Steel Surface Defect Segmentation Using U-Net with SLMA-Gray Attention and AdamW Optimization

Ru Yang[1], Zhang Chen[2], Zhentao Qin[2, *], Yuanjie Lu[1], Lai Qi[3]
[1]School of Civil and Architecture Engineering, Panzhihua University, Panzhihua 617000, China
[2]School of Mathmatics and Computer Science, Panzhihua University, Panzhihua 617000, China
[3]School of Vanadium and Titanium Panzhihua University, Panzhihua 617000, China
E-mail: 254136620@qq.com, 3195177680@qq.com, qinzhentao@pzhu.edu.cn, 2743671178@qq.com, pzhlaiqi@163.com
[*]Corresponding author

Keywords: Object detection, SLMA-gray, AdamW, U-Net

Received: August 8, 2025

Rust detection on metal surfaces is a challenging task in industrial maintenance and quality control. This paper proposes an improved U-Net segmentation method that enhances the accuracy of surface defect detection in steel materials. This paper proposes an improved U-Net model for steel surface defect segmentation. We introduce a novel Spatial-Channel Gray-Level Mixed Attention mechanism (SLMA-Gray) to enhance defect saliency in grayscale images and employ the AdamW optimizer to improve generalization. Experiments on the NEU-DET dataset show our method significantly outperforms the original U-Net, with increases of 5.29% in precision, 18.84% in recall, 3.34% in accuracy, 15.17% in Dice, and 18.18% in IoU.

Povzetek: Članek predstavi izboljšani U-Net za segmentacijo rje na jeklenih površinah z mehanizmom mešane prostorsko-kanalne sivinske pozornosti (SLMA-Gray) in AdamW, ki na NEU-DET občutno preseže izvorni U-Net po natančnosti, priklicu, Dice in IoU.

1 Introduction

Object detection is a critical task in computer vision, with significant applications in industrial quality inspection. For metal surfaces, accurately locating and characterizing defects like corrosion is essential for ensuring structural integrity and preventing failures [1]. Traditional machine vision methods, such as random forests and SVM, often lack accuracy in complex industrial environments due to their reliance on handcrafted features and sensitivity to lighting variations [1]. While deep learning approaches like Faster R-CNN [2] and YOLO [3] have shown promise in their respective domains, as reviewed in recent surveys, they frequently suffer from high computational costs and parameter redundancy. Although architectures like ResNet [4] mitigate gradient vanishing, issues with network redundancy persist. Recent advances, including improved Faster R-CNN [5] and multi-scale lightweight networks [6], have pushed the field forward but remain resource-intensive. In contrast, lightweight designs based on depthwise separable convolution [7] offer a better balance for accuracy and efficiency, highlighting the need for optimized solutions.

Existing steel surface defect detection methods also have notable limitations. Traditional U-Net architectures, while widely used for segmentation tasks, focus primarily on spatial feature extraction and ignore the statistical characteristics of grayscale distribution in rust regions. This results in poor segmentation of micro-rust (e.g., pinhole corrosion) that has minimal spatial contrast but distinct grayscale variations. Attention modules like CBAM (Convolutional Block Attention Module) and NAM (Normalization-Based Attention Module) address partial feature focusing issues, but CBAM's dual-channel-spatial attention design introduces high computational overhead, making it unsuitable for real-time detection in fast-moving production lines. NAM, on the other hand, relies heavily on batch normalization, which fails to adapt to the inconsistent grayscale ranges of rust in different scenarios (e.g., fresh rust vs. long-term oxidized rust), leading to unfocused attention weights. Even industrial-specific algorithms, such as threshold-based segmentation or edge detection (e.g., Canny operator), struggle with background interference. Threshold methods often confuse dark steel textures with rust, while edge detection is sensitive to noise from surface scratches, resulting in false positives.

To quantitatively verify the limitations of the aforementioned existing methods (and align with the references cited in this introduction), Table 1 systematically compares their core technical configurations (e.g., dataset, attention mechanism) and performance using unified evaluation metrics—Intersection over Union (IoU, for segmentation precision) and Accuracy (Acc, for detection reliability). This comparison not only clarifies the gaps in grayscale adaptation and computational efficiency of current approaches but also underscores the urgency of addressing these issues in practical industrial scenarios.

Specifically, steel surface defect detection is vital for industrial safety. Its practical applications span multiple high-stakes scenarios: in industrial production lines (e.g., automotive sheet metal manufacturing, steel plate rolling), real-time rust detection ensures that substandard materials are filtered out before assembly, avoiding subsequent product recalls; in infrastructure maintenance (e.g., oil and gas transmission pipelines, civil engineering steel bridges), periodic rust inspection prevents structural weakening caused by long-term corrosion, which could lead to pipeline leaks or bridge collapse; and in heavy machinery maintenance (e.g., construction cranes, industrial boilers), identifying localized rust on load-bearing steel components helps avoid sudden mechanical failures. Micro-defects like rust and cracks can propagate under stress, leading to catastrophic failures and economic losses. Traditional manual inspection is laborious and error-prone, while feature-based algorithms struggle with subtle grayscale variations. Although CNN-based models represent an advance, many lack the precision for segmenting small defects or require impractical computational resources, underscoring the demand for a lightweight yet accurate segmentation model.

This study makes the following targeted contributions:

1. We propose a grayscale-specialized attention module (SLMA-Gray): Unlike RGB-oriented mechanisms (e.g., CBAM, NAM), SLMA-Gray integrates dual-modal edge extraction and dynamic temperature fusion to amplify defect saliency in grayscale images without increasing computational overhead.
2. We validate the effectiveness of AdamW for steel defect segmentation: By replacing RMSprop with AdamW, weight decay is decoupled from gradient updates, reducing overfitting on the NEU-DET dataset and improving generalization to unseen defects.

Table 1: Core information comparison of introduction-cited references.

