Anomaly Detection of TP4056 Charging Module Based on Improved SimpleNet

Shu Chen, Rouyi Fan, Xiaofeng Li*
School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China
**Corresponding Author*

Abstract: This study suggests an industrial anomaly detection technique based on an enhanced SimpleNet to fulfill the surface defect detection requirements of the TP4056 charging module in the realm of industrial manufacturing. First, the algorithm adds a Multi-Head Attention (MHA) module to SimpleNet's discriminator. By concentrating on important details within picture features, this module can improve the model's capacity to distinguish intricate features and precisely identify a range of minute flaws in the TP4056 charging module. Additionally, residual connections are used to optimize the gradient propagation path. This preserves original information while injecting attention-guided feature interactions by adding the original features to the attention correction term. A dropout module, which successfully avoids overfitting and enhances the model's generalization ability, is presented in consideration of the model's capacity for generalization. In order to optimize the anomalous feature production process and enable the model to learn anomaly feature representations under various noise situations, SimpleNet's anomaly feature generator simultaneously implements a dynamic noise correction technique. The VISA-PCB4 dataset was used to verify the model's efficacy. According to experiments, the enhanced model performs better on image-level AUROC, pixel-level AUROC, and anomaly pixel-level AUROC (PRO) measures, achieving 95.2%, 97.3%, and 88.8%, respectively. When compared to the original model, the anomaly pixel-level AUROC rose by 9.8%, greatly improving the model's capacity to identify anomalous pixels and offering a practical way to check the quality of charging modules.

Keywords: Industrial Quality Inspection; SimpleNet; Attention Mechanism; Dynamic Noise

1. Introduction

Finding data points or patterns in a significant amount of data that do not follow the expected typical behavior is known as anomaly detection, or outlier detection [1]. Deep learning is being used more and more in industrial anomaly detection, particularly in equipment maintenance and production lines. Deep learning may efficiently discover possible abnormalities or faults by automatically extracting and learning information from massive datasets through the construction of intricate neural network models [2].

Many portable electronic gadgets employ the TP4056 charging module, which is an essential part of lithium battery management systems. Ensuring the quality of the TP4056 module is crucial since its performance and safety have a direct impact on the device's lifespan and user safety. Nevertheless, the TP4056 charging module may acquire a number of surface flaws throughout manufacturing, including improper soldering or missing solder joints [3]. These flaws may result in safety events in addition to affecting the module's regular operation. As a result, identifying these surface flaws effectively and precisely has become crucial for production quality control.

Manual inspection is the mainstay of traditional fault detection techniques. This approach is ineffective and vulnerable to human error, despite its ability to produce high-precision detection results. Deep learning-based automated detection techniques have steadily emerged as a research hotspot, particularly in the field of image processing, as deep learning technologies have advanced [4]. Accurate defect identification is made possible by the automatic feature extraction and classification of images through the use of deep learning

techniques, specifically Convolutional Neural Networks (CNN) [5]. The lack of faulty samples is one of the difficulties these approaches encounter, though. Insufficient faulty image samples during deep learning model training can cause overfitting, which lowers the detection accuracy of the model.

Numerous deep learning-based anomaly detection techniques have been developed to solve these issues. For example, techniques such as PatchCore [6] and PaDiM [7] have achieved good detection performance and have partially solved the sample scarcity issue. These approaches do have some disadvantages, though; for instance, the PatchCore model uses a lot of memory when processing, and PaDiM has a lot of computational complexity issues, which limits its use in real-world scenarios. As an unsupervised learning model, SimpleNet does not require tagged anomalous samples; instead, it trains using just normal samples. By creating artificial anomalous features in the feature space and employing a discriminator to separate normal and abnormal features, it finds and locates anomalies [8]. Its successful implementation in industrial settings with limited anomaly samples is made possible by this unsupervised learning methodology.

In order to increase the effectiveness and precision of anomaly detection, this study suggests an upgrade plan for the SimpleNet model. Among the specific enhancements are 1) adding a multi-head attention mechanism to the discriminator to enhance feature interactions and the model's capacity to identify important features; 2) optimizing deep feature fusion through residual connections, which preserves low-level features and injects high-level features; 3) adding a dropout layer to prevent overfitting during model training, which enhances the model's capacity for generalization; and 4) creating a dynamic noise adjustment strategy that can adaptively modify the noise level in response to various training phases and data properties, thus enhancing the caliber of anomaly features that are produced.

