

Research on Improvements to Industrial Defect Detection Algorithms Based on YOLOv8

Xiaoxin Zheng¹, Weijie Song¹, Chaohui Guo¹, Haoyu Nie¹, Kaiming Liu¹

¹Artificial Intelligence Research Institute, Xiamen University

¹School of Infomatics, Xiamen University

36920241153286, 36920241153247, 36920241153206, 23320241154655, 23320241154654

Abstract

Industrial defect detection is a crucial component in achieving industrial automation. Single-stage object detection algorithms, represented by the YOLO series, have found significant applications in industrial inspection. To address the complexity and variability of defect detection tasks and further enhance detection accuracy and efficiency in complex industrial environments, we propose an improved industrial defect detection algorithm based on YOLOv8. Specifically, we introduce a Cross-Scale Partial Multi-Scale Feature Aggregation module (CSP-PMSFA). This module utilizes efficient Partial Convolution to extract multi-scale feature information from partial channels, enabling the fusion of features across different scales, thereby improving computational efficiency and model representation capabilities. Additionally, we propose a Lightweight Shared Convolution Detection Head (LSCD), which minimizes the number of parameters and computational cost of the detection head while maintaining detection precision. Finally, we incorporate an enhanced focal powerful Intersection over Union (F-PIoU) loss function to improve the accuracy and robustness of the model in detecting defects of varying scales and shapes, thereby enhancing its generalization ability. Experimental results demonstrate that the proposed defect detection algorithm performs well on multiple industrial defect detection datasets, achieving superior performance across several evaluation metrics compared to baseline model YOLOv8, thus offering improved detection capabilities.

Introduction

In the field of industrial automation, defect detection technology is key to ensuring product quality and safety. Traditional detection methods, such as manual visual inspection and basic image processing techniques, often fail to meet the high standards of modern industry due to low efficiency, insufficient accuracy, and lack of real-time performance. With the rise of deep learning technology, object detection algorithms based on deep learning have become the new standard for industrial defect detection, providing faster detection speed and higher accuracy (Zhang, Ding, and Yan 2011; Lee, Kim, and Kim 2019).

Currently, deep learning object detection techniques are mainly divided into two categories: single-stage and two-

stage algorithms. Two stage algorithms such as the R-CNN series, which first generate candidate regions and then classify and regress them, have high detection accuracy. However, the inference time is long and difficult to meet the real-time requirements of industrial production (Redmon et al. 2016). Single stage algorithms such as YOLO series directly perform object detection on images, with fast detection speed and high accuracy, and have been widely used in industrial detection (Redmon and Farhadi 2018a; Bochkovskiy, Wang, and Liao 2020; Redmon and Farhadi 2017). Although existing single-stage and two-stage object detection algorithms have achieved certain results in industrial defect detection, there are still some problems. For example, the complexity and inference time of two-stage algorithms limit their use in real-time applications, while the detection accuracy of single-stage algorithms in handling small targets and complex backgrounds still needs to be improved.

In this article, aiming at the visual inspection of industrial defect detection, an enhanced defect detection algorithm based on YOLOv8 is proposed. The algorithm leverages a Partial Multi-Scale Feature Aggregation mechanism, which applies efficient partial convolution to the input, enabling the extraction of multi-scale features. The extracted features are subsequently integrated with the original input through a 1x1 convolution and residual connections, effectively preserving the original feature information while incorporating new multi-scale information. This approach enhances the model's feature representation capabilities. Furthermore, a Lightweight Shared Convolutional Detection Head (LSCD) is introduced, significantly improving the detection head's performance in both localization and classification tasks. This innovation reduces the number of model parameters without compromising detection accuracy, making the model more suitable for deployment on edge devices. In addition, the algorithm integrates an improved Focal Powerful IoU (F-PIoU) loss function, which enables more accurate bounding box regression and accelerates model convergence, thereby improving the overall efficiency and effectiveness of defect detection.

The primary contributions of this paper are summarized as follows:

- A Partial Multi-Scale Feature Aggregation module (PMSFA) is proposed, which improves multi-scale feature extraction.

