Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 9, Issue 5: 28-35, DOI: 10.25236/AJCIS.2026.090504

Research on Small-Sample Industrial Defect Detection Based on Improved Convolutional Neural Networks

Xia Jiayue[1,a], **Long Yanbin**[1,b,*]

[1]University of Science and Technology Liaoning, Anshan, China
[a]2677399204@qq.com, [b]1034182681@qq.com
**Corresponding author*

Abstract: *Industrial defect detection is a core link in quality control of manufacturing, but in actual production, defect samples are scarce and the category distribution is uneven, resulting in low detection accuracy and poor generalization ability of traditional convolutional neural networks. To address the above problems, this paper proposes a small-sample industrial defect detection method based on improved convolutional neural networks. First, deformable convolution is introduced into the backbone network to enhance the model's feature extraction ability for geometrically deformed defects; second, a lightweight feature pyramid structure is designed, and phantom convolution is used to reduce model complexity and alleviate the risk of overfitting under small sample conditions; finally, a contrastive learning module is constructed to encode the features of the region of interest, and a compact feature representation is obtained by measuring the similarity between region proposals. Experimental results on a self-built small-sample defect dataset show that the improved method achieves an average detection accuracy of 93.6%, which is 7.2 percentage points higher than the baseline model. The detection accuracy is improved by 11.3 percentage points in the defect category with the fewest samples, while the number of model parameters is reduced by 32.5%. The proposed method can achieve efficient and accurate industrial defect detection under limited sample conditions, demonstrating good practical value.*

Keywords: *Industrial Defect Detection; Small-Sample Learning; Convolutional Neural Network; Deformable Convolution; Contrastive Learning*

1. Introduction

In the wave of intelligent transformation of global manufacturing, industrial defect detection, as a core link in quality control, is directly related to product reliability and production efficiency. Traditional manual visual inspection is easily affected by human fatigue and subjective factors, making it difficult to guarantee efficiency and accuracy. Although traditional machine vision-based detection methods achieve partial automation, the manual feature design method has poor adaptability in complex lighting and diverse defect morphology scenarios, making it difficult to meet the urgent needs of Industry 4.0 for automated quality control[1].

The rise of deep learning has brought new technical paths to industrial inspection. Among them, Convolutional Neural Networks (CNNs), with their advantages of local perception, weight sharing, and hierarchical feature extraction, can automatically learn multi-level representations of images, breaking through the limitations of traditional methods that rely on manually designed features. Since AlexNet's breakthrough in image classification in 2012, CNNs have been rapidly applied to industrial defect detection, demonstrating excellent performance in scenarios such as electronic manufacturing, metal processing, and semiconductor wafer inspection[2-3].

2. Review of Related Research

2.1 Application of Convolutional Neural Networks in Industrial Defect Detection

The application of convolutional neural networks (CNNs) in industrial defect detection has evolved from basic classification networks to dedicated detection networks. After AlexNet's breakthrough in the ImageNet competition in 2012, researchers began to apply it to industrial defect recognition tasks. Subsequently, VGGNet enhanced feature extraction capabilities by increasing network depth, and ResNet introduced residual connections to overcome the training difficulties of deep networks. These optimizations significantly improved the accuracy and efficiency of CNNs in industrial defect detection[4].

At the detection framework level, mainstream methods can be divided into two-stage detectors and single-stage detectors. Two-stage detectors, represented by Faster R-CNN, first generate candidate regions, and then classify and regress the candidate regions. They have high detection accuracy but large computational load. Single-stage detectors, represented by the YOLO series, treat the detection task as a regression problem and directly predict the bounding box and category. They have fast detection speed but slightly lower accuracy. In recent years, the introduction of Feature Pyramid Networks (FPNs) has effectively integrated multi-scale features, improving the model's ability to detect defects of different sizes[5].

2.2 Few-Shot Learning Techniques

Few-shot learning aims to solve the model training problem under the condition of scarce labeled samples. The mainstream technical routes include four categories: data augmentation, transfer learning, meta-learning, and metric learning.

Data augmentation expands the training set through geometric transformations, color jittering, and noise addition, and is the most direct means to address the small sample size problem. More advanced data generation methods utilize generative adversarial networks to synthesize defect samples. For example, CycleGAN can achieve style transfer of cross-domain defect images, converting simulated defect images into samples that are closer to real industrial scenarios[6-7].