We experimented on the VISA-PCB4 dataset to confirm the efficacy of the suggested approach. According to experimental data, the enhanced SimpleNet model outperforms conventional techniques in the TP4056 charging module defect detection task in terms of accuracy and robustness. In particular, it enhances the detection of anomalous pixels in defects, effectively addresses problems with current approaches, and lowers the likelihood of false detections (such as those in reconstruction-based methods [9-11]).

2. Improved SimpleNet Model

2.1 Procedure

This research suggests an enhanced SimpleNet model to tackle the problem of identifying minute surface flaws on the TP4056 charging module. In order to optimize the anomaly feature creation process, this model designs a dynamic noise adjustment technique and incorporates a multi-head attention module and residual connections during the feature discrimination stage. In contrast to conventional industrial anomaly detection techniques, the MHA module uses adaptive feature interaction weight computation and cross-dimensional information fusion to enable self-focusing on defect-sensitive regions instead of relying on intricate feature pyramid [12] building. While the attention correction term concentrates on local anomalous regions, enabling the model to learn at deeper levels, residual connections allow deep features to retain high similarity with the input. By using a training-epoch adaptive noise production mechanism instead of the fixed intensity mode, the dynamic noise method improves the model's representation of progressive errors while lowering the costs associated with manual parameter tuning.

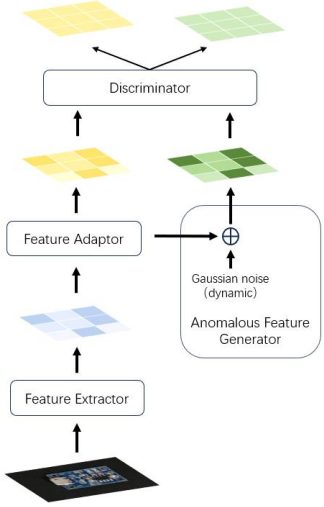

Figure 1. Process Diagram for the Enhanced SimpleNet Model

The enhanced SimpleNet model greatly improves the detection accuracy of circuit flaws such as microcracks and cold solder connections. Figure 1 depicts the enhanced

model's total network flow. The discriminator component of SimpleNet is the primary area of improvement in this research. Figure 2 depicts the enhanced SimpleNet discriminator's network design.

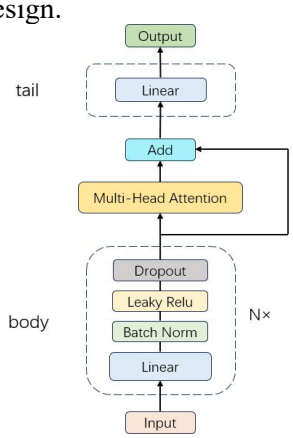

Figure 2. The Enhanced SimpleNet Discriminator Model's Architecture

2.2 MHA Module and SA Mechanism

Given the complexity and unpredictability of the different signal data throughout the charging process, MHA is expertly employed in the SimpleNet discriminator while dealing with the anomaly detection task for the TP4056 charging module. Key regions of aberrant signal features can receive additional attention weights from MHA. This greatly increases the accuracy and dependability of TP4056 charging module anomaly detection by strengthening the ability to extract abnormal pattern features and successfully filtering out interference from normal operating states or other non-critical factors on the detection results.

In order to provide a clearer explanation of the MHA module's operation, this study first presents the Self-Attention (SA) mechanism [13]. The SA technique can create global dependencies and increase the image's receptive field because the queries and keys originate from the same collection of elements. SA can capture more contextual information and has a wider receptive field than CNN. Query (Q), Key (K), and Value (V) vectors are first created using linear transformations depending on the input feature matrix, as indicated in formula (1). Following scaling (to avoid gradient vanishing), the correlation scores between the Q and K vectors are then computed using the dot product to determine the attention weights. Formula (2) illustrates how to do a weighted sum on the V vectors using the resulting attention weights. The output of SA is the outcome following the weighted sum. When processing data, SA enables the model to dynamically allocate various weights to each point, thereby capturing internal relationships. The model has the ability to automatically focus on feature regions associated with anomalies, like microcracks or distortions near solder joint edges. By avoiding straightforward local pattern recognition and capturing intricate interactions between features, this approach improves the modeling capacity for long-range dependencies.

