

Master's Thesis Presentation for the 146th Dissertation Evaluation

A Study on Efficient Crack Segmentation Using Evidence-Preserving Receptive-Field Expansion

증거 보존형 수용장 확장을 활용한
효율적 균열 분할 연구

Daehyeon Kim

rlaeogus031128@gmail.com

5th Sep, 2025



중앙대학교
CHUNG-ANG UNIVERSITY



Contents

- 1. Introduction**
- 2. Related Work**
- 3. Proposed Method**
- 4. Experimental Settings**
- 5. Overall Experimental Results**
- 6. In-Depth Analysis**
- 7. Conclusion and Contribution**
- 8. Reference**

Introduction

Crack Segmentation

- Crack segmentation is the task of isolating only crack regions from given images of infrastructure surfaces, enabling remote inspection that improves safety while standardizing condition assessment to reduce human subjectivity in maintenance [1, 2].
- After fully convolutional networks (FCNs) dominated crack segmentation by chiefly modeling local cues, recent methods prioritize architectures that capture long-range dependencies with local detail, often via residual blocks or attention mechanisms [2, 3].

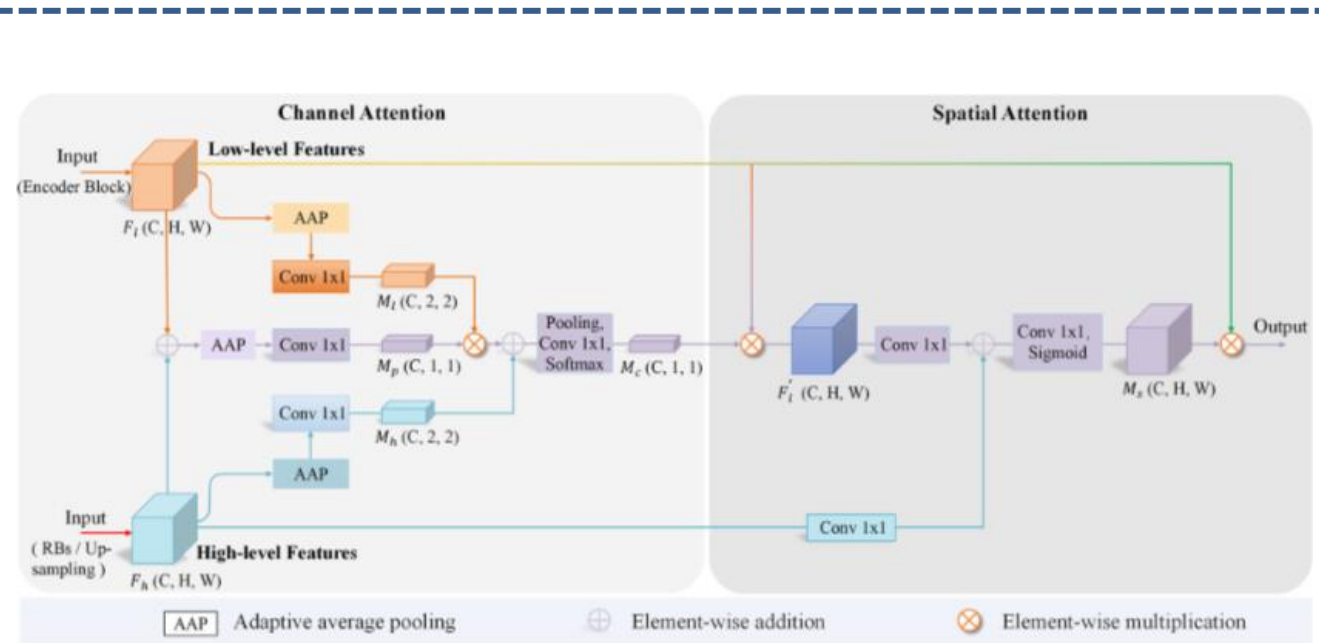


Figure 1. Attention based residual-connection block

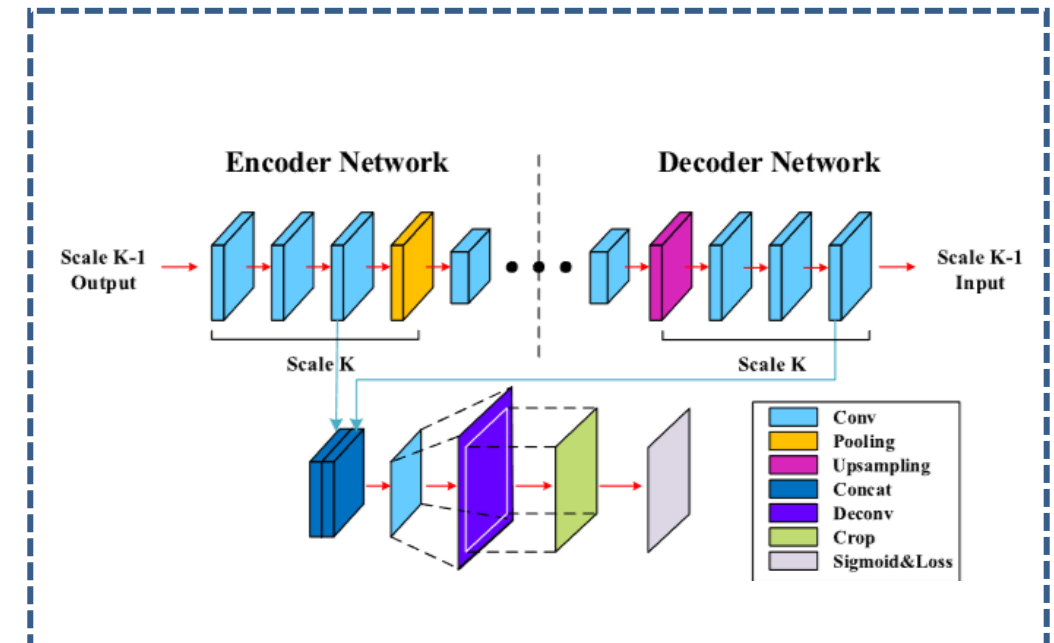


Figure 2. FCN based Network

Introduction

Importance of Efficiency on Crack Segmentation

- For inspections in harsh, hard-to-access settings that rely on drones or on-device systems with tight resources, it is crucial that crack segmentation models keep parameter counts small and maintain strong computational efficiency to remain deployable [5].
- In safety-critical inspections, computational efficiency is only meaningful if high predictive performance is preserved, which requires strictly suppressing false cases while consistently maintaining thin-crack continuity for stable and reliable outputs.

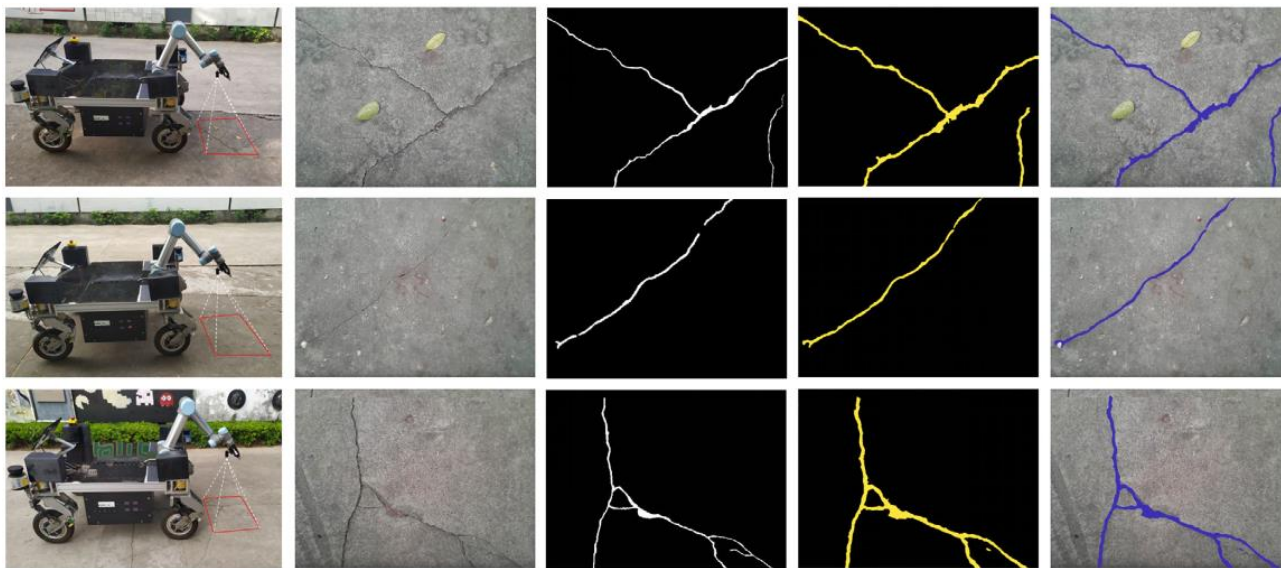


Figure 3. Crack segmentation with on-device system

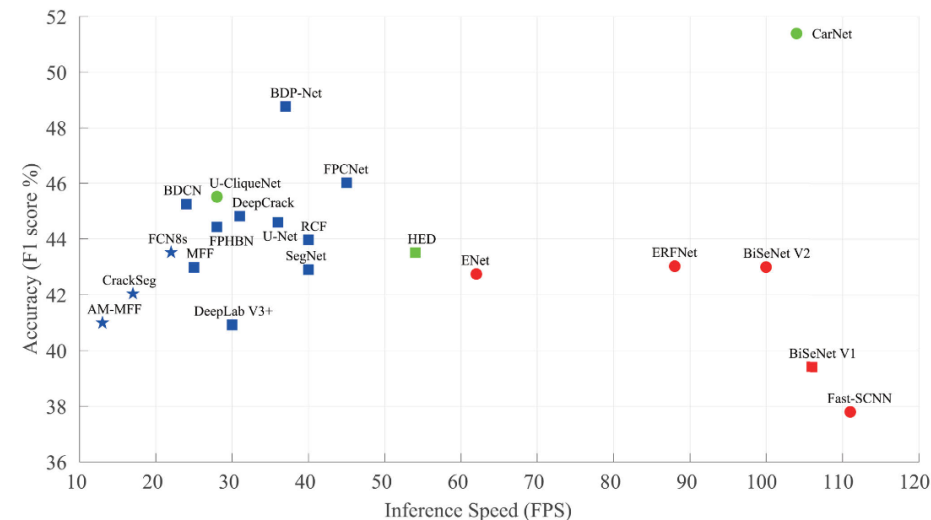


Figure 4. Model accuracy v.s. inference efficiency

Introduction

Objective and Standard Procedure

- Efficient crack segmentation takes an image containing cracks as input and produces a pixel-level segmentation mask that reliably and consistently isolates only the crack regions as output while meeting tight compute and memory budgets for edge deployment.
- The network learns crack-specific features from labeled masks by minimizing a segmentation loss during supervised training and, at inference, efficiently and reliably encodes rich multi-scale context from surface images to decode a pixel-wise binary crack mask.

