Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 11 Issue 42 (March 2026) PP. 16-31

JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)
www.gaexcellence.com/jistm

DEFECT ENHANCEMENT DETECTION METHOD IN COMPLEX BACKGROUNDS BASED ON MIXED ATTENTION

Changyin Cheng[1]*, Mohammad Nazir Ahmad[2]

[1]Kuala Lumpur University of Science and Technology, Kajang, Selangor Darul Ehsan, Malaysia
243925343@s.klust.edu.my https://orcid.org/0009-0003-9457-0978
[2]Kuala Lumpur University of Science and Technology, Kajang, Selangor Darul Ehsan, Malaysia
Nazir@s.klust.edu.my https://orcid.org/0000-0003-3639-1157

Article Info:

Article history:
Received date: 24.12.2025
Revised date: 05.01.2026
Accepted date: 25.01.2026
Published date: 01.03.2026

To cite this document:

Changyin. C., & Ahmad, M. N., (2026). Defect Enhancement Detection Method in Complex Backgrounds Based on Mixed Attention. *Journal of Information System and Technology Management, 11* (42), 16-31.

DOI: 10.35631/JISTM.1142002

Abstract:

This paper proposes a defect enhancement detection method based on Mixed attention, aiming to address the difficulty of feature extraction caused by complex background interference in metal surface defect detection. The core of this method is the design of a lightweight Mixed Attention Module (MAM). This module integrates channel attention and spatial attention mechanisms in parallel, working collaboratively at multiple levels: channel attention adaptively recalibrates channel feature responses by modeling the interdependencies between feature channels, emphasizing feature maps related to defects; spatial attention focuses on key spatial locations in the feature maps, generating spatial weight masks to highlight defect regions and suppress texture and noise interference from irrelevant backgrounds. Simultaneously, the module employs an efficient structural design, achieving effective capture and fusion of multi-scale contextual information without introducing significant computational overhead, thereby enhancing the discriminative representation of defect features in complex backgrounds. Experimental results demonstrate that this method achieves significant improvements in detection accuracy (mAP) on the NEU-DET and GDUT-DET public metal surface defect datasets.

Keyword:

Attention Mechanism, Feature Enhancement, Defect Detection, Mixed Attention, Lightweight Network.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

Introduction

Metal surface defect detection is crucial in intelligent manufacturing. However, due to the complex production environment and varied metal surface textures, acquired images often suffer from severe background interference, making it difficult to distinguish defects from the background. Traditional methods are easily affected by interference during feature extraction, leading to decreased detection accuracy. Attention mechanisms, by adaptively adjusting feature weights, can enhance useful information and suppress noise, demonstrating good performance in defect detection in recent years. However, existing attention modules often rely on global pooling and fully connected structures, resulting in information loss, large parameter counts, and insufficient modeling of long-range dependencies. Attention mechanisms can automatically learn to adjust the weights of each element of the input features, thereby suppressing interference and enhancing useful information. Several excellent attention modules have been designed, but existing attention mechanisms still have some problems that need to be addressed. The main issue is that attention mechanisms typically use only single-scale global pooling operations when acquiring channel features. This method only obtains partial global information, losing much spatial structural information, which is not conducive to comprehensively judging the importance of each channel of the original features.

Despite the promising application of attention mechanisms in metal surface defect detection, several fundamental research gaps persist in existing methods:

a) Incomplete feature characterization: Most attention modules rely on single-scale global pooling, which leads to the loss of spatial structural and positional information. This limitation hinders the accurate distinction and localization of defects against complex, textured backgrounds.
b) Parameter inefficiency: The prevalent use of fully connected layers or complex multi-branch structures significantly increases model complexity, contradicting the practical need for lightweight, deployable solutions in industrial settings.
c) Suboptimal attention fusion: While integrated modules (e.g., CBAM, BAM) combine channel and spatial attention, their sequential or parallel designs often lack a collaborative, multi-source feature enrichment mechanism. They fail to simultaneously leverage diverse information streams—global statistics, structural cues, and precise location embeddings—for a holistic feature recalibration.

To bridge these gaps, this paper proposes a novel Mixed Attention Module (MAM) and a corresponding defect enhancement detection method for complex backgrounds. The fundamental outcome of this research is not merely another attention variant but a systematic lightweight framework that integrates channel and spatial attention in a concatenated manner. MAM is designed to effectively enhance valid information and suppress interfering noise within features, thereby optimizing feature quality and improving detection accuracy. By integrating and refining multi-scale, multi-type features, it achieves superior discriminative power with minimal computational overhead. Crucially, we demonstrate that embedding this

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31
carefully designed module into existing, never-before-tried basic detection backbones (such as ResNet-50 within Faster R-CNN, TOOD, etc.) yields consistent and significant performance gains, proving the universal applicability and effectiveness of our approach. The main contributions of this paper are as follows:

a) A Mixed attention module consisting of channel attention and spatial attention concatenated is proposed.
b) A multi-scale channel feature extraction mechanism is introduced, fusing global, structural, and positional information.
c) Cascaded dilated convolutions are used to replace fully connected layers to achieve lightweight long-range dependency modeling.
d) The effectiveness and generalization ability of MAM are verified on multiple public datasets.

Literature Review

Overview of Attention Mechanisms

Attention mechanisms, through adaptive weight adjustment, enhance useful information in features and suppress interference, and are widely used in computer vision tasks. Depending on their scope, they can be divided into channel attention and spatial attention.

