EB-UNet++: An enhanced crack segmentation network combining EfficientNet-B2 and UNet++ with boundary extraction module

Phan Thi Hai Hong[1*], Truong Thi Thu Hang[2], Dang Van Giap[3], Ta Huu Vinh[4]

[1]Institute of Information and Communications Technology, Le Quy Don Technical University, 236 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam;
[2]Institute of Information Technology and Electronics, Academy of Military Science and Technology, 17 Hoang Sam, Nghia Do, Hanoi, Vietnam;
[3]Telecommunications University, 101 Mai Xuan Thuong, Bac Nha Trang, Khanh Hoa, Vietnam;
[4]Advanced Technology Center, Le Quy Don Technical University, 236 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam.
[*]Corresponding author: hongpth@lqdtu.edu.vn

Received 10 Aug. 2025; Revised 26 Sep. 2025; Accepted 10 Oct. 2025; Published 30 Oct. 2025.
DOI: https://doi.org/10.54939/1859-1043.j.mst.IITE.2025.148-159

ABSTRACT

Pavement crack detection is a crucial task in intelligent transportation systems and infrastructure maintenance. However, accurate segmentation of cracks remains challenging due to their irregular shapes, low contrast against the background, and varying lighting or surface conditions. In this study, we propose EB-UNet++, a novel deep learning architecture designed to enhance crack segmentation performance. EB-UNet++ integrates the powerful feature encoding capabilities of EfficientNet-B2 into the UNet++ encoder structure, enabling more efficient and robust multi-scale feature extraction. To further refine the crack boundaries and suppress false detections, we incorporate a Boundary Extraction Module into the network. Experimental results on benchmark pavement crack datasets demonstrate that EB-UNet++ outperforms several state-of-the-art models in both segmentation accuracy and boundary delineation, achieving higher IoU and F1-scores. The proposed architecture shows strong potential for practical deployment and scalability in automated road inspection and infrastructure monitoring systems.

Keywords: Pavement crack detection; Crack segmentation; Boundary extraction module (BEM); Road surface inspection; Multi-scale feature extraction.

1. INTRODUCTION

Cracks in road pavements are early indicators of structural deterioration and potential safety hazards, which can lead to costly maintenance and even serious traffic accidents if not detected in time. Timely and accurate detection of these cracks is essential for effective road maintenance and infrastructure management. Traditional manual inspection methods, however, are labor-intensive, time-consuming, and prone to human error. In response, computer vision techniques-particularly those based on deep learning-have emerged as promising tools for automating crack detection with higher precision and efficiency.

Among deep learning approaches, encoder–decoder architectures such as Unet [16] and its variants have been widely adopted in crack segmentation due to their ability to capture both global context and fine-grained structural details. Nonetheless, crack segmentation remains challenging in real-world scenarios due to issues such as low contrast between cracks and background, high levels of noise, and the irregular, fine-grained nature of cracks. These challenges often lead to false detections and blurred or discontinuous boundaries in segmentation results.

To tackle these issues, researchers have increasingly focused on semantic segmentation methods at the pixel level, which enable precise localization and delineation of cracks. In 2020, Ren et al. proposed CrackSegNet [17], which employed a fully convolutional network with dilated convolutions and pyramid pooling, outperforming traditional UNet models in terms of F1-score and mIoU on tunnel crack datasets. Building on this, Lin et al. (2023) introduced DeepCrackAT [12], which integrated attention mechanisms to fuse multi-scale features from convolutions in an encoder–decoder structure, achieving competitive results on concrete surface datasets but still facing limitations when segmenting cracks with highly variable widths. Meanwhile, hybrid architectures that combine CNNs with Transformers have demonstrated further improvements. CrackFormer [11] incorporated self-attention modules inspired by Transformers into a SegNet-based framework, improving multi-scale feature learning and computational efficiency. More recently, MixSegNet [23], introduced by Zhou et al. (2024), leveraged a hybrid CNN–Transformer architecture built upon UNet to jointly capture local textures and long-range dependencies, achieving state-of-the-art performance on the Cracks-APCGAN benchmark. Similarly, MambaCrackNet [6], introduced by Han et al. (2024), fused Vision MLPs with CNN components to enhance the detection of fine cracks, although its dual-branch architecture increased computational complexity and inference time.