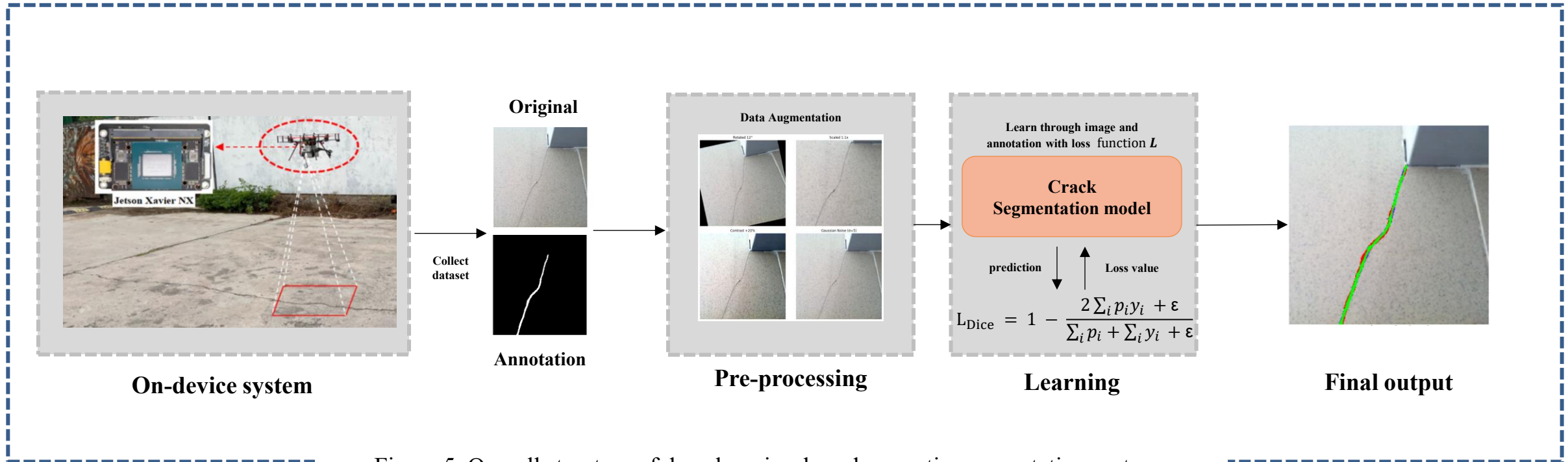


Figure 5. Overall structure of deep learning-based semantic segmentation systems

Introduction

Ideal Goal and Challenge

- Ideal crack segmentation produces topology-faithful masks that preserve pixel evidence, keeping thin cracks continuous with accurate widths and endpoints so hairline fractures remain connected and measurable for reliable quantification and assessment.
- An efficient long-range modeling design is crucial for crack segmentation because learning the continuity of elongated cracks requires aggregating information over a wide receptive field, a process that is conventionally computationally expensive.

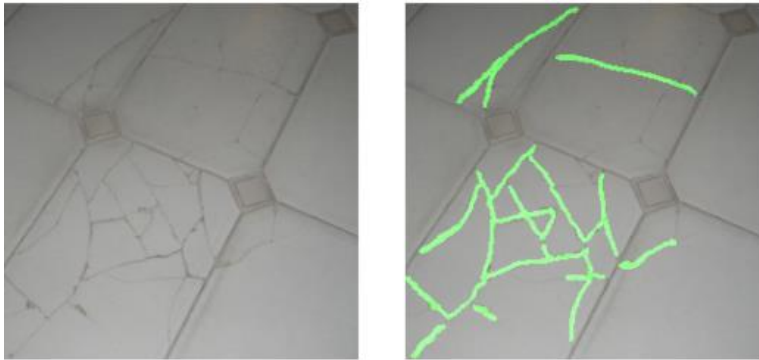


Figure 6. Ideal segmentation map

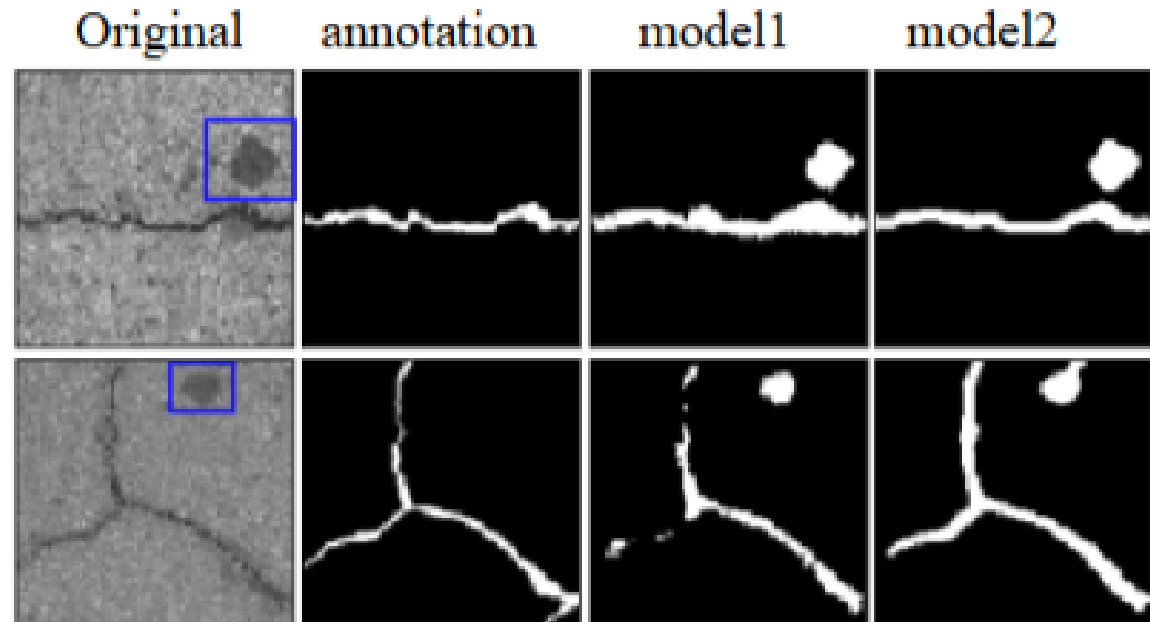


Figure 7. Cases where crack cues are overwhelmed by background texture

Related Work

RHACrackNet

- RHACrackNet utilizes depthwise separable convolutions and a residual-connection module called the Hybrid Attention Block (HAB) for efficiency, while incorporating spatial and channel attention to preserve fine grained crack continuity and width [3].
- HAB merges low/high features with channel–spatial attention; computational complexity drops substantially via depthwise operations, but undersized early receptive fields starve fusion of global cues, causing a disproportionate overall accuracy loss [3].