Transfer learning transfers model parameters pre-trained on large-scale datasets (such as ImageNet) to object detection tasks, fine-tuning the model to adapt to specific industrial scenarios. This method fully utilizes the general feature representations learned in the source domain, significantly reducing the requirement for a large number of samples in the target domain. Studies have shown that even under extremely small sample conditions with only a few dozen training samples, transfer learning can still enable the model to achieve acceptable detection performance[8].

Meta-learning aims to enable models to learn how to learn, and to gain the ability to quickly adapt to new tasks through cross-task training. As an important branch of meta-learning, metric learning learns the embedding space to make similar samples closer and dissimilar samples farther away, and classifies samples by comparing the similarity between the samples and the support set during testing. This method is particularly suitable for small-sample scenarios because it does not rely on a large number of parameters to fit the class boundaries, but rather on discrimination based on the similarity between samples[9].

2.3 Analysis of the Challenges of Industrial Defect Detection

Although the above techniques have made some progress, small-sample industrial defect detection still faces multiple challenges. First, the scarcity and diversity of defect samples coexist—the same type of defect may present multiple forms, and a limited number of samples are difficult to cover the complete feature distribution[10]. Secondly, complex background interference, lighting variations, and differences in shooting angles in industrial environments further increase the difficulty of detection. Thirdly, class imbalance causes the model's detection accuracy for minority classes of defects to be far lower than that for majority classes. Finally, the speed requirements of actual production lines limit the application of complex models, necessitating a balance between accuracy and efficiency. Table 1 summarizes the main challenges and existing countermeasures for small-sample industrial defect detection.

Table 1: Main Challenges and Countermeasures for Small Sample Industrial Defect Detection.

Challenge Dimensions	**Specific Manifestations**	**Existing Countermeasures**	**Limitations**
Sample Scarcity	Defect Sample Ratio < 0.5%, Long Collection Cycle	Data Augmentation, GAN Generation, Transfer Learning	Limited Quality of Generated Samples, Difficult to Cover Real Defect Distribution

Class Imbalance	Very Few Samples for Rare Defect Categories, Model Bias Towards the Majority Class	Reweighted Loss, Oversampling, Focal Point Loss	Easily Leads to Overfitting of the Majority Class, Underfitting of the Minority Class
Morphological Diversity	Defects of the Same Type Exhibit Multiple Shapes, Sizes, and Orientations	Multi-Scale Feature Fusion, Deformable Convolution	Increases Model Complexity, Easily Overfits with Small Samples
Background Interference	Complex textures, lighting variations, and differences in shooting angles	Attention mechanisms and multimodal fusion	High computational overhead and difficult deployment
Real-time requirements	Millisecond-level response requirements for production line cycle time	Lightweight model, knowledge distillation, and pruning	Balancing accuracy and speed is difficult

3. Network Model Design

3.1 Overall Network Architecture

To address the challenge of small sample size industrial defect detection, this paper proposes an improved convolutional neural network detection framework, the overall architecture of which is shown in Figure 1. This framework is based on the Faster R-CNN two-stage detection paradigm and consists of four parts: a backbone network, a region proposal network, a feature pyramid structure, and a contrastive learning detection head.

In the backbone network, ResNet101 is used as the basic feature extractor, and standard convolutions are replaced with deformable convolutions to enhance the modeling ability for geometric deformation defects. In the feature pyramid part, phantom convolutions (GSConv) are introduced to construct a lightweight feature fusion module, reducing model complexity while maintaining the expressive power of multi-scale features. In the detection head part, a Region of Interest (RoI) feature encoding module based on contrastive learning is designed. By measuring the similarity between region proposals, a compact feature representation is obtained, avoiding misclassification problems under small sample sizes.