These advances reflect a clear research trend: the design of increasingly deeper and more hybridized architectures-combining CNNs, Transformers, and MLPs-to enhance segmentation performance under challenging real-world conditions. However, two gaps remain insufficiently addressed: (i) the need for efficient multi-scale feature extraction that balances accuracy with computational cost, and (ii) still difficult to clearly outline thin and irregular cracks, and current models often misclassify them.

In this paper, we propose EB-UNet++, a novel deep learning architecture designed to improve pavement crack segmentation. To summarize, our main contributions can be outlined in three key points as follows:

- We introduce EB-UNet++, an enhanced UNet++ architecture that integrates EfficientNet-B2 as the encoder to enable efficient and expressive multi-scale feature extraction.
- We incorporate a Boundary Extraction Module (BEM) to guide the network in learning boundary-aware features, leading to more precise segmentation of narrow and irregular crack edges.
- We validate the proposed method on public crack datasets, where EB-UNet++ outperforms other models in terms of segmentation accuracy and boundary localization performance.

The remainder of the paper is organized as follows: Section 2 reviews related work on crack segmentation. Section 3 explains the methodology, while section 4 presents the experimental and results. Lastly, section 5 concludes the paper and suggests potential directions for future research.

2. RELATED WORK

2.1. UNet and UNet++ in crack segmentation

Unet [16], originally developed for medical image segmentation (Ronneberger et al. 2015), has become a foundational architecture for segmentation tasks due to its multi-level feature extraction and encoder–decoder skip connections, which help mitigate information loss. In crack segmentation, UNet has been widely adopted thanks to its strong performance. Several studies have shown that UNet outperforms fully convolutional networks (FCNs [18]) on small-scale crack datasets. For instance, Liu et al. (2019) demonstrated that UNet achieved more accurate segmentation results than FCNs on small crack datasets [2]. Similarly, Huyan et al. (2020) proposed CrackU-Net, which combines the strengths of FCNs and UNet, significantly reducing false positives on a dataset of 3,000 smartphone-captured images [8].

Beyond UNet, the UNet++ architecture (Zhou et al., 2018) [24] has received considerable attention due to its improved segmentation performance. UNet++ introduces a nested encoder–

decoder structure with dense skip connections, which effectively helps bridge the semantic gap between encoder and decoder features. As a result, the network becomes more trainable, and segmentation accuracy is enhanced. Experimental results have shown that UNet++ can increase the average IoU by approximately 3–4% compared to the original UNet across various datasets.

Recent studies have further exploited this architecture to better capture multi-scale crack information. For example, EfficientU-Net by Park et al. (2022) integrates UNet++ with an EfficientNet backbone to improve segmentation quality [9]. Overall, models based on UNet/UNet++, sometimes referred to as “nested U-structures”, remain central to crack segmentation research due to their ability to effectively aggregate features for fine-grained objects such as cracks.

2.2. Using EfficientNet as encoder in segmentation networks

Within UNet and its variants, integrating stronger CNN backbones has been shown to significantly improve feature extraction capabilities. EfficientNet [20], introduced by Tan et al. (2019), is a family of CNNs optimized via compound scaling, offering an excellent balance between accuracy and computational efficiency. This architecture has therefore been widely adopted as an encoder in segmentation networks to exploit pretrained deep weights.

Integrating EfficientNet as an encoder enables the extraction of deeper and more discriminative features without substantially increasing model complexity, which is particularly beneficial for crack segmentation tasks involving fine structures in noisy backgrounds. Park et al. (2022) proposed EfficientU-Net, a UNet++-based model that employs EfficientNet-B4 as the encoder, enhancing multi-level feature extraction [9]. In this architecture, EfficientNet functions as the encoder while the UNet++ decoder and dense skip connections ensure effective feature aggregation. Experimental results showed that EfficientU-Net achieved faster convergence and higher segmentation accuracy than models based on traditional backbones such as VGG or ResNet.