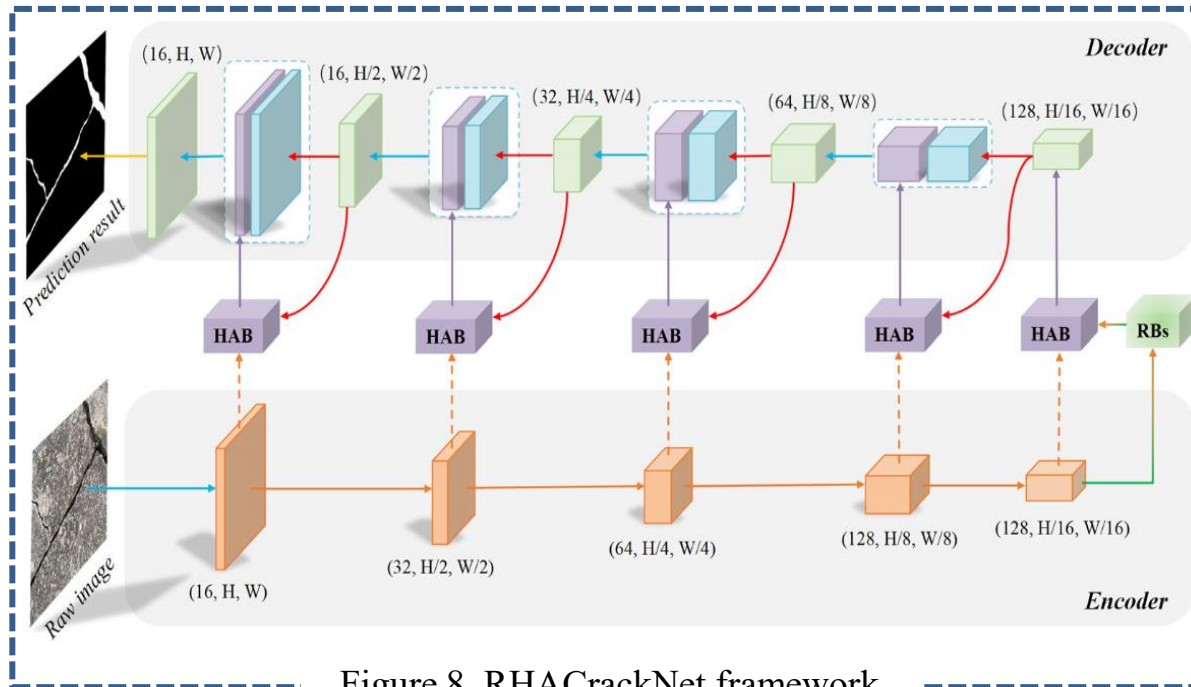


Figure 8. RHACrackNet framework

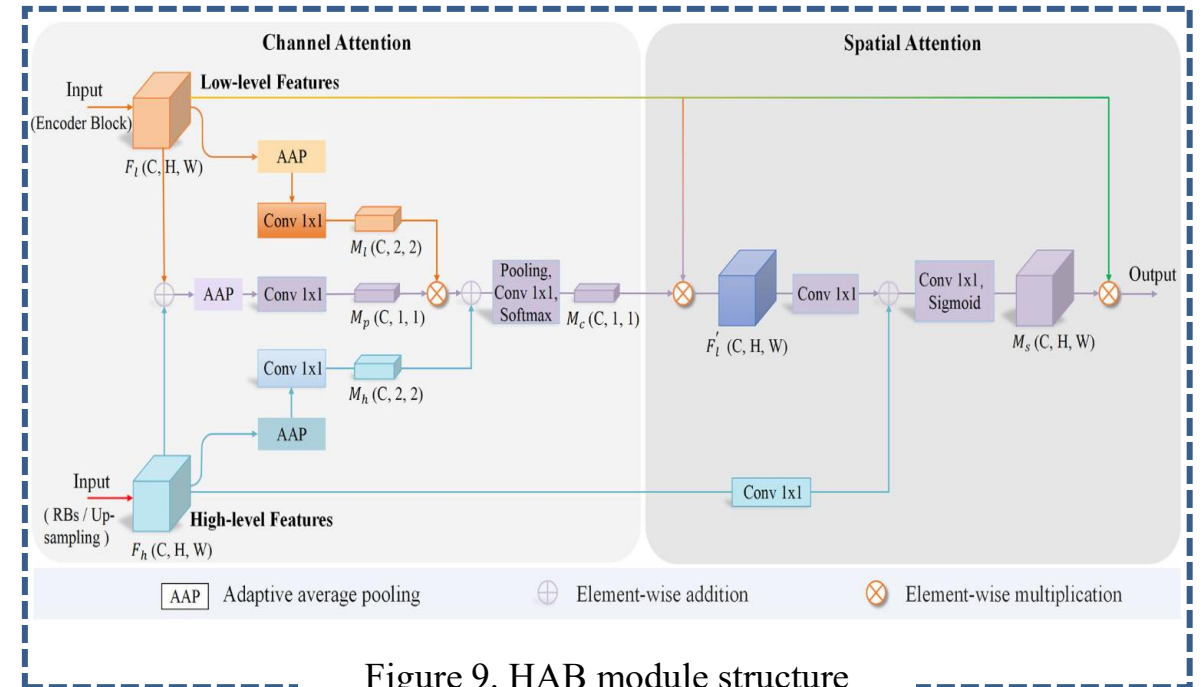


Figure 9. HAB module structure

Related Work

CarNet

- CarNet employs an olive-shaped encoder that combines residual blocks (RB) and downsampling blocks (DB) for efficient feature extraction, and a lightweight decoder that fuses multi-scale features via an up-sampling feature-pyramid block [5].
- Two consecutive stride-2 DBs cut cost by rapidly shrinking input image resolution, but the aggressive reduction in early stages yields an outsized thin-crack accuracy drop and fragmentation due to absence of global context at the early encoders [5].

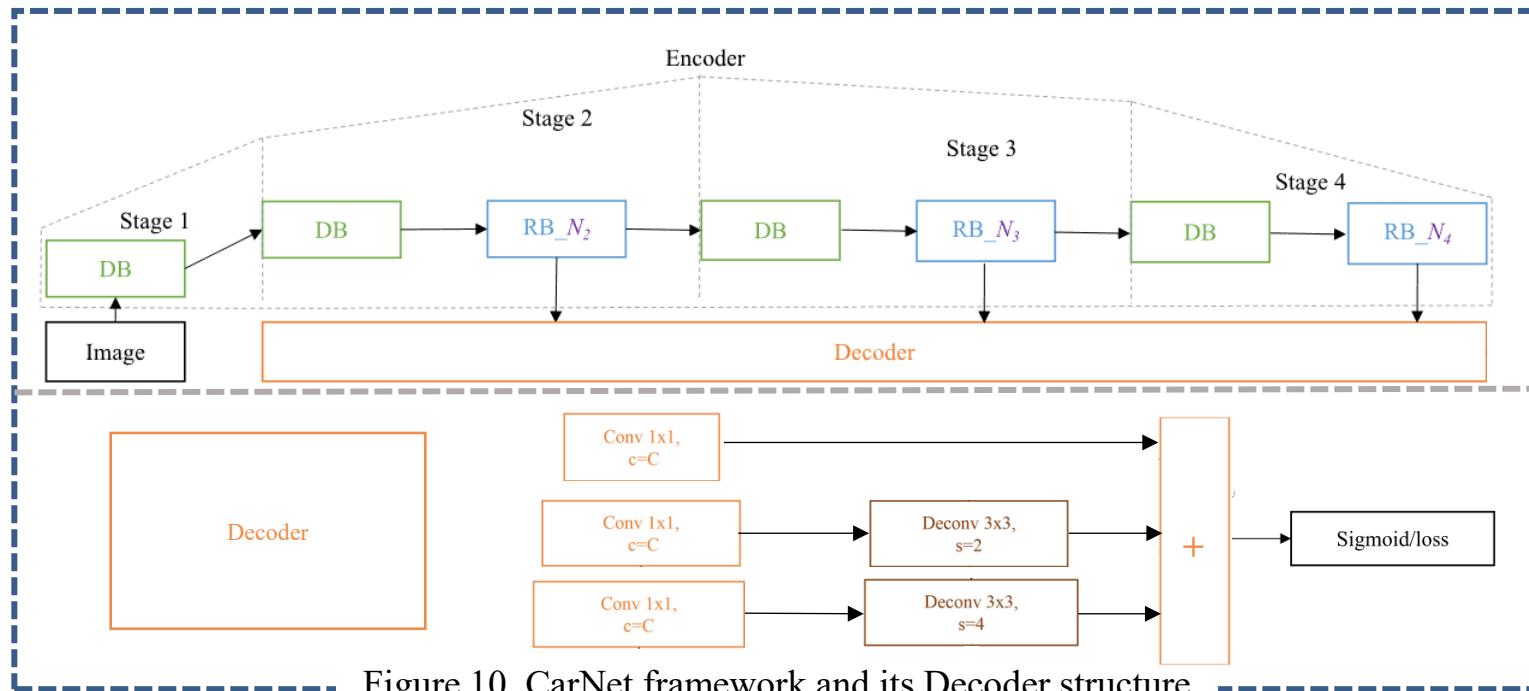


Figure 10. CarNet framework and its Decoder structure

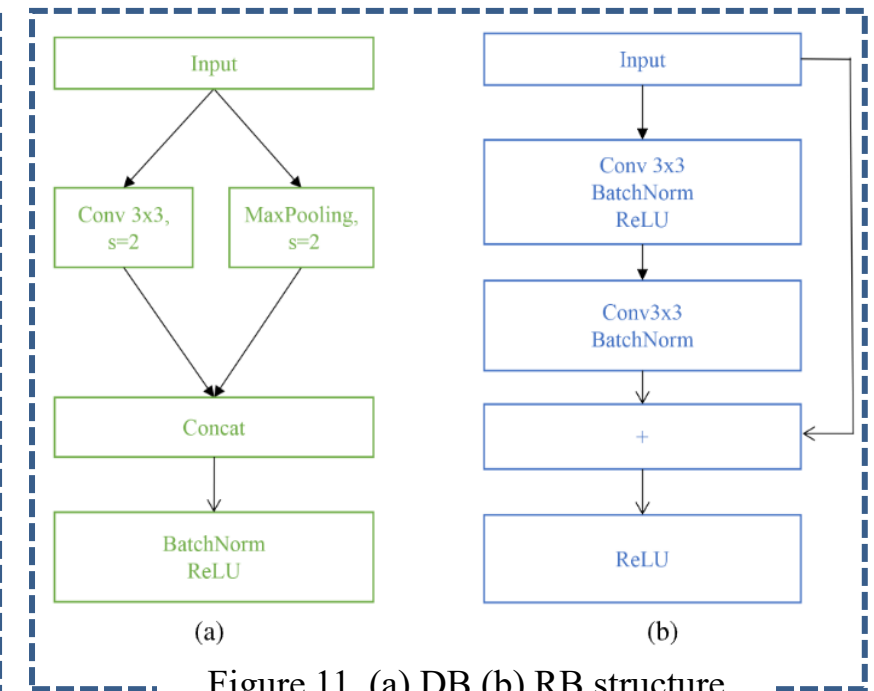


Figure 11. (a) DB (b) RB structure

Related Work

CrackFormer II

- CrackFormer II embedded self-attention modules into a convolutional framework and evolved into powerful dual-encoder networks that combine the local feature extraction strengths of convolutional networks with the global context modeling of Transformers [6].
- Max-pooling at every stage especially at early encoder with later scaling-attention or side fusion keeps early cost low yet leaves faint cues vulnerable on noisy backgrounds and spends compute for little return, degrading accuracy beyond the budget saved [6].

