Scientific and Social Research, 2025, Volume 7, Issue 9
https://ojs.bbwpublisher.com/index.php/SSR
Online ISSN: 2981-9946
Print ISSN: 2661-4332

CW-HRNet: Constrained Deformable Sampling and Wavelet-Guided Enhancement for Lightweight Crack Segmentation

Dewang Ma*

College of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China

**Author to whom correspondence should be addressed.*

Abstract: This paper presents CW-HRNet, a high-resolution, lightweight crack segmentation network designed to address challenges in complex scenes with slender, deformable, and blurred crack structures. The model incorporates two key modules: Constrained Deformable Convolution (CDC), which stabilizes geometric alignment by applying a tanh limiter and learnable scaling factor to the predicted offsets, and the Wavelet Frequency Enhancement Module (WFEM), which decomposes features using Haar wavelets to preserve low-frequency structures while enhancing high-frequency boundaries and textures. Evaluations on the CrackSeg9k benchmark demonstrate CW-HRNet's superior performance, achieving 82.39% mIoU with only 7.49M parameters and 10.34 GFLOPs, outperforming HrSegNet-B48 by 1.83% in segmentation accuracy with minimal complexity overhead. The model also shows strong cross-dataset generalization, achieving 60.01% mIoU and 66.22% F1 on Asphalt3k without fine-tuning. These results highlight CW-HRNet's favorable accuracy-efficiency trade-off for real-world crack segmentation tasks.

Keywords: Crack segmentation; Lightweight semantic segmentation; Deformable convolution; Wavelet transform; Road infrastructure

Online publication: September 9, 2025

1. Introduction

Crack detection is a critical task in road infrastructure maintenance, as road cracks are among the most common and hazardous pavement defects [1]. Accurate segmentation directly impacts the development of maintenance strategies and the long-term safety of transportation systems. Traditional manual inspection, while reliable to some extent, suffers from high labor costs, low efficiency, and strong subjectivity, making it increasingly unsuitable for modern, intelligent maintenance workflows [2].

Before the advent of deep learning, crack segmentation primarily relied on handcrafted features and classical machine learning methods. Typical approaches included edge operators, texture statistics, and descriptors such

as HOG or LBP, combined with classifiers like SVM, random forest, or K-means clustering [3]. While effective under low-noise and homogeneous conditions, these methods lacked robustness and generalizability in the face of complex materials, lighting conditions, and diverse crack shapes, due to their limited representational capacity.

With the rise of deep learning, convolutional neural networks have become the dominant paradigm for crack detection and segmentation. End-to-end pixel-wise training significantly improves robustness under complex imaging conditions [4]. Contextual modeling via feature pyramids and multi-scale fusion/attention strategies enhances the network's adaptability to cluttered backgrounds [5]. Architectures such as UNet, which utilize encoder-decoder frameworks with skip connections, effectively integrate fine-grained details and high-level semantics, thereby improving crack boundary sharpness and structural continuity [6]. Recent works further optimize the consistency and efficiency of feature learning by introducing dense feature aggregation and multi-level skip connections [7].

Despite these advances, two major challenges remain: First, in terms of geometric alignment, traditional convolutions with fixed sampling locations struggle to adapt to the irregular, slender, and bifurcated nature of cracks. While deformable convolutions introduce learnable offsets for enhanced shape modeling, unconstrained offsets may lead to instability, such as excessive deformation or irrelevant region sampling, resulting in distorted features and training difficulty. To address this, the study proposes Constrained Deformable Convolution (CDC), which introduces a tanh-based offset limiter and a learnable scaling factor to adaptively control sampling magnitudes, thereby achieving stable and precise alignment of complex crack structures [8]. Second, in the frequency domain, repeated downsampling in CNNs tends to erase high-frequency details, leading to blurred boundaries and poor texture representation. Purely spatial convolutions also struggle to jointly capture low-frequency global topology and high-frequency fine edges. To bridge this gap, we design the Wavelet Frequency Enhancement Module (WFEM), which decomposes feature maps into low-frequency (LL) and high-frequency (LH/HL/HH) subbands via Haar wavelets. Each subband undergoes lightweight convolutional projection and cross-subband residual modeling for information interaction, followed by an inverse wavelet transform to reconstruct the full-resolution feature map. This enables the network to preserve global topology while enhancing boundary and texture fidelity.