Channel Attention Module

The Style-based Recalibration Module (SRM) improves upon the Sense Module (SE) in two dimensions: feature compression and activation(Ge et al., 2021) . Its structure, shown in Figure 1, integrates global average pooling and global standard deviation pooling to enhance the capture of global features. Simultaneously, it replaces the traditional fully connected structure with a lightweight channel-based fully connected layer, thereby reducing computational complexity. However, the method used by SRM to generate the final channel weights is relatively simple, making it difficult to effectively establish complex dependencies between channels(Jeong et al., 2017). This limitation, to some extent, affects the module's performance improvement potential.

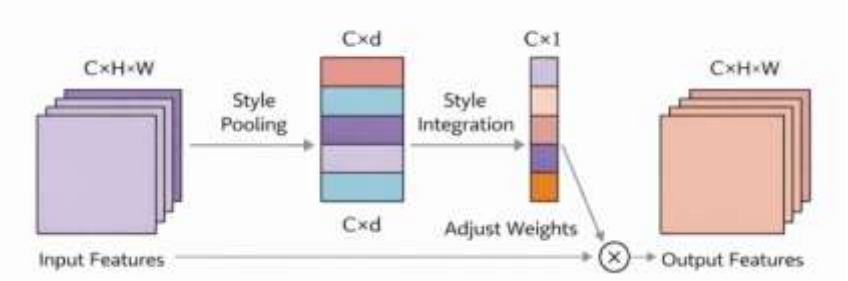

Figure 1: Schematic Diagram of SRM Attention Module

To address the spatial information loss caused by the SE module's reliance on only global features, the attention module SPA-Net introduces a multi-scale channel information modeling mechanism. Its structure, as shown in Figure 2, captures both local and global contextual semantic information through a spatial pyramid composed of three resolutions of adaptive average pooling. Subsequently, the one-dimensional channel features extracted from each branch are concatenated along the channel dimension, embedding spatial structure information into the channel representation. Finally, a fully connected network is used to establish inter-channel dependencies and generate attention weights. However, this concatenation combined with fully connected processing significantly increases the module's parameter count.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

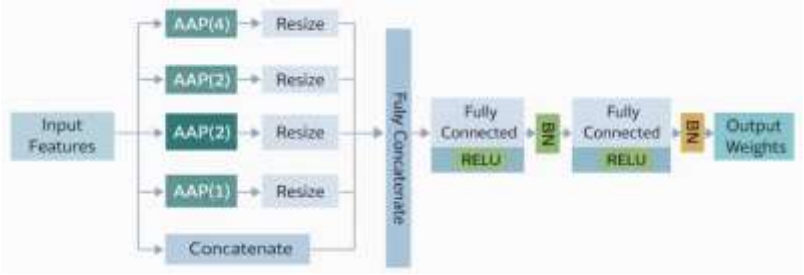

Figure 2: Schematic Diagram of SPA Attention Module

In summary, the aforementioned channel attention modules optimize the channel information of the original features from different dimensions, playing a crucial role in extracting high-quality features and laying an important foundation for the development of attention mechanisms. Numerous studies have successfully applied these channel attention modules to various metal surface defect detection tasks, significantly improving recognition accuracy(Hussain, 2023). However, existing channel attention mechanisms still have significant limitations: on the one hand, they primarily focus on global or large-scale semantic information, often ignoring detailed spatial location cues, which is detrimental to the accurate spatial localization of metal surface defects; on the other hand, while utilizing rich channel information, they often struggle to effectively control the number of parameters, limiting their application potential in lightweight scenarios.

Integrated Attention Module

The integrated attention mechanism adds a spatial attention mechanism to the channel attention mechanism. It not only focuses on "what" the target is, which is beneficial for classifying metal surface processing defects, but also focuses on "where" the target is located, which is beneficial for locating metal surface processing defects. A brief analysis of some representative integrated attention modules follows.

The Integrated Attention Module (CBAM) can be viewed as a spatial attention module chained onto the SE channel attention module(Du et al., 2019) , as shown in Figure 3. The spatial attention module first performs average pooling and max pooling along the channel direction to generate two single-channel feature maps. These two maps are then concatenated along the channel dimension and fused through convolution. Finally, a spatial weight map is generated using the sigmoid activation function. This mechanism effectively enhances the feature response of the target region while suppressing irrelevant background information, achieving adaptive recalibration of features in the spatial dimension.

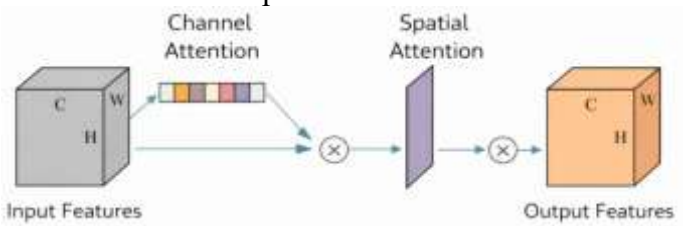

Figure 3: Schematic Diagram of CBAM Attention Module

The attention module BAM adopts a parallel structure to fuse channel and spatial information(Wang et al., 2024), as shown in Figure 4. The input features are processed through

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

parallel channel attention branches and spatial attention branches to generate corresponding channel weights and spatial weights. The two weights are then multiplied to obtain the final attention weight map, thereby achieving coordinated adjustment of feature information in both channel and spatial dimensions(Zhao et al., 2023).

