Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

AW-HRNet: A Lightweight High-Resolution Crack Segmentation Network Integrating Spatial Robustness and Frequency-Domain Enhancement

Dewang Ma[1],Tong Lu[2]

[1]College of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China
[2]College of Environment & Safety Engineering, Fuzhou University, Fuzhou, 350108, Fujian, China

Abstract: The study presents AW-HRNet, a lightweight high-resolution crack segmentation network that couples Adaptive residual enhancement (AREM) in the spatial domain with Wavelet-based decomposition–reconstruction (WDRM) in the frequency domain. AREM introduces a learnable channel-wise scaling after standard 3 × 3 convolution and merges it through a residual path to stabilize crack-sensitive responses while suppressing noise. WDRM performs DWT to decouple LL/LH/HL/HH sub-bands, conducts lightweight cross-band fusion, and applies IDWT to restore detail-enhanced features, unifying global topology and boundary sharpness without deformable offsets. Integrated into a high-resolution backbone with auxiliary deep supervision, AW-HRNet attains 79.07% mIoU on CrackSeg9k with only 1.24M parameters and 0.73 GFLOPs, offering an excellent accuracy–efficiency trade-off and strong robustness for real-world deployment.

Keywords: Crack segmentation; Lightweight model; Wavelet decomposition and reconstruction; Feature enhancement

Online publication: November 3, 2025

1. Introduction

Cracks on road surfaces are among the most visible and direct manifestations of pavement degradation during service [1]. Accurately segmenting the spatial distribution and geometric morphology of cracks not only provides a reliable quantitative basis for pavement condition assessment, maintenance prioritization, and lifecycle cost control, but also contributes to improved road safety and reduced potential risks [2]. Although traditional manual inspection methods can achieve a certain level of reliable judgment based on expert experience, they often suffer from inefficiency and lack of consistency when applied to large-scale road networks. Moreover, their detection performance is highly susceptible to external factors variations in lighting, weather conditions, and image acquisition parameters [3]. Therefore, developing an automated, scalable, and verifiable crack segmentation framework has become a research topic of both engineering relevance and scientific significance [4].

Prior to the widespread adoption of deep learning, crack detection primarily relied on handcrafted feature extraction methods combined with traditional machine learning classifiers [5]. Typical approaches involved using

edge detectors and texture filters to extract statistical features such as Histogram of Oriented Gradients (HOG) or Local Binary Patterns (LBP). Preliminary segmentation was often performed via thresholding and morphological operations, followed by classification using models like Support Vector Machines (SVM) or Random Forests [6]. While these methods could perform reasonably well under controlled imaging conditions with simple backgrounds, their robustness to shadows, stains, surface material variations, and complex textures was limited, and their generalization capabilities were weak [7]. Furthermore, they were highly sensitive to hyperparameter tuning and struggled to capture the thin, curved, and branched topology of cracks, resulting in significant performance degradation in cross-scene applications [8].

With the advancement of Convolutional Neural Network, crack segmentation has evolved into an end-to-end pixel-level prediction task, enabling joint optimization of feature extraction and classification, and significantly enhancing the model's adaptability to real-world complex road conditions [9]. Encoder-decoder architectures (e.g., U-Net and its variants) leverage skip connections to combine shallow spatial details with deep semantic features during decoding, which helps preserve spatial continuity of cracks [10]. Multi-scale context aggregation modules (e.g., Feature Pyramid Networks) have been introduced to improve the representation of cracks with varying widths and shapes [11]. In addition, techniques such as channel and spatial attention mechanisms, dense connections, and deep supervision have further optimized gradient flow and feature reuse, effectively enhancing crack boundary clarity and topological connectivity [12]. Overall, deep learning-based methods have shown clear advantages over traditional approaches in maintaining segmentation continuity, improving edge accuracy, and increasing robustness to environmental variations.

Despite the significant progress, crack segmentation still faces two core challenges.

(1) Conventional convolutions with fixed receptive fields lack adaptability to geometric deformations, making them ineffective in capturing the thin, curved, and branching nature of cracks [13].