Similarly, other studies have reported consistent improvements in both IoU and F1-score when EfficientNet is integrated into segmentation networks [9, 14, 19]. These findings highlight a growing trend: combining UNet++ with EfficientNet, as proposed in EB-UNet++, is a promising strategy that unites powerful feature encoding with optimal skip connection design.

2.3. Boundary-Aware techniques for improving crack segmentation

The thin and elongated nature of cracks makes segmentation highly dependent on accurate boundary detection. Although deep multi-level models achieve good coverage of crack regions, their outputs often suffer from blurred or fragmented edges due to the imbalance between large background areas and narrow crack regions, which hinders boundary-level optimization during training.

To address this issue, Saining et al. [22] proposed a multi-loss strategy that incorporates edge detection: the network first learns to extract crack edges using a boundary-specific loss function and then fuses these with multi-scale features to improve overall segmentation. More recently, Hu et al. (2025) introduced the Boundary-Body Coherence Network, which jointly optimizes both boundary and body features of cracks using an adaptive self-attention mechanism. This approach significantly improved segmentation accuracy, particularly at the boundaries of fine cracks [7].

Other works have also incorporated boundary-aware loss functions and edge refinement modules [10, 15, 21], to ensure sharper and more continuous predictions along crack edges.

In summary, related studies have demonstrated the rapid progress in crack segmentation of deep learning approaches, evolving from the effective application of traditional architectures such as UNet and UNet++ to the integration of stronger backbones like EfficientNet and specialized modules such as BEM (boundary extraction mechanisms) for boundary refinement. These advancements provide a strong foundation for the proposed EB-UNet++ architecture, which leverages the combined strengths of UNet++ and EfficientNet-B2, along with boundary-aware techniques to enhance pavement crack segmentation performance.

3. PROPOSED METHOD

3.1. Overview of EB-UNet++ architecture

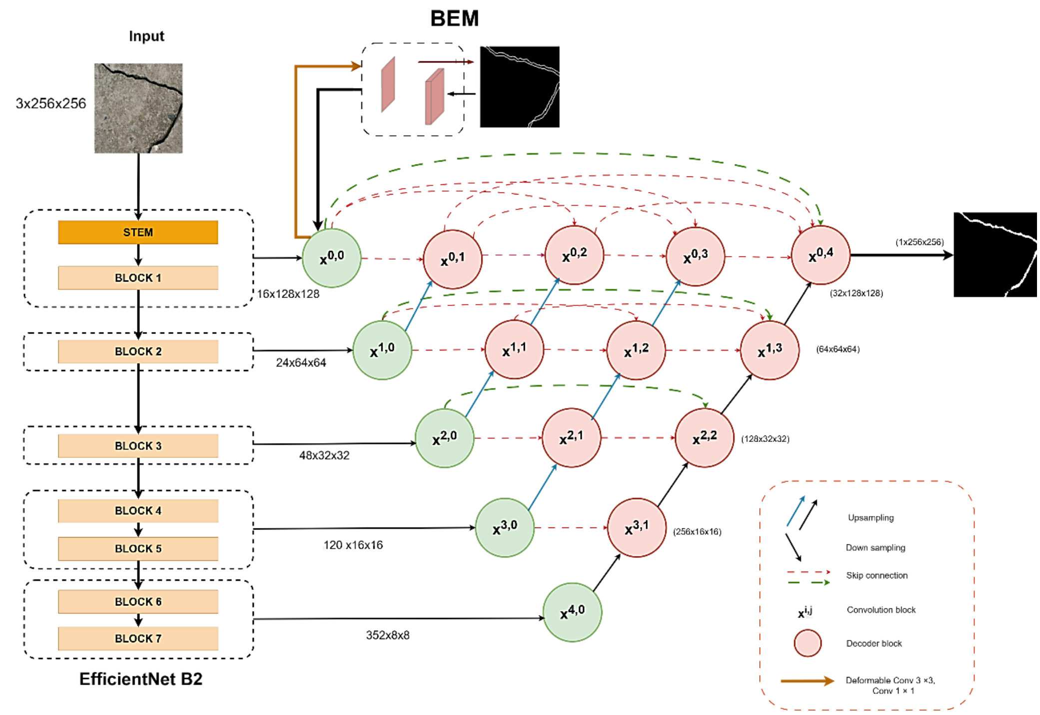

Figure 1. *Proposed EB-UNet++ architecture.*

In this study, we propose EB-UNet++, a novel encoder–decoder-based deep neural network architecture tailored for crack segmentation in pavement images. The proposed architecture builds upon the UNet++ framework, which is known for its nested and dense skip connections that enable effective multi-scale feature fusion and improved segmentation accuracy. EB-UNet++ introduces two key enhancements to the original UNet++: the integration of EfficientNet-B2 as the encoder backbone and the incorporation of a BEM to enhance edge-awareness in the early stages of feature processing.