- A Lightweight Shared Convolutional Detection Head (LSCD) is employed, enhancing detection performance while reducing model complexity.
- The improved focal powerful IoU loss is put forward to tackle the performance bottlenecks from sample imbalance and inaccurate regression.

The rest of this article is organized as follows. The second part reviewed relevant industrial defect detection methods, and the third part elaborated on the proposed enhanced algorithm based on YOLOv8 in detail. The fourth part presents the experiment and discussion. Finally, the fifth section summarizes this article.

Related Work

The YOLO series has made significant advancements in industrial defect detection due to its balance between fast inference and high detection accuracy. The original YOLO established the foundation for real-time object detection, while YOLOv3 introduced multi-scale feature extraction to improve small defect detection. YOLOv4 enhanced feature representation by incorporating CSPDarknet53 as its backbone, effectively reducing computational overhead and improving robustness (Bochkovskiy, Wang, and Liao 2020; Redmon and Farhadi 2017, 2018b; Bochkovskiy 2020). These models have been widely applied to tasks such as fabric defect detection, PCB fault detection, and weld inspection (He 2020; Jiang 2019). Lightweight versions, such as YOLOv4-Tiny and YOLOv5-Tiny, were further developed to address resource constraints, ensuring high inference speed while maintaining acceptable detection accuracy (Zhang 2021). YOLOv5, with its modular design, facilitates customization and deployment for specific tasks, and introduces new modules such as Focus and Cross-Stage Partial Networks (CSPNet), further improving detection efficiency (Wang and Liao 2020; Wang 2020). Subsequent versions, including YOLOv6, YOLOv7, and YOLOv8, introduced architectural optimizations and advanced training strategies to further enhance detection efficiency and accuracy (Zheng 2022).

Defect detection in industrial applications often involves handling defects of varying scales, such as cracks, dents, and deformations. Multi-scale feature fusion techniques, including Feature Pyramid Networks (FPN) and Path Aggregation Networks (PANet), have been proven effective in improving detection across different defect sizes. FPN combines high-level semantic features with low-level spatial features, while PANet introduces path aggregation to facilitate efficient feature transmission (Lin 2017). In the YOLO framework, Spatial Pyramid Pooling (SPP) and Adaptive Spatial Feature Fusion (ASFF) enable effective multi-scale feature extraction, preserving spatial details and dynamically fusing features to enhance detection performance in complex scenarios (Liu 2018; He 2015; Liu 2019).

For real-time industrial defect detection, lightweight models are essential to meet the constraints of limited computational resources and high processing speeds. Architectures such as MobileNet, ShuffleNet, and GhostNet have been developed to address these issues. MobileNet utilizes depth-

wise separable convolutions to reduce parameters and computational costs, while ShuffleNet and GhostNet optimize feature extraction through channel shuffling and ghost modules, respectively (Howard 2017; Zhang 2018; Han 2020). These lightweight architectures have been integrated into YOLOv4-Tiny and YOLOv5-Tiny, enabling efficient defect detection in real-time industrial applications (Hong 2021; Ultralytics 2021, 2020).

Intersection over Union (IoU) is a core metric for evaluating and optimizing object detection models. However, traditional IoU has limitations when dealing with small or irregularly shaped defects. Several improvements have been proposed to address these issues. Generalized IoU (GIoU) introduces penalties for non-overlapping bounding boxes, helping the model refine predictions (Rezatofighi 2019). Distance IoU (DIoU) considers the distance between the center points of the predicted and ground truth boxes, accelerating convergence and improving bounding box localization accuracy (Zheng 2020). Complete IoU (CIoU) further enhances this by incorporating aspect ratio consistency, ensuring that the predicted boxes align with the target not only in position but also in shape (Wang 2021). These improved IoU metrics have been integrated into various defect detection systems, enhancing the localization of small and irregular defects, and making the models more robust in different industrial scenarios. In the context of defect detection, improved IoU metrics such as GIoU, DIoU, and CIoU are crucial for accurately localizing defects with blurred or irregular shapes, such as cracks, dents, and wear (Xu 2022).

