International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 6, December, 2024, pp. 3193-3201
Journal homepage: http://iieta.org/journals/ts

Image Recognition Technology in Supply Chain Management Based on Deep Learning

Jikai Huang

Beijing Huayu Engineering Co., Ltd., Beijing 100120, China

Corresponding Author Email: HJK03341@163.com

https://doi.org/10.18280/ts.410635

Received: 9 May 2024
Revised: 28 October 2024
Accepted: 5 November 2024
Available online: 31 December 2024

Keywords:
supply chain management, quality control, deep learning, image recognition, lightweight model, defect detection

ABSTRACT

With the rapid development of globalization and intelligent manufacturing, quality control issues in supply chain management have become increasingly prominent. In particular, how to efficiently and accurately detect product defects has become a key factor in improving product quality and production efficiency in manufacturing. Traditional manual inspection methods face challenges such as low accuracy, inefficiency, and high costs. Deep learning, as a powerful tool for automated inspection, has gradually become an important solution due to its advantages in image recognition. However, existing deep learning-based image recognition methods still have limitations in supply chain management applications, such as high computational resource demands and a lack of adaptability for lightweight products. These issues hinder the widespread adoption and application in real-world production environments. This paper aims to address these problems by proposing a lightweight product defect detection approach for supply chain management quality control and a corresponding deep learning model structure. First, the defect detection approach based on deep learning ensures efficient defect recognition while reducing the computational resource requirements. Secondly, a lightweight model structure is designed to optimize algorithm performance, making it suitable for quality control in real-time production environments. This research is expected to provide an efficient and practical technological solution for quality control in supply chain management and promote the application of deep learning technology in the industrial sector.

1. INTRODUCTION

With the advancement of globalization and digitalization, supply chain management plays an increasingly important role in the operation of modern enterprises [1-5]. In particular, in the manufacturing industry, product quality control directly affects a company's market competitiveness and brand reputation. Traditional quality control methods often rely on manual inspection and experience-based judgment, which face issues such as low efficiency, poor accuracy, and high costs [6-9]. In recent years, with the rapid development of deep learning technology, the application of image recognition in automated quality inspection has gradually become a new trend. By analyzing product images through deep learning models, defects can be identified efficiently and accurately, enabling real-time monitoring, which offers great potential for quality control in supply chain management.

In supply chain management, timely detection and handling of product defects is a key factor in ensuring quality and production efficiency. As production scale expands and product varieties increase, how to achieve precise quality control while ensuring efficient production has become a significant challenge for enterprises. Deep learning-based image recognition technology, as an innovative solution, can effectively improve the accuracy and efficiency of product defect detection, reduce the need for human intervention, and be widely applied in various production environments [10-14]. Therefore, researching how to combine deep learning with quality control in supply chain management and exploring new technological paths has important theoretical and practical significance for enhancing the competitiveness of enterprises.

However, there are still some shortcomings in the application of existing deep learning-based image recognition technology in supply chain management. First, traditional deep learning models often require large computational resources, which may not be ideal under the real-time requirements of production sites [15-18]. Second, most existing models are designed for large-scale and complex defect detection problems and lack adaptability for lightweight products, resulting in detection accuracy and efficiency that cannot meet specific needs [19-23]. In addition, many existing methods are overly complex in their model design, making it difficult to deploy them widely in supply chain management. Given these issues, designing a deep learning model that is both efficient and lightweight is a key problem that needs to be solved.

This paper mainly focuses on quality control in supply chain management and proposes two key research contents. First, the paper discusses an approach for lightweight product defect detection, aiming to optimize deep learning models, reduce computational resource requirements, and improve detection efficiency. Second, the paper designs a lightweight defect detection model structure for supply chain management quality control, aiming to improve algorithms and architectures to make them more suitable for practical production applications. Through these two studies, it is expected to provide an efficient, practical, and sustainable technological solution for quality control in supply chain management, promote the widespread application of deep learning in the industrial sector, and enhance enterprise production efficiency and product quality.

2. LIGHTWEIGHT PRODUCT DEFECT DETECTION APPROACH FOR SUPPLY CHAIN MANAGEMENT QUALITY CONTROL

In supply chain management, especially in the manufacturing field, product quality control is a key link in ensuring product qualification rates and production efficiency. However, traditional manual inspection and rule-based automated inspection methods often face issues such as high costs, low efficiency, and difficulty adapting to complex production environments. With breakthroughs in deep learning technology in the field of image recognition, its application in defect detection has gradually become an efficient solution. However, existing deep learning models usually require large computational resources and complex hardware support, which is not feasible in many production sites, especially when high real-time and efficiency requirements are needed for quality control. To address this issue, the lightweight deep learning model proposed in this paper aims to reduce the computational complexity and resource consumption of the model while ensuring high accuracy, making it more widely applicable for quality control in supply chain management.

Traditional deep learning models, especially deep neural networks for image recognition tasks, typically have high computational complexity and large parameter sizes. This often leads to slow processing speeds and high hardware resource consumption in large-scale production scenarios, thereby affecting overall production efficiency. Particularly in supply chain management, rapid defect detection and real-time feedback are crucial, as any delay can lead to production line stoppage or the expansion of quality problems. To meet the strict requirements for efficiency, real-time performance, and resource consumption in actual production environments, this paper focuses on designing a lightweight product defect detection model for supply chain management quality control. By optimizing the model structure and reducing computation and memory consumption, the lightweight deep learning model can significantly improve processing speed, reduce dependence on high-performance hardware, and better meet the practical demands of production environments.