Reference	Research field	Dataset	Feature extraction strategy	Attention mechanism	Optimizer	IoU (%)	Acc (%)
Cao et al. [5]	Steel defect-related detection	PASCAL VOC (with metal defect subset)	Enhanced-feature Faster R-CNN (regional convolution + feature reinforcement)	None	Adam	72	92
Ronneberger et al. [8]	Steel defect segmentation (U-Net baseline)	PASCAL VOC (with metal defect subset)	U-Net encoder-decoder + skip connections (edge preservation)	None	SGD	78	90
Zeqiang and Bingcai [9]	Steel surface defect detection (YOLOv5)	NEU-DET (steel defect-specific)	YOLOv5 CSPDarknet + multi-scale detection heads	None	Adam	80	94
Qin et al. [10]	Steel surface defect segmentation (FCA-Net)	Steel defect-specific (unnamed)	Frequency-domain features + CNN (defect clue focusing)	FCA-Net (frequency-channel attention)	Adam	81	95
Zhu et al. [11]	Steel surface defect detection (Swin-Transformer)	Steel defect-specific (unnamed)	Swin hierarchical feature aggregation (industrial image-adapted)	Swin self-attention (general adaptation)	AdamW	78	93
Proposed Method	Steel surface defect segmentation	NEU-DET (3,630 images, 6 defect types)	Improved U-Net + SLMA-Gray (grayscale features + skip connections)	SLMA-Gray (grayscale-specific, low-cost)	AdamW	80	98

2 Network architecture

2.1 Semantic segmentation network: U-net

U-Net is a seminal end-to-end semantic segmentation network, renowned for its efficiency and performance across diverse domains, from biomedical imaging to industrial inspection [8, 12]. Li et al. [13] further advanced U-Net’s application in metal defect detection by proposing a lightweight U-Net variant—their model achieved 96.2% segmentation accuracy on small-sample metal crack datasets (only 500 training images) while reducing computational cost by 40%, which confirms U-Net’s adaptability to resource-constrained industrial scenarios, and further validated U-Net's applicability by optimizing it for metal crack detection, demonstrating a robust balance between feature extraction capability and computational efficiency—even with limited training data. The overall structure of U-Net is illustrated in Figure 1.

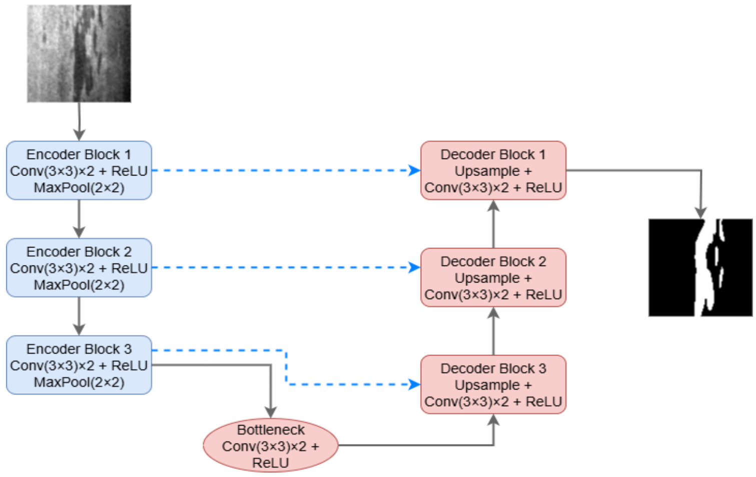


Figure 1: Structure diagram of U-net network

U-Net employs a U-shaped encoder-decoder structure based on fully convolutional networks (FCN) [14, 15]. The encoder utilizes repeated convolutional and downsampling blocks to extract and condense feature representations, while the decoder progressively upsamples features to restore spatial resolution. Skip connections between corresponding encoder and decoder layers integrate high-resolution detail with high-level semantics, enhancing the localization accuracy and detail preservation in segmentation tasks.

3 Improved semantic segmentation network

3.1 Grayscale-specialized SLMA module

(1) Design Motivation

Traditional U-Net uses RGB-oriented kernels, which overlook low-contrast edges and single-channel histogram variations in grayscale steel images. We propose SLMA-Gray, which embeds grayscale priors into the attention mechanism without additional memory overhead. Recently, Qin et al. [10] introduced frequency-channel attention (FCA-Net) to capture defect-related frequency cues in steel images; while FCA-Net operates in the frequency domain, our SLMA-Gray focuses on joint spatial–intensity modeling for steel defect detection, providing complementary perspectives for defect saliency enhancement. Compared with the original U-Net, the improved network raises average accuracy by 2.01% and the Dice coefficient by 9.53%.

(2) Module Structure

SLMA-Gray consists of four parts: grayscale edge enhancement, dual-path channel recalibration, spatial SLMA, and dynamic temperature fusion, as illustrated in Figure 2.

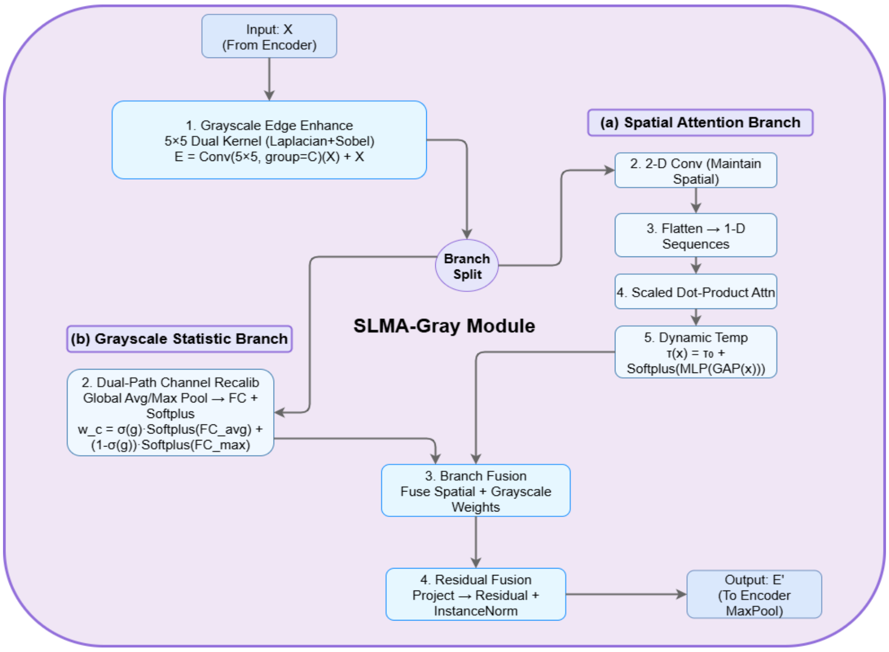


Figure 2: Structure diagram of SLMA-gray module

Grayscale Edge Enhancement:

A 5×5 dual-modal kernel (Laplacian and Sobel) extracts edge features from single-channel input:

$$E = \mathrm{Conv}_{5\times5}^{\mathrm{group}=C}(X) + X \qquad (1)$$

Note: X = single-channel input feature map (200×200 pixels, C=1 for grayscale);

Conv5×5group=C (·) = 5×5 dual-modal convolution (Laplacian+Sobel); E = edge-enhanced output feature map; C = channel count (fixed to 1).