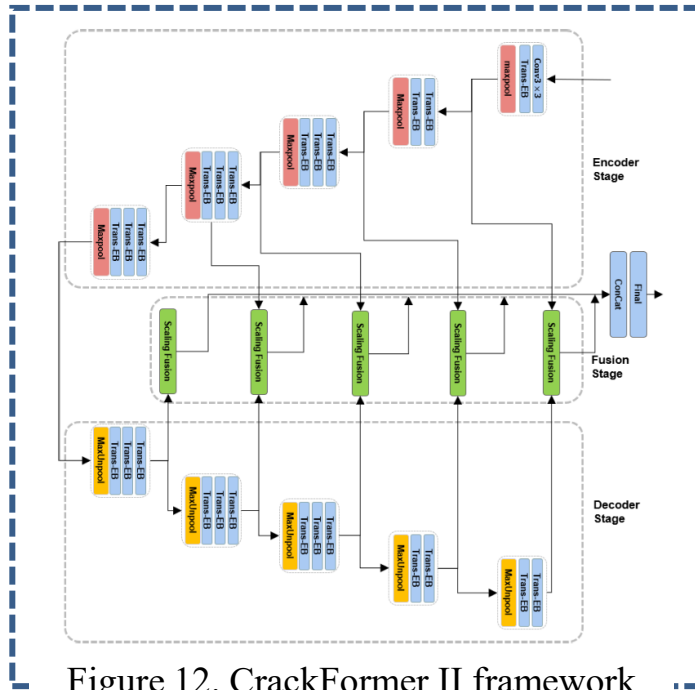


Figure 12. CrackFormer II framework

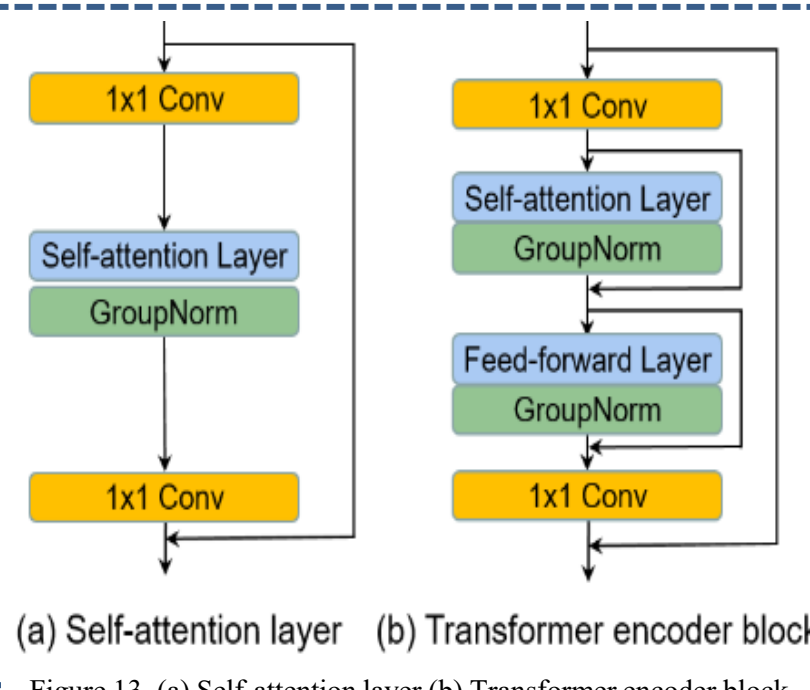


Figure 13. (a) Self-attention layer (b) Transformer encoder block

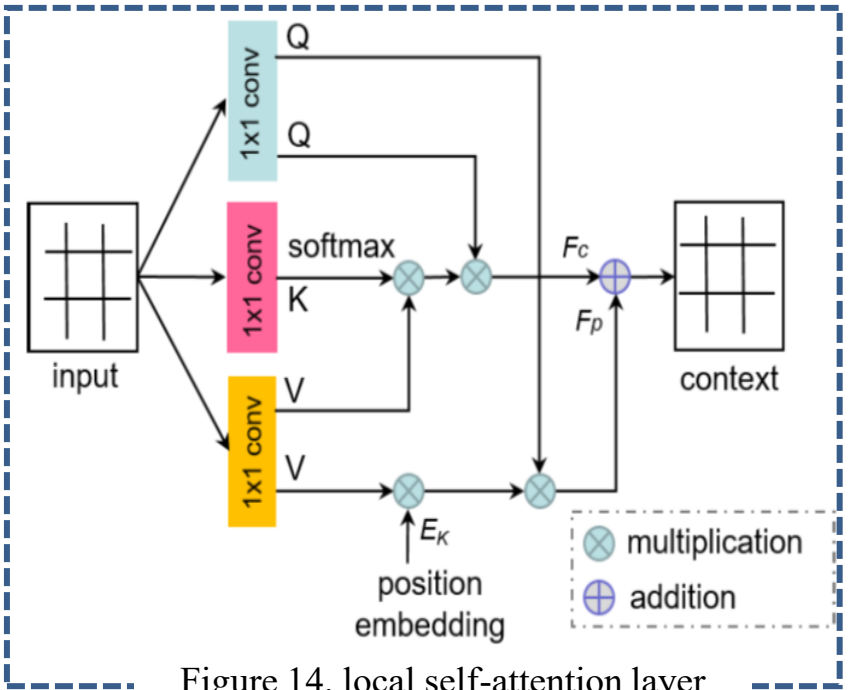


Figure 14. local self-attention layer

Related Work

CSNet

- CSNet employs group convolutions with channel shuffling and attention mechanisms, adopts DeepLabv3+ as its backbone, and also uses attention mechanisms which rely on DenseASPP for multi-scale processing under tight budgets and latency constraints [7].
- DenseASPP with post-downsampling attention strengthens boundaries late; large dilation rates induce gridding and miss hairline cracks, so the accuracy penalty often outweighs efficiency gains, especially for thin cracks in noisy, textured images [7].

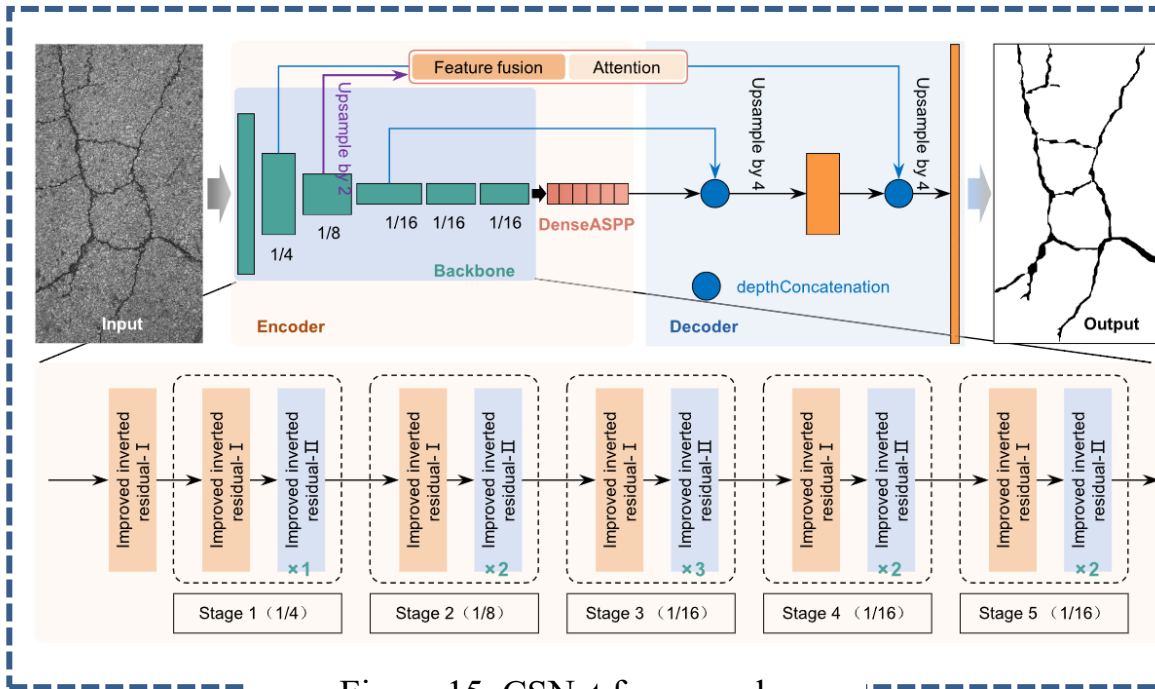


Figure 15. CSNet framework

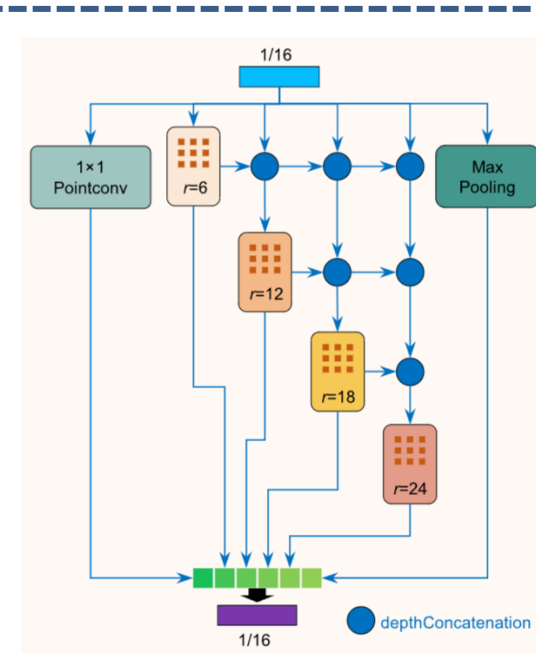


Figure 16. DenseASPP module

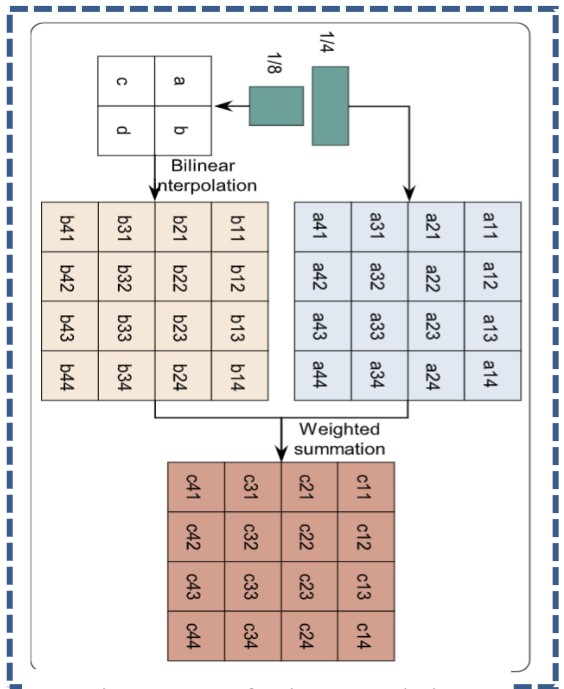


Figure 17. fusion module

Related Work

LMNet

- A four-stage lightweight convolutional neural network with three max-pooling steps, wide early Parallel Feature Modules, pixel-wise dilation-and-attention (PDAM), and a Feature-Reduction–Concatenation module (FRCM) before bilinear up-sampling [8].
- Multiple downsamples followed by PDAM and FRCM keep parameters small, but long, thin trajectories stay quite inconsistent, especially under noisy, textured scenes—an accuracy hit typically disproportionate to the reduced cost in practice [8].