2.1 Feature extraction based on spatial pyramid pooling (SPP)

For product defect detection tasks in supply chain management, especially when dealing with defects of various shapes and sizes, traditional convolutional neural networks (CNN) often struggle to fully extract features at different scales. To optimize the feature extraction process, reduce computational workload, and improve model inference speed, this paper employs the module in the lightweight product defect detection model for supply chain management quality control, replacing the CSP1_X in the backbone network. SPP allows for feature extraction from both small and large defects by using pooling operations at different scales, making the model more robust when handling various types of defects. Additionally, this module reduces reliance on high-level convolutional features, lowering the model's computational complexity and thus reducing inference time, which is critical for real-time monitoring and rapid feedback in large-scale production environments.

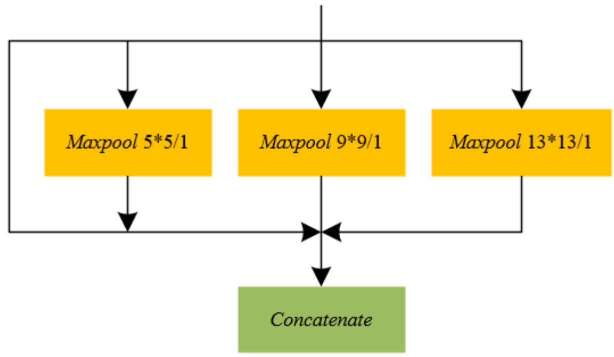

Figure 1. SPP module network structure

Figure 1 shows the network structure of the SPP module. The basic principle of the SPP module is to extract feature information at different scales through multi-scale pooling operations, thereby improving the accuracy of defect detection and simplifying the network structure. Unlike traditional single sliding window pooling methods, the SPP uses multiple pooling kernels of different sizes, such as 5×5, 9×9, and 13×13, for max pooling operations. This multi-scale design allows for a more comprehensive capture of defect features at different sizes within the image. By pooling different scaled regions of the input feature map and performing corresponding padding operations according to the kernel size, the pooled feature map maintains the same size as the original input feature map, avoiding issues with feature map size mismatches. After the fusion of these intermediate feature maps with the original feature maps, multi-scale local and global feature information is retained, which helps improve the accuracy of product defect recognition, particularly when facing defects of varying shapes and sizes, effectively enhancing the model's ability to recognize different types of defects.

In the lightweight product defect detection model for supply chain management quality control, the above-mentioned SPP module offers two significant advantages. On one hand, SPP provides multiple levels of receptive fields, which can effectively enhance the recognition of small targets and defects that are similar to the background. In product defect detection, defect points often resemble the surrounding background or lines, making detection difficult. Through multi-scale pooling, SPP can capture feature information at different scales, enabling more accurate identification of these subtle defects that are similar to the background. This multi-scale feature fusion helps improve the robustness and accuracy of the model in complex environments. On the other hand, the SPP module, while maintaining detection accuracy, adopts a lighter network structure, significantly reducing computational load. In practical production environments for supply chain management, product defect detection models must not only achieve high accuracy but also meet real-time and efficient requirements. SPP simplifies the original network structure and reduces reliance on deeper feature extraction layers, enabling the model to perform efficient inference with lower computational resources, ensuring quick response and low latency. These two advantages make SPP an ideal choice for lightweight product defect detection models aimed at supply chain management quality control, as it improves detection accuracy while optimizing the model's operational efficiency, meeting the dual requirements for speed and accuracy in industrial production. Let the feature output of the SPP network layer be denoted by a_{OP}, and the concatenation operation be denoted by *CAT*. The fusion process of the SPP network is given by the following equation:

$$a_{OP} = CAT\begin{pmatrix} a, POOL_{5\times5}(a), \\ POOL_{9\times9}(a), POOL_{13\times13}(a) \end{pmatrix} \tag{1}$$

2.2 Introduction of the CBAM attention mechanism

Product defects typically occupy small areas in images and are easily confused with the surrounding background, especially in complex production environments. Capturing these subtle defects accurately is a significant challenge. To enhance the model's attention to key defect regions, this paper introduces the Convolutional Block Attention Module (CBAM) attention mechanism in the lightweight product defect detection model for supply chain management quality control. CBAM optimizes feature maps adaptively by introducing attention mechanisms in both the channel and spatial dimensions, focusing on the most distinguishable parts of the image. In the channel domain, CBAM assigns larger weights to more relevant feature channels, thereby enhancing the feature information related to defects. In the spatial domain, CBAM further emphasizes the features of the defect area and weakens background noise, ensuring that the model can accurately detect small and subtle defects. Figure 2 shows the structure diagram of the CBAM attention mechanism.