Based on the above discussion, developing a comprehensive defect detection method remains a promising avenue of research. As discussed earlier, challenges such as handling varying defect scales in industrial detection, lightweight algorithm deployment, and improving IoU for better detection performance are critical tasks that need to be addressed. Therefore, the proposed YOLOv8-based defect detection algorithm focuses on three key issues: multi-scale feature processing, lightweight detection heads, and improved bounding box loss.

METHODOLOGY

This chapter introduces our approach to implementing industrial defect detection. We begin by analyzing the overall framework of the network model. Then, we provide a detailed explanation of the implementation process for the CSP-PMSFA and LSCD modules, as well as the mechanism of the F-PIoU loss function.

Overall Framework

The overall framework of the proposed network structure is illustrated in Figure 1. First, all images are resized to a resolution of 640×640, followed by data augmentation operations such as geometric transformations, color adjustments, noise addition, and morphological operations. Subsequently, the input images are downsampled twice using Conv_1 and Conv_2 operations, each consisting of a 3×3 convolution layer with batch normalization and SiLU activation. During the downsampling process, additional gradient

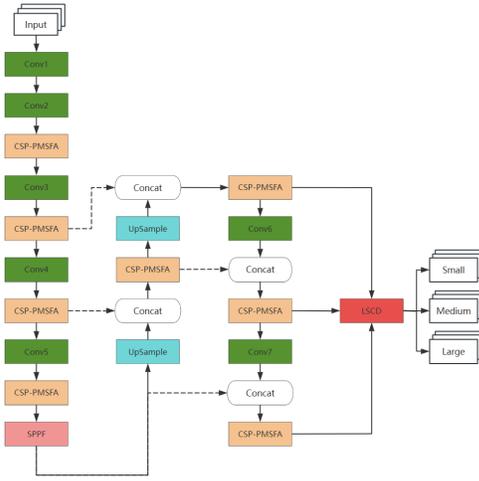


Figure 1: Overall Framework.

flow branches are implemented in parallel to enrich gradient information while maintaining a lightweight model. This approach aims to enhance the network’s contextual learning capability continuously.

In the improved algorithm, all C2f modules in the original network structure are replaced with CSP-PMSFA modules, which enhance the network’s ability for multi-scale feature extraction and feature fusion. This ultimately improves the model’s representation of multi-scale information.

Furthermore, the Spatial Pyramid Pooling-Fast (SPPF) module provides robust capability to capture contextual information, aiding in addressing scale variation challenges in defect detection (Zhu 2021).

Following this, multi-scale feature maps are integrated to fully leverage low-level information and consolidate both local and global features (Lin 2017). Finally, a lightweight shared convolutional detection head (LSCD) is employed to scale the feature maps for different target sizes, enabling the identification and localization of defects of varying dimensions. This approach minimizes parameter requirements while reducing accuracy loss.

Cross Stage Partial Networks-Multi-Scale Feature Aggregation(CSP-PMSFA)

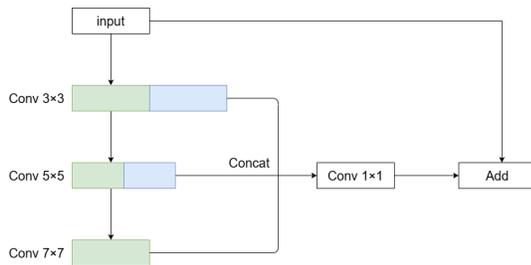


Figure 2: CSP-PMSFA Framework.

In the network, we replace the C2f module with the CSP-PMSFA module. The structure of the PMSFA module is

shown in Figure 2. Since the final feature map derived through downsampling contains multi-scale information but inevitably loses some information from the original feature map, we propose the CSP-PMSFA module to preserve some of the original information from each scale’s feature map.

The PMSFA module retains part of the feature map information at each scale after convolution, while the remaining part continues with the convolution operations. Finally, the multi-scale feature maps are fused, followed by a 1×1 convolution and a residual connection with the original input to produce the output. The CSP-PMSFA module further extends the PMSFA module by adopting a cross-stage stacking approach. By stacking multiple PMSFA modules, the network can extract more complex and higher-level features, thereby enhancing the model’s representational capacity.

