어텐션 매커니즘 기반 심층 컨볼루션 뉴럴 네트워크를 사용한 산업용 불량 칩 검사

Industrial Defective Chip Inspection using Deep Convolutional Neural Network with Attention Mechanism

Min Je Kim
Sejong University
Seoul, South Korea
minjekim3797@gmail.com

Altaf Hussain
Sejong University
Seoul, South Korea
altafh3797@gamil.com

Muhammad Munsif
Sejong University
Seoul, South Korea
munsif3797@gmail.com

Sangil Yoon
Sejong University
Seoul, South Korea
sangilyoon3797@gmail.com

Mi Young Lee
Sejong University
Seoul, South Korea
miylee@sejong.ac.kr

Sung Wook Baik*
Sejong University
Seoul, South Korea
sbaik@sejong.ac.kr

Abstract

The identification of anomalies in industrial settings poses a significant challenge, especially when there is a lack of negative samples and when the anomalous regions are small. Although existing computer vision methods have automated this task to some extent, these approaches struggle to extract salient features for inspecting defective chips. To tackle this problem, a deep learning-based framework is proposed for detecting anomalies in industrial settings. The framework utilizes a fine-tuned backbone convolutional neural network model and incorporates an enhanced attention mechanism. The attention module generates discriminative feature maps along two dimensions: channel and spatial. This is achieved by processing intermediate features obtained from the backbone model. These attention maps are then multiplied with the input feature map to dynamically enhance the relevant features. Extensive experiments demonstrate the effectiveness of our proposed method in maintaining a high level of detection accuracy for industrial product inspections. Consequently, our results conclude a suitable solution for optical chip inspection systems in industrial settings.

Keywords: *Industrial defect detection, convolutional neural network, deep learning, industrial automation.*

1. Introduction

Identification of objects or events that deviate from a normal pattern or other objects known as anomaly [1]. Anomaly Detection (AD) has several uses in a variety of industrial applications, such as the manufacturing sector, where activities involving visual examination are essential for guaranteeing the quality of products. However, it can be difficult to produce a large enough labeled dataset for conventional supervised classification approaches [2]. Therefore, developing accurate computer vision techniques is an extremely challenging task for automatic defect detection in industrial settings. In the early days, researchers developed Patchwise inspection [3] and Downsizing [4] techniques for automatic industrial defect recognition. However, it achieved limited performance when the anomaly is too small that it might have been lost during downsizing. Deploying patches for inspection also increases the risk of missing anomalies that are larger than the patch size.

In [5] suggested PatchCore, that detects an image-level anomaly and achieved lower accuracy for complex scenarios. [6] utilized a deep residual-based triplet network approach to train the CNN on surface texture for anomaly detection. Usually, for accurate defect plates detection a large CNN network is required and enough training samples for efficient training. This issue is addressed by using efficient training strategy for salient features selection and assists the users by visualizing the defected region of industrial plates during testing. The main contributions of this research are as follows:

1. We proposed a deep learning-based model for industrial defect plates detection that explores the advantages of attention module for detecting abnormal regions.
2. Our study demonstrates how the attention mechanism boosts the capability of pre-trained CNNs by extracting more distinctive features and improving their performance even with limited training data.
3. We perform detailed ablation study of various backbone CNN models

- * Correspondence author: Sung Wook Baik

with/without attention mechanism.

2. Proposed Method

This section introduces the proposed method for anomaly recognition. To facilitate comprehension, a visual representation of the method is provided in Figure 1.

2.1 Proposed Features Optimizer and Extractor

The proposed approach used Deep Learning (DL) method that incorporates ResNet50 and convolutional block attention module (CBAM) to enhance the accuracy and efficiency of anomaly recognition tasks. The input data is fed to ResNet50 to extract intermediate features. As illustrated in Figure 1, the ResNet50 architecture is composed of multiple layers of convolutional and pooling operations. Rather than considering the entire ResNet50 architecture, the intermediate features from the last convolution layer are examined, and each activation map is considered as a single feature map containing diverse information of the input image. To optimize the obtained intermediate features from the previous module are then passed through the CBAM module as shown in Figure 1 (features refiner) to learn more robust and enhanced features. CBAM consists of two main mechanisms: the channel attention module (CAM) and the spatial attention module (SAM).

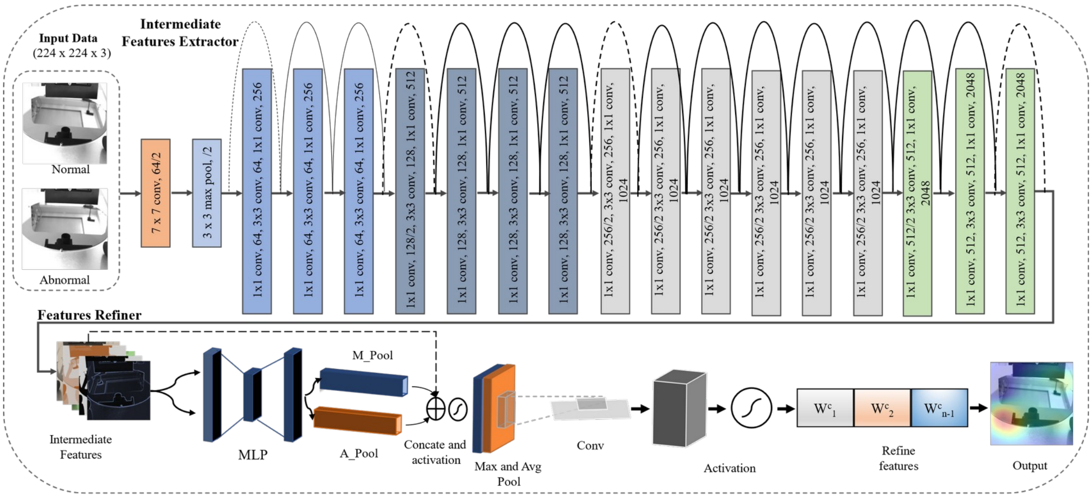

Figure 1. The proposed framework comprises of three main blocks: Input data, intermediate features extractor, and CBAM features refiner.