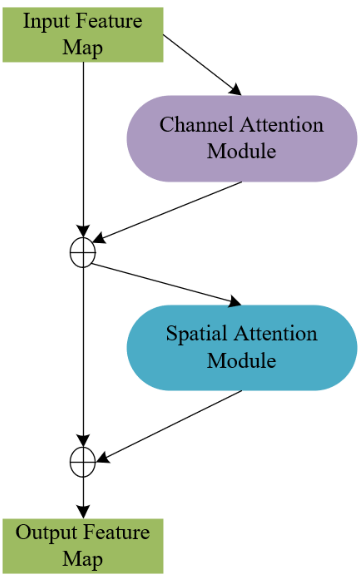

Figure 2. CBAM attention mechanism structure

The core principle of the channel attention module in the CBAM attention mechanism is to calculate attention weights for each feature channel, enhancing the channels related to defects while suppressing less important channel information, thus optimizing the feature map representation. Figure 3 shows the structure of the channel attention module in the CBAM attention mechanism. During the workflow, the channel attention module first applies max pooling and average pooling operations on the input feature map to obtain the maximum and average values for each channel, then computes the mean and standard deviation for each channel through normalization to generate a set of statistical information for each channel. These statistical details are then processed by a multi-layer perceptron (MLP) to further improve feature extraction efficiency and generate the fused feature map. Afterward, the module computes the attention weights for each channel using the sigmoid activation function. These weights are multiplied with the original feature map to enhance important channels and suppress irrelevant channels. This process ensures that the model can focus on the channels that best represent defect features, particularly when dealing with subtle defects that are similar to the background, improving detection accuracy. Additionally, CBAM reduces computational complexity by compressing the spatial dimensions of the input feature map, ensuring the efficiency of the channel attention module and avoiding excessive computational burdens during large-scale data processing. Let the channel attention module be represented by L_z, the input feature information be denoted by D, and the sigmoid function be denoted by δ. The MLP weights Q_0 and Q_1 are shared, and the calculation process is given by the following equation:

$$\begin{aligned} L_Z(D) &= \delta\begin{pmatrix} MLP(AVPool(D)) \\ +MLP(MAXPool(D)) \end{pmatrix} \\ &= \delta\begin{pmatrix} Q_1\left(Q_0\left(D^z_{AVG}\right)\right) \\ +Q_1\left(Q_0\left(D^z_{MAX}\right)\right) \end{pmatrix} \end{aligned} \tag{2}$$

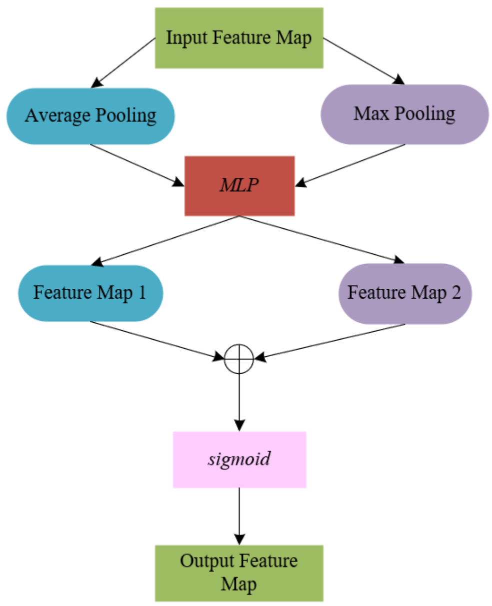

Figure 3. Structure of the channel attention module in CBAM attention mechanism

In the model, the spatial attention module of the CBAM attention mechanism aims to assess the importance of various spatial locations in the feature map, further enhancing the representation of defect regions. Figure 4 illustrates the structure of the spatial attention module in the CBAM attention mechanism. In the workflow, the spatial attention module first performs global max pooling and average pooling on the input feature map, obtaining the maximum and average values for each spatial location, respectively, thus reducing the spatial dimensions of the feature map and extracting overall spatial features. Then, the pooled feature maps are concatenated and passed through a convolution layer to compute the attention weights for each spatial location. These spatial attention weights reflect the importance of various locations in the feature map, particularly in distinguishing between defect and background regions. Using the sigmoid activation function, the generated spatial attention weights are element-wise multiplied with the original feature map, enhancing defect-related regions and suppressing irrelevant regions, allowing the model to focus on key areas of defects and improve detection accuracy. Let the spatial attention operation be represented by L_T, and the calculation process for the spatial attention module is given by:

$$\begin{aligned} L_T(D) &= \delta\left(d^{7\times7}\left(\left[AVPool(D);MAXPool(D)\right]\right)\right) \\ &= \delta\left(d^{7\times7}\left(\left[D^t_{AVG};D^t_{MAX}\right]\right)\right) \end{aligned} \tag{3}$$

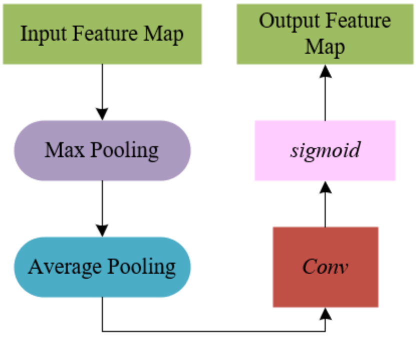

Figure 4. Structure of the spatial attention module in CBAM attention mechanism

The spatial attention module combined with the channel attention module sequentially optimizes both channel and spatial information in the feature map, helping the model more effectively detect small defects in images. This combination is particularly useful for supply chain management, as it enables more precise identification of defects that are complex in shape and resemble the background, significantly improving quality control detection accuracy.

2.3 Introduction of residual networks to facilitate feature fusion

Although SPP effectively expands the receptive field and improves detection capability for small targets, it can weaken the feature extraction ability of the model, especially when dealing with medium to large targets. To address this issue, this paper introduces the idea of residual networks and adds weighted branches to the shallow convolution features. Specifically, weighted branches are added before the 38×38 and 19×19 output layers in the Neck module, aiming to extract low-level features from the shallow convolution layers of the backbone network and fuse them with higher-level features. Let the weighted feature output be denoted by ADD_{OP}, and the feature fusion network before the introduction of the branches be denoted by $D(a)$. The fusion formula is as follows:

$$ADD_{OP} = D(a)\cdot J + a\cdot(1-J) \tag{4}$$

$$Q_{IN} = T \tag{5}$$

$$Q_{LI}\times T = H \tag{6}$$

$$J = \frac{1}{1+e^{-H}} \tag{7}$$

By fusing shallow and deep features and reconstructing the features using the CSP2_1 structure, the model can still achieve near or even superior detection accuracy compared to the original network, even with weaker feature extraction capabilities. This feature fusion strategy effectively compensates for the detection deficiencies of traditional methods when faced with complex defect images, allowing the model to provide more accurate detection results for defects of various sizes and types.