(2) Although deformable convolutions introduce spatial flexibility, they can suffer from excessive deformation and offset drift during offset estimation, leading to boundary distortions, topological structure damage, and training instability [14].

To address the first issue of geometric adaptability, the study proposes a robust feature enhancement convolution module called AREM (Adaptive Residual Enhancement Module). This module introduces a learnable channel-wise scaling factor following standard convolution operations, combined with a residual enhancement mechanism. It adaptively calibrates channel responses, strengthens crack-related features, and suppresses noise. Without sacrificing feature map resolution, this approach significantly improves the stability and noise resistance of feature representations, enabling a more accurate depiction of complex crack morphologies. To tackle the second issue of multi-scale feature modeling, the study designs a wavelet-based module called WDRM (Wavelet Decomposition and Reconstruction Module). This module employs Haar wavelet transforms to decompose feature maps into low-frequency (LL) and high-frequency (LH, HL, HH) sub-bands. After lightweight convolution and residual interaction across sub-bands, the inverse wavelet transform reconstructs the enhanced feature map. This allows unified modeling of low-frequency topological constraints and high-frequency detail enhancement, effectively mitigating the high-frequency information loss caused by repeated downsampling.

In summary, this paper introduces a segmentation network named AW-HRNet, which integrates the AREM and WDRM modules into a high-resolution lightweight architecture from two complementary perspectives: robust feature modeling in the spatial domain and high–low frequency decoupling in the frequency domain. The proposed network effectively addresses the twin challenges of unstable feature representation and loss of high-frequency details, offering a practical and efficient solution for accurate crack segmentation.

2. Literature review

This study primarily focuses on two research directions: the design of lightweight neural networks and crack segmentation methods. In the following, the study reviews represent progress in both areas and emphasize how their integration addresses the challenges of real-world deployment and complex scenarios.

2.1. Lightweight neural networks

To meet the demands of edge computing and mobile deployment, lightweight neural networks have emerged as a key trend in the computer vision community [15]. Early representative models include SqueezeNet, which compresses parameters using the Fire module; the MobileNet series, which decouples spatial filtering and channel mixing via depthwise separable convolutions to significantly reduce computational cost [16]; ShuffleNet, which enhances inter-group information exchange through channel shuffling [17]; and GhostNet, which generates "ghost features" using cheap linear operations as a low-cost substitute for standard convolutions. These methods effectively control model size and inference latency while maintaining accuracy.

In recent years, specialized lightweight strategies for low-level and structured vision tasks have emerged. For example, LSNet draws inspiration from the human visual system and proposes a hybrid paradigm of "large kernel perception + small kernel aggregation": global context is captured using large kernel depthwise convolutions, while fine-grained local features are fused using small kernels to achieve a better accuracy–efficiency trade-off. Another line of research is represented by SCSegamba, which builds a lightweight structure-aware network based on state-space models (SSM) and the Mamba architecture. This approach compresses parameters using low-rank decomposition and dynamic gated bottleneck convolutions, while leveraging structure-aware state-space modules to model topological relationships between pixels, thus enhancing semantic continuity. DSAN, on the other hand, combines deformable convolutions with spatial attention, proposing DSCN (Deformable Strip Convolution with Single-Axis Constraint). By limiting offsets to a single axis and using linear interpolation without modulation masks, it maintains geometric flexibility while further reducing parameter count and computational overhead.

2.2. Crack segmentation methods

As a critical task in road defect detection and infrastructure health monitoring, crack segmentation has evolved from traditional image processing methods to end-to-end deep learning frameworks [18]. Early methods relied on edge or texture operators (e.g., Canny, Sobel, Gabor) to extract statistical features such as HOG and LBP, followed by thresholding and morphological processing. Classification was then performed using models such as SVMs or Random Forests. These methods performed reasonably well under controlled imaging conditions and simple backgrounds, but were highly sensitive to lighting changes, shadows, material variations, and complex textures, leading to poor generalization performance.