In summary, this paper addresses two critical limitations—unstable geometric alignment and loss of high-frequency details—through the integration of CDC and WFEM in a high-resolution, lightweight architecture, providing a principled foundation for the design and evaluation of the proposed network in subsequent sections.

2. Related work

Deformable convolution augments standard convolution by introducing learnable spatial offsets that relax the constraint of a fixed sampling grid, thereby increasing the network's capacity to model geometric variation. By predicting offsets with an auxiliary branch and applying bilinear interpolation at displaced locations, the effective receptive field becomes content-adaptive rather than purely grid-bound. This property has made deformable operators highly effective in dense prediction tasks—particularly object detection and semantic/instance segmentation—where target shapes may be elongated, multi-scale, or discontinuous. In essence, the kernel is no longer tied to a rigid lattice; it can bend toward informative structures and away from distractors, improving coverage of thin filaments, junctions, and tortuous boundaries.

In crack segmentation, this adaptivity is especially valuable: cracks are typically slender, irregular, and

branched, with low contrast and significant texture interference from surrounding materials. Vanilla deformable convolutions, however, come with a notable caveat. When the offset field is unconstrained, the model may push sampling points far beyond the neighborhood where the underlying features are reliable. Such excessive displacements or drifts into irrelevant regions lead to distorted local evidence, noisy gradients, and training instability. The risk is amplified along crack boundaries, where subtle misplacement can blur edges or fragment topology.

To counter these effects, we introduce Constrained Deformable Convolution (CDC), which explicitly regularizes the offset magnitude during generation. Concretely, CDC applies a tanh-based limiter followed by a learnable scaling factor s, mapping raw offsets into a bounded range that still permits meaningful deformation. The tanh operation suppresses extreme values symmetrically, while s adapts the allowable offset scale to local statistics and task difficulty. This adaptive upper bound curbs erratic sampling without reverting to a rigid grid, striking a practical balance between flexibility and stability.

The resulting offset field is smoother, better conditioned, and less prone to outliers, which in turn improves boundary alignment and feature fidelity near thin, branching structures. Empirically, CDC stabilizes optimization, reduces artifacts linked to offset explosion, and yields cleaner gradients for the backbone. In downstream decoding, features produced by CDC exhibit crisper edges and more coherent topology, enabling the overall network to maintain high-resolution detail while remaining robust to geometric variability and background clutter.

3. Wavelet transform and frequency-domain feature fusion

The wavelet transform offers a distinctive balance between spatial locality and frequency resolution, enabling simultaneous representation of structural context and detailed variations. Recently, wavelet-based approaches have gained attention in deep learning as both an alternative to conventional downsampling/upsampling layers and as a feature enhancement tool [9]. Compared to max-pooling or strided convolutions, which often discard fine details, wavelet decomposition retains richer structural cues by explicitly separating low-frequency and high-frequency components. This property is particularly beneficial for segmentation tasks, where subtle edge continuity and texture fidelity are critical for accurate predictions. By preserving edges and textures through multiple downsampling stages, wavelet-based modules allow networks to sustain fine-grained representation capacity that standard CNN architectures often fail to maintain [10]. Typical strategies in prior work have leveraged the discrete wavelet transform (DWT) to replace pooling operations, and its inverse counterpart (IDWT) to replace upsampling layers. Another common design decomposes feature maps into a low-frequency subband (LL) that captures overall structure and several high-frequency subbands (LH, HL, HH) that emphasize edges and fine patterns. These components are then recombined in various ways to enrich feature hierarchies and strengthen boundary localization. Despite their promise, most existing implementations suffer from two notable limitations. First, decomposition is often restricted to low-resolution stages, thereby neglecting early and mid-level features where much of the fine-grained information resides. Second, subbands are frequently handled in a simplistic manner, such as direct concatenation, without explicit modeling of inter-subband dependencies. As a result, these approaches may underutilize complementary relationships between frequency components, leading to boundary blurring, detail loss, and limited robustness.