3. LIGHTWEIGHT PRODUCT DEFECT DETECTION MODEL FOR SUPPLY CHAIN MANAGEMENT QUALITY CONTROL

The proposed lightweight product defect detection model builds upon the YOLOv5s model framework. The structure of this model mainly consists of an input layer, backbone network, neck network, and output layer. Considering the quality control requirements in the supply chain, especially when processing product images with small defects, this paper replaces the original CSP1_X module in the backbone network with the SPP module. This design enhances the receptive field through multi-scale pooling operations, enabling the model to better capture defects of various scales, particularly small defects, and extract more comprehensive features. Furthermore, the introduction of the SPP module not only strengthens feature extraction capabilities but also simplifies the network structure, thereby reducing computational load and meeting the requirements of a lightweight model. Figure 5 shows the network structure of the lightweight product defect detection model for supply chain management quality control.

To further enhance the detection accuracy of small target defects, this paper introduces the CBAM attention mechanism in front of the 76×76 prediction channel. By weighting the channel and spatial dimensions of the feature map, CBAM effectively strengthens the features related to defects while suppressing irrelevant or background noise interference. This is especially beneficial for product quality control, as it helps the model more accurately detect small and subtle defects that are difficult to notice. This strategy not only improves detection accuracy but also ensures computational efficiency, avoiding excessive complex calculations. Additionally, weighted residual branches from the backbone network are introduced before the 38×38 and 19×19 prediction channels, effectively promoting multi-level feature fusion. With the design of the residual structure, the model is able to establish deeper connections between features at different levels, enhancing the network's learning capability and improving object detection accuracy. Finally, feature fusion is carried out through Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN), followed by concatenation and convolution operations to obtain the final output feature map. This ensures that the model can effectively recognize and detect product defects in the supply chain management process at multiple levels and scales, providing more precise quality control results.

