Enhanced Image Segmentation-Based Detection Technique for X-ray Film Images of Weld Seams

Ruixiang Li[1], Shanwen Zhang[1,*], Lei Huang[2], Mingda Yang[1], Chengyu Hu[1]

[1]School of Electronic Information, Xijing University, Xi'an, China
[2]Tubular Goods Research Institute of CNPC, Xi'an, 710077, China
[*]Corresponding author

Abstract: *In industrial pipeline systems, quality monitoring of steel pipes and welds is a critical component to ensure safe operation. Utilizing deep learning techniques to analyze X-ray images can efficiently identify weld defects, such as porosity, cracks, and inclusions, significantly enhancing the accuracy and efficiency of non-destructive testing. In response to the existing weld defect detection models' insufficient feature extraction and lack of handling diversity, this paper introduces a weld defect identification method based on the Hierarchical Attention Fusion Network (HAFNet). Initially, a Dilated Hierarchical Attention Mechanism (DHAM) is employed to capture multi-scale global and local information, thereby enhancing the focus on key features of different scales and effectively addressing the issue of large intra-class variability and small inter-class differences in defects. Subsequently, a Residual Fusion Module (RFM) is introduced, which adaptively learns the feature weights of different encoding layers and fully utilizes contextual information during the decoding phase to suppress the complex background interference of weld images. Finally, through a Multi-Level Feature Fusion Module (MFFM), the decoded network's multi-layer features are strengthened by a fusion mechanism, enhancing the interaction and complementarity between different levels of features, reducing the model's sensitivity to noise and non-critical information, and further enhancing the model's recognition accuracy and robustness.*

Keywords: *Weld Defect Detection, Dilated Hierarchical Attention Mechanism, Residual Fusion Module, Multi-Level Feature Fusion Module*

1. Introduction

1.1. Background and significance of weld identification technology

Welding technology is a critical component of modern manufacturing, significantly impacting product performance and quality. Due to the complexity of the welding process and the variability of parameters, various weld seam defects often occur, affecting the strength, performance, and lifespan of products. Hence, quality inspection of welded products is crucial. Currently, non-destructive testing methods such as X-ray, ultrasonic, magnetic particle, and liquid penetrant inspection are widely used, but these traditional methods have drawbacks like slow speed, low accuracy, and high dependence on manual evaluation. With the advancement of artificial intelligence and machine learning, image processing techniques based on deep learning have demonstrated strong potential in detecting weld seam defects, offering an efficient and accurate method that reduces the reliance on manual evaluation.

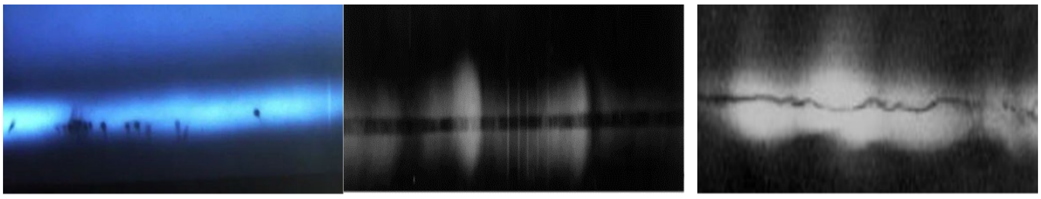

Figure 1: Illustrates the complex background of weld seam defect images.

Despite numerous attempts at weld seam defect detection over the past few years, accurately extracting features from images remains a challenge. The first challenge is the complex background of weld seam images. For example, as shown in Figure 1, images include effects from X-ray equipment settings, steel pipe surface quality, the use of image quality indicators, and the interaction between defects.

To mitigate this, attention mechanisms have been widely applied in defect detection tasks. For instance, Zhang Shuai and colleagues utilized a convolutional attention module to focus on important features by integrating spatial and channel information. Cheng Song and others introduced the SELayer, emphasizing the extraction of features around porosity and incomplete penetration defects by integrating global spatial and channel information. However, these methods focus only on global or local feature information, failing to reasonably allocate between the overall and detailed features of the image [1-3].

The second challenge involves the diversity and morphological differences of defects within weld seam images. Current defect detection methods primarily focus on extracting multi-scale features, often employing a pyramid approach that integrates feature maps at various resolutions without exploring the higher-order relationships between different scale features.

To address these issues, we propose a Hierarchical Attention Fusion Network (HAFNet), which leverages various feature information and their interrelations. Specifically, for the first challenge, this paper introduces a Residual Fusion Module (RFM) and a Multi-level Feature Fusion Module (MFFM). RFM utilizes fusion weights generated by a Sigmoid function (ranging between 0 and 1) to adaptively learn the feature weights of different encoding layers, thereby achieving an effective fusion of feature maps from the decoding stage with corresponding feature maps from the encoding stage, resulting in a weighted fused feature map. MFFM begins by guiding the optimization of low-level features using high-level features, adjusting the high-level features to the same dimension as low-level features through upsampling for integration, and then uses the optimized low-level features to refine high-level features, adjusting low-level features to the same dimension through max pooling for integration. Subsequently, features from various levels are concatenated and further extracted through a convolution module, while introducing low-level features as residuals to compensate for potential loss of detail information during processing. In response to the second challenge, this paper proposes a Dilated Hierarchical Attention Mechanism (DHAM). DHAM employs dilated convolutions with varying kernel sizes to effectively capture local features of weld seam defects at different receptive fields, integrating local and surrounding features to achieve a global feature representation. Through embedded channel and spatial attention mechanisms, this method further analyzes and utilizes global and local features to enhance the accuracy of defect detection.