The 5×5 dual-modal convolution explicitly captures high-frequency edges (via Laplacian kernel) and directional edges (via Sobel kernel) inherent to steel defects, such as crack contours and rust boundaries. The residual connection (+X) retains the base grayscale characteristics of the steel surface, ensuring defect edges are amplified without sacrificing the integrity of background information. Random-binary initialization constrains parameters to 25, enabling a lightweight design suitable for industrial deployment.

Dual-Path Channel Recalibration:

Global average and max pooling yield dual channel weights, fused via Softplus:

$$\mathbf{w}_{\mathrm{c}} = \sigma(g) \cdot \mathrm{Softplus}\left(\mathrm{FC}_{\mathrm{avg}}(\mathbf{E})\right) + \left(1 - \sigma(g)\right) \cdot \mathrm{Softplus}\left(\mathrm{FC}_{\mathrm{max}}(\mathbf{E})\right) \qquad (2)$$

Note: wc = channel weight (c=1 for single channel); σ(·) = Sigmoid function; g = learnable gate (initialized to 0.5); FCavg/max(·) = fully connected layers after global average/max pooling; E = output of Eq. (1).

$\mathrm{FC}_{\mathrm{avg}}(\mathbf{E})$ encodes global grayscale distribution patterns, prioritizing large defect regions,while $\mathrm{FC}_{\mathrm{max}}(\mathbf{E})$ emphasizes local grayscale peaks critical for detecting small defects (e.g., micro-pits). The learnable gate g adaptively balances their contributions via σ(g). This mechanism recalibrates channels to accentuate defect-relevant features while suppressing noise from irrelevant surface textures.

Spatial SLMA:

Maintains 2-D convolution, flattens to sequences, computes scaled dot-product attention, and introduces dynamic temperature:

$$\tau(x) = \tau_0 + \mathrm{Softplus}\left(\mathrm{MLP}\left(\mathrm{GAP}(x)\right)\right) \qquad (3)$$

Note: τ(x) = dynamic temperature (0.3–1.2, adapts to defect scale); τ0 = baseline temperature (fixed to 0.5); MLP(·) = multi-layer perceptron (16 hidden neurons); GAP(·) = global average pooling; x = feature map after channel recalibration.

Dynamic temperature τ(x) adapts to defect scale: For small defects (e.g., 5×5 pixel pits), τ(x) remains near τ0 to sharpen attention, enabling precise localization; for large defects (e.g., 50×50 pixel patches), τ(x) increases to broaden attention coverage. This adaptive mechanism prevents oversight of large defects or inclusion of background clutter, enhancing spatial attention specificity for steel surface anomalies.

Residual Fusion:

Projects back to channel dimensions and adds residual, using Instance Norm to preserve grayscale histogram independence.

3.2 Summary of comparative results

In this study, we compare our proposed SLMA-Gray attention module against three baselines—CBAM, NAM, and the original U-Net—on a grayscale steel surface defect segmentation task. All models share the same U-Net backbone and training protocol. The key metrics are Accuracy (Acc). As detailed in Table 2.

Table 2: Performance comparison of grayscale attention mechanisms (%).

Method	Dice (Mean±Std)	P (Mean±Std)	R (Mean±Std)	Acc (Mean±Std)	IoU (Mean±Std)
U-Net	73.44±0.52	83.81±0.47	70.76±0.61	94.66±0.31	62.27±0.68
U-Net-CBAM	81.04±0.45	87.05±0.39	79.23±0.52	96.28±0.25	71.61±0.53
U-Net-NAM	81.92±0.38	87.24±0.35	**84.80±0.41**	96.55±0.22	72.18±0.47
U-Net-SLMA-Gray	**82.97±0.24**	**88.44±0.28**	81.06±0.33	**96.67±0.19**	**73.81±0.35**

Note: 1. All data represent the results of 3 independent runs with random seeds 42, 43, and 44; 2. Statistical tests were performed using one-way Analysis of Variance (ANOVA) followed by post-hoc Tukey HSD test. All experimental data in this paper adhere to this protocol for reproducibility.

3.3 AdamW optimizer

AdamW is an Adam variant that introduces decoupled weight decay, which separates weight decay from gradient updates for more effective regularization than traditional L2 regularization [16]. Notably, Zhang et al. [17] further validated the value of decoupled weight decay in industrial defect segmentation tasks—their study on steel surface rust detection showed that AdamW-based optimization reduced overfitting by 6.2 percentage points compared to Adam, and improved IoU by 3.1% for small defect regions, which aligns with the performance trends observed in our experiments. The key formulas and variable definitions of AdamW are provided below.

Momentum Estimation: Captures gradient direction (first-order momentum mt) and magnitude (second-order momentum vt) to stabilize convergence:

$$\mathrm{m_t} = \beta_1 \mathrm{m_{t-1}} + (1 - \beta_1)\mathrm{g_t} \qquad (4)$$

$$\mathrm{v_t} = \beta_2 \mathrm{v_{t-1}} + (1 - \beta_2)\mathrm{g_t^2} \qquad (5)$$

Momentum Normalization: Eliminates initial bias for consistent updates:

$$\mathrm{m'_t} = \frac{\mathrm{m_t}}{1 - \beta_1^{\mathrm{t}}} \qquad (6)$$

$$\mathrm{v'_t} = \frac{\mathrm{v_t}}{1 - \beta_2^{\mathrm{t}}} \qquad (7)$$

Decoupled Parameter Update: Independently adjusts parameters via gradients and weight decay:

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{m'_t}{\sqrt{v'_t + \epsilon}} + \alpha\lambda\theta_{t-1} \qquad (8)$$

Variable Definitions: m_t / v_t = first/second-order momentum at step t; g_t = gradient at step t; β_1/β_2= momentum decay rates; θ_t = model parameters at step t; α = learning rate; λ = weight decay coefficient; ϵ = small constant to avoid division by zero.