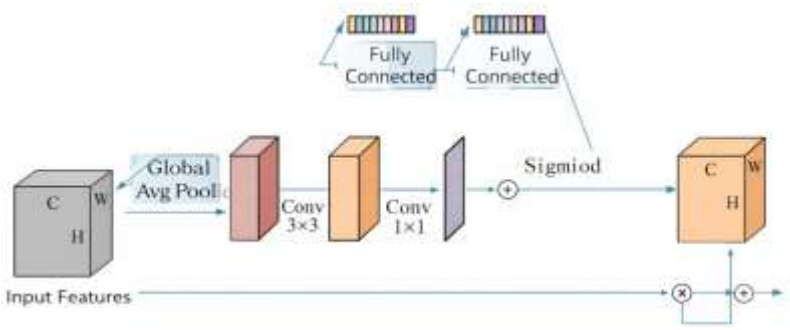

Figure 4: Schematic Diagram of BAM Attention Module

The above analysis shows that the integrated attention module adjusts features from both channel and spatial dimensions, which can more comprehensively enhance useful information and suppress interference. Therefore, it is generally more effective in improving the detection accuracy of metal surface defects than using channel attention alone (Li et al., 2022). However, integrated attention inevitably introduces more parameters, with the fully connected layer in the channel attention part being the main source of the increased parameter count. Furthermore, although some attention modules have achieved good results by introducing channel features containing multiple types of information to comprehensively evaluate the importance of each channel(Zong et al., 2023), the commonly used fully connected processing method further exacerbates the increase in parameter count. Therefore, how to design an efficient and lightweight feature processing mechanism to achieve a balance between accurate analysis of channel importance and parameter efficiency while enriching channel feature information remains a key problem that needs to be solved.

Application Of Attention Mechanism in Defect Detection

In recent years, attention mechanisms have shown outstanding advantages in feature extraction. By suppressing interference information and enhancing key information, they have become an effective tool for dealing with the interference of complex backgrounds on metal processing surfaces (Ma et al., 2019). Numerous studies have confirmed that integrating attention mechanisms into metal surface defect detection models can significantly improve detection accuracy. For example, Cha et al. (Ma et al., 2024) introduced the SE attention module into Faster R-CNN to improve the detection performance of wheel hub surface defects; Block et al. (Liang et al., 2024) used an improved AW-CBAM attention module in the YOLOv5 backbone network to enhance the feature extraction capability of aero-engine surface defects. Akhyar et al. (Akhyar et al., 2023) integrated ECA attention into YOLOv7 to effectively suppress the interference of complex backgrounds on steel surfaces; similar studies on enhancing feature discriminativeness through attention mechanisms have emerged, fully verifying its practical value and development potential in the field of metal surface defect detection.

Existing research has clearly demonstrated that attention mechanisms can effectively suppress interference information during feature extraction, thereby significantly improving the accuracy of metal surface defect detection. Based on this understanding, this paper aims to embed attention mechanisms into basic detection networks to optimize their detection

performance(Wang et al., 2025). After comprehensively analyzing the advantages and limitations of existing mainstream attention modules, this paper designs a novel Mixed Attention Module (MAM). This module integrates channel attention and spatial attention in a concatenated manner. Its core design idea draws on a variety of classic attention mechanisms: First, inspired by modules such as SRM and SPA, MAM extracts eight channel features to comprehensively evaluate and adjust the importance of each channel; second, drawing on the idea of CA attention modules(Redmon & Farhadi, 2018), it embeds positional information in the height and width directions into channel features to enhance spatial perception; furthermore, inspired by SRM and ECA modules(Szegedy et al., 2017), it uses one-dimensional dilated convolutions with different dilation rates to process multi-channel features in parallel, thereby replacing the traditional fully connected operation. This method can not only simultaneously establish short-range and long-range channel dependencies, avoiding information loss caused by channel compression, but also significantly reduce the number of parameters, achieving a balance between efficiency and performance.

The Need for Specialized Attention in Basic Detection Networks

Despite the proven benefits of attention mechanisms, a critical gap remains existing basic detection networks (e.g., Faster R-CNN, YOLO, TOOD with backbones like ResNet) are fundamentally ill-equipped for metal defect detection in complex backgrounds. These networks lack the intrinsic ability to:

a) Dynamically distinguish subtle defects from highly textured, noisy backgrounds.
b) Adaptively integrate multi-scale contextual information crucial for varied defect sizes.
c) Efficiently suppress pervasive background interference during feature extraction.

This limitation necessitates the embedding of specialized attention modules. However, prior works predominantly employ generic attention modules (e.g., SE, CBAM) as universal enhancements. While beneficial, they are not optimized for the specific challenges of metal surface inspection and often introduce efficiency trade-offs.

Therefore, the key innovation of this work is not merely embedding attention, but embedding a purpose-designed module—the Mixed Attention Module (MAM). MAM is novel because it is specifically engineered to compensate for the aforementioned shortcomings of basic networks:

a) Its multi-source channel attention directly addresses the lack of discriminative feature prioritization.
b) Its lightweight, cascaded-dilated design efficiently captures long-range context without the parameter cost of traditional modules.
c) This targeted combination of capabilities, optimized for complex metal backgrounds, has not been explored in previous attempts to enhance basic detectors.

Thus, MAM represents new knowledge: a task-optimized, lightweight attention plugin that fundamentally upgrades the capability of standard detection backbones for this demanding application.

Research Gap and Motivation

The reviewed literature establishes the value of attention mechanisms but reveals a clear research gap: there is a lack of a lightweight, yet comprehensively attentive module that efficiently fuses multi-scale channel features (global, structural, positional) with refined spatial

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

attention without incurring substantial parametric costs. Existing modules often trade off richness of information for efficiency, or vice-versa.