1.2. Related Research Status

Since the 21st century, the widespread adoption of computer technology has propelled research in the field of deep learning. In weld seam defect detection, methods based on deep learning focus on optimizing and improving traditional models, enabling more effective use of data, and providing greater learning capabilities and robustness.

Neural networks are categorized into two types:

(1) BP neural networks: For example, Ding Xiaodong and others developed a weld seam defect classification model based on a three-layer BP neural network, utilizing seven feature parameters to identify surface defects in weld seams. Wang H and others employed a BP neural network model optimized by genetic algorithms, using weld seam depth and width as inputs and welding speed, laser power, etc., as outputs, achieving a defect detection accuracy of 97%. Li Tangdong and others applied wavelet denoising and BP neural networks for defect analysis and recognition, offering feasibility and identification for the qualitative analysis of weld seam defects.

(2) Deep learning: Hou and others used deep convolutional neural networks (DCNN) to build models that extract deep features, achieving an accuracy of 97.2% through learning patches cropped from X-ray images, providing strong support for the intelligent evaluation of X-ray images. Lin and others utilized an intra-frame attention strategy to reduce false positives, and inter-frame DCNNs to extract features from suspicious defect areas and obtain deep learning feature vectors, calculating the similarity of suspicious defect areas to track casting defects across continuous frames, thus eliminating false positives after defect tracking. Chen Yanfei and others addressed the weight bias update issue in the MobileNet model by introducing a residual structure and ReLU activation function, applying transfer learning during training to achieve optimal results with less computation. With deeper research into deep learning object detection models, the use of improved models has become a new focus, including single-stage models like SSD and YOLO, and two-stage models like Faster R-CNN. Wang Zhuyun and others improved the SSD network with atrous convolution, enlarging the receptive field of feature points without changing the feature map size, achieving an accuracy of 92.62% in identifying the location of circumferential welds, spiral welds, and defects. Li Yanfeng and others improved the YOLO network by using feature pyramids,

reducing network depth, introducing skip connection convolution blocks, and the K-means algorithm, reaching a defect identification rate of 87.9%. Wang Rui and others introduced a lightweight inverted residual structure in the YOLO-M model, reducing computational load and employing a multi-scale prediction mechanism for different defect features; grid-cross augmentation enhanced positive sample information in images, accelerating network convergence during training. Zhang K and others built a weld point dataset, applied data augmentation and transfer learning, and replaced the backbone network of Faster R-CNN with ResNet-101, achieving an average detection accuracy of 84% for six types of weld seam defects. Tang Maojun also replaced the Faster R-CNN backbone network with ResNet-101 and optimized the anchor boxes of the Faster R-CNN model using the K-means algorithm, enhancing the model's precision by 2.4% through image augmentation, introduction of multi-scale detection networks, and DCR decoupled classification refinement structure[4-6].

1.3. The Main Content of This Paper

Due to the multi-scale variations, diversity, and complex background interference in weld seam defect images, the current challenges and difficulties in weld seam defect detection research are as follows:

(1) Weld seam defect images suffer from complex background issues. The complexity of the backgrounds in weld seam defect images primarily manifests in the diversity of defects and their different grayscale and texture characteristics in X-ray images. This not only increases the difficulty of accurately extracting defect information from complex backgrounds but also reduces recognition accuracy due to the interaction between defects. Moreover, the same type of defects may exhibit different image characteristics under different welding conditions, posing additional challenges to defect identification and classification.

(2) There are significant variations in the scale of defects in weld seam defect images, along with a wide variety of types and distribution patterns. Defect types include porosity, cracks, inclusions, and lack of fusion, each presenting different characteristics in X-ray images. Defects are unevenly distributed across the weld seam; some may be spread throughout the seam, while others are concentrated in specific areas. Furthermore, the positions of defects vary, with some on the surface and others hidden internally. Such scale differences and distribution complexities require the detection system to have high flexibility and precision to identify, classify, and accurately locate various defects, significantly increasing the technical challenges.

This paper proposes a weld seam defect identification method based on the Hierarchical Attention Fusion Network (HAFNet), aiming to address the challenges posed by multi-scale variations, diversity, and complex background interference in weld seam defect images, as well as the existing models' shortcomings in feature extraction and handling diversity. Initially, a Dilated Hierarchical Attention Mechanism (DHAM) is employed to capture multi-scale information, enhancing the focus on key features across different scales and effectively addressing the issue of significant intra-class variability and minor inter-class differences in defects. Subsequently, a Residual Fusion Module (RFM) is introduced, which adaptively learns the feature weights of different encoding layers using fusion weights generated by a Sigmoid function and fully utilizes contextual information during the decoding phase to suppress the complex background interference of weld images. Finally, through a Multi-level Feature Fusion Module (MFFM), the fusion of multi-layer features from the decoding network is strengthened, enhancing the interaction and complementarity between features at different levels, reducing the model's sensitivity to noise and non-critical information, thereby further improving the model's recognition accuracy and robustness. Through embedded channel and spatial attention mechanisms, this method further analyzes and utilizes global and local features to enhance the accuracy of defect detection.