The self-attention mechanism serves as the foundation for MHA. Applying several SA techniques on the input data at once is its main concept. By learning complementing feature patterns and perceiving information at different scales, multiple attention heads can successfully improve the model's perception of details, which in turn improves the expressive capability of the model [13]. This allows for a variety of views for the final prediction. The final output is obtained by concatenating and transforming the findings from several self-attention heads through a linear layer. MHA increases the model's sensitivity in identifying minute abnormalities or defects, including minute manufacturing errors or minor variations in images.

The output feature matrix $X \in \mathbb{R}B\times D$ from the discriminator's body part must first be transformed as input before the multi-head attention mechanism can be added. Adapting the input to the format needed by the multi-head attention mechanism is the aim of this transformation. In this case, D stands for the feature dimension, or the length of the feature vector for each sample, and B for the batch size, or the number of data samples processed each time. A new tensor $X' \in \mathbb{R}B\times 1\times D$ is created by adding a dimension (of size 1) to the second dimension, which is the sequence dimension. In order to prepare for further attention computations, the input data has now been converted to a format that can be fed into the multi-head attention mechanism.

The Query (Q), Key (K), and Value (V) matrices needed for each attention head are initially obtained by linearly transforming the input matrix in the multi-head attention method. The dimension of each attention head is dk = D/N, where N is the number of attention heads. Learnable parameter matrices $WiQ \in \mathbb{R}D\times dk$, $WiK \in \mathbb{R}D\times dk$, and $WiV \in \mathbb{R}D\times dk$ are used

to conduct multiplication operations on the input original matrix in order to construct the Q, K, and V matrices through linear transformation. The transformation formula for the i-th attention head is:

$$Q_i = X' W_i^Q,\ K_i = X' W_i^K,\ V_i = X' W_i^V \quad (1)$$

The dot product is used to assess how comparable the query Q and the key K are. In order to help the model identify which keys are more pertinent given a query, the dot product result calculates the degree of matching between each query and each key. The attention output for each head is then determined by doing a weighted sum on the Vi matrix based on this similarity. The formula can be written as follows:

$$Attention_i = Softmax\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (2)$$

The outputs of each attention head are first concatenated to create a larger-dimensional vector in the fusion process of multiple attention head outputs. By combining the knowledge that each attention head has independently acquired, this procedure improves the model's comprehension of various pieces of information. After that, the concatenated vector is transformed linearly—that is, it is mapped through a weight matrix W0—to become the final attention output. Maintaining the information's efficacy and expressiveness while integrating the multifaceted data gathered by the many attention heads is the aim of this operation. In the end, the model may weight and aggregate the input information from various viewpoints using this multi-head attention technique, enhancing overall performance.The formula is:

$$Z=Concat(Attention_1,\ldots,Attention_4)W_0 \quad (3)$$

2.3 Dropout Module and Residual Connection

By directly connecting the input to the output, residual connections maintain low-level features in the network and prevent feature information from being lost as network depth increases [14]. The expressive power and learning efficacy of the model are increased by this mechanism, which allows high-level information to better integrate with low-level features by reinforcing low-level features in deeper layers and directly injecting high-level features into the network through residual connections. By eliminating common problems in deep network training, such as gradient vanishing and information decay, this method allows the network to transmit and process information more effectively, improving performance and stability throughout network training. Its formula can be written as follows:

$$X_{out} = X + Z \quad (4)$$

The MHA module and residual connections work in concert. While the MHA module enhances defect-related features by suppressing unnecessary regions through the self-attention mechanism, residual connections enable the model to optimize features in a progressive correction fashion.

Furthermore, a Dropout module is added to the SimpleNet discriminator's body section in this work. It can effectively reduce overfitting [15] by randomly masking parts of the network's neurons. This works particularly well when training data is scarce, enhancing the model's robustness and generalization capabilities.