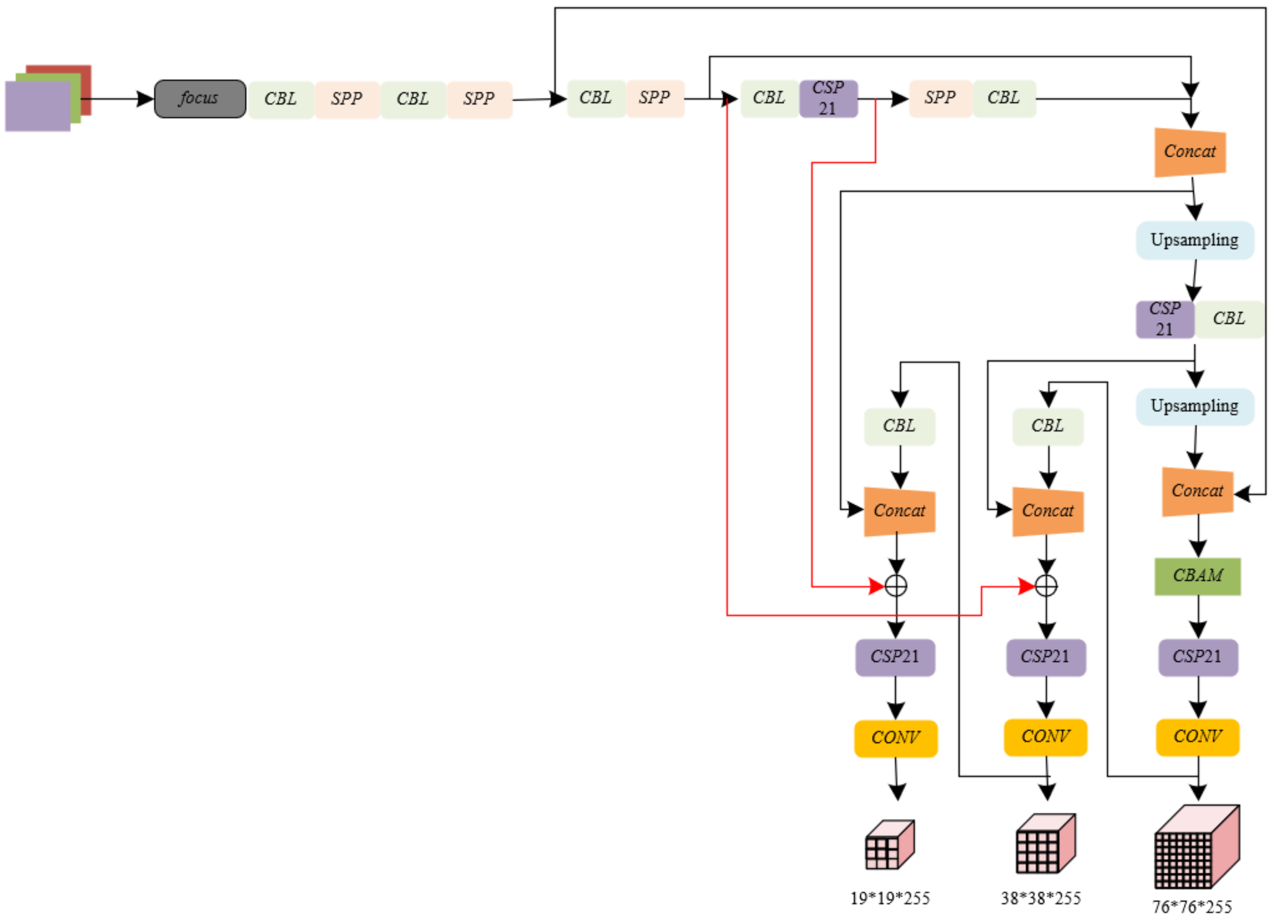

Figure 5. Lightweight product defect detection model network structure for supply chain management quality control

For the loss function, this paper adopts the original YOLOv5s loss structure, which includes classification loss, bounding box regression loss, and confidence loss. The classification loss is computed using binary cross-entropy, which is effective for classification tasks. For each class in the product defect images, the classification loss calculates the difference between the predicted result and the true label using the binary cross-entropy function, thus helping the model learn to assign defects to the correct categories more accurately. As defect types can vary greatly, optimizing the classification loss helps improve classification accuracy, ensuring that different defects can be quickly and accurately identified in product quality control. Let the weighting coefficients be denoted by x, y, and z. The overall YOLOv5s loss is defined as:

$$LOSS = x * LOSS_{FL} + y * LOSS_{HG} + z * LOSS_{ZX} \tag{8}$$

Assume that the number of categories in the dataset is represented by V, the probability of the predicted category after activation is denoted by e_u, and the true value of the predicted category is represented by e^*_u. The classification loss is defined as follows:

$$LOSS_{FL} = -\sum_{v=1}^{V} e_u^* \log(e_u) + (1 - e_u *) \log(1 - e) \tag{9}$$

The bounding box regression loss uses the Generalized Intersection over Union (GIoU) loss. Compared to traditional Intersection over Union (IoU) loss, GIoU loss is more robust. In product defect detection within supply chain management, the targets may appear in various shapes and sizes, and particularly when the defects are small or in complex backgrounds, traditional IoU loss might not accurately reflect the distance between the predicted and true bounding boxes. GIoU loss addresses this by introducing an external box to complement the IoU. It not only solves the issue of non-intersecting boxes but also provides a valid loss value even when the predicted box is inside the true box. This makes GIoU loss more effective for regression training in more complex images, providing more precise localization information, especially for small target defects such as component surface flaws. Let the ground truth box be denoted by A, the predicted box by B, and the IoU calculation process is given by:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{10}$$

The external box range is denoted by Z, and the area formed by concatenating the true and predicted boxes is T. The GIoU calculation process is:

$$GIoU = IoU - \frac{|Z - T|}{Z} \tag{11}$$

The confidence loss is also calculated using binary cross-entropy, similar to the classification loss. The primary task of confidence loss is to assess whether the predicted bounding box contains an object and quantify the probability of the object's presence. In product defect detection, especially in production lines, some defects may be difficult to detect or mistakenly identified as background noise. Therefore, precise confidence estimation is critical. By optimizing confidence loss, the model can better identify regions with defects and effectively distinguish between the background and the objects, thereby avoiding false positives or missed detections. Overall speaking, by combining YOLOv5s's classification loss, GIoU regression loss, and confidence loss, the lightweight product defect detection model proposed in this paper achieves higher precision while ensuring the stability and convergence speed of the training process. This model meets the real-time detection, rapid feedback, and high-precision demands of supply chain management, ultimately improving the efficiency and reliability of quality control.

4. EXPERIMENTAL RESULTS AND ANALYSIS

From Table 1, it can be seen that there are certain differences in the Average Precision (AP) results of various detection models for different types of defects. For traditional deep learning models, Faster RCNN's AP results are stable across different defect types, with a maximum of 91.2%, but it slightly lags in some specific defect types (e.g., welding or connection defects with 88.9%). The SSD model performs better in surface defects, shape and size deviations, and material defects, achieving 92.5%. YOLO series models (including YOLOv3, YOLOv4, YOLOv5, etc.) show fluctuations in performance across different defect types. YOLOv4 and YOLOv5s demonstrate better detection accuracy, with 92.6% and 97.8%, respectively. The YOLOX model performs excellently with an overall AP of 98.9%, and performs well across various defect types, particularly reaching 99.9% for surface defects and assembly defects. The proposed model in this paper achieves an overall AP value of 98.8%, comparable to YOLOX, and demonstrates balanced and excellent performance across various defects, especially for shape and size deviations, functional defects, and assembly defects, which are close to or above 97%, further improving detection efficiency and accuracy.