2. Design of Dual-encoding Multi-scale Attention Network

In this paper, U-Net is utilized as the backbone, and upon this network model, a Residual Fusion Module, a Dilated Hierarchical Attention Mechanism, and a Multi-level Feature Fusion Module are integrated to enhance the model's performance in the task of weld seam defect image detection. The structure of the Hierarchical Attention Fusion Network is illustrated in Figure 2.

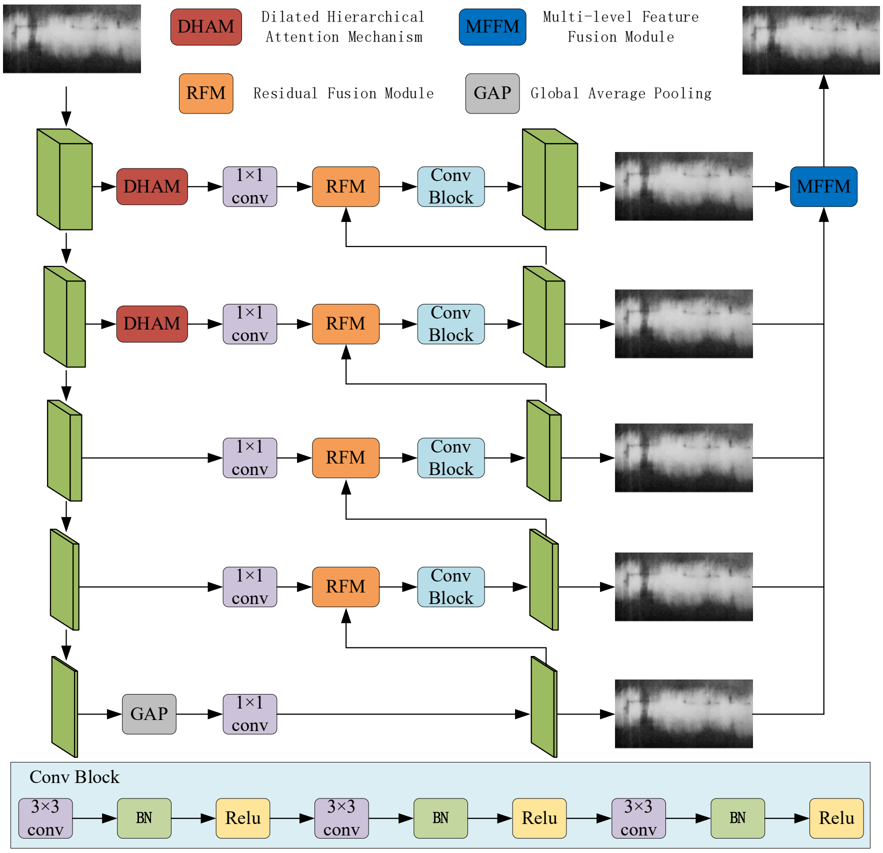

Figure 2: The structure of the Dual-encoding Multi-scale Attention Network.

2.1. Dilated Hierarchical Attention Mechanism (DHAM)

During the convolution process, larger convolution kernels are more suitable for capturing larger objects, while smaller kernels are better suited for capturing smaller objects. In the context of steel pipe weld X-ray images, using a convolution kernel of a single size does not effectively capture the necessary information features. Therefore, a method similar to the Receptive Field Block (RFB) is adopted, utilizing receptive fields of different sizes to establish various branches that complement each other. Consequently, a Dilated Hierarchical Attention Mechanism is proposed, with its structure illustrated in Figure 3.

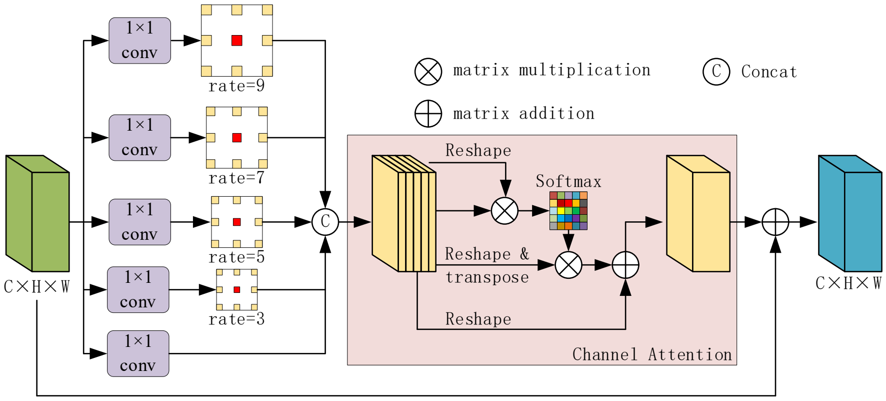

Figure 3: The structure of the DHAM.