The CSP-PMSFA module integrates the design philosophy of CSPNet (Cross Stage Partial Network). Through cross-stage connections, the network maintains a strong flow of information between different stages, enhancing the model’s depth while optimizing feature reuse. This design also alleviates the vanishing gradient problem by preserving short paths between the early and later layers via residual connections, thereby facilitating the training of deeper networks.

Moreover, CSP-PMSFA considers computational efficiency by employing grouped convolutions to reduce computational and parameter overhead. Grouped convolutions are particularly effective for improving efficiency when processing larger inputs. By progressively reducing the number of channels, the model focuses on extracting more complex features while maintaining low computational costs.

Lightweight Shared Convolutional Detection Head(LSCD)

We replace the detection head in the baseline model with the LSCD detection head. The structure of the LSCD module is shown in Figure 3. In the LSCD module, all Batch Normalization layers are replaced with Group Normalization layers, as demonstrated in the FCOS (Tian et al. 2019) to enhance localization and classification performance. During the process of lightweight optimization, reducing the number of parameters can lead to a decline in feature extraction performance. However, the use of Group Normalization (GN) mitigates this issue by minimizing accuracy loss while maintaining a reduced parameter count.

Specifically, the downsampled feature maps p3, p4, and p5 do not share parameters during GN operations but do share parameters afterward. For detection and classification tasks, the targets belong to the same category, differing only in scale. Parameter sharing leverages feature maps from different scales to generate prediction results, enabling multi-scale object detection. Parameter sharing is achieved by applying the same convolution operations and weights to input feature maps of different scales using a shared convolution layer.

While using shared convolutions, to address the inconsistency in target scales detected by each detection head, the Scale layer is employed to adjust the features of different scales by learning scaling factors. The parameters of

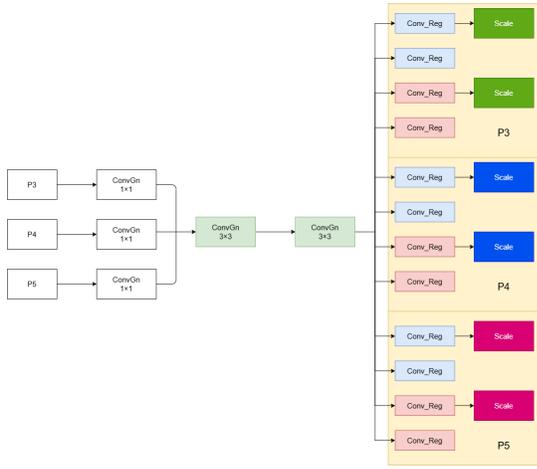


Figure 3: LSCD Framework.

the Scale layer participate in predictions alongside boundary box predictions, class scores, and other outputs of feature maps at various scales. For instance, if the targets in a particular scale are smaller, the Scale factor may increase the output of that scale, enhancing the model’s ability to detect small objects. Conversely, for larger targets, the output of that scale may be reduced.

Focaler-Powerful-IoU Loss

The F-PIoU introduces a penalty factor adaptive to the target size and a gradient adjustment function based on the quality of the anchor box (Liu 2024). It no longer relies on the diagonal length of the smallest enclosing rectangle but instead directly calculates based on the dimensions of the target box. This approach ensures that the gradient does not degenerate to zero and provides a reasonable penalty for anchor and target boxes of different relative sizes, effectively guiding the anchor box to regress along an efficient path. The mathematical expression is as follows:

$$P = \left[\frac{dw_1 + dw_2}{w_{target}} + \frac{dh_1 + dh_2}{h_{target}} \right] / 4 \quad (1)$$

$$PIoU = 1 - (1 - IoU) - \left(1 - e^{-P^2} \right) \quad (2)$$

Where, dw_1, dw_2, dh_1, dh_2 represent the absolute values of the distances between the corresponding edges of the predicted bounding box and the target bounding box, h_{target} and w_{target} denote the width and height of the target bounding box.