With the advancement of deep learning, crack segmentation has entered a new stage of joint optimization for accuracy and efficiency. On one hand, encoder–decoder architectures (e.g., U-Net and its variants) utilize skip connections to reintegrate fine details during upsampling, improving boundary and fine-structure preservation. On the other hand, multi-scale context modeling enhances adaptability to cracks of various scales and complex backgrounds. Additionally, techniques such as channel/spatial attention, dense connections, and deep supervision contribute to optimizing gradient flow and connectivity representation. Recent research has advanced along two parallel directions, lightweight design and feature modeling, to meet the demand for efficient and deployable solutions. For instance, HMBCNet achieves lightweight modeling through the decoupling of downsampling, the multi-branch dilation, and the cascade fusion strategy. DFPA-Net employs dual-level partial convolutions

and multi-scale feature attention, showcasing extreme lightweight potential in targeted scenarios. Meanwhile, CrackdiffNet introduces a novel generative prior based on diffusion models, which maps segmentation masks to synthetic images and incorporates skeleton-based features to estimate crack width.

3. Methodology

3.1. Overall network architecture

AW-HRNet follows a dual-branch encoder–multi-scale enhancement–progressive decoder paradigm, designed to preserve high-resolution features while effectively integrating global semantics with fine-grained crack details (**Figure 1**). The input is first passed through shallow convolution and normalization layers to obtain features at a unified scale.In the dual-branch encoder, the high-resolution path stacks multiple AREM modules to enhance feature stability and robustness through residual convolution and channel-wise scaling mechanisms. Meanwhile, the low-resolution path expands the receptive field via downsampling to aggregate global context, and interacts with the high-resolution branch to supplement semantic information [19]. To further mitigate detail loss caused by downsampling, the high-resolution branch is followed by the WDRM module. This module first applies a Discrete Wavelet Transform (DWT) to decompose the features into LL, LH, HL, and HH sub-bands, enabling frequency-domain modeling of low-frequency topological constraints and high-frequency detail capture. Then, lightweight convolution and residual enhancement are applied to perform cross-subband information interaction, which is then followed by an Inverse DWT (IDWT) to reconstruct the features back to the original scale, thereby achieving a unified representation of low- and high-frequency information.

In the decoder stage, a lightweight segmentation head serves as the main output branch, complemented by two auxiliary heads for deep supervision. The final predictions are upsampled to the input resolution using bilinear interpolation, and a 1×1 convolution maps the features to a pixel-wise crack probability map, which can be jointly optimized with the task-specific loss function [20].

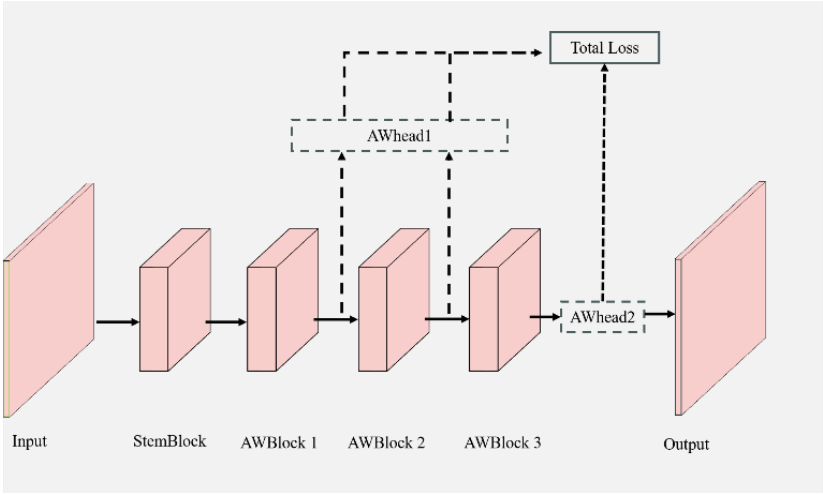

Figure 1. Overall Architecture of AW-HRNet.