To reduce the computational load, a convolution with a 1x1 kernel is used to decrease the number of channels. Dilated convolutions are used to concatenate these branches and standardize the number of feature channels to 64. The kernel size of the dilated convolution is (2b−1)×(2b−1) with a dilation rate of (2b−1), where b=5, …,1.However, the features obtained from different receptive fields of various branches cannot distinguish subtle features, and directly connecting these features would lead to spatial inconsistency. Therefore, embedding a Channel Self-Attention Mechanism (CAM) addresses the aforementioned issue. The channel self-attention map is calculated using the original features $F \in R^{C\times H\times W}$. First, F is reshaped into $R^{C\times N}$, where N=H×W. Matrix multiplication is then performed between the transposed F and the original feature F. The channel self-attention map, $M \in R^{C\times C}$ is obtained through applying the Softmax function.

$$I_{m_{i,j}} = \frac{\exp(F_i \otimes F_j)}{\sum_{j=1}^{C} \exp(F_i \otimes F_j)} \tag{1}$$

where $I_{m_{i,j}}$ measures the influence of the j-th channel on the i-th channel. Furthermore, after transposing M and F, matrix multiplication is performed, and their result is reshaped into $R^{H\times W\times C}$. The result is multiplied by the parameter λ, and then element-wise addition is performed with F to obtain the output result $C \in Q^{H\times W\times C}$:

$$C_i = \lambda \sum_{j=1}^{C} (I_{m_{i,j}} \times A_j) \oplus F_i \tag{2}$$

where λ is a parameter with an initial value of 0, which can be automatically adjusted during the training process.

From equation (2), it can be seen that to obtain the output feature of each channel, a weighted summation of the features across all channels and the original features is conducted. To compensate for the loss of feature information details and to enhance the network's feature transmission capability, a skip connection layer is incorporated. This approach retains the detailed information of lower-level features and combines them with higher-level features, thereby obtaining a richer and more accurate feature representation.

2.2. Residual Fusion Module (RFM)

To enable the network to fully leverage contextual information during the decoding phase and mitigate the interference from complex backgrounds in weld seam images, a Residual Fusion Module (RFM) is proposed, with its structure illustrated in Figure 4. Unlike conventional methods that employ element-wise summation or concatenate multiple features for fusion, RFM can adaptively learn the feature weights of different encoding layers. This module primarily consists of two steps: proportional scaling and adaptive fusion, effectively enhancing the network's capability to integrate and refine multi-level feature representations.

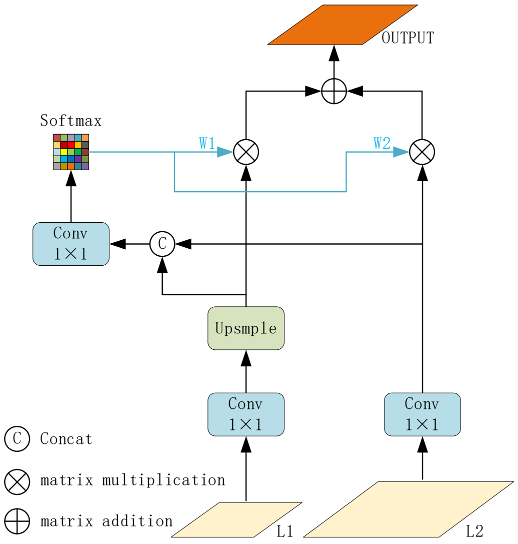

Figure 4: The structure of the RFM.

The feature representation at the i-th layer($i \in \{1,2\}$)in the encoder is denoted as X_i,corresponding to L1 and L2 in Figure 3. Since features at different levels possess varying resolutions and channels, they need to be standardized to the same size before fusion. The operation of resizing the feature X_n to the dimensions of feature X_i is denoted as X_n^i. To unify L1 and L2 to the same size (same resolution and channel count), convolution with a 1×1 kernel and upsampling by a factor of 2 are employed to redistribute the features. The specific operation is shown in equation (3):

$$X_1^2 = f_C(f_u(X_1))$$
$$X_2^2 = X_2 \tag{3}$$

where f_c represents a convolutional layer with a kernel size of 1 and a stride of 1, while f_u denotes upsampling with a scale factor of 2. X_n^i signifies the operation of resizing the features from the n-th layer to match the dimensions of the i-th layer.

The output of the RFM is as shown in equation (4):

$$Out = \varpi_1 X_1^i + \varpi_2 X_2^i \tag{4}$$

where ϖ_1 and ϖ_2 are trainable parameters representing the feature weights obtained through resizing. Here, ϖ_1 and ϖ_2 maintain distinct training weights at different positions within the decoding layer to accommodate the requirements of different scale fusion strategies. Furthermore, it holds that $\varpi_1 + \varpi_2 = 1$, where $\varpi_1, \varpi_2 \in [0,1]$. ϖ_n is defined as:

$$\varpi_n = \frac{e^{\lambda n}}{e^{\lambda_1} + e^{\lambda_2}} \tag{5}$$

where $\mathrm{n} \in 1,2$. ϖ_1 and ϖ_2 is defined using λ_1 and λ_2 as a control parameter and the Softmax function, and it can be learned through standard backpropagation.

2.3. Multi-Level Feature Fusion Module (MFFM)

The Multi-level Feature Fusion Module achieves the final output by integrating the output features from five distinct decoding layers, with its architecture depicted in Figure 5. This module enhances the detail and accuracy of the final output feature map by fusing features extracted at different levels of the network, thereby bolstering the network's robustness against the inherent variations and complexities in weld seam defect images.