Therefore, the core motivation of this work is to answer: Can we design an attention module that fundamentally enhances feature representation for defect detection by synergistically combining diverse information sources, while maintaining a lightweight profile suitable for integration into any existing detection network? The proposed MAM is our affirmative answer, providing new knowledge in the form of: (a) an 8-branch multi-information channel feature extraction strategy, and (b) a cascaded dilated convolution-based lightweight fusion mechanism to replace parameter-heavy FC layers.

Methodology

Overall Structure of Mixed Attention Module (MAM)

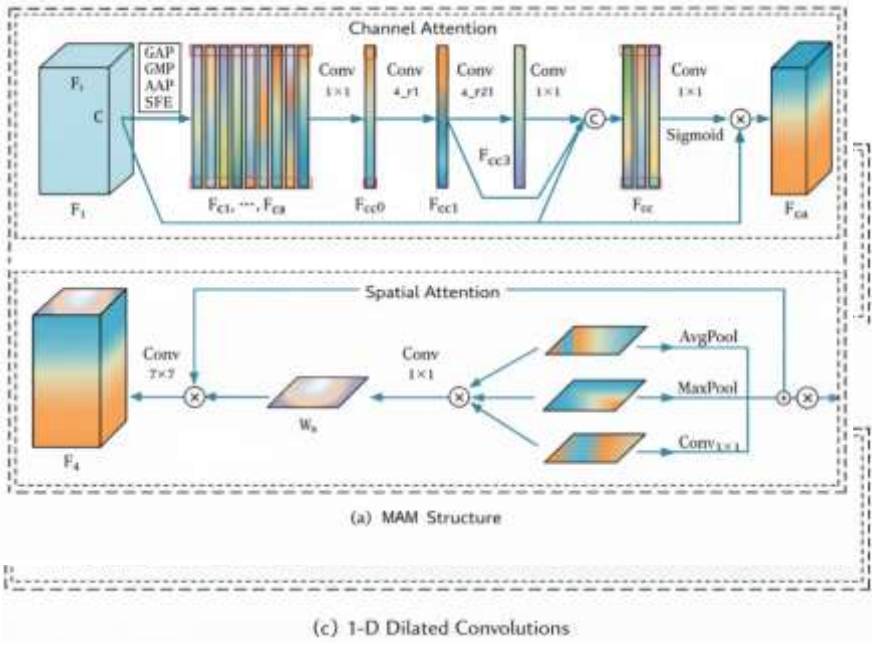

Figure 5: MAM Attention Module

MAM is composed of channel attention and spatial attention concatenated, as shown in Figure 5. The channel attention part extracts features from 8 channels, while the spatial attention part adds a 1×1 convolution to CBAM to enhance spatial feature learning.

Channel Attention Module

Multi-Channel Information Acquisition

Analysis of existing representative attention modules reveals that classic methods such as SRM, SPA, and CBAM often optimize attention regulation by increasing the number of channel features. The MAM module employs four methods to extract eight channel features to comprehensively regulate the importance of each channel in the input feature set. First, global average pooling and global max pooling are used to extract channel features Fc1 and Fc2, which possess global characteristics. These features excel at identifying large targets but lack

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

spatial structure information. To embed spatial information, MAM introduces 2×2 adaptive average pooling to generate a feature map containing spatial structure, which is then uniformly divided into four one-dimensional features Fc3-Fc6, thus obtaining features with coarse structural information, beneficial for detecting medium-sized targets. Furthermore, inspired by the CA attention mechanism, MAM uses a designed strip feature extraction module to perform one-dimensional average pooling and max pooling along the height and width directions of the feature map, generating channel features Fc7-Fc8. This embeds precise location information into the channel representation, enhancing the ability to perceive the spatial location of targets.

Figure 5(b) illustrates the structure of the MFE module. This section uses feature Fc7 extraction as an example to explain the process: First, one-dimensional pooling is performed along the width direction of the input feature to obtain a two-dimensional feature of size C×H, where the height H varies with the input image size. To facilitate subsequent processing, one-dimensional average pooling is used to compress this feature to a fixed size C×M (M is a preset parameter), and then further compressed through one-dimensional convolution to finally generate the one-dimensional channel feature Fc7, which has embedded positional information in the height direction. Similarly, by applying one-dimensional average pooling and max pooling along the height and width directions respectively, feature Fc8 can be obtained. In the application of this chapter, when embedding the MAM module into the last two convolutional groups of ResNet-50, the value of M is set to 8 and 4 respectively, depending on the input image size.

Lightweight Processing Of Multi-Channel Information

After obtaining eight sets of channel features (Fc1-Fc8), using a method similar to SPA attention—that is, first concatenating the multi-channel features and then processing them with a fully connected layer—would introduce a significant increase in the number of parameters. Therefore, how to efficiently and lightweightly integrate these features becomes a key issue. MAM employs a lightweight design to process the above features, ultimately generating channel attention weights that fuse multi-source information. Specifically, Fc1-Fc8 are first concatenated along the feature dimension to form a feature of size C×8, and then compressed into a single-channel comprehensive feature Fcc0 through a one-dimensional convolution (Conv1×1). This process can be represented as:

$$F_{cc0} = Conv_{1\times1}(Cat(F_{c1}, F_{c2}, \dots, F_{c8})) \quad (3\text{-}1)$$