2.4 Dynamic Strategy for Noise Adjustment

Gaussian noise with a set standard deviation ($\sigma = 0.015$) is used during training in the original SimpleNet model. At some training phases, this fixed noise level might not adjust adequately to changes in feature distribution. The model's features may be somewhat coarse in the early phases of training, and noise has little impact. However, a fixed noise intensity may make it more difficult for the model to learn complicated features as training goes on and its features gradually improve. Consequently, a fixed noise standard deviation hinders the model's ability to adjust to the demands of several training phases. This study suggests a dynamic noise adjustment solution to address this issue. This approach enables the noise standard deviation to change dynamically as the number of training epochs increases, improving the model's ability to adapt to the shifting feature distribution as it learns. In particular, we created a technique that, when written as follows, causes the noise standard deviation to rise linearly with the training epochs.The formula is:

$$\sigma(t) = \sigma_0\left(1 + \alpha \cdot \frac{t}{T}\right) \quad (5)$$

Where t is the current training epoch, T is the total number of training epochs, σ_0 is the initial noise intensity, and α is the adjustment coefficient, which controls the noise intensity's increasing amplitude.

In the early phases of training, this technique

keeps the noise level lower, enabling the model to learn more steadily and without undue disruptions. The noise intensity progressively rises with training depth, improving the model's capacity to recognize intricate patterns and preventing overfitting in later training phases. This dynamic noise adjustment lessens the detrimental effects of ineffective noise on feature learning while providing the model with more acceptable noise interference at various training phases, enhancing its robustness and generalization capacity.

3. Process of Experiments

3.1 Experimental Setting

Table 1 displays the experimental setup used in this work.

Table 1. Overview of the Experimental Setting

Name	Specifications
Operating System	Windows11
Development Language	Python3.9
Development Framework	Pytorch2.1.1+cu118
GPU	GeForce RTX 3050
CPU	AMD Ryzen 5 5600H
CUDA	12.8

3.2 Experimental Parameters and Dataset

With 10,821 high-resolution color photos encompassing 12 objects across 3 domains (9,621 normal samples and 1,200 anomaly samples), the Visual Anomaly (VisA) collection is the largest industrial anomaly detection dataset to date [16]. The PCB4 sub-dataset from the VisA dataset is used in the studies in this publication. Among the 1105 photos in this dataset are 904 normal images, which serve as the training dataset for the model's performance in typical scenarios. The test dataset is made up of the remaining portion; precisely, it includes 100 aberrant and 101 normal photos.

Weight decay is set to 0.00001, and the discriminator and feature adapter learning rates are set at 0.0001 and 0.0002, respectively. Each dataset's training epochs are set to 160, its batch size to 4, and its image size to 288. The embedding dimension is set to 1024, the number of neurons in the discriminator's hidden layer is set to 1024, the number of discriminator layers is set to 2, the initial noise intensity is set to 0.015, Dropout is set to 0.2, the number of heads in the multi-head attention layer is set to 4, and the feature extraction backbone network is set to WideResnet50.

3.3 Metrics for Model Evaluation

The Area under the Receiver Operator Characteristic Curve (AUROC), also known as I-AUROC, is used to assess the model's performance in detecting abnormalities across complete images at the image level. By creating a ROC curve using the model's anomaly score for the entire image and calculating the area under this curve, I-AUROC is determined. We calculate the image's anomaly score by taking the highest anomaly score across all of the image's areas. The model's capacity to identify if an image is anomalous is measured by I-AUROC; a number nearer 1 denotes a higher ability of the model to differentiate between normal and anomalous images. Let S be the anomaly score, or the negative value of the discriminator's output. The picture anomaly score can then be calculated using this formula:

$$S_{AD}(x_i) = max(S) \quad (6)$$

The model's capacity to accurately identify anomaly locations is measured by computing the area under the ROC curve produced from the scores of each pixel in the anomaly map. This is known as the pixel-level anomaly localization metric. P-AUROC is the name of this measure. P-AUROC, as opposed to I-AUROC, concentrates on the anomaly score of every pixel in the picture. A higher number means that the model can more precisely detect each aberrant pixel in the image and locate anomaly zones.