Furthermore, the F-PIoU integrates a non-monotonic focusing mechanism that adjusts the gradient magnitude through a specific functional form to accommodate anchors of varying qualities, which can be described as follows:

$$q = e^{-P} \quad (3)$$

$$u(x) = 3x \cdot e^{-x^2} \quad (4)$$

$$L_{F-PIoU} = u(\Lambda q) \cdot (1 - PIoU) \quad (5)$$

Here, $u(x)$ is a non-monotonic function. q is a function based on the differences between the predicted bounding box and the target bounding box, which measures the quality of the anchor box, and q ranges from $(0, 1]$. $u(\Lambda q)$ represents the attention function. Λ is the hyperparameter that controls the behavior of the attention function

Experimental Verification

In this section, we describe the experimental setup and evaluate the performance of the proposed network using two publicly available defect detection datasets.

Datasets Description

In this study, we have meticulously considered data analysis and utilized two datasets to assess the proposed model, including NEU-DET and GC10-DET

NEU-DET dataset		GC10-DET dataset	
Category	Amount	Category	Amount
/	0	Punching_hole	329
/	0	welding_line	513
Rolled_scale	300	Crescent_gap	265
Patches	300	Water_spot	354
Crazing	300	Oil_spot	569
Pitted_surface	300	Silk_spot	884
Inclusions	300	Inclusion	347
Scratches	300	Rolled_pit	85
/	0	Crease	74
/	0	Waist_folding	150
Total	1800	Total	3570

Table 1: Description of The Number of Categories in GC10-DET and NEU-DET Datasets

1. NEU-DET: As shown in Table 1, the NEU-DET dataset comprises six common surface defects found on hot-rolled steel strips, which contains a total of 1800 grayscale images, with 300 samples for each defect type, saved in grayscale at an original resolution of 200*200 pixels.
2. GC10-DET: As shown in Table 1, the GC10-DET dataset comprises 10 surface defect types, which features 3570 grayscale images of steel plate surface defects from real industrial settings. To verify the effectiveness of our method, we divided the NEU-DET and GC10-DET datasets into training and testing sets in a 9:1 ratio.

Experimental Environment

1. Experimental Settings: During the model training process, we maintained consistency in input image size, data augmentation methods, learning rate, and batch size. We conducted experiments on the NEU-DET and GC10-DET defect detection datasets, using 300 epochs and a batch size of 16 for both datasets. All experiments were conducted under the PyTorch deep learning framework on an NVIDIA RTX 4080 GPU.

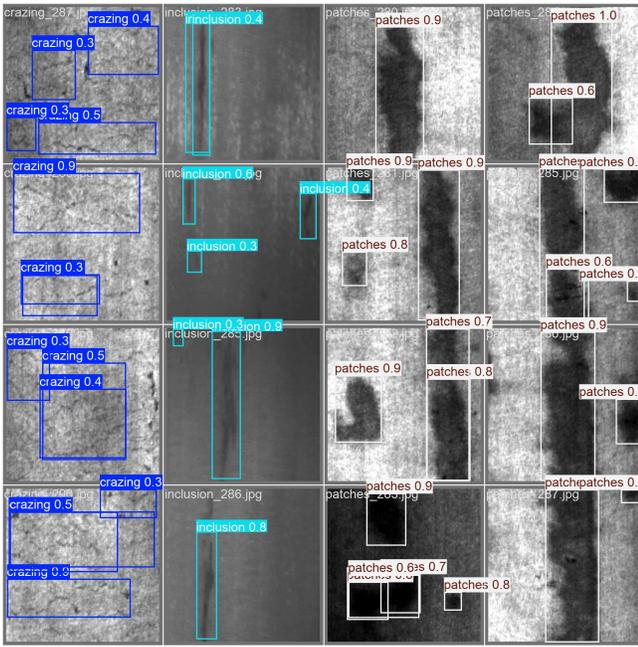


Figure 4: Improved Network Predicts Results 1 on NEU-DET.