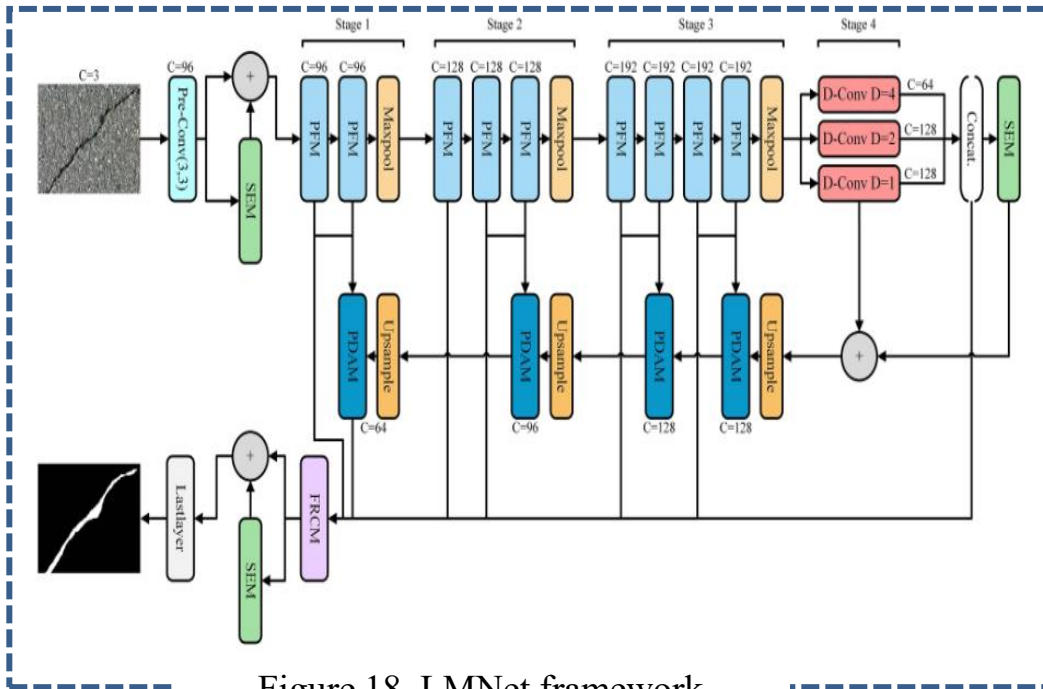


Figure 18. LMNet framework

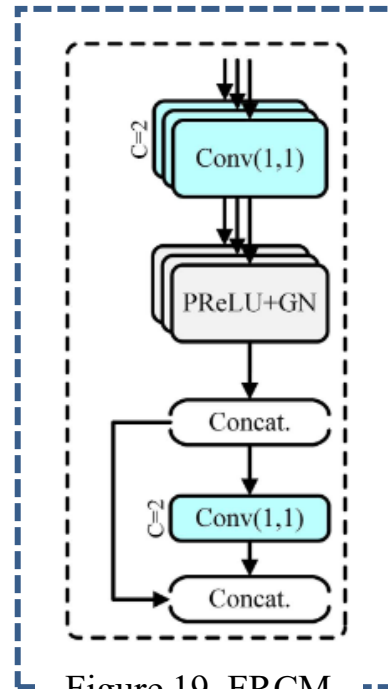


Figure 19. FRCM

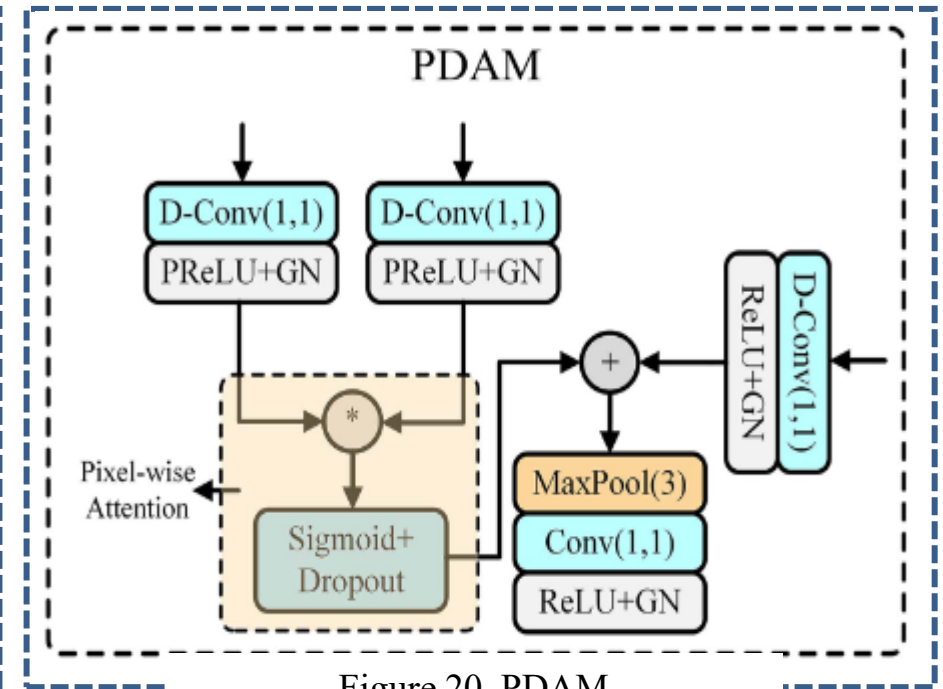


Figure 20. PDAM

Common Drawback of Conventional Methods

- Existing models downsample early to shrink resolution, then compensate for information loss with add-on modules. While this pipeline is efficient, it has limited receptive fields and misses long-range context in early stages where evidence is intact.
- As a result, accuracy degrades because early evidence is prematurely discarded before global context can be integrated, leading to missed faint segments and endpoints, broken long-range connectivity, and frequent false detections on crack-like textures.

Table 1. Summary and Drawback

Method	Venue	Summary	Common drawback
RHACrackNet	Computer-Aided Civil and Infrastructure Engineering	Leverages depthwise-separable convolutions and a residual Hybrid Attention Block with spatial/channel attention to maintain crack continuity.	Existing encoders downsample early and confine large-rate dilations to deep and low-resolution stages. This can trim compute and reduce gridding, but it weakens early receptive fields, degrades long-range context in the early stage where evidence is most intact, and ultimately hurts accuracy.
CarNet	IEEE Transactions on Intelligent Transportation Systems	Combines an olive-shaped RB/DB encoder with a lightweight decoder and an up-sampling feature-pyramid block for efficient multi-scale fusion.	
CrackFormer II	IEEE Transactions on Intelligent Transportation Systems	Integrates self-attention into a dual-encoder CNN–Transformer framework to couple local feature extraction with global context modeling.	
CSNet	Automation in Construction	Builds on DeepLabv3+ using grouped-shuffled convolutions, attention modules, and DenseASPP to integrate multi-scale contextual information.	
LMNet	IEEE Transactions on Intelligent Vehicles	Stacks a four-stage lightweight CNN with early Parallel Feature Modules, PDAM, and FRCM, followed by bilinear up-sampling.	

Proposed Method

Notation and Definition

- The crack dataset is formally defined as $\mathcal{D} = \{(I_n, M_n)\}_{n=1}^N$, where $I_n \in \mathbb{R}^{H \times W \times 3}$ is an RGB image and $M_n \in \{0,1\}^{H \times W}$ is a binary mask (1 = crack, 0 = background). Here N denotes dataset size and, i indexes samples (1,, N) to pair each image with its mask.
- Crack Segmentation aims to learn $\mathcal{F} : \mathbb{R}^{H \times W \times 3} \rightarrow [0, 1]^{H \times W}$, parameterized by θ , that outputs a per-pixel crack score map $P = \mathcal{F}(I; \theta)$, with θ^* obtained by minimizing a chosen loss L over $\mathcal{D}$ under given evaluation measures protocols during training.

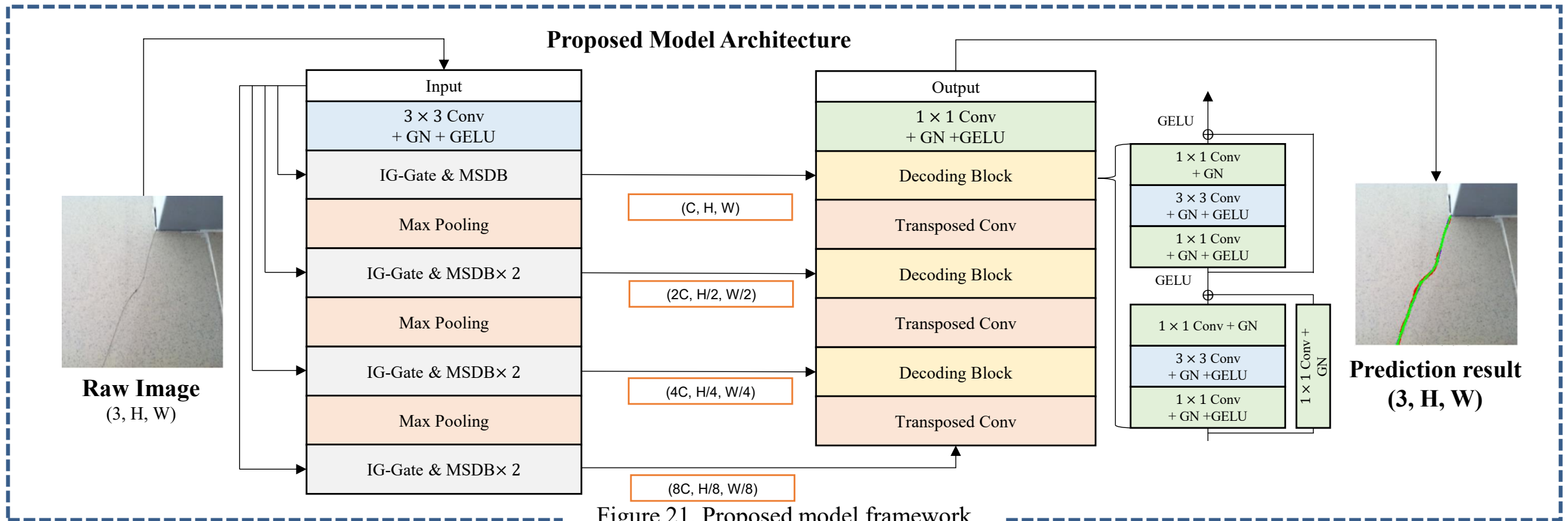
Table 2. Notations and definitions

Symbol	Name	Description
$\mathcal{D}$	Dataset	Overall Dataset
I_n	Crack Image	n-th Crack Input Image
M_n	Binary mask	n-th binary ground-truth mask (1=crack, 0=background)
N	Dataset Size	Number of samples in the Dataset
$\mathcal{F}$	Segmentation Function	Mapping $\mathcal{F} : \mathbb{R}^{H \times W \times 3} \rightarrow [0, 1]^{H \times W}$ producing per-pixel crack score map P.
L	Loss function	One minus the Dice overlap between prediction and ground truth.
TP	True positives	Number of pixels where both the ground truth and the prediction are crack
FP	False positives	Number of pixels where the ground truth is background but the prediction is crack
FN	False negatives	Number of pixels where the ground truth is crack but the prediction is background