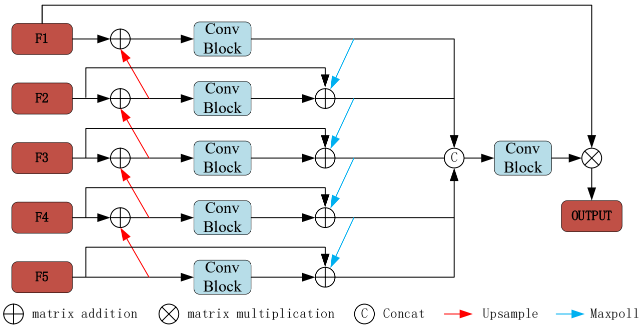

Figure 5: The structure of the MFFM.

First, optimize low-level features using high-level features, then utilize the optimized low-level features to refine high-level features in a reverse manner. Afterward, concatenate the optimized features and refine them through convolutional blocks for feature enhancement. Finally, multiply the refined features with low-level features to preserve detailed information while supplementing semantic information. The specific process is illustrated in equations (6), (7), and (8):

$$F_i' = Up(F_{i+1}) \oplus F_i \tag{6}$$

$$F_{i+1}'' = Max(Conv(F_i')) \oplus Conv(F_{i+1}') \oplus F_{i+1} \tag{7}$$

$$OUT = Conv(Cat(F_1'', F_2'', F_3'', F_4'', F_5''))\otimes F_1 \tag{8}$$

where Up represents upsampling, Max represents max pooling, Conv represents for convolutional layer, Cat(,) represents feature concatenation, $\oplus$ represents matrix addition, and $\otimes$ represents matrix multiplication.

3. Experiments

All experiments in this study were conducted using a workstation server equipped with an NVIDIA RTX3090 24G graphics card. Python version 3.8.1 and PyTorch version 1.11.0 were utilized, with model weights initialized using the Kaiming parameter initialization method. Image preprocessing techniques such as rotation, horizontal flipping, and vertical flipping were applied. Due to hardware limitations, all images were resized to 256×256 for faster training. During training, a batch size of 4 was employed, and the Ranger algorithm with a learning rate of 0.0001 and weight decay of 0.0005 was chosen as the optimizer. The learning rate adjustment was performed using the Cosine Annealing with Warm Restarts (CAWR) algorithm, presenting a cosine decay curve[7-9].

3.1. Dataset

There is currently no publicly accessible dataset of X-ray film images for weld seam defect segmentation tasks to our knowledge. Therefore, the sample images used in this study were provided by a professional steel pipe manufacturing company. The company primarily focuses on producing large-diameter, thick-walled, high-grade spiral welded steel pipes for long-distance transportation of oil and natural gas. To ensure product quality, the company employs an advanced Digital Radiography (DR) X-ray imaging detection system during the quality inspection stage of the production process. This detection system captures X-rays penetrating the steel pipe weld seam using high-sensitivity digital detectors and converts the acquired information into electrical signals, which are then input into the computer system. In the computer, digital signal processing techniques are applied to enhance the images, adjust contrast and brightness, among other processes, to generate high-quality imaging results.

The obtained weld seam images during the inspection process are output and saved in the 16-bit DCM format with a size of 512×512 pixels. The entire dataset comprises 2000 weld seam images, covering various defect types such as cracks, pores, lack of fusion, incomplete penetration, slag inclusions, and burn-through. Strict adherence to the following principles was maintained during the image collection process to ensure the quality and diversity of the dataset, thus supporting the development of efficient and accurate defect detection models:

(1) Diversity of weld seam defect types and morphological differences.

(2) Variability in image quality due to factors such as X-ray equipment settings and material thickness variations.

(3) Contrast between defects and background.

(4) Interactions between defects.

The entire dataset comprises 2000 weld seam images, encompassing a wide range of defect types such as cracks, pores, lack of fusion, incomplete penetration, slag inclusion, and burn-through as shown in Figure 6. Given that the number of samples of weld seam defect images may be insufficient to support the training of complex neural network models, data augmentation is employed to simulate various interferences that may occur during actual operations, thereby preventing overfitting issues. By applying operations such as image rotation, horizontal flipping, vertical flipping, sharpening, brightness adjustment, and contrast adjustment, each original image is expanded into 10 images, thus increasing the entire dataset to 20,000 images. To further enhance the model's generalization ability and ensure evaluation accuracy, the entire dataset is randomly divided into training, validation, and testing sets in a 3:1:1 ratio.

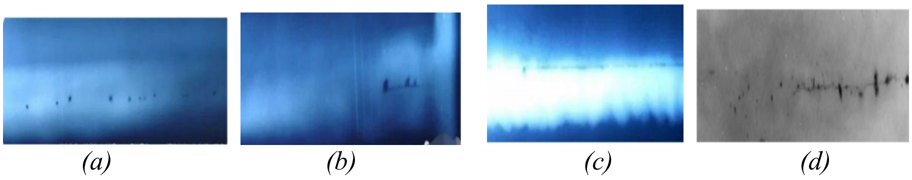
(a) *(b)* *(c)* *(d)*

Figure 6: The structure of the RFM. (a)porosity. (b)incomplete penetration. (c)lack of fusion. (d)cracks.