To address these shortcomings, we propose the Wavelet Frequency Enhancement Module (WFEM). In WFEM, input features are decomposed using fixed Haar wavelets into one LL and three HF subbands. Each

subband is then projected through lightweight 1×1 convolutions with normalization and nonlinearity, ensuring compactness and recalibration of channel responses. The processed subbands are concatenated along the channel dimension, followed by cross-subband residual modeling, which explicitly enables information flow and interaction across LL and HF components. Finally, the refined features are reconstructed using IDWT to restore spatial resolution. This design simultaneously preserves low-frequency topological connectivity and enhances high-frequency boundary sharpness and texture richness, effectively overcoming the deficiencies of pure decomposition or naive concatenation strategies.

4. Methodology

4.1. Overall network architecture

CW-HRNet follows a dual-path encoder–multi-scale fusion–progressive decoder design to preserve high-resolution representations while effectively integrating global semantics with fine crack details, as illustrated in **Figure 1**. The input is first processed by shallow convolutions and normalization to obtain unified-scale features. In the dual-path encoder, the high-resolution branch stacks multiple CDC layers, which progressively apply constrained geometric adaptation to sampling positions. This stabilizes the alignment of elongated, branched, and tortuous crack boundaries while retaining shallow textures and fine-grained structures. In parallel, the low-resolution branch enlarges the receptive field to capture global context and complements the high-resolution pathway through cross-scale fusion and semantic interaction.

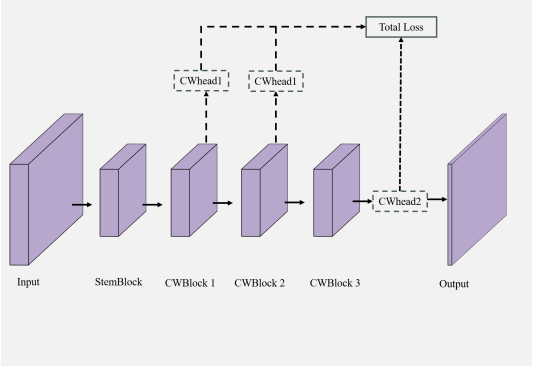

Figure 1. Overall architecture of CW-HRNet

To compensate for high-frequency loss caused by downsampling, the high-resolution branch incorporates the Wavelet Frequency Enhancement Module (WFEM) after CDC. Specifically, Haar wavelets are employed to decompose features into LL/LH/HL/HH subbands. Each subband undergoes lightweight projection and cross-subband residual modeling to enable effective information interaction, followed by inverse DWT reconstruction. This design achieves "topology preservation in low frequency and boundary strengthening in high frequency", thereby enhancing feature representation in the frequency domain.

During the decoding stage, cascaded convolutions and progressive upsampling gradually restore spatial resolution, while skip connections mitigate detail degradation. A final 1×1 convolution maps features into pixel-wise crack probability maps, optimized jointly with a standard binary classification loss.

In summary, CW-HRNet introduces targeted modifications to the classical high-resolution framework along two complementary dimensions: CDC constrains offset drift and improves boundary alignment stability, while

WFEM explicitly models cross-subband dependencies to recover high-frequency details. Their synergy enables the network to achieve a superior balance between accuracy and efficiency, while maintaining robustness in complex crack segmentation scenarios.

4.2. Deformable convolution enhancement module

In real-world scenarios, cracks often appear irregular, slender, and branched. Standard convolutions, constrained by fixed sampling locations, struggle to adapt to such complex geometries. Although the original deformable convolution introduces learnable offsets, the absence of proper constraints may lead to excessive deformations and boundary drift, causing training instability. To address this, the study proposes the Constrained Deformable Convolution (CDC) module. As illustrated in **Figure 2**, the offset generation stage replaces conventional convolutions with an OffsetConvBlock, which progressively extracts geometric cues and enhances the discriminability and stability of offset prediction. Furthermore, we design an Offset Regulation Module (ORM), which imposes adaptive constraints on the offset magnitude by combining a tanh-based limiter with a learnable scaling factor. This mechanism suppresses structural distortion at the source by preventing extreme sampling.