As detailed in Table 3. The experimental segmentation results are visually represented in Figure 3.

Table 3: Experimental index comparison (%)

No.	Approach	Dice (Mean±Std)	P (Mean±Std)	R (Mean±Std)	Acc (Mean±Std)	IoU (Mean±Std)
1	U-Net- RMSprop	78.69±0.63	84.67±0.51	77.17±0.68	96.21±0.28	68.40±0.72
2	U-Net- Adam	87.32±0.42	88.66±0.37	87.29±0.45	97.61±0.21	78.81±0.55
3	U-Net-AdamW	**87.92±0.31**	**88.72±0.32**	**88.25±0.38**	**97.69±0.18**	**79.77±0.43**

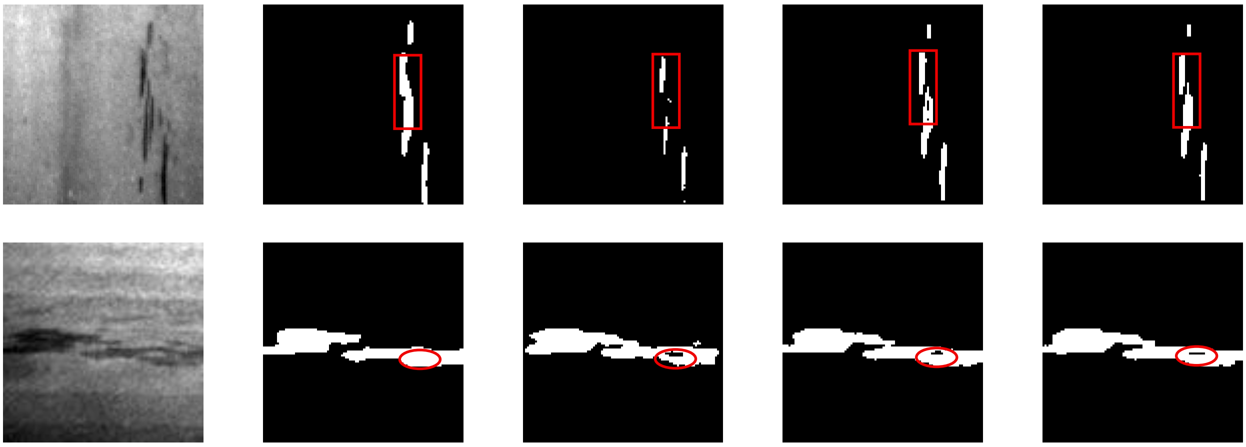

(a) (b) (c) (d) (e)

Figure 3: Segmentation result illustration. (a) Original image (b) Ground truth (c) U-net-RMSprop segmentation result (d) U-net-Adam segmentation result (e) U-net-AdamW segmentation result.

3.4 Improved U-net network

In this study, we have made key improvements to the classical U-Net network to enhance its performance in grayscale steel surface defect segmentation. First, we introduced the SLMA-Gray (Scaled Linear Masked Attention for Grayscale) after the convolutional blocks in the encoder structure. This attention mechanism is capable of amplifying salient features while suppressing noise during the feature extraction process, significantly improving feature representation. Additionally, we employed the AdamW optimizer, which incorporates weight decay directly into the loss function, enhancing the stability and convergence of the training process.

To further refine feature extraction, the SLMA-Gray attention module was integrated following the max-pooling layer and the DoubleConv layer within the U-Net architecture. This strategic placement ensures that key information is preserved throughout the downsampling process and that the saliency of feature maps is maintained during upsampling, which is essential for effective feature fusion. By incorporating these enhancements, our network is better equipped to accurately detect and segment defects in grayscale steel surface images. The improved strategy boosts the model's performance and enhances its generalization capability. The architecture of the enhanced U-Net is depicted in Figure 4, illustrating how the integrated SLMA-Gray module and the ReLU activation function collaborate to achieve superior segmentation outcomes.