3.2. Evaluation Metrics

The F_β, MAE, S_m are used as assessment metrics to measure the performance of the proposed method:

$$Recall = \frac{TP}{TP+TN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$F_\beta = \frac{(1+\beta^2)\cdot Presion\cdot Recall}{\beta^2\cdot Presion\cdot Recall} \tag{11}$$

$$MAE = \frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H}|S(i,j)-G(i,j)| \tag{12}$$

$$S_m = \alpha\times S_0 + (1-\alpha)\times S_r \tag{13}$$

where, TP represents correct predictions of positive samples, TN represents correct predictions of negative samples, FP represents incorrect predictions of negative samples as positive, and FN represents incorrect predictions of positive samples as negative. Recall measures the proportion of TP in positive samples, Precision measures the proportion of TP in positive predictions. MAE is used to evaluate the difference between predicted image S and label G. F_β is the weighted harmonic mean of precision and recall. S_m is an index for assessing the structural similarity difference between predicted images and labels.

3.3. Comparison with Other Defect Image Segmentation Methods

To validate the advantages of the proposed method, we first conducted a comparison with various traditional weld seam image defect segmentation methods on the same dataset. These compared methods include: SegNet, which employs a unique encoder-decoder architecture for image segmentation, with the decoder utilizing the pooling indices from the encoder to recover image details; U-Net, which uses a symmetric structure and skip connections, combining high-resolution features from the encoder with features from the decoder to enhance the accuracy of image segmentation; R2Net, which captures complementary salient information from different feature layers for more accurate detection of salient objects in images; RANet, equipped with recurrent aggregation of deep features, fully leveraging complementary salient information captured at different layers to enhance the detection accuracy of salient objects; and PoolNet, a novel bidirectional information transmission model that integrates multi-level features for salient object detection. Figure 7 presents the comparison of prediction results obtained after conducting experiments on a homemade X-ray film weld defect image dataset with these methods.

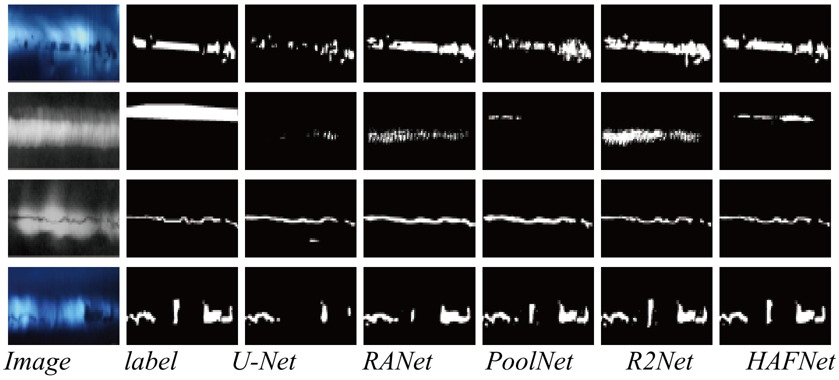

Figure 7: Visual Comparison of Experimental Results.

From Figure 7, it can be observed that compared to U-Net, and PoolNet, HAFNet demonstrates better ability to suppress environmental interference and preserve defect integrity. This is attributed to the proposed Dilated Hierarchical Attention Mechanism (DHAM), which effectively extracts multi-scale features by obtaining different receptive fields on different branches. However, features obtained from different receptive fields cannot distinguish subtle features, and directly connecting these features may lead to spatial inconsistency. Therefore, a channel attention mechanism is employed to improve multi-scale information representation and suppress background interference. In addition, PoolNet and R2Net lack adaptability to complex backgrounds, making it almost impossible to distinguish defects from the background, resulting in inaccurate defect localization. It is worth noting that the results of RANet indicate better detection performance, but some defects may appear to be stuck together due to incorrect boundary recognition. To better integrate multiple features and reduce the impact of complex environmental backgrounds, a residual fusion module is proposed. Unlike other methods that use element-wise summation or joint multi-feature fusion methods, this module can adaptively learn the feature weights of different encoding layers, thereby better integrating multiple features and avoiding the influence of complex environmental backgrounds. In the fourth and fifth images, the morphology and dispersion of defects vary, but there are some common characteristics. However, R2Net, RANet, PoolNet, and LVNet fail to clearly describe the defect contours. HAFNet addresses the issue of mutual influence among defects in the weld seam defect image dataset by introducing the Multi-level Feature Fusion Module (MFFM) to focus on the correlation between features at different levels and enhance the interaction and complementarity between features through fusion mechanisms. Therefore, for defects of different shapes and sizes, HAFNet demonstrates better detection performance. On the other hand, PoolNet's feature aggregation module also exhibits issues of missed detections and false alarms when dealing with scale changes. By comparing the images, it can be seen that compared to other models, HAFNet demonstrates stronger capability to capture effective features. It not only clearly delineates defect boundaries but also suppresses other interferences, giving it a significant advantage in defect segmentation tasks[10-12].

Table 1: Comparative Experimental Results.