Where Cat represents the feature concatenation operation, and Conv1×1 represents a 1×1 convolution. Inspired by the ECA attention mechanism, this method uses one-dimensional convolution instead of fully connected layers to process feature Fcc0, avoiding information loss caused by channel compression and significantly reducing the number of parameters. Unlike the ECA module, which uses only a single convolution kernel, MAM cascades three one-dimensional dilated convolutions with a kernel size of 4 and different dilation rates to process Fcc0 step by step in order to simultaneously establish short-range and long-range channel feature dependencies. As shown in Figure 5(c), the dilation rates of these three dilated convolutions are set to 1, 4, and 21, respectively, and their sampling range covers channel dependencies from local to global. At the same time, this cascaded structure effectively avoids feature undersampling problems that may be caused by grid effects through carefully designed dilation rate configurations. After Fcc0 is processed step by step by the above dilated convolutions, features Fcc1, Fcc2, and Fcc3 are obtained in sequence, and their calculation process is shown below:

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

$$F_{cc1} = Conv_{4\_r1}(F_{cc0}) \quad (3\text{-}2)$$

$$F_{cc2} = Conv_{4\_r4}(F_{cc1}) \quad (3\text{-}3)$$

$$F_{cc3} = Conv_{4\_r21}(F_{cc3}) \quad (3\text{-}4)$$

Where Conv4_r4 represents a 4-kernel, 4-dilated one-dimensional convolution, and so on. Finally, features Fcc1, Fcc2, and Fcc3 are concatenated and input into a 1×1 convolution, processed using the Sigmoid function to obtain the channel attention weights Wc. This calculation process can be expressed as:

$$W_c = Sig(Conv_{1\times1}(Cat(F_{cc1}, F_{cc2}, F_{cc3}))) \quad (3\text{-}5)$$

Here, Sig represents the Sigmoid function, and Cat corresponds to the symbol © in the diagram, indicating feature concatenation. The Sigmoid function is a commonly used activation function. Because its output value is between (0, 1), it is frequently used in attention mechanisms. The expression for the Sigmoid activation function is:

$$Sigmoid(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3\text{-}6)$$

After obtaining the channel attention weights Wc, they can be used to adjust the importance of each channel of the input feature. The calculation process is as follows:

$$F_{ca} = F_i \otimes W_c \quad (3\text{-}7)$$

Here, Fca represents the output feature after channel attention adjustment, Fi is the original input feature, and ⊗ represents element-wise multiplication. At this point, the channel attention part of the MAM module is complete. This module extracts rich channel features, comprehensively evaluates the importance of each channel, and utilizes a lightweight cascaded dilated convolutional structure to model both local and long-range channel dependencies, thereby achieving adaptive adjustment of the input feature channel weights.

Spatial Attention Module

The spatial attention module aims to enhance the location information in the features, focusing on solving the problem of "where is the target?". As shown in Figure 5, a spatial attention module is cascaded after the channel attention in the MAM to further supplement the spatial location information. This module is mainly an improvement on the spatial attention part (SAM) in CBAM. To control the overall module complexity, this chapter does not make significant changes to the SAM structure, but only adds a 1×1 convolutional layer to extract additional spatial information, thereby improving the representational ability of spatial attention. After concatenating the three spatial features, they are fused through a 7×7 convolutional layer, and the spatial attention weights Ws are generated by the Sigmoid function. The calculation process can be expressed as follows:

$$W_s = Sig(Conv_{7\times7}(Cat(AvgPool(F_{ca}), MaxPool(F_{ca}), Conv_{1\times1}(F_{ca})))) \quad (3\text{-}8)$$

Where AvgPool is average pooling and MaxPool is max pooling. After obtaining the spatial attention weights Ws, they can be used to adjust the importance of features in the spatial location direction to highlight the target location. The calculation process is as follows:

$$F_a = F_{ca} \otimes W_s \quad (3\text{-}9)$$

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

Here, Fa represents the final output feature after adjustment by the complete MAM attention module. At this point, the entire computational process of the MAM module is complete. This module, by concatenating channel attention and spatial attention, collaboratively optimizes the original input from both feature semantics and spatial location dimensions, effectively suppressing background interference while enhancing key information. Furthermore, MAM introduces only a very small number of parameters while achieving the above functions, maintaining the model's lightweight characteristics.

Experiments

Dataset and Evaluation Metrics

This chapter validates the proposed method using two public datasets for metal surface defects: the NEU-DET dataset for hot-rolled steel surface defects and the GDUT-DET dataset for aluminum profile surface defects. These datasets will be briefly introduced below.

NEU-DET

The dataset primarily used in this study is the NEUDET dataset for hot-rolled steel strip surface defects, provided by Northeastern University. This database contains 1800 images of 200×200 pixels each, representing six different types of surface defects, with 300 samples for each type. These six typical surface defects are: rolled-in scale (RS), patches (Pa), crazing (Cr), pitted surface (PS), inclusion (In), and scratches (Sc). NEU-DET was randomly divided into a training set and a test set, with the training set containing 1270 images and the test set containing 530 images.

GDUT-DET

This dataset comes from the training set of the semi-final round of the 2018 Guangdong Industrial Intelligent Manufacturing Big Data Innovation Competition for aluminum profile surface defect recognition. The dataset contains 3000 images, covering a total of 10 defect types: non-conductive (bdd), scratches (ch), corner defects (jwld), orange peel (jp), defects (ld), jetting (pl), paint bubbles (qp), pitting (qk), discoloration (zs), and dirt spots (zd). GDUT-DET was split into training and testing sets in a 7:3 ratio for training and testing the detection network, respectively.

