese)

[2] LIU Wenzhao, ZHANG Dan. Lightweight steel surface defect detection model based on machine vision[J/OL]. Computing Technology and Automation, 2024, 43(3): 43-49. DOI:10.16339/j.cnki. jsjsyzdh. 202403008. (In Chinese)

[3] Tao Yufeng, Jiang Lin, Da Mei. Algorithm for steel surface defect detection based on RCD-YOLO[J/OL]. Control Engineering of China: 1-9. DOI:10.14107/j.cnki.kzgc.20240426. (In Chinese)

[4] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-Based Learning Applied to Document Recognition[J/OL]. Proceedings of the IEEE, 1998, 86: 2278-2324. DOI:10.1109/5.726791.

[5] LECUN Y, BENGIO Y, HINTON G. Deep learning [J/OL]. Nature, 2015, 521(7553): 436-444. doi:10.1038/nature14539.

[6] HE Y, SONG K, MENG Q, et al. An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features[J/OL]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1493-1504. doi:10.1109/TIM.2019.2915404.

[7] ULTRALYTICS. yolo11 🚀 new [EB/OL]. [2025-01-10]. https://docs.ultralytics.com/zh/models/yolo11.

[8] QIN D, LEICHNER C, DELAKIS M, et al. MobileNetV4 - Universal Models for the Mobile Ecosystem [A/OL]. arXiv, 2024[2025-01-10]. http://arxiv.org/abs/2404.10518. DOI:10.48550/arXiv.2404.10518.