Model	MAE	F_β	S_m
U-Net	0.0355	0.7912	0.8426
RANet	0.0303	0.8143	0.8467
PoolNet	0.0251	0.8199	0.8689
R2Net	0.0202	0.8416	0.8823
HAFNet	0.0143	0.8517	0.8981

Table 1 presents the results of the comparative experiments. Compared to U-Net, HAFNet achieves enhancements of 0.0605 and 0.0555 in F_β and S_m respectively, with a reduction in MAE by 0.0212. Compared to RANet, HAFNet exhibits increases of 0.0374 and 0.0514 in F_β and S_m, respectively, along with a decrease in MAE by 0.016. Furthermore, compared to PoolNet, it demonstrates superior performance on the dataset, with F_β and S_m increasing by 0.0318 and 0.0292, respectively, while MAE decreases by 0.0108. Notably, HAFNet outperforms R2Net, showing increases of 0.0101 and 0.0158 in F_β and S_m, respectively, while MAE decreases by 0.0059.

3.4. Ablation Study

The proposed method consists of three key components: DHAM, RFM, and MFFM. Therefore, we conducted ablation experiments to evaluate the importance and contribution of each component. The baseline model adopts the same encoder-decoder architecture and optimization strategy. Figure 8 provides visualizations of feature maps, demonstrating the effectiveness of our proposed components intuitively. It can be observed that compared to the baseline network, our proposed method can reduce background interference, create more discriminative features, and provide more accurate segmentation results.

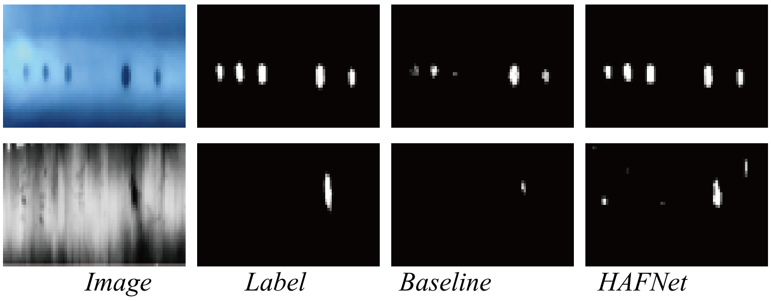

Figure 8: Visualization results of ablation study.

To better illustrate the roles of different components, we gradually added the DHAM, RFM, and MFFM modules to the baseline model. To minimize the influence of other factors, we selected MAE, F_β, and S_m as evaluation metrics. Figure 9 shows the results obtained by progressively adding the DHAM, RFM, and MFFM modules to the baseline model.

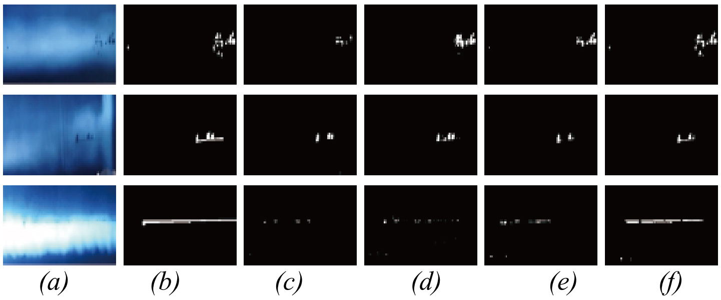

Figure 9: The results obtained by progressively adding the three modules DHAM, RFM, and MFFM to the baseline model. (a)Image; (b)Label; (c)Baseline; (d)Baseline+ DHAM; (e) Baseline+ DHAM+ RFM; (f) Baseline+ DHAM+ RFM+ MFFM.

Through Figure 9, it can be observed that using the Backbone alone does not accurately detect all targets. However, with the addition of the DHAM, RFM, and MFFM modules to the Backbone, detection performance significantly improves. With the DHAM, the model can extract more multi-scale feature information and segment the target contour region, albeit susceptible to background interference. After incorporating the RFM, the model gains rich global and detail feature information, thereby delineating clear contours of the target region. Finally, with the introduction of the MFFM, by fully integrating and utilizing low-level and high-level feature information, background interference is further suppressed. Table 2 presents the comparative results of the ablation experiments for HAFNet's performance.

Table 2: Quantitative Evaluation of Ablation Studies of Different Components.

Backbone	DHAM	RFM	MFFM	MAE	F_β	S_m
√				0.0355	0.7912	0.8426
√	√			0.0299	0.8155	0.8653
√	√	√		0.0210	0.8299	0.8798
√	√	√	√	0.0143	0.8517	0.8981

From Table 2, it can be observed that Backbone + GBAM + DSAM + RFM achieves the best performance in terms of MAE, F_β, and S_m metrics. In contrast, the performance of using Backbone alone is the poorest. The introduction of DHAM facilitates the extraction of effective multi-scale features. It enables the extraction of effective multi-scale features with different receptive fields on different branches, while incorporating channel attention to improve multi-scale information representation and suppress background interference. Compared to the backbone, there is a decrease of 0.0056 in MAE, while F_β and S_m increase by 0.0243 and 0.0227, respectively. Subsequently, RFM is introduced to fuse multi-scale feature information, adaptively learn the feature weights of different encoding layers, and further suppress the influence of complex background environments. From Table 2, it is evident that compared to the model without RFM, there is a decrease of 0.0089 in MAE, while F_β and S_m increase by 0.0144 and 0.0145, respectively. Finally, with the addition of MFFM, by strengthening the fusion of multi-layer features of the decoding network, enhancing interaction and complementarity between features at different levels, reducing the model's sensitivity to noise and non-critical information, further improvements in model recognition accuracy and robustness are achieved. Simultaneously, there is a decrease of 0.0067 in MAE, while F_β and S_m increase by 0.0218 and 0.0183, respectively. Thus, it is concluded that GBAM, DSAM, and RFM effectively enhance the model's detection.