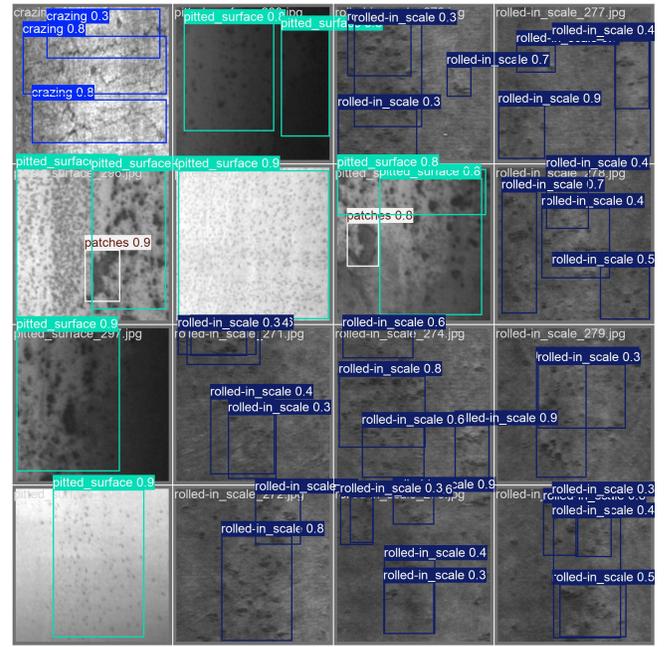


Figure 5: Improved Network Predicts Results 2 on NEU-DET.

2. Evaluation Metric: To comprehensively evaluate the performance of the proposed modal, we employ several commonly used metrics for performance assessment, including Precision, Recall, mAP50, mAP50:95, where Precision gauges the model’s accuracy in identifying true positive instances among all positive predictions, while Recall indicates its ability to capture all relevant positive samples within the dataset and AP denotes Average Precision. mAP50 represents AP over IoU at 0.5, and mAP50:95 represents AP over IoU from 0.5 to 0.95 with an interval of 0.05.

Performance Evaluation

We compared our improved network with the baseline model YOLOv8 on two datasets. The comparison results are shown in the Table 2. Our network outperforms the baseline model in terms of precision, mAP50, and mAP50:95. On the NEU-DET dataset, these metrics improved by 0.07, 0.006, and 0.015, respectively, while on the GC10-DET dataset, the improvements were 0.016, 0.017, and 0.022, respectively. However, our network showed a decrease in recall, with reductions of 0.014 and 0.022 compared to the baseline model on the NEU-DET and GC10-DET datasets, respectively.

Data/Metrics		Precision	Recall	mAP50	mAP50:95
NEU-DET	base	0.712	0.713	0.755	0.387
	ours	0.782	0.699	0.761	0.402
GC10-DET	base	0.399	0.414	0.379	0.167
	ours	0.414	0.392	0.396	0.189

Table 2: Baseline Model Comparison

From a visual perspective, Figure 4 and 5 illustrate the comparative experimental results of the proposed improved network and the baseline model on the NEU-DET and GC10-DET datasets. Overall, our improved network demonstrates enhanced performance.

Ablation Experiments

In the rigorous process of validating the effectiveness of the proposed improved network core components, we conducted an ablation study, using the NEU-DET dataset as the benchmark for our experiments. The results are recorded in the Table 3. The network, integrated with CSP-PMSFA, LSCD, and FPIoU, showed performance improvements. This highlights the network’s ability to accurately and comprehensively detect defects within a certain threshold range, particularly in multi-scale object detection. However, the network’s evaluation on the recall metric showed a continuous decline. To address this issue, we analyzed the following aspects:

1. The CSP-PMSFA module, by progressively fusing features through convolutions of different scales, may lead to the loss of certain low-level information in the high-level feature maps, thereby affecting the detection accuracy of small-scale objects.
2. Replacing the original detection head with the LSCD detection head reduced the size of the detection head, but this also somewhat decreased the detection accuracy, especially for defects in small-scale or complex background regions, ultimately leading to a decline in recall.