In summary, AW-HRNet enhances the classical high-resolution framework in both the spatial and frequency domains: the AREM module improves the robustness and stability of spatial features, while the WDRM module leverages explicit DWT/IDWT transformations to decompose and reconstruct high- and low-frequency sub-bands,

thereby enabling implicit multi-scale feature enhancement. Working in synergy, these two modules allow the network to simultaneously model global topology and capture fine-grained crack structures without introducing explicit deformable offsets. As a result, the network achieves a better trade-off between accuracy and efficiency, along with greater robustness in complex scenarios.

3.2. Adaptive residual enhancement module

In complex environments, cracks are often affected by noise, and their edge details can easily become confused with background textures, which frequently leads to instability in local feature representation [21]. Traditional convolution operations, with fixed kernel weights, lack the ability to adaptively adjust channel responses, often resulting in feature bias and weakened details, thereby compromising the modeling of crack continuity. To address this, the study proposes a robust feature enhancement module called AREM (Adaptive Residual Enhancement Module). This module jointly models residual convolution and learnable scaling factors to effectively improve the robustness and adaptability of feature extraction (**Figure 2**).

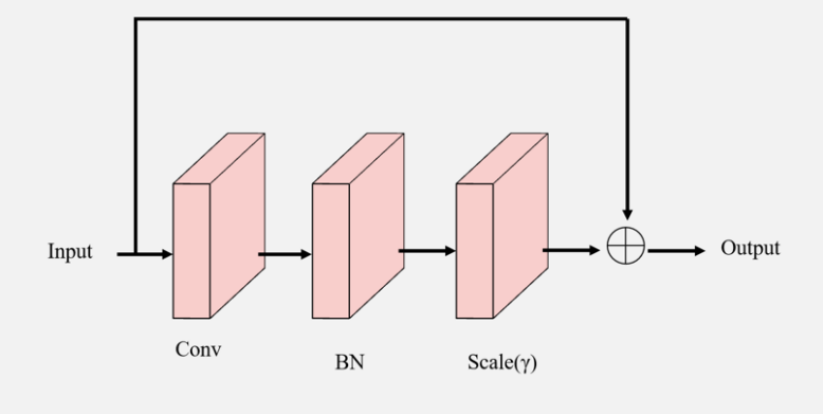

Figure 2. Structure of the AREM Module.

More specifically, the AREM module consists of convolution, batch normalization, and nonlinear activation layers, with a learnable channel-wise scaling factor introduced at the end to enable adaptive feature modulation. A residual connection fuses the enhanced features with the input features, which not only facilitates stable gradient propagation but also enhances feature robustness, all without changing the spatial resolution or the number of channels in the backbone [22]. The _ScaleModule, which introduces the learnable channel-wise scaling factors, Γ is designed to amplify crack-sensitive channels while suppressing noise-sensitive ones, thereby mitigating detail attenuation and feature bias caused by fixed sampling patterns. AREM is deployed along the high-resolution path of the backbone network, aiming to stabilize feature representations after multi-scale interactions. In collaboration with the frequency-domain WDRM module, AREM provides fine-grained and robust enhancement in the spatial domain, allowing the network to achieve more stable crack representations and better generalization—without introducing explicit deformable offsets.

$$F(X) = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}_{3\times3}(X)\right)\right) \tag{1}$$

$$U = \Gamma \mathbf{e}\ F(X) \tag{2}$$

$$\mathrm{Y} = \mathrm{X} + \mathrm{U} \tag{3}$$

Let ⊙ denote the standard element-wise multiplication. Accordingly, the branch feature is defined as . Given the input feature , a channel-wise scaling factor is introduced in parallel and applied for feature re-scaling. This mechanism enhances crack-sensitive channels while suppressing noise-sensitive ones, thereby mitigating feature degradation caused by fixed sampling. The final output is expressed as .