This paper uses the MS COCO accuracy evaluation metric for object detection to assess the accuracy of each detection network. The most important metric is average precision (AP). In object detection, the intersection-over-union (IoU) ratio is generally used to represent the accuracy of target location prediction.

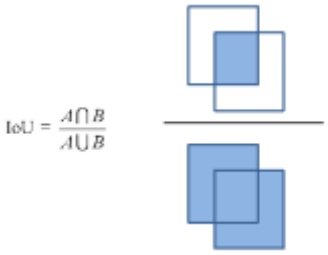

Figure 6: IoU Schematic Diagram

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

Figure 6 illustrates the IoU, representing the overlap between the bounding boxes (BBoxes) predicted by the detection network and the original ground truth (GT), i.e., the ratio of their intersection to their union. A higher IoU value indicates more accurate target localization by the BBox. Generally, an IoU $\geqslant 0.5$ is considered a detected target.

Experimental Setup

To ensure the fairness and reproducibility of the network comparisons, all experiments in this paper were conducted using the PyTorch deep learning framework and the OpenMMLab open-source algorithm platform. The experimental hardware uniformly used a Tesla V100 32G GPU as the training platform. During training, the optimizer used the stochastic gradient descent algorithm by default, with an initial learning rate of 0.002 and a phased decay strategy: the learning rate was reduced to one-tenth of its original value at the 26th and 30th training epochs. The batch size was fixed at 2, and the total number of training epochs was 40, meaning the training set was traversed 40 times.

Comparison Experiment of Detection Networks

Table 1: Enhancement Effect of MAM On the Detection Network

Detection Network	Backbone	AP	AP_{50}	APS	APM	APL	Model Size (M)
YOLOv3	Darknet-53	28.4	66.2	33.8	23.7	27	470.02
SSD	VGG-16	31.8	69.1	20	26.7	39.6	191.7
GA-Faster R-CNN	ResNet-50	39.1	75.5	34.4	32.5	46.6	319.65
Faster R-CNN	ResNet-50	38.4	73.1	28.8	32	45.8	315.1
TOOD	ResNet-50	37.8	71.5	26.4	30.9	48.3	246.76
MAM-TOOD	ResNet-50	38.8	72.9	35.6	31.5	49.1	246.9
DDOD	ResNet-50	39	72.5	31	32.1	46.8	248.65
MAM-DDOD	ResNet-50	39.8	74.2	33.9	32.7	47.1	248.93

To verify the effectiveness of the designed Mixed Attention Module (MAM), this section embeds it into several classic detection networks (Faster R-CNN and TOOD) and conducts experiments on the NEU-DET surface defect dataset. The experiments follow the default training settings, with YOLOv3 trained for 240 epochs. The detection results of each network on the NEU-DET dataset are shown in Table 1.

Experimental results after applying the MAM module to the TOOD network show that its core accuracy (AP) improved from 37.8% to 38.8%, an increase of 1.0 percentage point; meanwhile, AP50 reached 72.9%, an increase of 1.4 percentage points, and all other accuracy indicators showed significant improvement. In terms of parameter quantity, the model only increased by 0.14M parameters after adding MAM (from 246.76M to 246.9M), a very small increase.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

Similarly, embedding MAM into the DDOD network systematically improved its average accuracy (AP) and the detection performance of defects at all scales (especially small-sized defects), with the APS increasing by 2.9 percentage points. These results consistently demonstrate that the MAM attention mechanism, by effectively enhancing key features and suppressing background interference, can significantly improve the performance of various detection networks on hot-rolled steel surface defect tasks, and is characterized by high parameter efficiency.

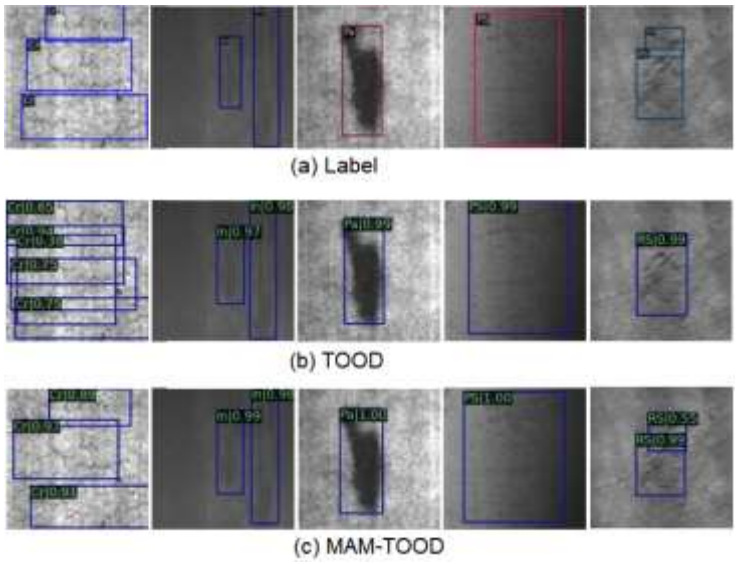

Figure 7: Examples Of Detection Before and After Using MAM On NEU-DET

Figure 7 shows several detection examples from the NEU-DET dataset, visually demonstrating the enhancement effect of the MAM module on detection performance. The selected hot-rolled steel images have complex backgrounds and significant interference, making it difficult to clearly distinguish defects from the background. Using TOOD as the detection network, compared with the original model, the figure clearly shows that MAM-TOOD can more accurately identify and locate various defects on the surface of hot-rolled steel.