The overall architecture of EB-UNet++ consists of three main components:

1. **Encoder:** EfficientNet-B2 is employed as the feature encoder, replacing the conventional convolutional blocks in UNet++. It extracts hierarchical feature maps from the input image at five different resolution levels (denoted as $X^{0,0}$ to $X^{4,0}$), capturing both low-level textures and high-level semantics while maintaining computational efficiency.
2. **BEM:** The output of the first encoder block ($X^{0,0}$) is fed into a dedicated boundary extraction branch to emphasize fine-grained spatial and edge information, which is then fused back into the main decoding stream.
3. **Decoder:** The decoder follows the standard UNet++ design with nested skip pathways and dense connections. These connections bridge encoder and decoder layers across multiple levels, allowing for refined feature propagation and reconstruction of the segmentation mask.

Figure 1 illustrates the overall pipeline of EB-UNet++. The input image is first passed through the EfficientNet-B2 encoder, which extracts multi-level feature representations. The lowest-level feature map ($X^{0,0}$), which retains high spatial resolution, is simultaneously processed by the BEM to enhance edge-specific details. The BEM-enhanced features are then merged back into the original stream and progressively decoded through the UNet++ decoder to generate the final segmentation output.

Unlike existing UNet++ variants that primarily focus on multi-scale fusion or attention mechanisms, EB-UNet++ introduces an explicit boundary-aware learning branch, enabling the network to refine edge features at the feature level rather than relying solely on loss-based supervision. This architectural novelty allows the model to achieve sharper delineation of cracks, particularly in thin and fragmented regions where other methods often fail.

3.2. EfficientNet-B2 as feature encoder

To enhance the feature extraction capability of the original UNet++ architecture [24], we adopt EfficientNet-B2 as the encoder backbone in our proposed EB-UNet++. EfficientNet, introduced by Tan and Le (2019) [20], is a family of convolutional neural networks that achieves state-of-the-art performance through a compound scaling method, which uniformly scales depth, width, and resolution of the network. Among its variants, EfficientNet-B2 offers a balanced trade-off between model accuracy and computational cost, making it well-suited for crack segmentation tasks where capturing fine-grained details is crucial but real-time processing is also desirable.