3.3. Wavelet decomposition and reconstruction module

Cracks often exhibit irregular, slender, and branched shapes, and frequently share high visual similarity with background textures. This results in significant differences in the representation of high-frequency details and low-frequency structures. Traditional convolutional networks rely on fixed sampling positions for feature modeling. After multiple downsampling stages, they tend to suffer from blurred high-frequency edges and insufficient global low-frequency structure, ultimately leading to broken crack contours and topological discontinuities. To address these issues, the study proposes the Wavelet Decomposition and Reconstruction Module (WDRM), which decouples and reintegrates high- and low-frequency information via Discrete Wavelet Transform (DWT) and Inverse DWT (IDWT), thereby enhancing the network's multi-scale feature representation capability (**Figure 3**).

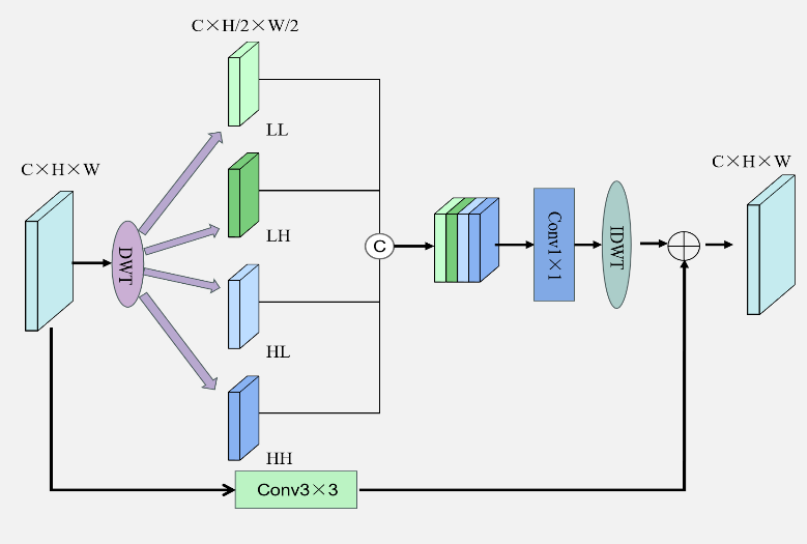

Figure 3. Structure of the WDRM Module.

Specifically, the input feature is first decomposed using fixed wavelet filters through DWT, producing four sub-bands: LL, LH, HL, and HH. The LL sub-band retains the overall structural information, while the three high-frequency sub-bands (LH, HL, and HH) capture boundary and texture details in different orientations. The four sub-bands are concatenated along the channel dimension and passed through a 1 × 1 convolution for feature compression and fusion. This is followed by a convolutional enhancement step to improve semantic consistency and robustness. After enhancement, the fused feature map is restored to the original resolution via a fixed IDWT, and then added to the input branch via residual connection, enabling feature reconstruction and enhancement.

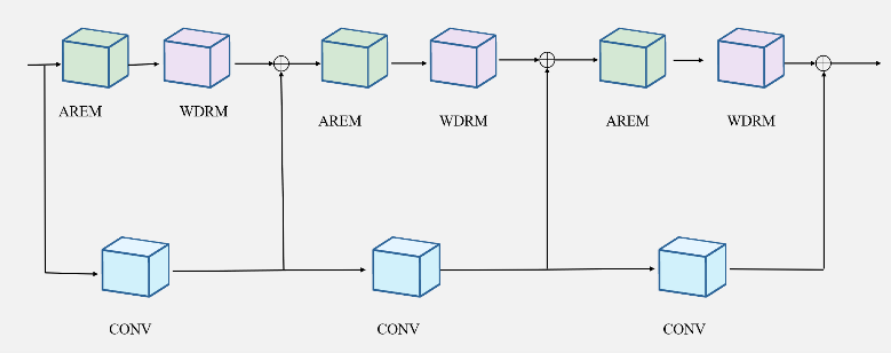

Figure 4. Structure of the AWBlcok.