The degree of overlap between the model-predicted anomaly region and the actual anomaly region is measured by a statistic called Per-Region Overlap (PRO). By computing the regional overlap rate between the anticipated anomaly area and the actual anomaly mask, this metric assesses how well the model locates anomaly regions. In particular, a higher PRO value means that the model is more accurate in identifying anomaly regions because there is a larger overlap between the model-predicted anomaly region and the actual anomaly region. In real-world applications, PRO works especially well for jobs that call for accurate localization of anomalous zones, which aids researchers in comprehending how well the model localizes in various locations. Let GTi stand for the actual region and Predi for the expected region. The formula used to calculate

PRO is:

$$PRO = \frac{1}{N}\sum_{i=1}^{N}\frac{|Pred_i \cap GT_i|}{|GT_i|} \quad (7)$$

3.4 Evaluation of Experimental Findings in Comparison

The Table 2 shows that enhanced SimpleNet demonstrates a distinctive balance of performance in anomaly detection and localization tasks, according to experimental results on the VisA-PCB4 datase. The enhanced model's pixel-level localization accuracy (P-AUROC) is quite near to the original model, despite a 4.4% drop in the image-level anomaly detection metric (I-AUROC) when compared to the original SimpleNet. At the same time, it significantly improves the Per-Region Overlap (PRO) statistic, which rises 9.8% from 79.0% to 88.8% in comparison to the original model. This fact suggests that in the fine-grained localization of anomalous locations, the enhanced model clearly has an advantage.

Table 2. Comparison of Model Results with the Original Model

Model	SimpleNet	Enhanced SimpleNet
I-AUROC	99.6%	95.2%
P-AUROC	97.5%	97.3%
PRO	79.0%	88.8%

The significant rise in PRO suggests that the enhanced model has more practical relevance in industrial fine inspection settings, even though the decrease in I-AUROC may be the result of a minor loss in the model's sensitivity to global abnormalities (such as recognizing large-scale structural faults).

The enhanced SimpleNet is especially well-suited for industrial situations where the accuracy of defect location is crucial, such as TP4056 charging module inspection, by giving up some of its worldwide detection capacity in exchange for more accurate anomaly localization capabilities.

Figure 3 displays the actual image of the TP4056 charging module on the left, the module's genuine anomaly mask in the middle, and the anomaly heatmap that the model predicted on the right. In contrast to the original model (upper image), the modified model (lower image) significantly improves the anomaly heatmap's performance, particularly in terms of localization accuracy. In particular, the overlap rate with the genuine anomaly mask is substantially greater in the anomaly heatmap produced by the enhanced model. By enabling more accurate detection of anomaly zones and lowering false positives and false negatives, this enhancement shows that the model's capacity to identify and locate solder joint faults has been much improved. The enhanced model's accuracy and resilience have also been enhanced, and it can now more effectively capture minor defect details, particularly when complex backgrounds and noise are present. In conclusion, the localization of the anomaly heatmap has been improved as a result of the model's optimization, significantly increasing the dependability of solder joint failure and other anomaly identification.

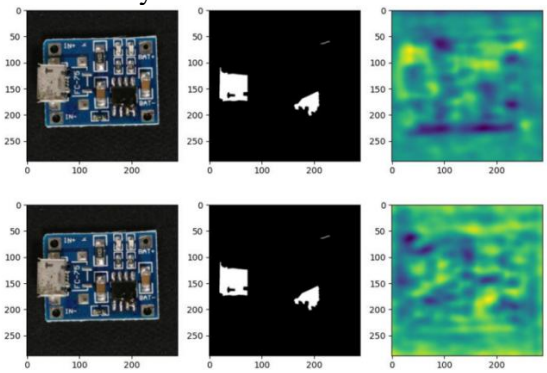

Figure 3. Comparison of Model Anomaly Detection Results

3.5 Study of Ablation

The SimpleNet model with only the MHA module added, the SimpleNet model with both MHA and residual connection added, and the complete model from this paper (SimpleNet with MHA, residual connection, and dynamic noise) were all trained and tested on the VisA-PCB4 dataset under the same hardware conditions. Table 3 displays the comparison of anomaly detection effects.