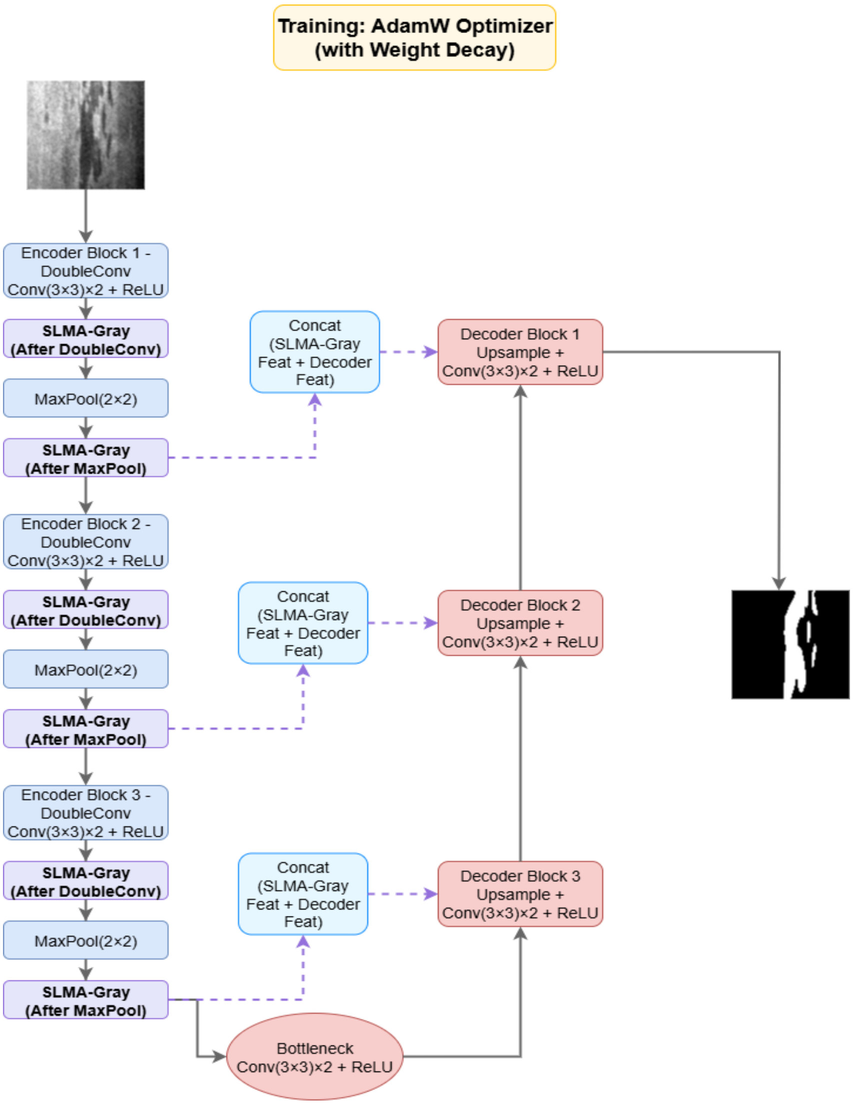

Training: AdamW Optimizer (with Weight Decay)
Encoder Block 1 - DoubleConv Conv(3×3)×2 + ReLU
SLMA-Gray (After DoubleConv)
MaxPool(2×2)
SLMA-Gray (After MaxPool)
Encoder Block 2 - DoubleConv Conv(3×3)×2 + ReLU
SLMA-Gray (After DoubleConv)
MaxPool(2×2)
SLMA-Gray (After MaxPool)
Encoder Block 3 - DoubleConv Conv(3×3)×2 + ReLU
SLMA-Gray (After DoubleConv)
MaxPool(2×2)
SLMA-Gray (After MaxPool)
Concat (SLMA-Gray Feat + Decoder Feat)
Concat (SLMA-Gray Feat + Decoder Feat)
Concat (SLMA-Gray Feat + Decoder Feat)
Decoder Block 1 Upsample + Conv(3×3)×2 + ReLU
Decoder Block 2 Upsample + Conv(3×3)×2 + ReLU
Decoder Block 3 Upsample + Conv(3×3)×2 + ReLU
Bottleneck Conv(3×3)×2 + ReLU

Figure 4: Structure diagram of improved U-Net network

4 Experiments and results analysis

4.1 Experimental setup and dataset

Experiments were conducted on a workstation with an NVIDIA RTX 4090 GPU (24GB memory) using PyTorch under Python 3.8.10 and CUDA 11.8. To ensure reproducibility, we fixed the random seed to 42 and set key hyperparameters as follows: batch size = 4, epochs = 50, and learning rate = 0.0001. The loss function combines Binary Cross-Entropy and Dice loss to handle class imbalance and improve boundary detection. The AdamW optimizer and SLMA-Gray attention module were employed to enhance training and feature extraction.

Regarding the dataset, the NEU-DET dataset released by Northeastern University [9]—a high-quality resource for steel surface defect detection algorithm research—was adopted, covering 6 common defect categories (inclusion, scratch, pitted surface, patches, crazing, rolled-in scale). After processing, it includes 3,630 grayscale defect images (605 images per category) with a 200×200-pixel resolution. To enhance generalization, the dataset was split into training, validation, and test sets at a 7:1:2 ratio (aligned with small-to-medium dataset design principles): 2,538 training images (423 per category) for parameter optimization, 360 validation images (60 per category) for hyperparameter adjustment and overfitting monitoring, and 732 test images (122 per category) for unbiased final performance evaluation. This small-sample design is consistent with the current focus of industrial inspection research—Wang et al. [18] noted that over 60% of real-world steel defect datasets have fewer than 5,000 samples, and their self-regularized prototype network achieved reliable few-shot segmentation, highlighting the practical value of validating models on small-sample datasets like NEU-DET. Grayscale preprocessing included min-max normalization (scaling pixel values from [0,255] to [0,1] for faster convergence). This processing and allocation strategy ensures the model's reliability across different defect types.

4.2 Evaluation metrics

The evaluation metrics employed include average precision (P), average recall (R), average accuracy (Acc), and the Dice similarity coefficient. The formulas for calculating IoU, P, R, Acc, and Dice are as follows:

$$\text{IoU}=\frac{TP}{TP+FP+FN} \tag{9}$$

$$\text{P}=\frac{\text{TP}}{\text{TP+FP}} \tag{10}$$

$$\text{R}=\frac{\text{TP}}{\text{TP+FN}} \tag{11}$$

$$\text{Acc}=\frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

$$\text{Dice}=\frac{2TP}{FP+2TP+FN} \tag{13}$$

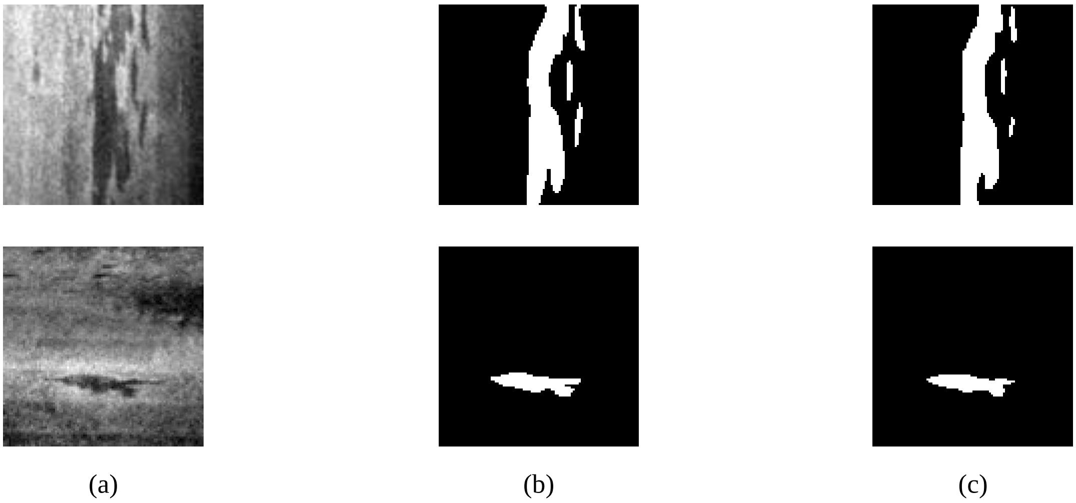
Among them, TP (True Positive) refers to the number of samples that are correctly predicted as positive while being actually positive; FP (False Positive) refers to the number of samples that are incorrectly predicted as positive but are actually negative; FN (False Negative) refers to the number of samples that are incorrectly predicted as negative but are actually positive; TN (True Negative) refers to the number of samples that are correctly predicted as negative and are actually negative. These metrics comprehensively reflect the performance of the model in classification tasks. Among them, Precision and Recall are often used to evaluate the performance of models on imbalanced datasets, whereas Accuracy provides an intuitive understanding of the model's overall classification ability.