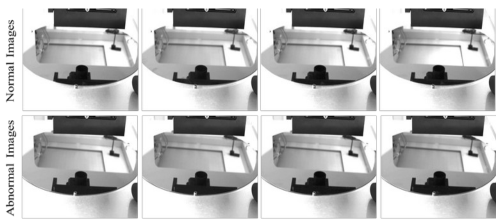

Figure 2.Visual sample from the proposed dataset.

CAM learns to selectively emphasize important channels in the feature maps by applying multi-layer perceptron's (MLP), average and max pooling, while SAM learns to selectively emphasize important spatial locations using channel wise 2D convolution and activation.

3. Experimental Results

3.1. Dataset

We have collected real-world data from industries, which represent both normal and abnormal scenarios. Each image is captured in Red Green Blue (RGB) format with dimensions of 640x640 pixels, and there are a total of 221 images representing abnormal scenarios and 921 images for normal scenarios. For training purposes, we have divided the input data into 70% for training, 20% for validation, and 10% for testing. The visual representation of the dataset are shown in Figure 2. The proposed method is implemented in python language with Pytorch library on window operating system with 24 GPU graphics card.

3.2. Results comparison

To evaluate the performance of the proposed method we have conducted multiple experiments on different approaches as shown in Table 1. Their results show that our method achieved higher 96% accuracy when compared to other models for defected plates detection as shown in Figure 3. The main reason behind higher performance is that we used transfer learning strategy which used the pretrained weights of the ResNet50 which is already trained on the huge ImageNet datasets containing 1000 classes. Additionally, for accurate defect plates detection the ResNet50 is empowered by CBAM attention to discard the non-salient information and learn the most discriminative features for industrial plates defect pattern detection.

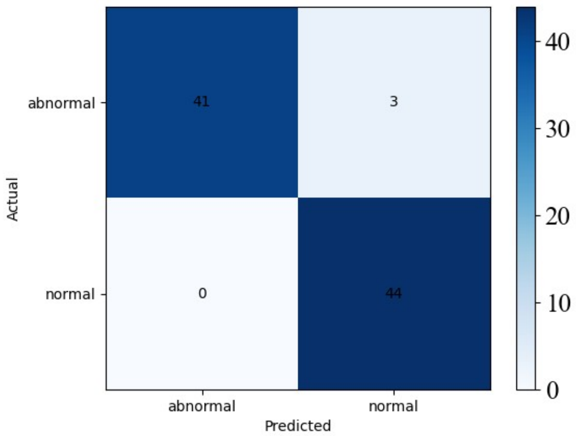

Figure 3. Performance evaluation using confusion matric.

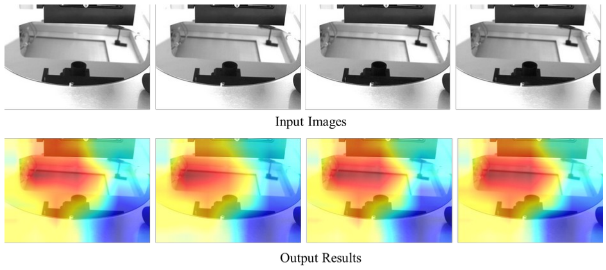

Figure 4. Output results of the proposed method on the testing images from abnormal class.

Furthermore, we have used the explainable artificial intelligence (XAI) modules of the proposed method for evaluating the industrial defects in plates which provides transparency into the decision-making, accountability for ensuring responsible and reliability in of the model for industrial applications. Figure 4 shows the visual results of the proposed model for interpretability, where they accurately highlight the defected regions of the abnormal images.

Table 1. Ablation study of various baseline backbone CNN models.

Method	**Accuracy (%)**	**Recall (%)**	**Precision (%)**
MobilenetV2	70	72	68
Resnet-50	82	84	79
Mobilenet V2 + CBAM	84	86	82
Resnet-50+CBAM	**96**	**100**	**90**

4. Conclusions

The detection of anomalies within industrial settings is a critical task that requires a reliable and accurate DL-based model. Our framework leverages a lightweight CNN model with enhanced attention mechanism to improve detection capabilities. The performance of the proposed framework is evaluated on real world industrial defective chips data and achieved higher performance for detecting anomalies in industrial settings.

Acknowledgment

This work was supported by the faculty research fund of Sejong University in 2022.
"Deep learning-based anomaly detection and predictive maintenance management system for industrial machinery" was supported by Seoul R&D Program (CY220049) on 2022.

References

1. Ahn, J.-Y. and G. Kim, Application of optimal clustering and metric learning to patch-based anomaly detection. Pattern Recognition Letters, 2022. 154: p. 110-115.
2. Saeedi, J., et al., Measurement and inspection of electrical discharge machined steel surfaces using deep neural networks. Machine Vision and Applications, 2021. 32: p. 1-15.
3. Kim, D., et al., FAPM: Fast Adaptive Patch Memory for Real-time Industrial Anomaly Detection. arXiv preprint arXiv:2211.07381, 2022.
4. Hoh, M., et al., A Generative Model for Anomaly Detection in Time Series Data. Procedia Computer Science, 2022. 200: p. 629-637.
5. Roth, K., et al. Towards total recall in industrial anomaly detection. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
6. Tayeh, T., et al. Distance-based anomaly detection for industrial surfaces using triplet networks. in 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). 2020. IEEE.