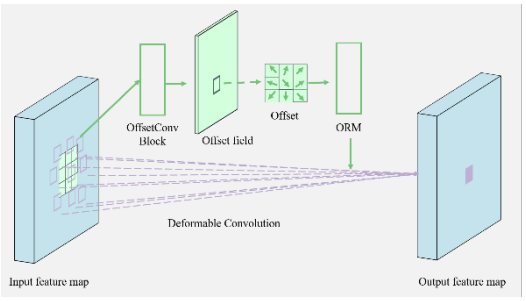

Figure 2. Structure of the constrained deformable convolution module

In the CDC, each convolutional sampling point k is associated with a learnable offset. The ORM constraint is formulated as:

$$\Delta p_k^{*}(x) = s \cdot \tanh(\Delta p_k(x)) \tag{1}$$

where tanh(·)compresses offsets into [−1,1] to suppress extreme values, and s is a learnable scaling factor that adaptively adjusts the offset magnitude. Based on the constrained offsets Δp∗(x), the deformable convolution can be expressed as:

$$y(p_0) = \sum_{k=1}^{K} w_k \cdot x\left(p_0 + p_k + \Delta p_k^{*}(x)\right) \tag{2}$$

where p0denotes the convolution kernel center, $p_k \in \mathbb{Z}^2$ represents the regular sampling grid, and x(·) indicates bilinear interpolation sampling [11].

Compared with the unconstrained DCN, the proposed CDC maintains geometric flexibility while significantly improving the stability of offset prediction and boundary alignment accuracy. When stacked in multiple layers within the high-resolution branch, CDC provides structurally coherent and edge-preserving geometric representations, thereby supplying the decoder with more reliable features and ultimately improving both segmentation accuracy and robustness [12].

4.3. High–low frequency feature decoupling and fusion module

Cracks often exhibit slender, tortuous, or even branched patterns, with a high degree of similarity to background textures. This leads to significant differences in the statistical distribution of high-frequency details and low-frequency structures. Conventional convolutional networks tend to lose high-frequency information after multiple downsampling operations, while low-frequency global semantics are insufficiently captured due to limited receptive fields [13]. As a result, boundaries become blurred, fine-grained textures are lost, and local topology is often disrupted. To overcome these limitations, we propose the Wavelet Frequency Enhancement Module (WFEM), which achieves joint enhancement of global and local features through frequency-domain decomposition and cross-subband modeling, as illustrated in **Figure 3**.

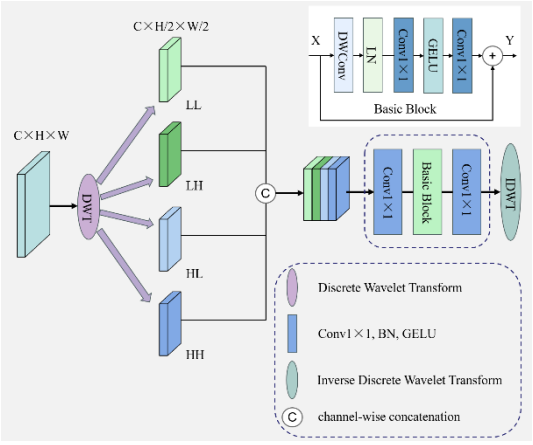

Figure 3. Structure of the wavelet frequency enhancement module

Specifically, WFEM employs fixed Haar wavelets to decompose input features via discrete wavelet transform (DWT), yielding a low-frequency subband (LL) and three high-frequency subbands (LH/HL/HH). The LL component encodes overall shape and connectivity, while the high-frequency subbands capture crack boundaries and texture details [14]. Each subband is then passed through a lightweight projection composed of 1×1 convolution + BatchNorm + GELU, which recalibrates channel responses and aligns bandwidth. The four subbands are concatenated along the channel dimension, followed by a Residual Cross-Subband Block and a subsequent 1×1 compression layer. This explicitly models dependencies and complementarity between LL and high-frequency components, mitigating bias caused by the dominance of a single subband, thereby improving boundary localization and topological consistency. Finally, the processed features are re-split into four subbands and reconstructed to the original resolution using inverse wavelet transform (IDWT), producing a unified representation that preserves global topology in low frequencies while strengthening boundary details in high frequencies.