Within the overall network, the LL sub-band in WDRM provides constraints on global connectivity, while the high-frequency sub-bands enhance boundary and texture representation. This effectively mitigates edge blurring and topological fragmentation caused by repeated downsampling. Ultimately, WDRM serves as a bridge between the spatial and frequency domains, complementing spatial-domain enhancement modules such as AREM. This integration allows the model to maintain high-resolution feature representations while simultaneously enhancing noise robustness and boundary precision, thereby significantly improving the overall accuracy and generalization capability in crack segmentation tasks (**Figure 4**).

4. Experiments

4.1. Datasets

To evaluate the effectiveness of the proposed model, the study adopts the CrackSeg9k dataset as the experimental benchmark [23]. CrackSeg9k is a medium-scale semantic segmentation dataset specifically constructed for crack detection and segmentation tasks. It contains approximately 8,751 high-quality crack images, covering surfaces of various materials such as concrete, ceramic, and brick, with a unified resolution of 400×400 pixels. The dataset is divided into two categories: crack and background. Under a fixed random seed, the images are split into training, validation, and test sets in a ratio of 7:1:2.

4.2. Experimental settings

All experiments are conducted on the CrackSeg9k dataset, with input image size fixed at 400 × 400. The training batch size is set to 56, and the total number of iterations is 100k. The study uses the SGD optimizer with a momentum of 0.9 and a weight decay of 5×10^{-4}. The initial learning rate is 0.01, which is decayed following the Polynomial Decay strategy with a power of 0.9, until reaching 0. In addition, a warm-up phase is applied during the first 2,000 iterations, starting from a learning rate of 1×10^{-5}. During validation, only normalization is applied. The OHEM Cross-Entropy Loss is used as the loss function, with the coefficient of the main branch set to 1.0, and the two auxiliary supervision branches both set to 0.5, in order to mitigate class imbalance and enhance training stability.

4.3. Evaluation Metrics

To comprehensively assess the performance of segmentation models with varying depths, the study employs four evaluation metrics: Precision (Pr), Recall (Re), F1-score, and mean Intersection-over-Union (mIoU). The definitions are as follows:

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

$$Re = \frac{TP}{TP + FN} \tag{5}$$

$$\mathrm{Fl} = 2 \cdot \frac{\mathrm{Pr} \cdot \mathrm{Re}}{\mathrm{Pr} + \mathrm{Re}} \tag{6}$$

$$mIoU = mean\left(\frac{TP}{TP + FP + FN}\right) \tag{7}$$

True Positive (TP) refers to pixels correctly identified as cracks, while False Positive (FP) denotes background pixels that are incorrectly classified as cracks. False Negative (FN) corresponds to crack pixels that are mistakenly predicted as background. The computational cost and model size are evaluated using two indicators: the number of floating-point operations (GFLOPs) and the number of parameters (Params).

4.4. Comparison with State-of-the-Art Models

This study focuses on designing a lightweight crack segmentation model and compares the proposed AW-HRNet with several representative methods (**Table 1**). Overall, classical networks such as UNet and PSPNet achieve high segmentation accuracy. For instance, UNet reaches an mIoU of 79.15%, but at the cost of 13.40M parameters and 75.87 GFLOPs. PSPNet similarly requires 21.07M parameters and 54.20 GFLOPs, making deployment costly in resource-constrained scenarios. In terms of lightweight architectures, BiSeNetv2 and DDRNet demonstrate advantages in inference efficiency, yet their accuracies remain insufficient, with mIoUs of 75.09% and 76.77%, respectively. HrSegNet-B16 achieves the best efficiency, requiring only 0.61M parameters and 0.66 GFLOPs, while delivering an mIoU of 78.02% and the highest ACC of 98.54%, highlighting its superior lightweight design. By comparison, AW-HRNet delivers an mIoU of 79.07% and an F1 score of 87.00%, with only a slight increase in complexity, representing a 1.05% improvement over HrSegNet-B16. Compared with UNet, AW-HRNet achieves nearly equivalent accuracy while requiring only about 1/11 of the parameters and 1/100 of the computational cost, clearly demonstrating its excellent accuracy–efficiency balance.