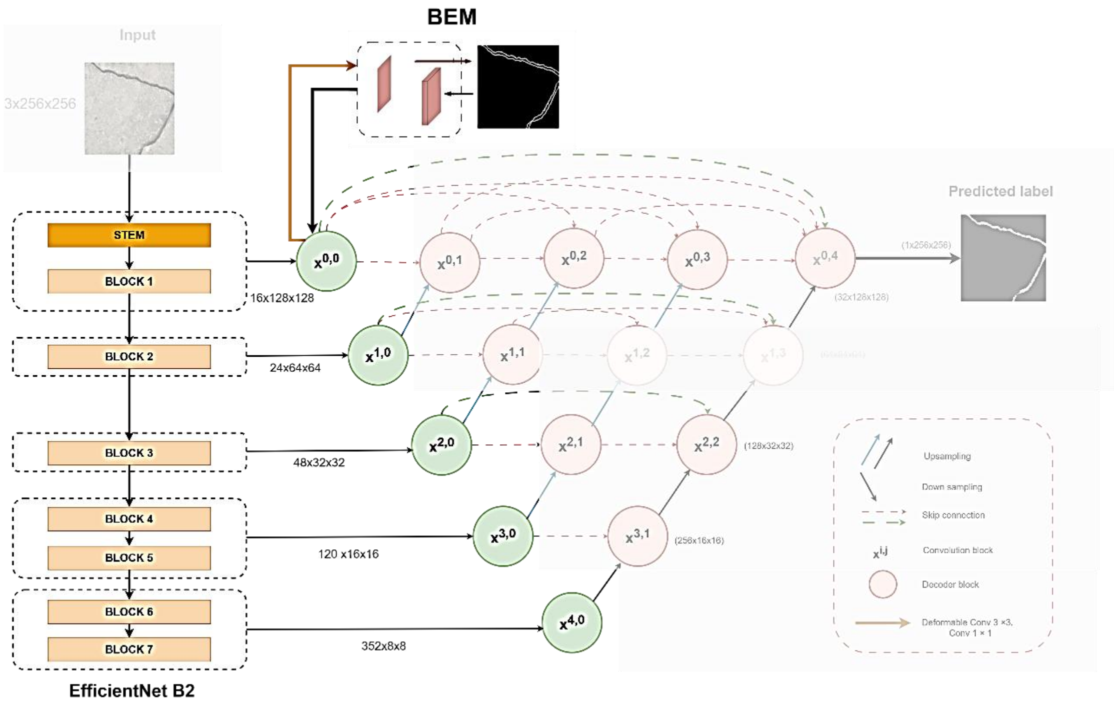

Figure 2. *Encoder architecture of the proposed EB-UNet++.*

In EB-UNet++, the standard encoder blocks of UNet++ are replaced by the feature extraction layers of EfficientNet-B2. As in figure 2, the encoder processes the input image through a series of inverted residual blocks, resulting in multi-scale feature maps at five different stages of abstraction. We denote these feature maps as $X^{0,0}, X^{1,0}, X^{2,0}, X^{3,0}$, and $X^{4,0}$, corresponding to different spatial resolutions from high ($X^{0,0}$) to low ($X^{4,0}$). This hierarchical representation captures both low-level edge textures and high-level semantic structures, which are essential for accurately identifying and localizing crack regions that often appear thin, fragmented, and low in contrast.

Moreover, EfficientNet-B2 leverages depthwise separable convolutions and squeeze-and-excitation (SE) blocks, which contribute to more efficient learning of channel-wise dependencies and improved feature sensitivity. These properties are particularly beneficial in crack detection, as cracks often exhibit subtle visual cues that require fine discriminative power across both spatial and channel dimensions.

The extracted encoder features are subsequently passed to the UNet++ decoder via dense skip connections. These connections preserve spatial information across scales and facilitate more precise reconstruction of the segmentation map. In addition, the high-resolution feature map from the earliest encoder stage ($X^{0,0}$) is also used as the input to BEM, which further enhances the model's sensitivity to crack boundaries.

By utilizing EfficientNet-B2 as the encoder, EB-UNet++ benefits from strong generalization capabilities and efficient feature representation, while maintaining a compact model size suitable for deployment in real-world crack inspection systems.

3.3. BEM

Accurate delineation of crack boundaries is challenging due to their thin, elongated, and often fragmented structures. To address this issue, we design a Boundary Enhancement Module aimed at explicitly learning and refining edge-related features during the early stages of the network.

The BEM is positioned immediately after the first encoder stage, which outputs the high-resolution feature map $X^{0,0}$. Since this stage retains the most spatial detail and contains low-level features such as textures, corners, and edges, it serves as an ideal input for boundary-aware learning.