Proposed Method

Graphical Abstract

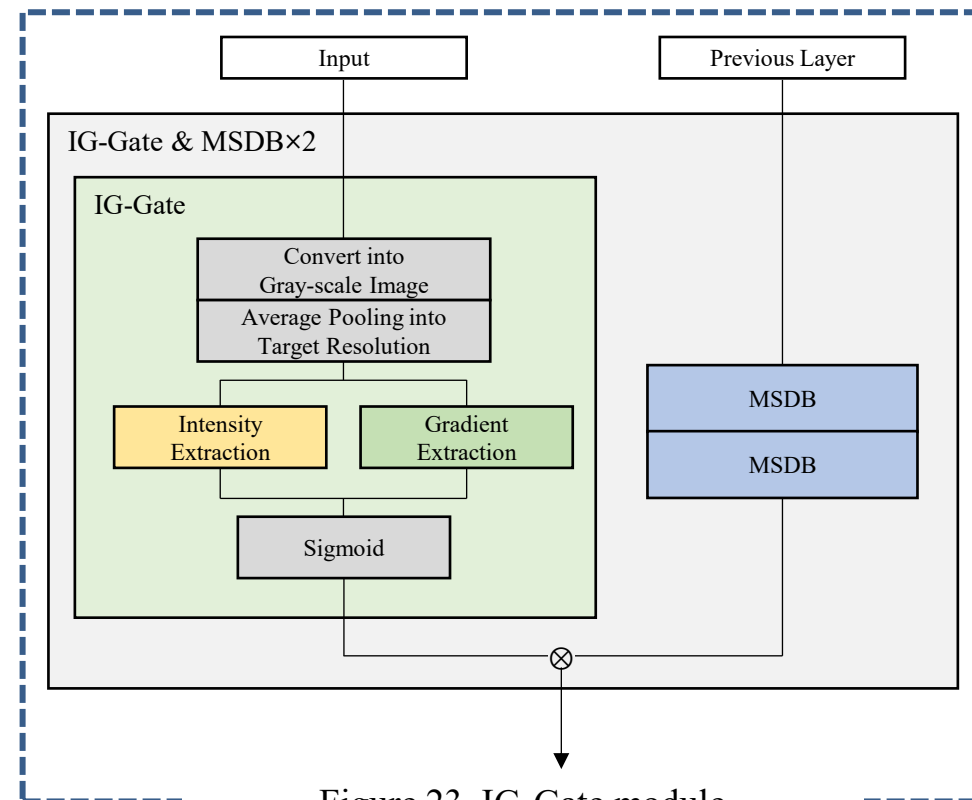
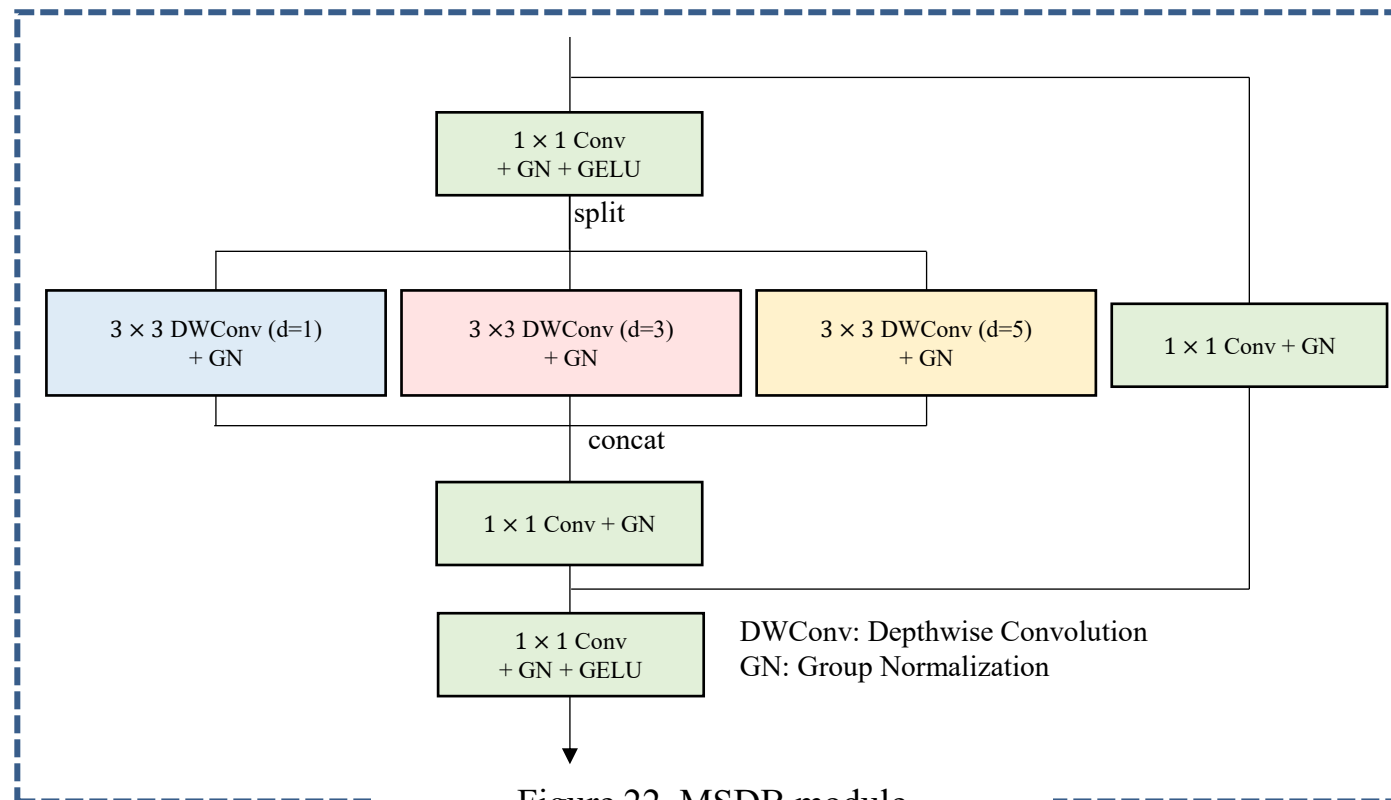
- We adopt an encoder–decoder architecture, where the encoder stacks Multi-Scale Dilated Blocks (MSDBs). Each MSDB efficiently expands the effective receptive field by using parallel dilated convolutions to capture multi-scale contextual information.
- After each MSDB in the encoder, a parameter-free IG-Gate is applied. Using stage-pooled grayscale intensity and Sobel cues from the raw input image, it multiplies the features to significantly reduce background mixing and reliably preserve crack continuity.



Proposed Method

Details

- MSDB Projects the input with a 1×1 conv, then runs three 3×3 depthwise branches at dilations 1, 3, and 5 for multi-scale context. Their outputs are normalized, concatenated, fused by a 1×1 conv, and added via a residual (with 1×1 projection if needed).
- IG-Gate Converts the input to grayscale, down samples to stage size, and combines intensity and gradient maps. A sigmoid produces a single-channel gate that element-wise modulates MSDB features, boosting evidence and suppressing background.



Experimental Settings

Dataset

- Ceramic comprises tile-surface, where repetitive tile patterns create segmentation ambiguity; Masonry dataset contains diverse brick whose periodic joints and mortar seams often trigger crack-like responses, producing strong crack–background ambiguity [9, 11].
- DeepCrack-237/537 are benchmark splits with manually annotated pixel-level masks across multiple scenes and scales, spanning concrete, asphalt, and other textured surfaces to evaluate generalization and continuity under varied backgrounds [10].

Table 3. Dataset summary

Dataset	Original Resolution	Input Resolution	# Images	Train/Val/Test
Ceramic [9]	$256 \times 256 \times 3$	$256 \times 256 \times 3$	100	60/20/20
DeepCrack237 [10]	$384 \times 544 \times 3$	$384 \times 640 \times 3$	237	142/47/48
Masonry [11]	$224 \times 224 \times 3$	$256 \times 256 \times 3$	240	144/48/48
DeepCrack537 [10]	$384 \times 544 \times 3$	$384 \times 640 \times 3$	537	316/105/106

[9] Junior, G.S.; Ferreira, J.; Millán-Arias, C.; Daniel, R.; Junior, A.C.; Fernandes, B.J.T. Ceramic Cracks Segmentation with Deep Learning. Appl. Sci. Switz. 2021, 11.
[10] Liu.; Yao, Y.; Lu, J.; Xie, X.; Li, R.; Li. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing 2019, 338, 139–153
[11] Dais, D.; Ihsan Engin Bal.; Smyrou, E.; Sarhosis, V. Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. Autom. Constr. 2021, 125, 103606

Experimental Settings

Evaluation Measures

- *Intersection of Union(IoU)* $IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{A \cap B}{A \cup B}$
- *F1 score* $F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
- *Precision*: fraction of predicted crack pixels that are truly crack, $TP/(TP+FP)$
- *Recall*: fraction of true crack pixels correctly detected, $TP/(TP+FN)$

Cross Validation

- Number of Iterations: 5
- Data Split Ratio (Train:Validation:Test) = 6:2:2

Proposed Model Variants

- Base model: Base variant uses **32 initial feature** channels and employs **1-2-2-2 configuration of MSDBs** across the four encoder stages.
- Small model: Small variant uses **28 initial feature** channels and employs **1-2-2-1 configuration of MSDBs** across the four encoder stages.
- Tiny model: Tiny variant uses **24 initial feature** channels and employs **1-1-2-1 configuration of MSDBs** across the four encoder stages.

Hyperparameter Settings

- Batch size: 4
- Epochs: 100
- Learning Rate: $1e-3$
- Optimizer: AdamW(Weighted Decay= $1e-4$)

Comparison model list

- | | | |
|----------------------|---------------|----------------|
| • RHACrackNet [3] | • CSNet [13] | • RSNet [18] |
| • CarNet [5] | • U-MPSC [14] | • DECSNet [19] |
| • CrackFormer II [6] | • DSNet [15] | • XYWNet [20] |
| • LMNet [8] | • PIDNet [16] | |
| • BLCDNet [12] | • DSUNet [17] | |

Experimental Results

Comparison to Conventional Methods

- Table 4 reports IoU for the proposed Base model versus a broad range of diverse state-of-the-art methods. With 32 initial channels and a 1–2–2–2 MSDB layout across four stages, it consistently achieves the best IoU on all four datasets with clear margins.
- The performance gains are statistically significant ($p < 0.05$) under paired t-tests against most competing models. Our model consistently outperforms CrackFormer II and DSUNet, including a +0.017 IoU improvement over DSUNet on the Masonry dataset.

Table 4. Experimental results

Model	Ceramic	DeepCrack237	Masonry	DeepCrack537
Proposed (Base)	0.304±0.024	0.793±0.015	0.650±0.017	0.768±0.012
BLCDNet	0.256±0.027▼	0.767±0.016▼	0.579±0.023▼	0.742±0.015▼
CSNet	0.188±0.027▼	0.717±0.013▼	0.537±0.008▼	0.693±0.015▼
CarNet	0.254±0.015▼	0.772±0.012▼	0.595±0.029▼	0.740±0.015▼
CrackFormer II	0.285±0.025	0.785±0.017	0.632±0.022	0.754±0.015▼
DSNet	0.214±0.018▼	0.760±0.019▼	0.611±0.023▼	0.739±0.016▼
PIDNet	0.107±0.015▼	0.687±0.052▼	0.552±0.049▼	0.689±0.017▼
DSUNet	0.301±0.037	0.779±0.010	0.633±0.023	0.751±0.014▼
RSNet	0.252±0.013▼	0.763±0.018▼	0.578±0.011▼	0.725±0.013▼
U-MPSC	0.235±0.017▼	0.683±0.030▼	0.529±0.033▼	0.617±0.029▼
DECSNet	0.265±0.019	0.774±0.013▼	0.629±0.015▼	0.741±0.017▼

*Best results are marked in **bold**. A Triangle(▼) indicates statistical significance at the 0.05 level based on paired t-tests against the proposed model

Experimental Results

Comparison to Conventional Methods

- Table 5 reports IoU for the proposed Small variant against a lightweight competitor on four benchmarks, under the identical training settings. With 28 initial channels and a 1–2–2–1 MSDB layout, the Small model attains the best IoU on all datasets by a clear margin.
- Table 6 reports IoU for the proposed Tiny variant against highly compact baselines on the same datasets, under the same protocol. Using 24 initial channels and a 1–1–2–1 MSDB layout, the Tiny model also leads on all four datasets despite tight compute budgets.