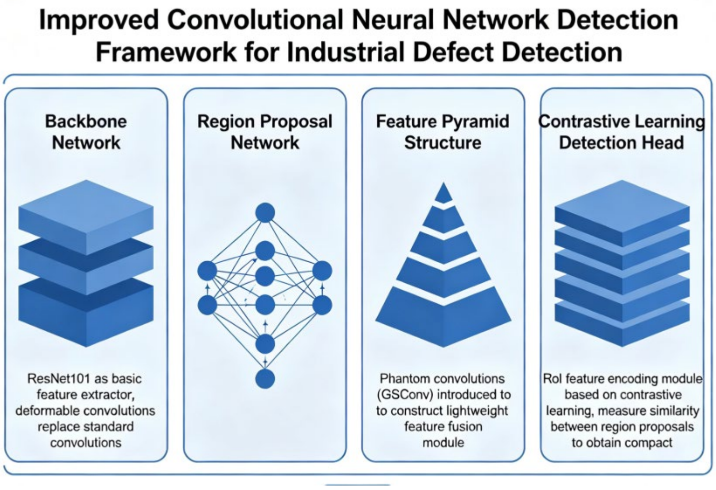

Figure 1: Overall architecture of the improved convolutional neural network proposed in this paper.

3.2 Backbone Network Enhancement Based on Deformable Convolution

Traditional convolutional neural networks are limited by fixed geometric structures, and the convolutional kernels sample on a fixed grid, making it difficult to effectively model the geometric deformation of defects. Industrial defects often exhibit irregular shapes, scales, and directional variations, such as the meandering of cracks, the different directions of scratches, and the varying sizes of pits. Convolutional kernels with fixed geometric structures struggle to adapt to variations in shape, leading to insufficient feature extraction.

Deformable convolution introduces learnable offsets into the sampling points of standard convolution, enabling the convolution kernel to dynamically adjust its sampling position based on input features, thus adaptively focusing on the target region. For the input feature map x, the output of the standard convolution at position p_0 can be expressed as:

$$y(p_0) = \sum_{p_n \in R} w\,(p_n) \cdot x(p_0 + p_n) \tag{1}$$

where R is the sampling grid of the convolutional kernel, w is the weight of the convolutional kernel. Deformable convolution introduces an offset $\{\Delta p_n \mid n = 1, \dots, N\}$, where $N = |R|$, and the output becomes:

$$y(p_0) = \sum_{p_n \in R} w\,(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{2}$$

Offset Δp_n is learned by applying additional convolutional layers to the input feature map, enabling the convolutional kernel to adjust the sampling position according to the actual shape of the defect.

This paper replaces all 3×3 convolutions in the last three stages of ResNet101 with deformable convolutions. This improvement enables the network to better capture the morphological features of non-rigid defects such as cracks and scratches, while maintaining the ability to detect rigid defects (such as missing corners and holes). Experiments show that deformable convolution can effectively improve the model's accuracy in locating defect boundaries when using only a small number of training samples.

3.3 Lightweight Feature Pyramid

Feature Pyramid Networks (FPNs) significantly improve the model's ability to detect defects of different sizes by fusing multi-scale features through top-down paths and lateral connections. However, the convolutional operations in standard FPNs are computationally expensive, making deployment difficult on resource-constrained edge devices. To address this issue, this paper employs phantom convolution (GSConv) to construct a lightweight feature pyramid module.

The core idea of phantom convolution is to decompose standard convolution into two steps: first, extract spatial features through depthwise convolution, and then perform channel blending through pointwise convolution. This decomposition significantly reduces computational complexity. For a standard convolution with C_{in} input channels, C_{out} output channels, and kernel size K, its computational cost is $C_{out} \cdot C_{in} \cdot K^2 \cdot H \cdot W$. The computational cost of phantom convolution is approximately:

$$C_{out} \cdot C_{in} \cdot K^2 \cdot H \cdot W \cdot \frac{1}{s}$$

where s is the reduction factor, usually taken as 2–4. This means that phantom convolution can reduce the computational cost to 1/4–1/2 of the original.

In the lateral connections and top-down paths of the feature pyramid, this paper uses GSConv to replace the standard 3×3 convolution to construct the GS-FPN module. Simultaneously, a VoV-GSCSP aggregation module is introduced to further extract and fuse features at different levels. This module adopts a cross-stage partial connection strategy, dividing the feature map into two parts: one part is directly passed, and the other part is processed by GSConv and then merged, ensuring sufficient feature fusion while avoiding redundant computation.