In summary, AW-HRNet achieves segmentation performance comparable to heavyweight models while maintaining extremely low computational cost, and shows consistent improvements over other lightweight methods, underscoring its strong potential for deployment in resource-constrained scenarios.

Table 1. Comparisons with state-of-the-art on CrackSeg9k.

Model	mIoU (%)	F1 (%)	ACC (%)	Params(M)	GFLOPs
UNet [10]	79.15	87.21	98.50	13.40	75.87
PSPNet [24]	76.78	85.39	98.14	21.07	54.20
BiSeNetv2 [25]	75.09	83.71	98.16	2.33	4.93
DDRNet [26]	76.77	85.45	98.40	20.18	11.11
HrSegNet-B16 [27]	78.02	86.18	98.54	0.61	0.66
AW-HRNet	79.07	87.00	98.41	1.24	0.73

4.5. Ablation study

To investigate the contribution of each module, the study conducts ablation experiments on the CrackSeg9k dataset using HrSegNet-B16 as the baseline, and progressively introduces AREM and WDRM (**Table 2**).

Table 2. Ablation study on CrackSeg9k (baseline: HrSegNet-B16)

Method	mIoU (%)	Params (M)	GFLOPs
HrSegNet-B16	78.02	0.61	0.66
+ AREM	78.64	0.72	0.72
+ WDRM	78.89	0.91	0.78
AW-HRNet	79.07	1.24	0.73

Introducing AREM on the HrSegNet-B16 baseline increases mIoU from 78.02% → 78.64% (+0.62 pp) with only +0.11M parameters and +0.06 GFLOPs. Adding WDRM further raises mIoU to 78.89% (+0.87 pp over baseline) with a modest footprint of 0.91M params and 0.78 GFLOPs. When both modules are combined in AW-HRNet, mIoU reaches 79.07% (+1.05 pp over baseline) while computation and model size remain low (1.24M, 0.73 GFLOPs). In summary, AREM strengthens spatial robustness via residual reinforcement and channel-wise scaling, whereas WDRM restores high-frequency details through wavelet decomposition–reconstruction; the two are complementary and jointly improve segmentation accuracy and stability.

5. Conclusion

In this work, the study addressed the challenges of fine-grained feature loss, boundary blurring, and insufficient lightweight design in crack segmentation by proposing AW-HRNet, a lightweight network built upon a high-resolution framework. Methodologically, the AREM module was introduced to enhance the stability and robustness of feature representation through residual convolution and channel-wise scaling, while the WDRM module was designed to explicitly model high- and low-frequency information in the frequency domain via wavelet decomposition and reconstruction. This enables the preservation of low-frequency topological integrity while strengthening crack boundaries and texture details. Working synergistically, the two modules allow the network to achieve an excellent balance between accuracy and efficiency, while extensive experiments across multiple datasets validated their effectiveness and robustness.

For future work, the study plans to explore the integration of learnable wavelet bases with Transformer architectures to further improve cross-scene generalization capability. In addition, the study aims to extend the proposed network to more complex engineering infrastructure scenarios, such as bridges, tunnels, and airport runways, thereby advancing the practical application of intelligent crack monitoring and maintenance.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Yuan Q, Shi Y, Li M, 2024, A Review of Computer Vision-Based Crack Detection Methods in Civil Infrastructure: Progress and Challenges. Remote Sensing, 16(16): 2910.

[2] Gavilán M, Balcones D, Marcos O, et al., 2011, Adaptive Road Crack Detection System by Pavement Classification. Sensors, 11(10): 9628–9657.

[3] Kim B, Cho S, 2018, Automated Vision-Based Detection of Cracks on Concrete Surfaces Using a Deep Learning Technique. Sensors, 18(10): 3452.

[4] Huang S, Chen H, Yan L, et al., 2025, A Review of the Progress in Machine Vision-Based Crack Detection and Identification Technology for Asphalt Pavements. Digital Transportation and Safety, 4(1): 65–79.

[5] Zawad M, Zawad M, Rahman M, et al., 2021, A Comparative Review of Image Processing Based Crack Detection Techniques on Civil Engineering Structures. Journal of Soft Computing in Civil Engineering, 5(3): 58–74.