Table 1. Average precision results of different product defect detection models

Model	***AP/%***					
	Surface Defects	**Shape and Size Deviations**	**Material Defects**	**Welding or Connection Defects**	**Functional Defects**	**Assembly Defects**
Faster RCNN	91.2	91.2	91.2	88.9	91.2	91.2
SSD	88.9	92.5	92.5	89.8	92.5	88.9
*YOLOv*3	91.5	88.9	88.9	88.9	88.9	85.6
*YOLOv*4	92.6	92.3	92.3	91.5	92.5	88.5
*YOLOv*5*s*	97.8	97.8	97.8	98.9	98.9	97.8
*YOLOv*5*m*	97.5	98.9	97.8	98.4	98.7	97.8
YOLOX	98.9	98.9	99.9	98.5	98.5	99.9
The proposed model	98.8	98.9	98.9	98.7	98.9	98.5

Table 2. Precision results of different product defect detection models

Model	***mAP*0.5(%)**	***mAP*0.75(%)**	***mAP*0.5-0.95(%)**
Faster RCNN	91.5	73.2	55.65
SSD	92.5	62.6	46.25
*YOLOv*3	87.9	41.2	38.95
*YOLOv*4	91.5	46.5	41.26
*YOLOv*5*s*	97.8	82.3	62.32
*YOLOv*5*m*	97.9	84.5	62.61
YOLOX	98.9	87.9	62.58
The proposed model	98.5	85.4	62.88

Table 3. Comparison of lightweight performance of different product defect detection models

Model	***Params*(*M*)**	***FLOPs*(*G*)**	***Inference*(*ms*/*img*)**
Faster RCNN	42.52	92.51	35.62
SSD	24.69	136.25	38.94
*YOLOv*3	62.38	78.95	32.61
*YOLOv*4	53.21	53.21	28.56
*YOLOv*5*s*	7.26	15.69	2.2
*YOLOv*5*m*	21.26	37.56	12.5
YOLOX	8.89	21.23	14.5
The proposed model	1.56	3.6	1.8

From the precision results in Table 2, it can be observed that the proposed model in this paper demonstrates superior performance across multiple evaluation metrics (mAP@0.5, mAP@0.75, mAP@0.5-0.95), especially at mAP@0.5 (lower threshold) and mAP@0.75 (higher threshold). Specifically, the proposed model achieves 98.5% in mAP@0.5, close to YOLOX (98.9%), and higher than YOLOv5s (97.8%) and YOLOv5m (97.9%). At mAP@0.75, the proposed model's precision is 85.4%, slightly lower than YOLOX (87.9%), but still higher than YOLOv5s (82.3%) and YOLOv5m (84.5%), indicating that it remains competitive under higher precision requirements. Finally, in the mAP@0.5-0.95 composite metric, the proposed model scores 62.88%, surpassing most other models (e.g., YOLOv4's 41.26%, YOLOv3's 38.95%) and is on par with YOLOX (62.58%). This demonstrates that the proposed model achieves very balanced and stable precision in the overall defect detection task, maintaining high detection performance across various precision requirements.

From Table 3, it can be seen that the model proposed in this paper shows excellent lightweight performance, significantly outperforming traditional deep learning models. Specifically, the proposed model has only 1.56M parameters, 3.6G FLOPs, and an inference speed of 1.8ms/image. These values are much lower than Faster RCNN (42.52M parameters, 92.51G FLOPs, 35.62ms inference time), SSD (24.69M parameters, 136.25G FLOPs, 38.94ms inference time), and other versions of YOLO (such as YOLOv5s with 7.26M parameters and YOLOX with 8.89M parameters, although YOLOv5s has an inference speed of 2.2ms, its parameters and FLOPs are still significantly higher than those of the proposed model). In particular, the proposed model achieves the shortest inference time of 1.8ms/image, demonstrating its advantage in real-time performance. Furthermore, the low FLOPs and parameter count indicate that the model not only competes with existing advanced models in terms of accuracy but also has a significant advantage in computational resource demand and inference efficiency.

Table 4. Experimental results of attention mechanism detector

Model	*AP*/%						*mAP*0.5(%)
	Surface Flaws	Shape and Size Deviations	Material Defects	Welding or Connection Defects	Functional Defects	Assembly Defects	
76×76	98.9	98.9	98.5	98.2	98.8	97.8	98.6
38×38	97.8	99.5	97.6	98.9	98.5	97.6	97.5
19×19	97.9	97.8	97.8	99.3	98.3	97.2	98.6

From Table 4, the experimental results show that the detector with the attention mechanism exhibits relatively high detection accuracy and stability with different input sizes. First, with different input resolutions, both the AP values and mAP@0.5 values remain at a high level. Specifically, for the 76×76 input size, the overall AP reaches 98.9%, and the AP values for all defect types are relatively balanced (Surface Flaws 98.9%, Welding or Connection Defects 97.8%, etc.). For smaller input sizes (38×38 and 19×19), although the overall AP and mAP@0.5 values slightly decrease, the detection accuracy for specific defect types remains high. Particularly, for welding or connection defects (98.9% for 38×38 input, 99.3% for 19×19 input) and shape and size deviations (99.5% for 38×38 input, 97.8% for 19×19 input), very stable performance is maintained. This suggests that the attention mechanism has good adaptability to different input sizes and can effectively improve the model's local feature learning and detection accuracy.

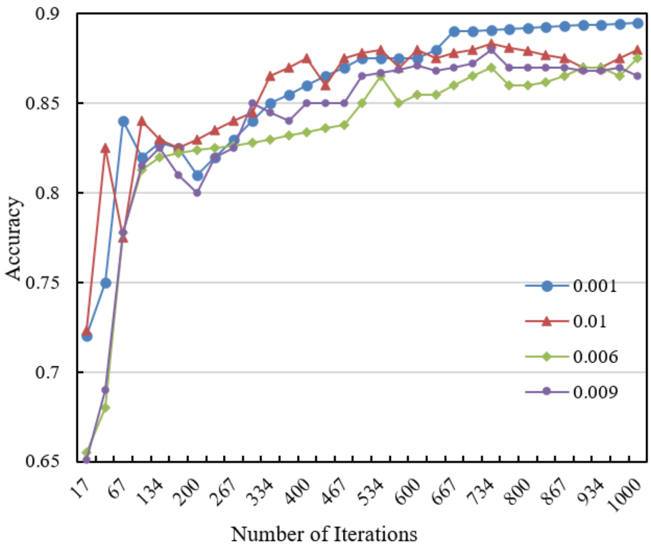

Figure 6. Learning rate comparison of the proposed model in product defect detection