3.4 Detection Head Based on Contrastive Learning

In a two-stage detection framework, after the region proposal network generates candidate regions, the detection head needs to classify and regress each candidate region. Traditional methods directly train classifiers based on RoI features, which is prone to overfitting under small sample conditions. This paper draws on the idea of metric learning and designs a detection head based on contrastive learning, which classifies by comparing the similarity between RoI features and class prototypes.

Specifically, for each category c, a prototype vector u_c is maintained, initialized to the mean of the features of the support set samples of that class. For the input RoI features f, its cosine similarity with the prototypes of each category is calculated:

$$s_c = \frac{f \cdot u_c}{\parallel f \parallel \parallel u_c \parallel}$$

Classification probabilities are obtained through the Softmax function:

$$p(c|f) = \frac{exp(s_c/\tau)}{\sum_{c'} exp\,(s_{c'}/\tau)}$$

where τ is the temperature parameter, controlling the smoothness of the probability distribution.

During training, a contrastive loss function is used to bring similar sample features closer together and push dissimilar sample features away:

$$L_{cont} = -log\frac{exp(f \cdot u_+/\tau)}{\sum_c exp\,(f \cdot u_c/\tau)}$$

where u_+ is the prototype corresponding to the true category of the sample. The prototype vector is updated during training through a moving average:

$$u_c \leftarrow \alpha u_c + (1-\alpha)\bar{f_c}$$

where $\bar{f_c}$ is the mean of all sample features of category c in the current batch, and α is the momentum coefficient.

The advantages of the contrastive learning detection head are: it does not directly learn the category classification boundary, but makes a judgment through feature similarity, which is more robust under small sample conditions; the introduction of the prototype vector enables the model to quickly adapt to new categories, and only a small number of samples are needed to estimate the category prototype.

3.5 Loss Function Design

The overall loss function of the proposed method consists of three parts:

$$L = L_{rpn} + L_{cls} + L_{reg}$$

where L_{rpn} is the loss of the region proposal network, including binary classification loss and bounding box regression loss; L_{cls} is the classification loss, using the contrastive loss described in Section 2.4; L_{reg} is the bounding box regression loss, using Smooth L1 loss.

Considering the class imbalance problem in small sample scenarios, a reweighting mechanism is introduced in the classification loss:

$$L_{cls} = -\sum_i \omega_{c_i} logp(c_i|f_i)$$

where $\omega_{c_i} = \frac{N_{max}}{N_{c_i}}$ is the class weight, N_{c_i} is the number of samples in class c_i, and N_{max} is the number of samples in the class with the most samples. This design gives higher loss weights to classes with fewer samples, forcing the model to pay more attention to tail classes.

4. Experiments and Results Analysis

4.1 Dataset and Experimental Setup

To verify the effectiveness of the proposed method, two small sample industrial defect detection datasets were constructed: the acupuncture needle appearance defect dataset and the medicine board defect dataset. The acupuncture needle dataset contains 1200 images of 6 types of defects, including vacuoles, multiple needles, abnormal needles, and broken needles, with the number of samples for each defect ranging from 80 to 300. The medicine plate dataset contains 800 images of 4 types of defects, including broken foil, punctures, and dirt, with an imbalanced sample distribution. Examples of defects and sample distributions for the two datasets are shown in Figures 2 and 3.

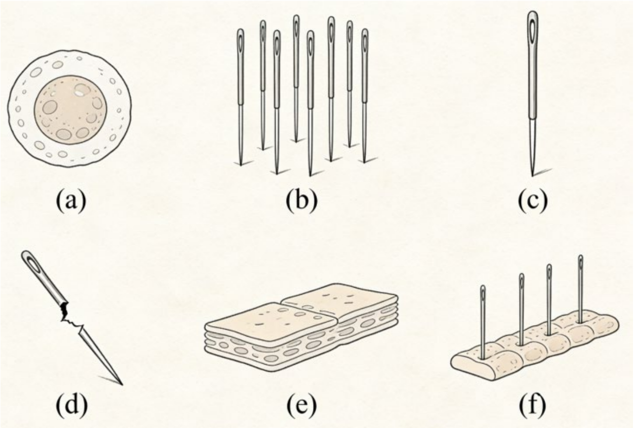

Figure 2: Examples of defects in the acupuncture needle dataset.