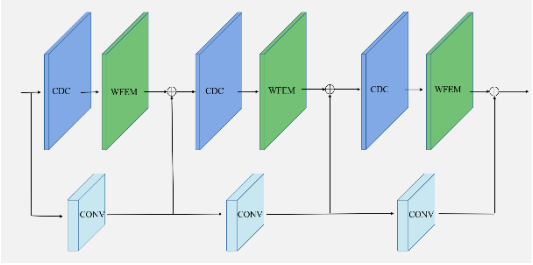

Figure 4. Structure of the CWBlock: Integration of CDC and WFEM modules

Within the overall network, WFEM is deployed in the high-resolution branch immediately after the CDC module: CDC first performs geometric alignment and stabilizes elongated or branched boundaries, after which WFEM restores and amplifies high-frequency details in the frequency domain while leveraging LL to enforce global connectivity. The synergy of these two modules enables CW-HRNet to maintain high-resolution representations while simultaneously achieving global topological modeling and fine-grained crack characterization, ultimately improving segmentation accuracy and robustness, as shown in **Figure 4**.

5. Experiments

5.1. Datasets

To evaluate the effectiveness of the proposed model, experiments are conducted on the CrackSeg9k dataset, representing mixed-scene cracks, and the Asphalt3k dataset, representing asphalt-specific cracks [15–16].

CrackSeg9k is a medium-scale semantic segmentation dataset designed for crack detection and segmentation tasks. It contains approximately 8,751 high-quality images with cracks, covering diverse materials such as concrete, ceramics, and bricks. Each image has a resolution of 400 × 400 pixels, with two defined classes: crack and background. Following a fixed random seed, the dataset is split into 70% training, 10% validation, and 20% testing subsets.

Asphalt3k is a domain-specific dataset focusing on asphalt pavement cracks, derived from the public dataset originally released by Yang. The study preprocesses the raw samples by cropping and organizing them into 3,000 image–annotation pairs, which are randomly divided into training/validation/testing sets at a ratio of 6:1:3. Unless otherwise specified, both training and evaluation are performed under a single-scale setting, where images are centrally cropped or padded to 400×400 pixels. The class definitions and annotation protocols remain consistent with the original datasets.

5.2. Evaluation metrics

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

$$Pr = \frac{TP}{TP + FP} \tag{3}$$

$$Re = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \cdot \frac{Pr \cdot Re}{Pr + Re} \tag{5}$$

$$mIoU = mean\left(\frac{TP}{TP + FP + FN}\right) \tag{6}$$

Where true positives (TP) denote correctly classified crack pixels, false positives (FP) represent background pixels incorrectly classified as cracks, and false negatives (FN) correspond to crack pixels misclassified as background.

In addition to segmentation accuracy, we also report GFLOPs (Giga Floating Point Operations) and Params (number of parameters) as measures of the model's computational complexity and size, respectively.

5.3. Comparison with state-of-the-art models

This study focuses on the design of lightweight crack segmentation models. The study compares the proposed CW-HRNet against a variety of representative approaches, including classic high-accuracy models (UNet, PSPNet, OCRNet, DeepLabV3+[19]), mainstream lightweight architectures (BiSeNetv2, STDCSet, DDRNet), as well as crack-oriented models (UNet with Focal Loss, U2CrackNet, RUCNet) [6, 17–25]. All models are trained from scratch under identical conditions to ensure fairness.

Table 1 reports the comparative results. The UNet family achieves relatively high accuracy but suffers from large parameter sizes and heavy computational cost. RUCNet attains 80.47% mIoU, yet requires 115.49 GFLOPs, making deployment in resource-constrained environments impractical. The HrSegNet series demonstrates superior efficiency; for example, the B48 variant achieves 80.56% mIoU with only 5.43M parameters and 5.60 GFLOPs, highlighting strong scalability [26].