Table 5. Experimental results on Small model

Model	Ceramic	DeepCrack237	Masonry	DeepCrack537
Proposed (Small)	0.288 ± 0.026	0.788 ± 0.013	0.644 ± 0.016	0.766 ± 0.016
RHACrackNet	0.268 ± 0.023	0.770 ± 0.022▼	0.626 ± 0.017▼	0.746 ± 0.014▼

*Best results are marked in **bold**. A Triangle(▼) indicates statistical significance at the 0.05 level based on paired t-tests against the proposed model

Table 6. Experimental results on Tiny model

Model	Ceramic	DeepCrack237	Masonry	DeepCrack537
Proposed (Tiny)	0.285 ± 0.041	0.789 ± 0.014	0.643 ± 0.022	0.763 ± 0.015
LMNet	0.257 ± 0.026	0.782 ± 0.017	0.632 ± 0.023	0.753 ± 0.015▼
XYWNet	0.278 ± 0.020	0.777 ± 0.017▼	0.625 ± 0.028▼	0.754 ± 0.018▼

*Best results are marked in **bold**. A Triangle(▼) indicates statistical significance at the 0.05 level based on paired t-tests against the proposed model

Experimental Results

Pareto Analysis

- On the Ceramic dataset, the Tiny variant matches the accuracy of much larger models, even under challenging texture variations. It does so with a fraction of the parameters, preserving efficiency without sacrificing stability during training and inference.
- Scaling from Tiny to Small to Base yields a predictable accuracy–complexity trade-off across datasets and practical budgets, as Pareto plots show. This lets practitioners choose a variant that fits their compute budgets for deployment on edge devices.

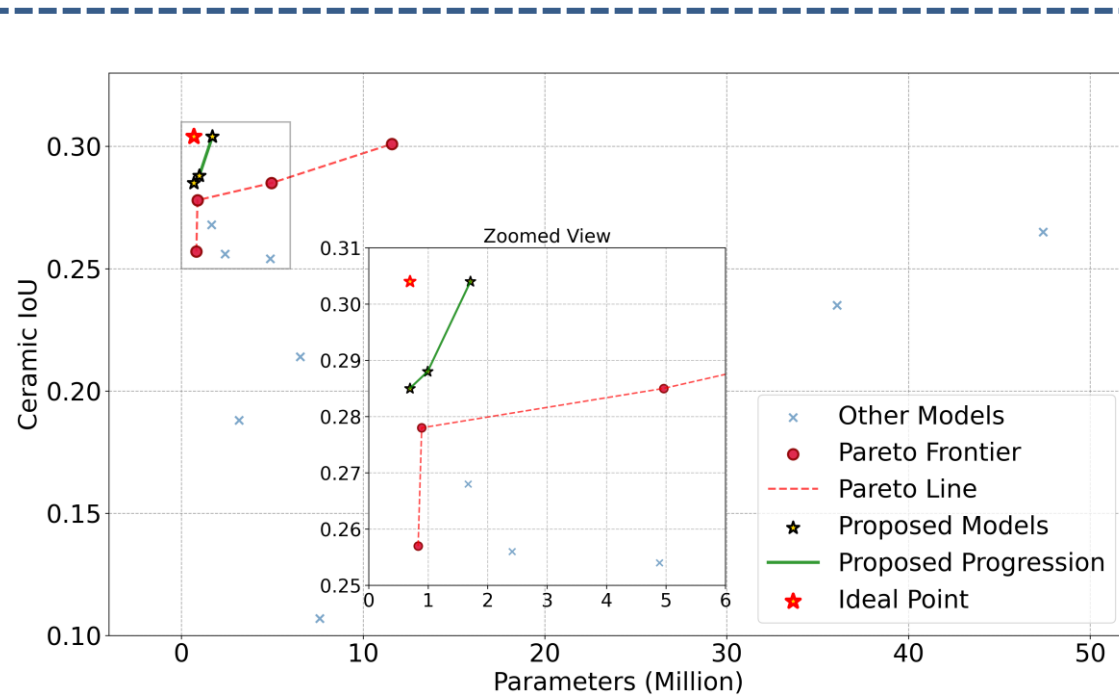


Figure 24. Pareto analysis of IoU vs. #Params on the Ceramic dataset

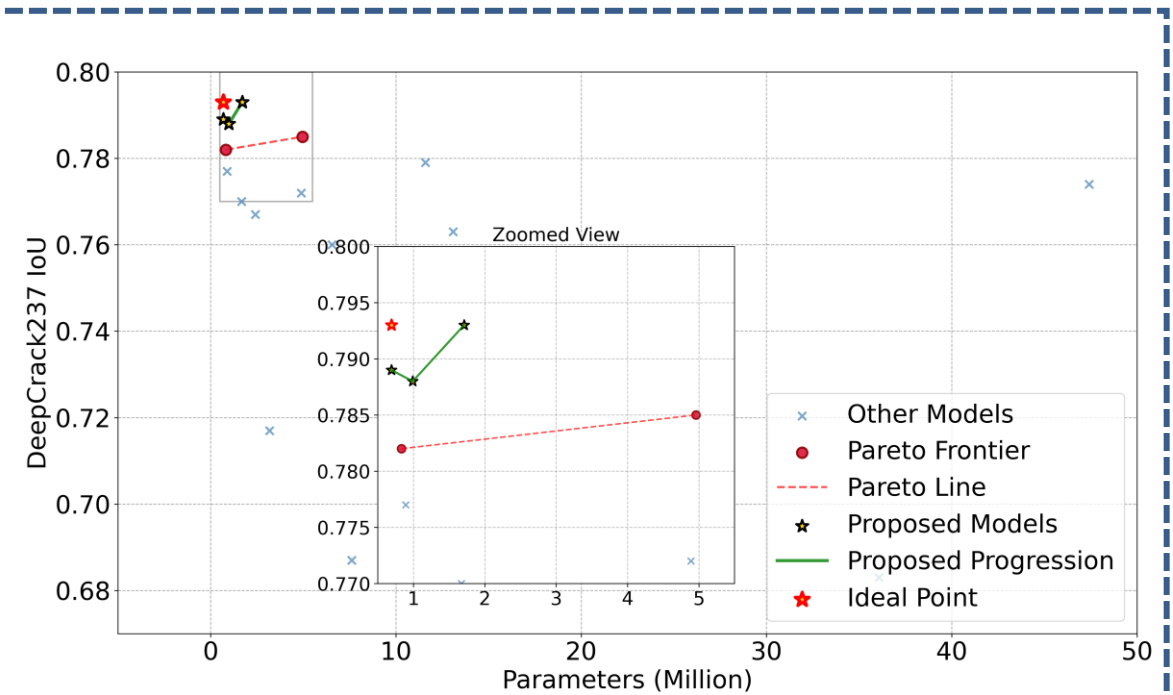


Figure 25. Pareto analysis of IoU vs. #Params on the DeepCrack237 dataset

Experimental Results

Pareto Analysis

- The Pareto frontier (circles and dashed line) clearly marks the best trade-offs among existing methods, highlighting efficiency at comparable accuracy and consistency. Our models sit beyond this front on efficiency, delivering higher accuracy per cost in FLOPs.
- The proposed progression (green) from Tiny→Small→Base steadily approaches the ideal point, yielding monotonic IoU gains for modest FLOPs increases. Zoomed views show Small and Tiny remain highly competitive even at low budgets for deployment.

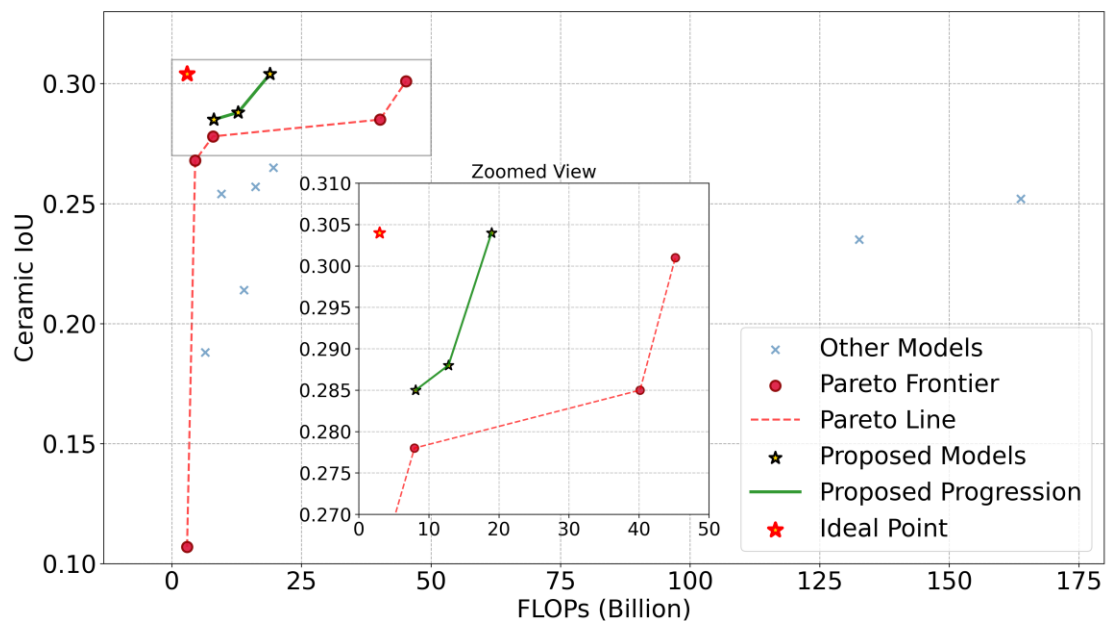


Figure 26. Pareto analysis of IoU vs. FLOPs on the Ceramic dataset

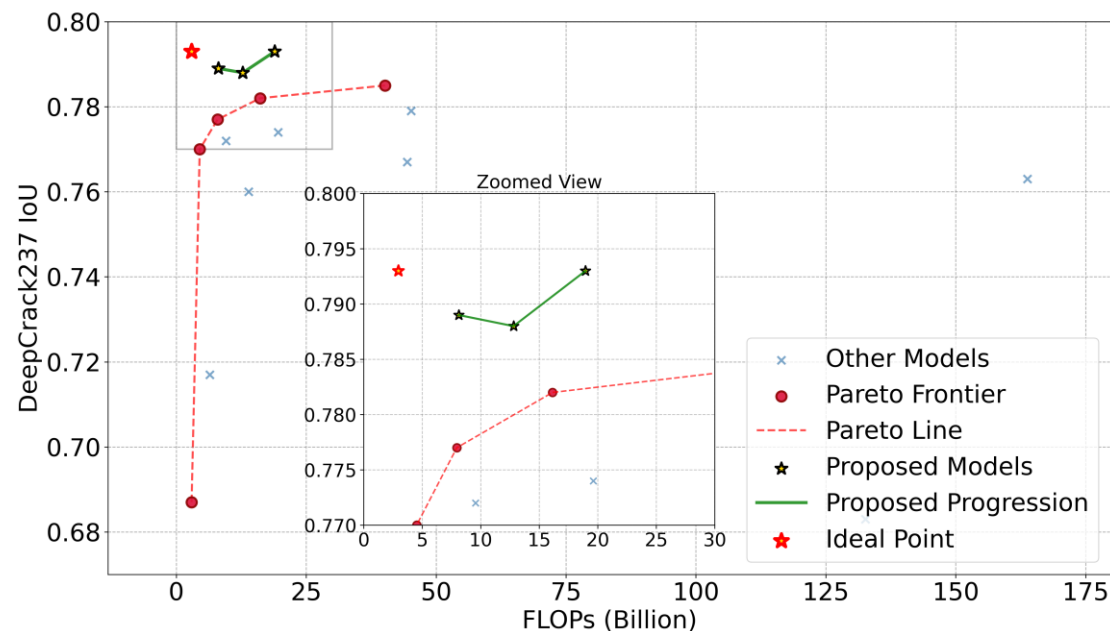


Figure 27. Pareto analysis of IoU vs. FLOPs on the DeepCrack237 dataset

Experimental Results

Ablation Study

- Enabling MSDB consistently raises IoU over the baseline by expanding the receptive field and capturing stronger context. IG-Gate also improves performance, with clear gains on Ceramic and Masonry by preserving low-level evidence that would otherwise be lost.
- With both MSDB and IG-Gate enabled, the model attains the top IoU on all four datasets. The two act synergistically because MSDB broadens context to learn crack traits while IG-Gate keeps learning guided by the most salient input evidence.