[6] Shi Y, Cui L, Qi Z, et al., 2016, Automatic Road Crack Detection Using Random Structured Forests. IEEE Transactions on Intelligent Transportation Systems, 17(12): 3434–3445.

[7] Koch C, Brilakis I, 2011, Pothole Detection in Asphalt Pavement Images. Advanced Engineering Informatics, 25(3): 507–515.

[8] Amhaz R, Chambon S, Idier J, et al., 2016, Automatic Crack Detection on Two-Dimensional Pavement Images: An Algorithm Based on Minimal Path Selection. IEEE Transactions on Intelligent Transportation Systems, 17(10): 2718–2729.

[9] Zhang L, Yang F, Zhang Y, et al., 2016, Road Crack Detection Using Deep Convolutional Neural Network. Proceedings of the IEEE International Conference on Image Processing (ICIP), 3708–3712.

[10] Ronneberger O, Fischer P, Brox T, 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 234–241.

[11] Lin T, Dollár P, Girshick R, et al., 2017, Feature Pyramid Networks for Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2117–2125.

[12] Hu J, Shen L, Sun G, 2018, Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7132–7141.

[13] Dai J, Qi H, Xiong Y, et al., 2017, Deformable Convolutional Networks. Proceedings of the IEEE International Conference on Computer Vision, 764–773.

[14] Zhu X, Hu H, Lin S, et al., 2019, Deformable ConvNets v2: More Deformable, Better Results. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9308–9316.

[15] Iandola F, Han S, Moskewicz M, et al., 2016, SqueezeNet: AlexNet-Level Accuracy with 50× Fewer Parameters and <0.5 MB Model Size. arXiv Preprint arXiv:1602.07360.

[16] Howard A, Zhu M, Chen B, et al., 2017, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv Preprint arXiv:1704.04861.

[17] Zhang X, Zhou X, Lin M, et al., 2018, ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6848–6856.

[18] Zawad M, Zawad M, Rahman M, et al., 2021, A Comparative Review of Image Processing Based Crack Detection Techniques on Civil Engineering Structures. Journal of Soft Computing in Civil Engineering, 5(3): 58–74.

[19] Yuan F, Lin Z, Tian Z, et al., 2025, Bio-Inspired Hybrid Path Planning for Efficient and Smooth Robotic Navigation: F. Yuan et al. International Journal of Intelligent Robotics and Applications, 2025: 1–31.

[20] Liang B, Yuan F, Deng J, et al., 2025, CS-PBFT: A Comprehensive Scoring-Based Practical Byzantine Fault Tolerance Consensus Algorithm. The Journal of Supercomputing, 81(7): 859.

[21] Zhang K, Yuan F, Jiang Y, et al., 2025, A Particle Swarm Optimization-Guided Ivy Algorithm for Global Optimization Problems. Biomimetics, 10(5): 342.

[22] Yuan F, Huang X, Jiang H, et al., 2025, An xLSTM–XGBoost Ensemble Model for Forecasting Non-Stationary and Highly Volatile Gasoline Price. Computers, 14(7): 256.

[23] Kulkarni S, Singh S, Balakrishnan D, et al., 2022, CrackSeg9k: A Collection and Benchmark for Crack Segmentation Datasets and Frameworks. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 179–195.

[24] Zhao H, Shi J, Qi X, et al., 2017, Pyramid Scene Parsing Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2881–2890.

[25] Yu C, Gao C, Wang J, et al., 2021, BiSeNet v2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. International Journal of Computer Vision, 129(11): 3051–3068.

[26] Hong Y, Pan H, Sun W, et al., 2021, Deep Dual-Resolution Networks for Real-Time and Accurate Semantic Segmentation of Road Scenes. arXiv Preprint arXiv:2101.06085.

[27] Li Y, Ma R, Liu H, et al., 2023, Real-Time High-Resolution Neural Network with Semantic Guidance for Crack Segmentation. Automation in Construction, 156: 105112.