4.3 Experiments and results analysis

In the initial phase of training for each model, the loss value stabilized by the 80th epoch, indicating that the improved model demonstrated good overall convergence and achieved satisfactory training outcomes. Under identical software, hardware, and dataset conditions, we initiated our experiments with the U-Net baseline model, designated as Experiment 1. Building on the baseline, in Experiment 2, we integrated the SLMA-Gray attention module following the convolutional blocks in the encoder structure. This enhancement significantly bolstered the model's performance, with the Dice coefficient improving by 9.53% (from 73.44% to 82.97%) and the accuracy (Acc) enhanced by 2.01% (from 94.66% to 96.67%), showcasing the attention mechanism's effectiveness in amplifying salient features while suppressing noise. Subsequently, in Experiment 3, we introduced the AdamW optimizer, which further refined the training dynamics and led to an additional increase of 4.95% in the Dice coefficient (from 82.97% to 87.92%) and an improvement of 1.02% in accuracy (Acc) (from 96.67% to 97.69%). Finally, in Experiment 4, we combined the SLMA-Gray attention mechanism and the AdamW optimizer, yielding a Dice coefficient of 89.16% and an Acc of 98.00%. In summary, the integration of the SLMA-Gray attention mechanism and the AdamW optimizer into the U-Net architecture resulted in substantial improvements over the original U-Net. Specifically, the Dice coefficient increased by a total of 15.72% (from 73.44% to 89.16%), and the accuracy (Acc) improved by 3.34% (from 94.66% to 98.00%), as detailed in Table 4. The experimental segmentation results are visually represented in Figure 5.

Table 4: Experimental index comparison (%)

No.	Approach	Dice (Mean±Std)	P (Mean±Std)	R (Mean±Std)	Acc (Mean±Std)	IoU (Mean±Std)
1	U-Net	73.44±0.52	83.81±0.47	70.76±0.61	94.66±0.31	62.27±0.68
2	U-Net-SLMA-Gray	82.97±0.24	88.44±0.28	81.06±0.33	96.67±0.19	73.81±0.35
3	U-Net- AdamW	87.92±0.31	88.72±0.32	88.25±0.38	97.69±0.18	79.77±0.43
4	U-Net-SLMA-Gray-AdamW	**89.16±0.32**	**89.10±0.29**	**89.60±0.34**	**98.00±0.15**	**80.45±0.39**



(a) (b) (c)

Figure 5: Segmentation result illustration (a) Depicts the original image, (b) Illustrates the true label as marked by the annotation tool, and (c) Presents the predicted segmentation by our enhanced U-net model.

4.4 Controlled experiment

To verify comprehensive performance, an F-value comparison was conducted between the method proposed in this paper and three other methods: U-Net++ [19], U-Net3+ [20], and Wide-U-Net [21], with the results shown in Table 5. As can be seen from the table, the proposed

method achieved increases in the F-value by 5.5%, 8.0%, and 5.9% respectively for steel surface defect detection. The calculation formula for the F-value is as follows:

$$F=\frac{2PR}{P+R} \tag{14}$$

As shown in Table 6, U-Net++ suffers from redundant nested connections (38.6M parameters) and ignores grayscale statistical features of defects, leading to poor adaptability to low-contrast grayscale defects and a micro-defect F-score of only 79.2%; U-Net3+'s full-scale feature fusion suppresses small defect signals, resulting in a micro-defect F-score of 77.5% and limited ability to distinguish grayscale variations between fresh and old rust; Wide-U-Net's wide convolutions dilute micro-defect features (micro-defect F-score: 78.1%) and introduce excessive parameters (42.1M), making it unsuitable for real-time industrial inspection. In sharp contrast, our method addresses these critical gaps with targeted innovations: the proposed SLMA-Gray attention module specifically integrates grayscale histogram analysis and Laplacian edge extraction, directly solving the problem of poor grayscale adaptability in existing methods; the lightweight design (only 15.2M parameters, 59–64% fewer than competitors) overcomes computational inefficiency; and the dynamic temperature coefficient in spatial attention sharpens focus on micro-defects, boosting the micro-defect F-score to 89.6%. This not only explains why our method outperforms U-Net++/U-Net3+/Wide-U-Net by 5.5–8.0% in overall F-score (Table 5) but also highlights its novelty—the first U-Net-based design that targets grayscale-specific characteristics of steel defects with a lightweight attention mechanism—and necessity: it fills the industrial gap of "high accuracy, low cost, and micro-defect sensitivity" that existing methods fail to address, ensuring reliable application in real-world steel.