Table 7. Ablation study of the MSDB and IG-Gate modules on four datasets.

Modules		IoU			
MSDB	IG-Gate	Ceramic	DeepCrack237	Masonry	DeepCrack537
✓	✓	0.295 ± 0.023	0.791 ± 0.014	0.650 ± 0.022	0.762 ± 0.011
✓	×	0.286 ± 0.007	0.786 ± 0.015	0.646 ± 0.016	0.762 ± 0.013
×	✓	0.291 ± 0.007	0.779 ± 0.009	0.617 ± 0.036	0.750 ± 0.022
×	×	0.266 ± 0.032	0.775 ± 0.015	0.632 ± 0.033	0.755 ± 0.017

*Best results are marked in **bold**. A Triangle(▼) indicates statistical significance at the 0.05 level based on paired t-tests against the proposed model

Experimental Results

Qualitative Analysis

- Other models such as XYWNet and RHACrackNet confuse mortar joints and brick textures with cracks, producing relatively more false positives. In both images, they light up vertical seams and textured faces, with corner thickening and boundary jitter.
- MSDB expands the effective receptive field pre-downsampling, separating true crack paths from brick seams. IG-Gate, via intensity/gradient cues, suppresses crack-like textures and preserves thin, continuous centerlines with accurate width/endpoints.

*In the figure, white marks true positives, green marks false positives, and red marks false negatives.

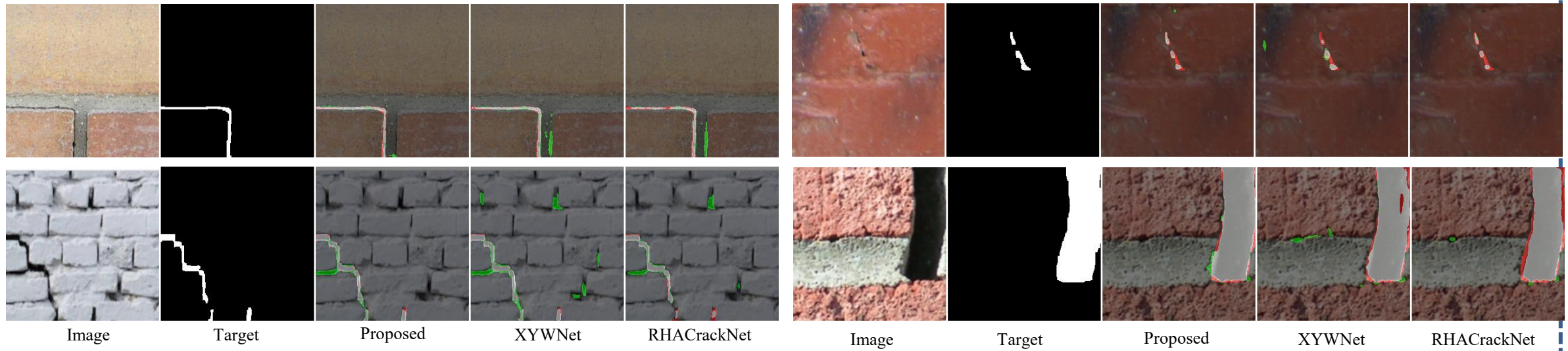


Figure 28. Qualitative comparison of segmentation outputs on the Masonry dataset

| Conclusion and Contribution

Conclusion

- Many lightweight crack-seg models place large-rate dilations only in deep, low-resolution stages. This lowers computational cost but weakens early receptive fields—long-range context is missed, background cues bleed in, and thin crack paths break.
- EP-REx introduces an early MSDB (parallel depthwise dilations) to widen context while keeping detail, plus an IG-Gate (from input grayscale and Sobel) to suppress background mixing. Together they preserve thin, continuous trajectories with higher consistency.
- Across four datasets over five runs, our model achieves best IoU—0.304/0.793/0.650/0.768 on Ceramic/DeepCrack237/Masonry/DeepCrack537 ($\pm 0.024/0.015/0.017/0.012$)—with paired $p < 0.05$ versus most baselines; on Masonry, it is +0.017 IoU over DSUNet.

Contribution

- In Department
 - Undergraduate Researcher, **AutoML Laboratory, Department of Artificial Intelligence, Chung-Ang University**, Seoul, Korea, 2024–2025
 - President, *3rd Student Council, Department of Artificial Intelligence, Chung-Ang University*, Seoul, Korea, 2023
- In International Journals
 - (Submitted) EP-REx: Evidence-Preserving Receptive-Field Expansion for Efficient Crack Segmentation
- In Domestic Conference
 - Enhancing Feature Fusion in Skip Connections for Efficient Crack Segmentation, *IEIE*, 2025

Thank you

Reference

- [1] Zou, Q.; Cao, Y.; Li, Q.; Mao, Q.; Wang, S. CrackTree: Automatic crack detection from pavement images. *Pattern Recognition. Lett.* 2012, 33, 227–238.
- [2] Zou, Q.; Zhang, Z.; Li, Q.; Qi, X.; Wang, Q.; Wang, S. Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE Trans. Image Process.* 2018, 28, 1498–1512.
- [3] Zhu, Guijie, et al. "A lightweight encoder–decoder network for automatic pavement crack detection." *Computer-Aided Civil and Infrastructure Engineering* 39.12 (2024): 1743-1765.
- [4] Zhang, Y.; Xu, Y.; Martinez-Rau, L.S.; Vu, Q.N.P.; Oelmann, B.; Bader, S. On-Device Crack Segmentation for Edge Structural Health Monitoring. *arXiv preprint arXiv:2505.07915* 2025
- [5] Li.; Yang, K.; Ma, J.; Wang, S.; Wang, B.; Tian, S.; Qi, Y.; Tian, Z.; Yingjie. Rethinking Lightweight Convolutional Neural Networks for Efficient and High-Quality Pavement Crack Detection. *IEEE Trans. Intell. Transp. Syst.* 2024, 25, 237-250.
- [6] Liu.; Yang, H.; Miao, J.; Mertz, X.; Kong, C.; Hui. CrackFormer Network for Pavement Crack Segmentation. *IEEE Trans. Intell. Transp. Syst.* 2023, 24, 9240–9252.
- [7] Liang, J.; Gu, X.; Jiang, D.; Zhang, Q. CNN-based network with multi-scale context feature and attention mechanism for automatic pavement crack segmentation. *Autom. Constr.* 2024, 164, 105482.
- [8] Al-maqtari.; Peng, O.; Al-Huda, B.; Al-Malahi, Z.; Maqtary, A.; Naseebah. Lightweight Yet Effective: A Modular Approach to Crack Segmentation. *IEEE Trans. Intell. Veh.* 2024, PP, 7961–7972.
- [9] Junior, G.S.; Ferreira, J.; Millán-Arias, C.; Daniel, R.; Junior, A.C.; Fernandes, B.J.T. Ceramic Cracks Segmentation with Deep Learning. *Appl. Sci. Switz.* 2021, 11.
- [10] Liu.; Yao, Y.; Lu, J.; Xie, X.; Li, R.; Li. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing* 2019, 338, 139–153
- [11] Dais, D.; 'Ihsan Engin Bal.; Smyrou, E.; Sarhosis, V. Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. *Autom. Constr.* 2021, 125, 103606
- [12] Lin, C.; Zhang, Z.; Peng, J.; Li, F.; Pan, Y.; Zhang, Y. A lightweight contour detection network inspired by biology. *Complex Intell. Syst.* 2024, 10, 4275–4291
- [13] Liang.; Gu, J.; Jiang, X.; Zhang, D.; Qipeng. CNN-based network with multi-scale context feature and attention mechanism for automatic pavement crack segmentation. *Autom. Constr.* 2024, 164, 105482.
- [14] Zhang, Z.; Wang, W.; Tian, X.; Luo, C.; Tan, J. Visual inspection system for crack defects in metal pipes. *Multimedia Tools Appl.* 2024, 83, 81877–81894
- [15] Guo, Z.; Bian, L.; Wei, H.; Li, J.; Ni, H.; Huang, X. DSNet: A Novel Way to Use Atrous Convolutions in Semantic Segmentation. *IEEE Trans. Circuits Syst. Video Technol.* 2025, 35, 3679–3692
- [16] Xu, J.; Xiong, Z.; Bhattacharyya, S.P. PIDNet: A Real-Time Semantic Segmentation Network Inspired by PID Controllers. In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*
- [17] Yu.; Zhang, Y.; Yu, Y.; Yue, J.; Jianwei. Lightweight decoder U-net crack segmentation network based on depthwise separable convolution. *Multimedia Syst.* 2024, 30, 295.
- [18] Ali.; AlJassmi, L.; Swavaf, H.; Khan, M.; Alnajjar, W.; Fady. RS-Net: Residual Sharp U-Net architecture for pavement crack segmentation and severity assessment. *J. Big Data* 2024, 11, 116.
- [19] Zhang.; Zeng, J.; Sharma, Z.; Alfarraj, P.K.; Tolba, O.; Wang, A.; Jin. A dual encoder crack segmentation network with Haar wavelet-based high–low frequency attention. *Expert Syst. Appl.* 2024, 256, 124950
- [20] Pang, X.; Lin, C.; Li, F.; Pan, Y. Bio-inspired XYW parallel pathway edge detection network. *Expert Syst. Appl.* 2024, 237, 121649.