Based on the learning rate comparison data in Figure 6, it can be observed that the learning rate of the proposed model changes over different iterations. Overall, as the number of iterations increases, the learning rate gradually stabilizes and flattens. In the initial iterations (e.g., after 17, 34, and 67 iterations), the learning rate is low (0.72, 0.75, and 0.84, respectively), indicating the model's gradual adaptation during the early stages of training. After more than 100 iterations, the rate of increase in the learning rate starts to slow down, and it stabilizes at around 0.875 after 500 iterations. Subsequent iterations (e.g., after 900 iterations) keep the learning rate at nearly 0.89, with minimal fluctuations or decay. Specifically, between the 100th and 1000th iterations, the learning rate changes stabilize and remain at a high level without significant fluctuations or reductions. This indicates that the model has gradually optimized during the training process and entered a more stable phase of training.

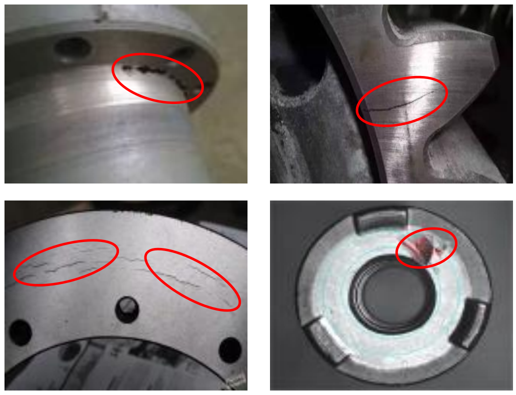

Figure 7. Product defect detection effect examples

From the experimental results shown in Figure 7, the lightweight defect detection model proposed in this paper demonstrates excellent results in detecting multiple defect types. Particularly, in quality control tasks within supply chain management, the model exhibits high detection accuracy and efficiency. Moreover, the mAP@0.5 value of the proposed model remains stable, consistently maintaining a high level at different stages of iteration, indicating that the model has good overall performance in multi-class defect detection. Especially in environments with limited computational resources, the model can maintain high accuracy and detection efficiency through optimization algorithms, meeting the requirements for lightweight and real-time detection.

5. CONCLUSION

This paper proposed a lightweight deep learning product defect detection method targeting quality control in supply chain management and designed a corresponding model architecture to optimize computational resource consumption and improve detection efficiency. By optimizing the deep learning model, this paper effectively reduced the computational resource demands while maintaining high detection accuracy, thus enhancing model operational efficiency, especially in resource-constrained real-world production environments. The experimental results show that the proposed lightweight defect detection model excels in detecting various defect categories, such as surface flaws, shape and size deviations, and welding defects, and can achieve efficient and accurate product defect detection. This provides a practical solution for quality control in supply chain management. Through rational architecture design and training strategies, the model is stable in various practical application scenarios, demonstrating excellent performance, meeting both the accuracy and speed requirements of industrial production. Overall, this research provides a new approach and practical method for quality control in supply chain management, particularly in the field of lightweight defect detection, with significant research value. Through continuous algorithm optimization and practical application validation, future research is expected to achieve greater breakthroughs in accuracy, efficiency, and applicability, advancing the development of intelligent quality detection technology in the industrial field.

However, this research also has certain limitations. First, although the proposed model demonstrates good performance in the experiments, its stability and real-time performance in extreme environments (such as large-scale data processing, multi-task parallelism, etc.) need further verification. Second, while the model performs well in detecting multiple defect types, there may still be limitations in recognizing certain special types of defects (such as microscopic defects or complex textures). Future research can focus on the following directions: on one hand, further optimizing the model architecture and algorithms to enhance its ability to detect complex and small defects; on the other hand, exploring more lightweight deep learning techniques, such as neural network quantization and knowledge distillation, to further reduce computational resource consumption and improve the model's applicability on different hardware platforms. Additionally, conducting larger-scale and longer-duration tests and validations with real production environment data will help further improve the model's robustness and practicality.

REFERENCES

[1] Kusrini, E., Miranda, S. (2021). Determining performance metrics of supply chain management in make-to-order small-medium enterprise using Supply Chain Operation Reference model (SCOR Version 12.0). Mathematical Modelling of Engineering Problems, 8(5): 750-756. https://doi.org/10.18280/mmep.080509

[2] Abdullahi, H.O., Mohamud, I.H. (2023). The Impact of ICT on supply chain management efficiency and effectiveness: A literature review. Journal Européen des Systèmes Automatisés, 56(2): 309-315. https://doi.org/10.18280/jesa.560216

[3] Ali, A.A.A., Mohamed Udin, Z., Abualrejal, H.M.E. (2023). The impact of humanitarian supply chain on non-government organizations performance moderated by organisation culture. International Journal of Sustainable Development and Planning, 18(3): 763-772. https://doi.org/10.18280/ijsdp.180312

[4] Xu, G.W., Ali, D.A., Bhaumik, A., Yang, M.S., Wang, N. (2023). The influence of digital transformation in enterprises on the dynamics of supply chain concentration: An empirical analysis of Chinese a-share listed companies. Journal of Organizations, Technology and Entrepreneurship, 1(2): 88-97. https://doi.org/10.56578/jote010202