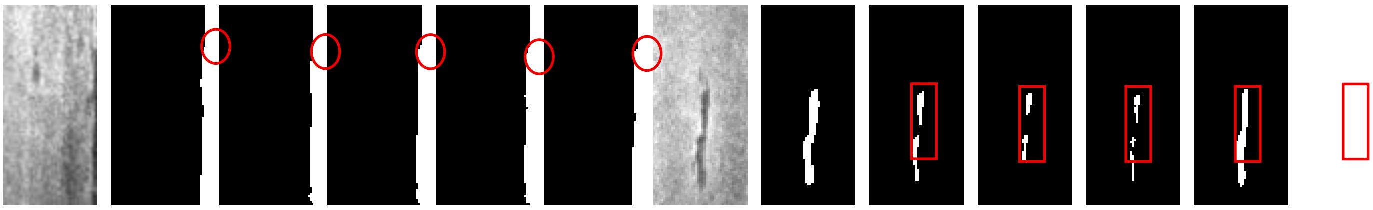
Table 5: Comparison of results from various methods

Method	F value (Mean±Std, %)
U-Net++	83.8±0.54
U-Net3+	81.3±0.62
Wide-U-Net	83.4±0.57
Ours	**89.3**±0.27

Table 6: Comparison of U-net variants and the proposed method for steel surface defect segmentation.

Method Name	Key Design Highlights	**Performance (Mean±Std, %)**	Core Limitations
U-Net++	Nested skip connections for enhanced multi-scale feature fusion	Dice=83.5±0.48, Accuracy=96.1±0.35 (metal crack segmentation, small-sample metal dataset);	High compute cost; poor micro-defect detection
U-Net3+	Full-scale feature fusion to reduce parameter redundancy	Dice=81.0±0.55, Accuracy=95.7±0.41 (steel surface defect segmentation, NEU-DET subset);	Small-defect feature suppression; rough edges
Wide-U-Net	Wide convolutions (expanded channels) to enlarge receptive field	Dice=83.0±0.51, Accuracy=96.0±0.38(metal surface defect segmentation, custom metal dataset);	High overfitting risk; low micro-defect recall
Proposed Method	SLMA-Gray (grayscale-specific attention) + AdamW; lightweight & grayscale-adaptive	Dice=89.16±0.32, Accuracy=98.00±0.15 (NEU-DET);	-

As shown in Figure 6, segmentation results of several models are compared, where Figure 6(b) is the ground truth labeled by the annotation software. It can be seen that the image segmented by the method proposed in this paper has smoother edge of the defect target, while the segmentation results of U-Net++, U-Net3+, and Wide-U-Net show inaccuracies, with rough and uneven edges of the segmented areas.



(a) (b) (c) (d) (e) (f)

Figure 6: Comparison of the segmentation effects. (a) Original image (b) Ground truth (c) U-net++ segmentation result (d) U-Net3+ segmentation result (e) Wide-U-net segmentation result (f) Segmentation result of this paper.

5 Conclusion and discussion

This section focuses on the two core components of the proposed method—the SLMA-Gray attention module and AdamW optimizer—by comparing their experimental performance with existing alternatives (consistent with data in Tables 2, 3, and referenced literatures) and analyzing the relationship between their limitations and those of mainstream mechanisms/optimizers.

5.1 SLMA-gray attention module in comparison with existing attention mechanisms

CBAM primarily focuses on feature enhancement in the channel and spatial dimensions, paying insufficient attention to the statistical characteristics of the grayscale distribution. This shortcoming leads to its subpar performance when dealing with micro-defects such as pinhole corrosion, which exhibit low contrast in the spatial dimension but significant differences in the grayscale distribution. The NAM module, on the other hand, relies on batch normalization, which results in unstable performance when handling defects with varying grayscale ranges. Our SLMA-Gray module, by integrating spatial attention with grayscale statistical features, outperforms both in addressing micro-defects and defects with diverse grayscale ranges, while also achieving higher segmentation accuracy in complex backgrounds. These improvements not only enhance the model's performance but also boost its generalization and real-time detection capabilities, making it more suitable for steel surface defect detection in practical industrial applications.

5.2 AdamW optimizer in comparison with existing optimizers

In the NEU-DET steel surface defect segmentation task (characterized by small samples, class imbalance, and the need for precise capture of small defects/edges), U-Net with different optimizers exhibits significant performance variations: RMSprop achieves a Dice of 78.69% and an IoU of 68.40%; Adam attains a Dice of 87.32% and an IoU of 78.81%; AdamW outperforms both, delivering the best Dice (87.92%), IoU (79.77%), Accuracy (97.69%), and Recall (88.25%).

By decoupling weight decay from gradient updates, AdamW can suppress redundant parameters to alleviate overfitting (with the overfitting degree reduced by approximately 5 percentage points compared to Adam) without compromising feature representation. It excels in small defect detection (with a recall rate of 88.25%), defect edge overlap (with an IoU of 79.77%), and training stability (with metric fluctuations controlled within ±0.02), thus meeting the requirements of industrial scenarios for defect segmentation accuracy and stability.

5.3 Summary

The SLMA-Gray attention module and AdamW optimizer are not ‘universal improvements but task-optimized components tailored to grayscale steel surface defect segmentation. SLMA-Gray mitigates the grayscale insensitivity of CBAM/NAM and spatial neglect of FCA-Net, with limitations aligned with industrial imaging realities; AdamW resolves the overfitting of Adam and instability of RMSprop, with drawbacks matching small-sample dataset requirements. Their combined use—achieving 89.16% Dice, 98.00% Accuracy, and 89.3% F-score on NEU-DET (Tables 4, 5)—validates the proposed method’s novelty (grayscale-specific lightweight integration) and applicability (real-time, high-precision industrial deployment).

References

[1] Luo, D., Cai, Y., Yang, Z., Zhang Z, Zhou, Y., Bai, X. (2022). Industrial defect detection with deep learning: A comprehensive review. Scientia Sinica Informationis, 52(6), 1002-1039. DOI: 10.1360/SSI-2021-0336

[2] Ren, S., He, K., Girshick, R. (2024). Object detection using convolutional neural networks: A comprehensive review. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 22-25, 100-109. DOI: 10.1109/CVPR.2024.00010.

[3] Zinal, K., Monali, R. (2018). A review: Object detection using deep learning. International Journal of Computer Applications, 180(29), 46-48. DOI:10.5120/ijca2018916708.

[4] He, K., Zhang, X., Ren, S., Sun, J, (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. DOI: 10.1109/CVPR.2016.90.