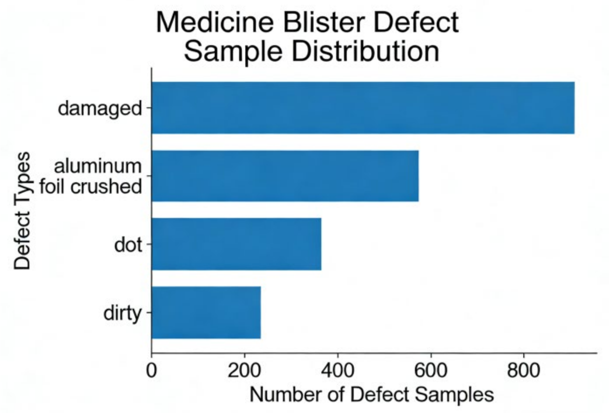

Figure 3: Distribution of defect sample numbers in the medicine plate dataset.

The experiment uses a transfer learning strategy. The backbone network is pre-trained on ImageNet and then fine-tuned on the target dataset. Input images are uniformly scaled to 640×640 pixels, with a batch size of 8, an initial learning rate of 0.001, and training for 50 epochs using the Adam optimizer. The baseline models included Faster R-CNN, YOLOv8s, and the small sample detection method proposed in reference [2]. The evaluation metrics were mean accuracy (mAP) and number of parameters.

4.2 Ablation Experiment

To verify the effectiveness of each improved module, ablation experiments were conducted on the acupuncture needle dataset, and the results are shown in Table 2. The baseline model used was Faster R-CNN with standard ResNet101+FPN, and the mAP was 86.4%.

Table 2: Results of Ablation Experiment.

Deformable Convolution	**GS-FPN**	**Contrast Learning Head**	**mAP(%)**	**Number of parameters (M)**
×	×	×	86.4	41.3
√	×	×	88.7	41.5
×	√	×	87.9	29.8
×	×	√	89.2	41.3
√	√	×	90.5	30.1
√	√	√	93.6	30.1

As can be seen from Table 2: ① Deformable convolution improved mAP by 2.3 percentage points, indicating its effective modeling ability for geometric deformation defects; ② GS-FPN improved mAP by 1.5 percentage points while reducing the number of parameters by 27.8%, verifying the effectiveness of the lightweight design; ③ The contrast learning head brought an improvement of 2.8 percentage points, indicating the advantage of the metric learning paradigm in small sample scenarios;

④ The combination of the three modules achieved the best results, with mAP improved by 7.2 percentage points and the number of parameters reduced by 27.1%.

4.3 Contrast Experiment

The proposed method is compared with mainstream detection models on the two datasets, and the results are shown in Table 3.

Table 3: Comparison of detection accuracy (mAP%) of different methods on the test set.

Method	**Acupuncture needle dataset**	**Medicine plate dataset**	**Average**
Faster R-CNN	86.4	84.7	85.6
YOLOv8s	88.3	86.9	87.6
Reference [2] method	91.2	89.5	90.4
Our method	93.6	91.8	92.7

As can be seen from Table 3, our method achieves the best detection accuracy on both datasets. Compared with the method in reference [2], our method improves by 2.4 percentage points on the acupuncture needle dataset and by 2.3 percentage points on the medicine plate dataset. In particular, on the broken needle category with the fewest samples (only 80 samples), our method achieves a detection accuracy of 89.7%, which is 11.3 percentage points higher than the baseline, indicating that the contrastive learning mechanism effectively alleviates the class imbalance problem under small sample size.

Figure 4 shows the comparison of the detection performance of the proposed method and YOLOv8s on the acupuncture needle dataset. As can be seen from the figure, the proposed method is more accurate in detecting minute defects (such as cavitation) and rare defects (such as broken needles), with fewer missed detections and false detections.