By contrast, CW-HRNet strikes a better balance between accuracy and complexity. With merely 7.49M parameters and 10.34 GFLOPs, it achieves 82.39% mIoU, 89.46% F1-score, 90.59% Precision, and 88.39% Recall, outperforming all competing methods. Compared with OCRNet, CW-HRNet improves mIoU by 1.49 percentage points while reducing parameters and computational cost by approximately 38% and 68%, respectively. Relative to HrSegNet-B48, CW-HRNet modestly increases complexity yet raises mIoU to 82.39%, significantly enhancing boundary delineation and fine crack representation.

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

Model	mIoU (%)	Pr (%)	Re (%)	F1 (%)	Params(M)	GFLOPs
UNet	79.15	89.82	84.75	87.21	13.40	75.87
PSPNet	76.78	87.57	83.33	85.39	21.07	54.20
BiSeNetv2	75.09	87.07	81.17	83.71	2.33	4.93
STDCSeg	78.48	88.24	84.60	86.65	8.28	5.22
DDRNet	76.77	89.10	82.10	85.45	20.18	11.11
OCRNet	80.90	88.26	88.58	88.41	12.12	32.40
DeeplabV3+	78.29	87.33	83.76	85.50	12.20	33.96
UNet(Focal Loss)	80.27	89.05	84.75	86.85	13.40	75.87
U2CrackNet	79.79	89.05	86.26	87.62	1.20	31.21
RUCNet	80.47	88.91	87.32	88.11	25.47	115.49
HrSegNet-B16	79.84	88.79	86.54	87.65	0.61	0.66
HrSegNet-B32	80.21	90.12	85.93	87.97	2.49	2.50
HrSegNet-B48	80.56	90.07	86.44	88.21	5.43	5.60
CW-HRNet	82.39	90.59	88.39	89.46	7.49	10.34

5.4. Ablation study

To investigate the contribution of each module to overall performance, we conduct ablation experiments on the CrackSeg9k dataset using HrSegNet-B48 as the baseline, and progressively introduce WFEM and CDC. The results are presented in **Table 2**.

Table 2. Ablation study

Method	mIoU (%)	Params (M)	GFLOPs
HrSegNet-B48	80.56	5.43	5.60
+ WFEM	81.64	5.53	8.72
+ CDC	81.71	6.11	7.21
CW-HRNet	82.39	7.49	10.34

As shown, incorporating WFEM into the baseline increases mIoU from 80.56% to 81.64%, a relative gain of 1.08 percentage points. The parameter count increases by only 0.10M, which is negligible; however, computational complexity rises considerably. This is mainly because WFEM introduces multi-subband parallel convolutions and IDWT reconstruction in the high-resolution branch, significantly expanding feature bandwidth and operator count, thereby incurring higher computational cost.

Further adding CDC raises mIoU to 81.71%, improving by 1.15 percentage points over the baseline, with an additional 0.68M parameters and 1.61 GFLOPs. The major overhead originates from the offset branch's convolutional prediction and bilinear interpolation sampling.

When WFEM and CDC are combined, performance reaches the best outcome, with mIoU improved to 82.39%, representing a 1.83 percentage point gain over the baseline. This is achieved at a complexity of 7.49M parameters and 10.34 GFLOPs, demonstrating a favorable balance between accuracy and efficiency.

In summary, the two modules exhibit complementary roles: WFEM focuses on modeling high-frequency boundaries and fine-grained textures, while CDC enhances geometric alignment and structural robustness. Their synergy significantly boosts both segmentation accuracy and stability of the proposed model.

5.5. Generalization ability evaluation

To assess the cross-dataset generalization capability of the proposed model, we train CW-HRNet on CrackSeg9k and directly transfer it to the Asphalt3k dataset for testing without any fine-tuning. The comparative results on Asphalt3k are reported in **Table 3**.

It can be observed that CW-HRNet achieves the best performance in both mIoU and F1 metrics, reaching 60.01% mIoU and 66.22% F1, outperforming all other methods. Considering the severe class imbalance inherent in road crack segmentation, overall accuracy tends to be overestimated and shows limited differences across models. Thus, mIoU and F1 are more reliable indicators of practical detection performance.