3.5. Sensitivity analysis of parameters experiment

Through sensitivity analysis experiments, the robustness of the HAFNet model is verified under different learning rate settings, as well as its tolerance to the application of different initialization weights. A series of experiments with different parameters were conducted, involving learning rates, optimizers, and initial weight values. These experiments were carried out on a custom-madedefects dataset to analyze the sensitivity of HAFNet. Initially, learning rates were set to 0.001, 0.0015, 0.0001, and 0.00015, and the sensitivity of HAFNet to the learning rate was analyzed by observing the experimental results. Next, sensitivity analysis of the initial weight values was performed using two initialization methods: Xavier and Kaiming. Additionally, two different optimizers, Ranger and Adam, were applied to analyze the sensitivity of HAFNet to different optimizers, while keeping other parameters constant. Table 3 presents the parameter sensitivity analysis of HAFNet.

As shown in Table 3, according to the results of parameter sensitivity analysis, it can be concluded that the learning rate, optimizer, and initialization weights have no significant impact on the performance of the proposed network model. This indicates that HAFNet is less affected by hyperparameter variations and exhibits strong robustness.

Table 3: Comparative Experimental Results.

	Parameters	MAE	F_β
Learning rate	0.001	0.0145	0.8521
	0.0015	0.0142	0.8516
	0.0001	0.0141	0.8518
	0.00015	0.0146	0.8520
Optimizer	Ranger	0.0142	0.8520
	Adam	0.0144	0.8519
Initialization	Xavier	0.0142	0.8517
	Kaiming	0.0143	0.8516

4. Conclusions

In this article, we introduce a HAFNet for defect image segmentation in weld seam X-ray film images. Thanks to the constructed feature extraction and fusion modules, along with attention mechanisms, the proposed method can accurately locate and segment targets of different categories while effectively suppressing noise interference. Within our proposed method, the DHAM is capable of extracting multi-scale global and local detail features, whereas the RFM and MFFM leverage contextual information to efficiently fuse features across different levels and reduce the interference of redundant feature information. Extensive experiments demonstrate that our method significantly outperforms other compared SSS image segmentation methods. Specifically, our method's precision, MAE, and F-measure reached 1.43%, 85.17%, and 89.81%, respectively. Ablation studies further validate the effectiveness of each designed component. In the future, designing more robust defect segmentation models with stronger generalization capabilities represents a promising direction for further research.

References

[1] Lin Z, Yingjie Z, Bochao D, et al. Welding defect detection based on local image enhancement[J]. IET Image Processing, 2019, 13(13):2647-2658.
[2] Xu Z, Wu M, Fan W. Sparse-based defect detection of weld feature guided waves with a fusion of shear wave characteristics [J]. Measurement, 2021, 174:109018.
[3] Abdelkader R, Ramou N, Khorchef M, et al. Segmentation of x-ray image for welding defects detection using an improved Chan-Vese model[J]. Materials Today: Proceedings, 2021, 42:2963-2967.
[4] Yan Z H, Xu H, Huang P F. Multi-scale multi-intensity defect detection in ray image of weld bead [J]. NDT & E International, 2020, 116:102342.
[5] Malarvel M, Singh H. An autonomous technique for weld defects detection and classification using multi-class support vector machine in X-radiography image [J]. Optik, 2021, 231:166342.
[6] Wang H, Li J, Liu L. Process optimization and weld forming control based on GA-BP algorithm for riveting-welding hybrid bonding between magnesium and CFRP [J]. Journal of Manufacturing Processes, 2021, 70:97-107.
[7] Park J-K, An W-H, Kang D-J. Convolutional Neural Network Based Surface Inspection System for Non-patterned Welding Defects [J]. International Journal of Precision Engineering and Manufacturing, 2019, 20(3):363-374.
[8] Miao R, Gao Y, Ge L, et al. Online defect recognition of narrow overlap weld based on two-stage recognition model combining continuous wavelet transform and convolutional neural network[J]. Computers in Industry, 2019, 112:103115.
[9] Xiao M, Yang B, Wang S, et al. A feature fusion enhanced multiscale CNN with attention mechanism for spot-welding surface appearance recognition [J]. Computers in Industry, 2022, 135:103583.
[10] Chen Z, Huang G, Lu C, et al. Automatic recognition of weld defects in ToFD D-scan images based on faster R-CNN[J]. Journal of Testing and Evaluation, 2020, 48(2):811-824.
[11] Zhang K, Shen H. Solder joint defect detection in the connectors using improved faster-rcnn algorithm [J]. Applied Sciences, 2021, 11(2):576.
[12] Jiang D, Li G, Tan C, et al. Semantic segmentation for multiscale target based on object recognition using the improved Faster-RCNN model[J]. Future Generation Computer Systems, 2021, 123:94-104.