Comparative Experiment of Attention Modules

To verify the effectiveness of the proposed MAM attention module, this section presents a comparative experiment with several representative and excellent attention modules. The experiment was conducted on the NEU-DET surface defect dataset, and Faster R-CNN, with ResNet-50 and FPN as the backbone, was used as the base detection network. Detailed comparison results are shown in Table 2.

Table 2: Attention Comparison Experiment

Detection Network	AP	AP_{50}	APS	APM	APL	Model Size (M)
Faster R-CNN	38.4	73.1	28.8	32.0	45.8	315.1
+ SE	38.9	75.2	29.5	32.5	48.5	333.3

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

+ CBAM	39.1	76	35.3	32.5	46.6	333.32
+ MAM	39.4	74.4	30.9	32.8	46.7	315.42

Based on the experimental results shown in Table 2, the basic detection network equipped with the MAM module achieved the best performance in terms of the core accuracy index (AP), reaching 39.4%. Meanwhile, the CBAM and SE attention modules also significantly improved the detection accuracy. However, in terms of model parameter count, MAM only introduced 0.32M additional parameters. In summary, compared to other mainstream attention modules, the proposed MAM module can significantly improve the detection accuracy of hot-rolled steel surface defects with only a slight increase in model complexity, demonstrating superior overall performance.

Ablation Test

To deeply analyze the MAM attention module and verify the effectiveness of its channel attention and spatial attention sub-modules, this section uses the classic Faster R-CNN as the base detection network and conducts ablation experiments on the NEU-DET dataset. The results are shown in Table 3. The experiments show that when only the channel attention module in MAM is used, the average precision (AP) increases from 38.4% to 39.1%, demonstrating that fusing multiple channel information can effectively optimize feature representation from the channel dimension. However, while APM and APL improve, APS slightly decreases. This may be because the channel features extracted by this module have a larger receptive field, making them more suitable for detecting large defects, but relatively insensitive to small defects. Further introducing spatial attention further improves the AP to 39.4%, an increase of 0.3 percentage points, and all other accuracy metrics are improved, with APS significantly increasing by 2.5 percentage points, effectively compensating for the insufficient detection capability for small defects when only channel attention is used.

Table 3: Ablation Experiments Of MAM

Channel Attention	**Spatial Attention**	**AP**	**AP_{50}**	**APS**	**APM**	**APL**
✗	✗	38.4	73.1	28.8	32	45.8
✓	✗	39.1	74.2	28.4	32.7	46.3
✓	✓	39.4	74.5	30.9	32.8	46.6

Generalization Experiment

To verify the generalization ability of the designed MAM module to enhance useful features and suppress interference information in complex backgrounds, this chapter uses the aluminum profile surface defect dataset GDUT-DET for testing. The experiment is based on the classic two-stage detection network Faster R-CNN, with all settings remaining consistent with the previous section except for the dataset. The results are shown in Table 4: After introducing the MAM module, all accuracy metrics of Faster R-CNN are significantly improved. Specifically, the average detection accuracy (AP) increases from 31.9% to 34.6%, an increase of 2.7 percentage points; AP50 increases by 4.1 percentage points. In terms of multi-scale detection

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

performance, MAM improves the detection accuracy for small, medium, and large-sized defects by 3.1, 3.7, and 3.1 percentage points, respectively. This result further demonstrates that the MAM module can effectively improve the detection performance of aluminum profile surface defects through feature enhancement and interference suppression mechanisms, exhibiting strong cross-dataset generalization ability.

Table 4: Generalization Experiments of MAM On The GDUT-DET Dataset

Detection Network	AP	AP_{50}	APS	APM	APL
Faster R-CNN	31.9	46.6	6.4	10.9	32.5
MAM-Faster R-CNN	34.6	50.7	9.5	14.6	35.6

To visually demonstrate the detection enhancement performance of the MAM module on the GDUT-DET dataset, Figure 8 provides a visualization of detection examples for some categories. The aluminum profile image in the figure contains various surface defects, and the background interference is significant, making it difficult to distinguish the defects from the background. Compared to the original Faster R-CNN, MAM-Faster R-CNN can more accurately identify and locate defects in the image, while showing significant improvement in both false negatives and false positives.

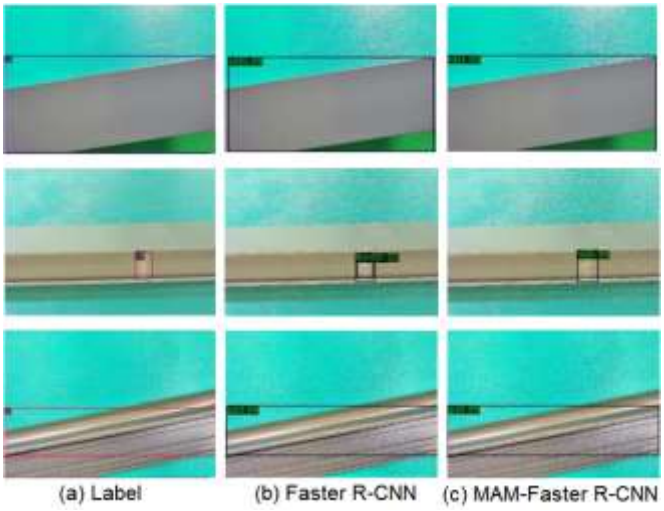

Figure 8: Detection Examples Before and After Using MAM

Conclusion

This paper addresses the fundamental challenge of extracting discriminative defect features under severe background interference in industrial inspection. To this end, we propose a defect enhancement detection method based on a novel Mixed Attention Module (MAM). The core contribution and new knowledge introduced by this work lie in the design of MAM itself—a lightweight, plug-and-play module with a unique dual-path architecture: (1) a multi-source channel attention path that concurrently extracts eight complementary features (encompassing global, structural, and positional information) for a holistic assessment of channel importance, and (2) a lightweight fusion mechanism that replaces parameter-heavy fully-connected layers

with a cascade of dilated 1D convolutions, enabling efficient modeling of both local and long-range channel dependencies.