These results demonstrate that CW-HRNet maintains stable structural recognition under challenging conditions such as complex textures and low-contrast backgrounds, highlighting its stronger cross-dataset generalization and robustness compared to competing approaches.

Table 3. Transfer to Asphalt3k

Model	BiSeNet	PSPNet	STDCSeg	U2CrackNet	HrSegNet	CW-HRNet
mIoU	55.10	54.24	55.53	54.80	58.27	60.01
F1	60.54	59.16	61.20	60.04	65.19	66.22
Params	2.33	21.07	8.28	1.20	5.43	7.49

6. Conclusion

This paper presents CW-HRNet, a lightweight crack segmentation network that integrates geometric adaptability with frequency-domain enhancement. In terms of methodological design, we introduce the Constrained Deformable Convolution (CDC), which employs a tanh-based limiter and a learnable scaling factor to effectively suppress offset drift, enabling stable alignment of slender and branched crack geometries. In parallel, we propose the Wavelet Frequency Enhancement Module (WFEM), which leverages Haar wavelet decomposition and cross-subband residual modeling to mitigate the high-frequency detail loss caused by convolutional downsampling. This design preserves low-frequency topological integrity while significantly strengthening boundary and texture representation.

Overall, the synergy between CDC and WFEM balances geometric modeling and frequency-domain enhancement, providing a new perspective for lightweight crack segmentation. In future work, the study plans to explore the integration of learnable wavelet bases with Transformer modules to further improve cross-scene adaptability. Moreover, the study aims to extend the model to broader infrastructure inspection tasks, such as bridges, tunnels, and airport runways, thereby advancing the intelligent maintenance of road and transportation engineering.

Disclosure statement

The author declares 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, 2024, 16(16): 2910.

[2] 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.

[3] Zawad MRS, Zawad MFS, Rahman MA, 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.

[4] Long J, Shelhamer E, Darrell T, 2015, Fully Convolutional Networks for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440.

[5] Lin TY, Dollar 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.

[6] 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. Springer International Publishing, Cham, 234–241.

[7] Huang G, Liu Z, Van Der Maaten L, et al., 2017, Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700–4708.

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

[9] Li Q, Shen L, 2022, Wavesnet: Wavelet Integrated Deep Networks for Image Segmentation. Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Springer Nature Switzerland, Cham, 325–337.

[10] Li Q, Shen L, 2022, Neuron Segmentation using 3D Wavelet Integrated Encoder–Decoder Network. Bioinformatics,

38(3): 809–817.

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

[12] 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.

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

[14] 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.

[15] 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. Springer Nature Switzerland, Cham, 179–195.

[16] Yang N, Li Y, Ma R, 2022, An Efficient Method for Detecting Asphalt Pavement Cracks and Sealed Cracks Based on a Deep Data-driven Model. Applied Sciences, 12(19): 10089.

[17] 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.

[18] Yuan Y, Chen X, Wang J, 2020, Object-contextual Representations for Semantic Segmentation. European Conference on Computer Vision. Springer International Publishing, Cham, 173–190.

[19] Chen LC, Papandreou G, Schroff F, et al., 2017, Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint, arXiv:1706.05587.

[20] 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.

[21] Fan M, Lai S, Huang J, et al., 2021, Rethinking Bisenet for Real-time Semantic Segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9716–9725.

[22] 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.

[23] Liu Z, Cao Y, Wang Y, et al., 2019, Computer Vision-based Concrete Crack Detection using U-net Fully Convolutional Networks. Automation in Construction, 2019(104): 129–139.

[24] Shi P, Zhu F, Xin Y, et al., 2023, U2CrackNet: A Deeper Architecture with Two-level Nested U-structure for Pavement Crack Detection. Structural Health Monitoring, 22(4): 2910–2921.

[25] Yu G, Dong J, Wang Y, et al., 2022, RUC-Net: A Residual-Unet-based Convolutional Neural Network for Pixel-level Pavement Crack Segmentation. Sensors, 23(1): 53.

[26] Li Y, Ma R, Liu H, et al., 2023, Real-time High-resolution Neural Network with Semantic Guidance for Crack Segmentation. Automation in Construction, 2023(156): 105112.