Method	Precision	Recall	mAP50	mAP50:95
base	0.712	0.713	0.755	0.387
base+PMSFA	0.756	0.704	0.764	0.399
base+LSCD	0.700	0.701	0.743	0.372
base+FPiOU	0.711	0.698	0.761	0.397
base+PMSFA+LSCD	0.730	0.674	0.742	0.377
ours	0.782	0.699	0.761	0.402

Table 3: Results of ablation experiments

Conclusion

This paper proposes an improved industrial defect detection algorithm based on YOLOv8. The algorithm achieves a remarkable balance between detection accuracy and model compactness, meeting the practical engineering requirements of industrial detection. The key components of the network include Cross Stage Partial Networks-Multi-Scale Feature Aggregation (CSP-PMSFA), a Lightweight Shared Convolutional Detection Head (LSCD), and an improved loss function, F-PiOU. Compared to the baseline model, the improved network demonstrates outstanding performance in terms of accuracy and efficiency. Furthermore, ablation studies were conducted to validate the effectiveness of the network components. However, the improved network still suffers from a relatively low recall, highlighting the need for further enhancements in its capability to detect small-scale objects.

References

Bochkovskiy, A.; Wang, C. Y.; and Liao, H. Y. M. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. .

Bochkovskiy, A. e. a. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. .

Han, K. e. a. 2020. GhostNet: More Features from Cheap Operations. In *CVPR*.

He, H. e. a. 2020. Fabric Defect Detection Based on YOLOv4 Network. *IEEE Access*.

He, K. e. a. 2015. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9): 1904–1916.

Hong, S. e. a. 2021. Real-Time Defect Detection for Fabrics Using MobileNet-Based Deep Learning. *Sensors*.

Howard, A. G. e. a. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. .

Jiang, K. e. a. 2019. Fault Detection of PCB Based on YOLOv3. *Electronics*.

Lee, J.; Kim, J.; and Kim, H. 2019. A Survey of Deep Learning-based Object Detection Methods for Industrial Applications. *Journal of Industrial Information Integration*, 13: 102–113.

Lin, T. e. a. 2017. Feature Pyramid Networks for Object Detection. In *CVPR*, 936–944.

Liu, C. e. a. 2019. Adaptive Spatial Feature Fusion for Object Detection. In *CVPR*.

Liu, K. e. a. 2024. Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism. *Neural Networks*, 170: 276–284.

Liu, S. e. a. 2018. Path Aggregation Network for Instance Segmentation. In *CVPR*.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.

Redmon, J.; and Farhadi, A. 2017. YOLO9000: Better, Faster, Stronger. In *CVPR*.

Redmon, J.; and Farhadi, A. 2018a. YOLOv3: An Incremental Improvement. .

Redmon, J.; and Farhadi, A. 2018b. YOLOv3: An Incremental Improvement. .

Rezatofghi, H. e. a. 2019. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In *CVPR*.

Tian, Z.; Shen, C.; Chen, H.; and He, T. 2019. FCOS: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9626–9635.

Ultralytics. 2020. YOLOv5-Tiny: Optimizing Speed and Accuracy for Industrial Use. .

Ultralytics. 2021. YOLOv4-Tiny: A Real-Time Object Detector for Edge Devices. .

Wang, C.; and Liao, H. Y. M. 2020. CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. In *CVPR Workshops*.

Wang, C. e. a. 2021. Complete IoU Loss: A Metric and Loss for Bounding Box Regression. In *CVPR*.

Wang, Y. e. a. 2020. YOLOv5: A Scalable Object Detection Model for Industrial Applications. .

Xu, H. e. a. 2022. Advanced Boundary Detection for Small and Irregular Defects Using CIOU. *Electronics*.

Zhang, J.; Ding, Y.; and Yan, H. 2011. A Survey of Image Defect Detection Methods. *Journal of Manufacturing Science and Engineering*, 133(6): 061011.

Zhang, X. e. a. 2018. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In *CVPR*.

Zhang, X. e. a. 2021. Weld Defect Detection Based on Improved YOLOv4. *IEEE Access*.

Zheng, C. e. a. 2022. YOLOv6: A Faster and More Accurate Object Detector. .

Zheng, Z. e. a. 2020. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. .

Zhu, X. e. a. 2021. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2778–2788.