The experimental results provide a clear answer to the fundamental research question: embedding MAM into existing basic detection networks (which have never been augmented with such a module) consistently and significantly improves their performance. This demonstrates that our work successfully improves existing detection methods by providing a superior, generalized feature enhancement tool. The evidence confirms that detection networks with MAM are superior to those without it, achieving notable gains in mean Average Precision (mAP), with particularly significant improvements for small defects (APS), on both the NEU-DET and GDUT-DET datasets.

In summary, by effectively enhancing salient features and suppressing background interference, MAM mitigates the feature degradation caused by complex backgrounds. The module introduces minimal parameters, facilitating network lightweighting and reduced computational resource demands. Specifically, on the NEU-DET and GDUT-DET datasets, the baseline detection network equipped with MAM achieved AP improvements of 1.1 and 2.8 percentage points, respectively, with other key metrics also showing substantial gains. Notably, MAM boosted the APS of the TOOD detector by 9.4 percentage points on NEU-DET, further validating its efficacy in detecting small-sized defects.

Acknowledgements:	The authors would like to express their sincere gratitude to Kuala Lumpur University of Science and Technology for providing the necessary resources and support throughout the course of this research. Special appreciation is extended to colleagues and peers who contributed valuable insights and constructive feedback, which greatly enhanced the quality of this paper.
Funding Statement:	No Funding
Conflict of Interest Statement:	The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the Journal of Information System and Technology Management (JISTM)
Ethics Statement:	This study did not involve any human participants, animals, or sensitive data requiring ethical approval. The authors confirm that the research was conducted in accordance with accepted academic integrity and ethical publishing standards.
Author Contribution Statement:	All authors contributed significantly to the development of this manuscript. Changyin Cheng was responsible for the conceptualization, methodology, and overall supervision of the study. Changyin Cheng handled data collection, analysis, and interpretation of results. Mohammad Nazir Ahmad contributed to the literature review, drafting, and critical revision of the manuscript. All authors read and approved the final version of the manuscript prior to submission.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 42 (March 2026) PP. 16-31

References

Akhyar, F., Furqon, E. N., & Lin, C.-Y. (2022). Enhancing precision with an ensemble generative adversarial network for steel surface defect detectors (EnsGAN-SDD). Sensors, 22(11), 4257. https://doi.org/10.3390/s22114257

Du, W., Shen, H., Fu, J., Zhang, G., & He, Q. (2019). Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning. NDT & E International, 107, 102144. https://doi.org/10.1016/j.ndteint.2019.102144

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. arXiv. https://doi.org/10.48550/arXiv.2107.08430

Hussain, M. (2023). YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines, 11(7), 677. https://doi.org/10.3390/machines11070677

Jeong, J., Park, H., & Kwak, N. (2017). Enhancement of SSD by concatenating feature maps for object detection. arXiv. https://doi.org/10.48550/arXiv.1705.09587

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv. https://doi.org/10.48550/arXiv.2209.02976

Liang, F., Zhao, L., Ren, Y., Wang, S., To, S., Abbas, Z., & Islam, M. S. (2024). LAD-Net: A lightweight welding defect surface non-destructive detection algorithm based on the attention mechanism. Computers in Industry, 161, 104109. https://doi.org/10.1016/j.compind.2024.104109

Ma, J., Hu, S., Fu, J., & Chen, G. (2024). A hierarchical attention detector for bearing surface defect detection. Expert Systems with Applications, 239, 122365. https://doi.org/10.1016/j.eswa.2023.122365

Ma, L., Xie, W., & Zhang, Y. (2019). Blister defect detection based on convolutional neural network for polymer lithium-ion battery. Applied Sciences, 9(6), 1085. https://doi.org/10.3390/app9061085

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv. https://doi.org/10.48550/arXiv.1804.02767

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1), 4278–4284. https://doi.org/10.1609/aaai.v31i1.11231

Wang, X., Ma, S., Wu, S., Li, Z., Cao, J., & Xu, P. (2025). Detection of surface defects in steel based on dual-backbone network: MBDNet-attention-YOLO. Sensors, 25(15), 4817. https://doi.org/10.3390/s25154817

Wang, Z., & Liu, W. (2024). Surface defect detection algorithm for strip steel based on improved YOLOv7 model. IAENG International Journal of Computer Science, 51(3), 308–317.

Zhao, C., Shu, X., Yan, X., Zuo, X., & Zhu, F. (2023). RDD-YOLO: A modified YOLO for detection of steel surface defects. Measurement, 214, 112776. https://doi.org/10.1016/j.measurement.2023.112776

Zong, R., Liu, Q., Wang, J., Jia, X., Qin, N., & Huang, D. (2024). A metal surface defect detection algorithm based on mixed supervised and cross stage partial darknet. In 2024 43rd Chinese Control Conference (CCC) (pp. 8106–8111). https://doi.org/10.23919/CCC63176.2024.10661652