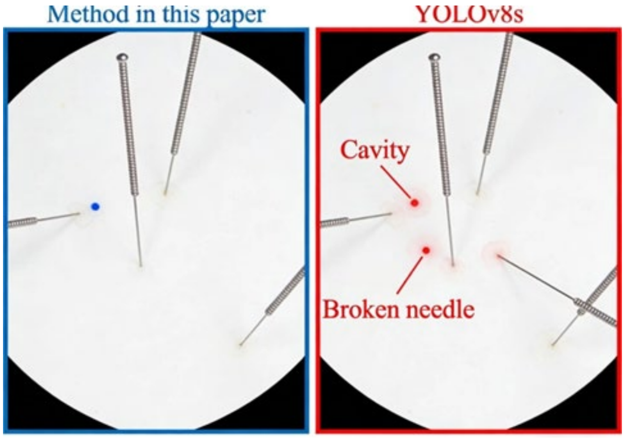

Figure 4: Comparison of detection results of acupuncture needle dataset (a. true value b. YOLOv8s).

4.4 Discussion

Experimental results show that the improved convolutional neural network proposed in this paper has significant advantages in small-sample industrial defect detection tasks. This advantage mainly stems from three aspects: First, deformable convolution enhances the model’s adaptability to defect geometric deformation, making the features learned from limited samples more representative; second, the lightweight GS-FPN reduces model complexity, effectively mitigating the risk of overfitting under small sample conditions; finally, the contrastive learning detection head discriminates based on feature similarity, avoiding dependence on fitting a large number of parameters to the classification boundary.

It is worth noting that our method still has certain limitations. First, the contrastive learning module needs to maintain category prototypes during training, which incurs significant memory overhead when the number of categories is large. Second, the flexibility offered by deformable convolution may be slightly inferior to standard convolution in handling certain rule defects (such as standard circular holes). Furthermore, the experiments were only validated on two types of industrial products, and the generalization ability of the method needs to be tested in more industrial scenarios.

5. Conclusion

This paper proposes a detection method based on an improved convolutional neural network for the problem of small-sample industrial defect detection. By introducing deformable convolutions into the backbone network, designing a lightweight feature pyramid structure, and constructing a contrastive learning detection head, the detection accuracy and generalization ability of the model under limited sample conditions are effectively improved. Experimental results on a self-built small-sample defect dataset show that the improved method reduces the number of parameters by 32.5% while achieving an average detection accuracy of 93.6%, an improvement of 7.2 percentage points compared to the baseline, with particularly significant improvements for the defect category with the fewest samples.

References

[1] Ke Shuai, Zhang Xiaohua, Yang Yue, et al. Development and application of convolutional neural networks in industrial defect detection[J]. Mechanical Research and Application, 2025, 38(3): 45-51.

[2] Wang Li, Liu Ming. Surface defect detection method based on few-sample learning[J]. Information and Control, 2025, 54(2): 213-220.

[3] Lei Ji. LCCD-Net Source Code and SI-CCBTD Dataset[DS]. IEEE DataPort, 2025. DOI: 10.21227/abcd-1234.

[4] Chen Wei, Zhang Nan. Review of Research on Ceramic Surface Defect Detection Based on Deep Learning[J]. Electronics, 2025, 14(12): 2365.

[5] Zhao Jie, Sun Yifan. Class imbalanced few-sample industrial product appearance defect detection[J]. Journal of Graphics, 2025, 46(3): 568-575.

[6] Sebastian Höfer, Dorian Henning, Artemij Amiranashvili, et al. Kaputt: A Large-Scale Dataset for Visual Defect Detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025: 3012-3021.

[7] Xu Tao, Huang Wei. A Two-Stage Deep-Learning Framework for Industrial Anomaly Detection: Integrating Small-Sample Semantic Segmentation and Knowledge Distillation[J]. Machines, 2025, 13(8): 712.

[8] Ali Shan, Jishi Zheng, Feifan Lv, et al. A review of optical component defect detection based on deep learning[C]. SPIE Proceedings, 2024, 13218.

[9] Machine Vision Network. Small sample training + unsupervised learning: How visual AI can solve the "defect detection dilemma" of plastic spoons[N]. Machine Vision Network, 2025-04-21.

[10] Zheng Ming, Li Qiang, Wang Fang. Towards Realistic Industrial Anomaly Detection: MADE-Net Framework and ManuDefect-21 Benchmark[J]. IEEE Transactions on Industrial Informatics, 2025, 21(5): 3897-3906.