The model performs best in the image-level detection task (I-AUROC 98.3%) when the multi-head attention mechanism is the only addition, according to the experimental data in Table 3. However, its pixel-level localization ability (P-AUROC 96.1%) and region overlap rate (PRO 81.1%) are much lower than those of the fully improved model. This suggests that while MHA lacks restrictions on local details, it improves the perception of global anomalies through improved feature correlation.

I-AUROC marginally declined to 97.7% after residual connections were added, but PRO sharply declined to 60.3%, suggesting that the straightforward residual design may worsen feature redundancy and impair localization

accuracy.

The model performs best in P-AUROC (97.3%) and PRO (88.8%) when MHA, residual connections, and dynamic noise reduction are added all at once, but I-AUROC (95.2%) suffers as a result. This suggests that the dynamic noise technique uses two-phase optimization to successfully balance localization and detection.

Table 3. Comparison of the Outcomes of the Ablation Experiment

Model	MHA	Residual Connection	Dynamic Noise	I-AUROC (%)	P-AUROC (%)	PRO (%)
Ours	✓	✓	✓	95.2	97.3	88.8
Other 1	✓	✓		97.7	89.4	60.3
Other 2	✓			98.3	96.1	81.1

4. Conclusion

This research suggests an unsupervised anomaly detection technique based on an enhanced SimpleNet with the goal of meeting the industry requirement for surface defect identification of the TP4056 charging module. The model's detection accuracy and localization capability for subtle defects are greatly enhanced by implementing a dynamic noise adjustment strategy to increase the adaptability of anomaly feature generation, combining residual connections to optimize the gradient propagation path, and introducing a Multi-Head Attention (MHA) mechanism to enhance feature interaction capability. According to experiments, the enhanced model performs 97.3% and 88.8% on the VISA-PCB4 dataset in terms of pixel-level AUROC and Per-Region Overlap (PRO) criteria, respectively. The PRO metric grew by 9.8% in comparison to the original model, confirming its usefulness in situations involving fine industrial inspection. By striking a balance between local detail restrictions and global anomaly perception, the enhanced model offers a more effective solution for defect localization in TP4056 charging modules, despite a modest drop in image-level AUROC.

Future research will concentrate on improving the model's global detection performance even more, investigating lightweight deployment options to increase real-time detection efficiency, and investigating multi-modal feature fusion methodologies to handle a variety of flaws in intricate industrial settings. This study offers a novel method for detecting industrial anomalies and has significant implications for enhancing the safety and quality of electronic component production.

References

[1] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM computing surveys (CSUR), 2009, 41(3): 1-58.

[2] Liu J, Xie G, Wang J, et al. Deep industrial image anomaly detection: A survey. Machine Intelligence Research, 2024, 21(1): 104-135.

[3] Wang Baoyu. Research on Solder Joint Defect and Component Detection Method Based on Deep Learning. Nanchang University, 2020.

[4] Zhong Zhuoran. Research and Application of Artificial Intelligence Image Processing Technology Based on Deep Learning. Internet Weekly, 2025, (14): 50-52.

[5] Guo Jialin, Zhi Min, Yin Yanjun, et al. A Survey of Hybrid Models of CNN and Visual Transformer in Image Processing. Journal of Computer Science and Exploration, 2025, 19(01): 30-44.

[6] Kim Y. Convolutional Neural Networks for Sentence Classification. Eprint Arxiv, 2014.

[7] Roth K, Pemula L, Zepeda J, et al. Towards Total Recall in Industrial Anomaly Detection. 2021. DOI:10.48550/arXiv.2106.08265.

[8] Defard T, Setkov A, Loesch A, et al. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. 2021.

[9] Liu Z, Zhou Y, Xu Y, et al. Simplenet: A simple network for image anomaly detection and localization. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 20402-20411.

[10] Gong D, Liu L, Le V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1705-1714.

[11] Ristea N C, Madan N, Ionescu R T, et al.

Self-supervised predictive convolutional attentive block for anomaly detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 13576-13586.

[12]Zavrtanik V, Kristan M, Skočaj D. Reconstruction by inpainting for visual anomaly detection. Pattern Recognition, 2021, 112: 107706.

[13]Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[14]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in neural information processing systems. 2017, 30.

[15]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[16]Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 2014, 15(1): 1929-1958.