[5] Sunmola, F., Burgess, P., Wertheim-Heck, S. (2024). Challenges and hotspots in food supply chain quality management practices. Total Quality Management & Business Excellence, 35(11-12): 1296-1330. https://doi.org/10.1080/14783363.2024.2367607

[6] Liu, Z., Xiao, J., Huang, Y. (2009). Proactive product quality control: An integrated product and process control approach to MIMO systems. Chemical Engineering Journal, 149(1-3): 435-446. https://doi.org/10.1016/j.cej.2009.02.007

[7] Song, Y., Lv, C., Liu, J. (2018). Quality and safety traceability system of agricultural products based on multi-agent. Journal of Intelligent & Fuzzy Systems, 35(3): 2731-2740. https://doi.org/10.3233/JIFS-169625

[8] Verdijck, G.J., Van Straten, G., Preisig, H.A. (2005). Optimisation of product quality and minimisation of its variation in climate controlled operations. Computers and Electronics in Agriculture, 48(2): 103-122. https://doi.org/10.1016/j.compag.2005.02.009

[9] Li, B., Wen, D., Shi, X. (2015). Research on product quality control in Chinese online shopping: Based on the uncertainty mitigating factors of product quality. Total Quality Management & Business Excellence, 26(5-6): 602-618. https://doi.org/10.1080/14783363.2013.865917

[10] Su, X., Wang, Y., Mao, J., Chen, Y., Yin, A., Zhao, B., Liu, M. (2022). A review of pharmaceutical robot based on hyperspectral technology. Journal of Intelligent & Robotic Systems, 105(4): 75. https://doi.org/10.1007/s10846-022-01602-7

[11] Jia, L.M., Wang, Y. (2022). Research on industrial production defect detection method based on machine vision technology in industrial internet of things. Traitement du Signal, 39(6): 2061-2068. https://doi.org/10.18280/ts.390618

[12] Khan, I.U., Aslam, N., Aboulnour, M., Bashamakh, A., Alghool, F., Alsuwayan, N., Alturaif, R., Gull, H., Iqbal, S.Z., Hussain, T. (2024). Deep learning-based surface defect detection in steel products using convolutional neural networks. Mathematical Modelling of Engineering Problems, 11(11): 3006-3014. https://doi.org/10.18280/mmep.111113

[13] Coskun, H., Yiğit, T., Üncü, İ.S., Ersoy, M., Topal, A. (2022). An industrial application towards classification and optimization of multi-class tile surface defects based on geometric and wavelet features. Traitement du Signal, 39(6): 2011-2022. https://doi.org/10.18280/ts.390613

[14] Sancho, L.F., Samoano, J.R., Álvarez, I., de Lera, A.F. (2008). Automated on-line inspection of surface scarfing in slabs and quality assessment based on novel image technologies. Revue de Métallurgie, 105(7-8): 383-389. https://doi.org/10.1051/metal:2008055

[15] Patel, K.K., Kar, A., Jha, S.N., Khan, M.A. (2012). Machine vision system: A tool for quality inspection of food and agricultural products. Journal of Food Science and Technology, 49: 123-141. https://doi.org/10.1007/s13197-011-0321-4

[16] Lu, Z., Zhang, Y., Xu, H., Chen, H. (2023). Fabric defect detection via a spatial cloze strategy. Textile Research Journal, 93(7-8): 1612-1627. https://doi.org/10.1177/00405175221135205

[17] Huang, L., Gong, A. (2023). Surface defect detection for no-service rails with skeleton-aware accurate and fast network. IEEE Transactions on Industrial Informatics, 20(3): 4571-4581. https://doi.org/10.1109/TII.2023.3327341

[18] Liu, H., Zhou, W., Kuang, Q., Cao, L., Gao, B. (2010). Defect detection of IC wafer based on two-dimension wavelet transform. Microelectronics Journal, 41(2-3): 171-177. https://doi.org/10.1016/j.mejo.2010.01.015

[19] He, S., Jian, Z., Liu, S., Liu, J., Fang, Y., Hu, J. (2024). PCSGAN: A perceptual constrained generative model for railway defect sample expansion from a single image.

IEEE Transactions on Intelligent Transportation Systems, 25(8): 8796-8806. https://doi.org/10.1109/TITS.2024.3368213

[20] Yu, J., Jiang, J., Fichera, S., Paoletti, P., Layzell, L., Mehta, D., Luo, S. (2024). Road surface defect detection—from image-based to non-image-based: A survey. IEEE Transactions on Intelligent Transportation Systems, 25(9): 10581-10603. https://doi.org/10.1109/TITS.2024.3382837

[21] Lee, Y., Lee, J. (2014). Accurate automatic defect detection method using quadtree decomposition on SEM images. IEEE Transactions on Semiconductor Manufacturing, 27(2): 223-231. https://doi.org/10.1109/TSM.2014.2303473

[22] Zhang, Y., Lefebvre, D., Li, Q. (2015). Automatic detection of defects in tire radiographic images. IEEE Transactions on Automation Science and Engineering, 14(3): 1378-1386. https://doi.org/10.1109/TASE.2015.2469594

[23] Xu, W., Liu, G., Wang, M. (2023). A deep neural network-based intelligent detection model for manufacturing defects of automobile parts. Journal of Circuits, Systems and Computers, 32(14): 2350236. https://doi.org/10.1142/S0218126623502365