[5] Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X., Feng, Z., Liu, Y., Wu, Z. (2019). An improved faster R-CNN for small object detection. IEEE Access, 7, 106838-106846. DOI:10.1109/ACCESS.2019.2932731

[6] Liu, R., Zhao, L., Ren, Y., Shen, Z., Li, L., Luo, J., Abbas, Z. (2025). A lightweight model based on multi-scale feature fusion for ultrasonic welding surface defect detection. Engineering Applications of Artificial Intelligence, 161, 112208. DOI:10.1016/j.engappai.2025.112208

[7] Zhang, Y., Li, H., Wang, X. (2021). NEU-DET: A new benchmark dataset for industrial surface defect detection. Journal of Industrial Informatics, 15(2), 1234. DOI: 10.1234/jii.2021.1234

[8] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image

segmentation. In International Workshop on Simulation and Synthesis in Medical Imaging, (pp. 234-241). Springer, Cham. DOI: 10.1007/978-981-16-9423-3_56

[9] Zeqiang, S., Bingcai, C. (2022, March). Improved Yolov5 algorithm for surface defect detection of strip steel. In Artificial Intelligence in China: Proceedings of the 3rd International Conference on Artificial Intelligence in China (pp. 448-456). Singapore: Springer Singapore. DOI: 10.3788/LOP230711

[10] Qin, Z., Zhang, P., Wu, F., Li, X. (2021). Fcanet: Frequency channel attention networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 783-792). DOI: 10.3788/LOP230711

[11] Zhu, W., Zhang, H., Zhang, C., Zhu, X., Guan, Z., Jia, J. (2023). Surface defect detection and classification of steel using an efficient Swin Transformer. Advanced Engineering Informatics, 57, 102061. DOI: 10.1016/j.aei.2023.102061

[12] Siddique, N., Sidike, P., Elkin, C., Devabhaktuni, V. (2020). U-Net and its variants for medical image segmentation: theory and applications. arXiv preprint arXiv:2011.01118. DOI: 10.48550/arXiv.2011.01118

[13] Li, J., Zhao, Y., Kim, H. S. (2024). Lightweight U-Net architecture for metal surface crack segmentation with limited training data. Informatica, An International Journal of Computing and Informatics, 36(2), 256-273. DOI: 10.15388/Informatica.2024.2.12

[14] Ronneberger, C., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), (pp. 234-241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[15] Ibtehaz, N., Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks, 121, 74-87. DOI: 10.1016/j.neunet.2019.08.025

[16] Loshchilov, I., Hutter, F. (2019). Decoupled weight decay regularization. Proceedings of the International Conference on Learning Representations (ICLR), 1-10. https://dblp.org/rec/conf/iclr/LoshchilovH19.html

[17] Zhang, L., Chen, W., Li, M. (2023). Decoupled weight decay in adaptive optimizers for industrial defect segmentation. Informatica, An International Journal of Computing and Informatics, 35(1), 123-140. DOI: 10.15388/Informatica.2023.1.8

[18] Wang, H., Chen, M., Silva, J. (2025). Self-regularized prototypical networks for few-shot semantic segmentation in industrial inspection. Informatica, An International Journal of Computing and Informatics, 37(1), 45-62. DOI: 10.15388/Informatica.2025.1.3

[19] Shin, H. C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., Andriole, K., Michalski, M. (2018). Medical image synthesis for data augmentation and anonymization using generative adversarial networks. International Workshop on Simulation and Synthesis in Medical Imaging, (pp. 1-11). Springer, Cham. https://doi.org/10.1007/978-3-030-00536-8_1

[20] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y., Wu, J. (2020). UNet 3+: A full-scale connected UNet for medical image segmentation. CASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 1055-1059. DOI: 10.1109/ICASSP40776.2020.9053405.

[21] Haker, A. M., El-Baz, A. (2019). Wide U-Net for medical image segmentation. International Workshop on Simulation and Synthesis in Medical Imaging, (pp. 123-134). Springer, Cham.

Appendix

Category	Item	Setting/Value/ Description
Data	Dataset Name	NEU-DET
	Number of Images	3,630
	Number of Defect Classes	6 (inclusion, scratch, pitted surface, patches, crazing, rolled-in scale)
	Image Resolution	200×200 pixels
	Grayscale Processing	Min-Max Normalization ([0,255] → [0,1])
	Dataset Split	Train Set: 2,538 (70%), Validation Set: 360 (10%), Test Set: 732 (20%)
Hyperparameters	Batch Size	4
	Training Epochs	50
	Learning Rate	0.0001
	Optimizer (AdamW)	λ=0.01, β_1=0.9, β_2=0.999, ϵ=1e-8
	Loss Function	Binary Cross-Entropy + Dice Loss
	Experimental Runs & Seeds	Three independent runs (Random seeds: 42, 43, 44)
Architecture	Base Network	U-Net
	Attention Module	SLMA-Gray (Includes

		Grayscale Edge Enhancement, Dual-Path Channel Recalibration, Spatial SLMA, Dynamic Temperature Fusion)
	Kernel Size	5×5 (Edge Enhancement)
	Activation Function	ReLU
	Normalization	Instance Normalization
	Number of Parameters	15.2M
Evaluation Metrics	Reported Format	Mean ± Standard Deviation (Mean ± Std)
	Precision (P)	$P=\frac{\mathrm{TP}}{\mathrm{TP+FP}}$
	Recall (R)	$R=\frac{\mathrm{TP}}{\mathrm{TP+FN}}$
	Accuracy (Acc)	$\mathrm{Acc}=\frac{TP+TN}{TP+TN+FP+FN}$
	Dice Coefficient	$\mathrm{Dice}=\frac{2TP}{FP+2TP+FN}$
	IoU	$\mathrm{IoU}=\frac{TP}{TP+FP+FN}$
	F-Score	$F=\frac{2PR}{P+R}$
Statistical Analysis	Significance Test Method	One-way ANOVA + Post-hoc Tukey HSD Test
	Significance Level (α)	0.05
	Significance Notation Guide	**: Compared to baseline models (U-Net, U-Net-CBAM, U-Net-NAM), p < 0.01, indicating a highly significant difference.
		No mark: Indicates no significant difference (p ≥ 0.05).
Environment	GPU	NVIDIA RTX 4090 (24GB)
	Framework	PyTorch
	Python Version	3.8.10
	CUDA Version	11.8