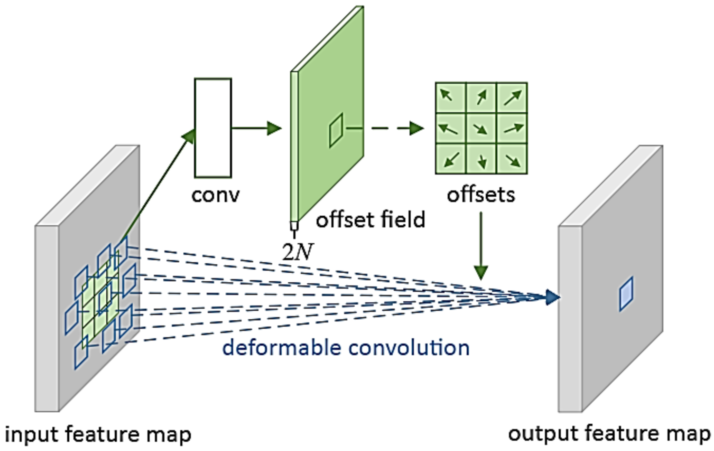

Figure 3. *3x3 Deformable convolution.*

The BEM in EB-UNet++ operates on the high-resolution feature map from the first encoder stage, where spatial details are best preserved. It employs a 3×3 Deformable Convolution (DConv) [3] as illustrated in figure 3, which dynamically adjusts its sampling locations to better capture irregular crack shapes and local geometric variations. This flexibility allows the network to learn complex and fine-grained boundary patterns that standard convolutions may fail to detect, thereby enhancing the segmentation of thin and fragmented cracks.

The boundary feature map is element-wise added to the original $X^{0,0}$, forming a refined feature $\tilde{X}^{0,0}$ that is both semantically rich and spatially precise. This refined feature map is then forwarded to the decoder and participates in skip connections within the UNet++ structure.

Unlike prior approaches that refine crack boundaries only at the loss level (e.g., boundary-aware loss functions), our BEM introduces a learnable module embedded directly into the architecture. This enables boundary enhancement at the feature representation level, resulting in sharper and more continuous crack edges.

3.4. Decoder and output segmentation

The decoder in EB-UNet++ adopts the nested, densely connected design of the original UNet++, enabling progressive upsampling and fusion of multi-scale encoder features for precise crack localization. The boundary-enhanced feature $\tilde{X}^{0,0}$ from the BEM is passed through the top-level decoding path, while other encoder outputs ($X^{1,0}, X^{2,0}, X^{3,0}, X^{4,0}$) from EfficientNet-B2 are integrated via skip connections to preserve hierarchical multi-scale information.

Compared to conventional UNet-style decoders that rely on simple concatenation of encoder features, the nested skip pathways in UNet++ are further reinforced by our boundary-refined features, allowing EB-UNet++ to better capture fine-grained crack structures across multiple scales. This design helps prevent the loss of spatial details typically observed in standard encoder–decoder pipelines.

In summary, EB-UNet++ combines boundary-refined features, dense skip connections, and multi-scale fusion to reconstruct segmentation masks with high coverage accuracy and precise boundary delineation. This joint design ensures that EB-UNet++ not only achieves superior crack coverage but also produces sharper and more continuous boundaries, addressing one of the key limitations of existing segmentation networks.

4. EXPERIMENTS AND RESULTS

4.1. Dataset

We evaluate the proposed EB-UNet++ on two widely used benchmark datasets: CrackVision12k and DeepCrack.

CrackVision12k [4] is a large-scale crack segmentation dataset consisting of over 12,000 images aggregated from 13 publicly available sources. It includes a wide variety of pavement types, crack shapes, lighting conditions, and background textures. This diversity makes CrackVision12k an ideal benchmark for evaluating the generalization ability of segmentation models across complex real-world scenarios.

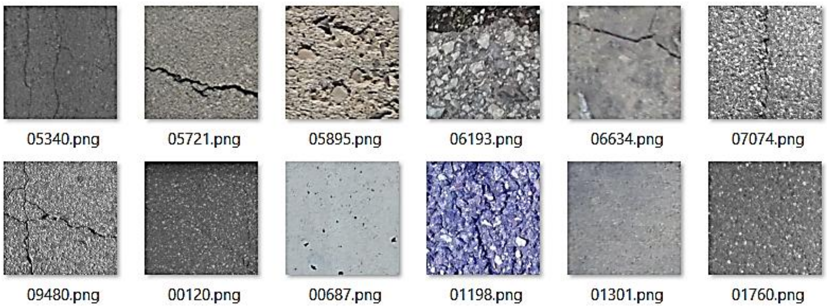

Figure 4. *Some sample images from the CrackVision12k dataset.*

DeepCrack [13] contains 537 high-resolution images of concrete and asphalt surfaces, each paired with pixel-level ground truth annotations. Although smaller in scale, DeepCrack is known for its high-quality, fine-grained annotations and challenging crack patterns, making it well-suited for assessing the boundary precision and detail sensitivity of segmentation models.

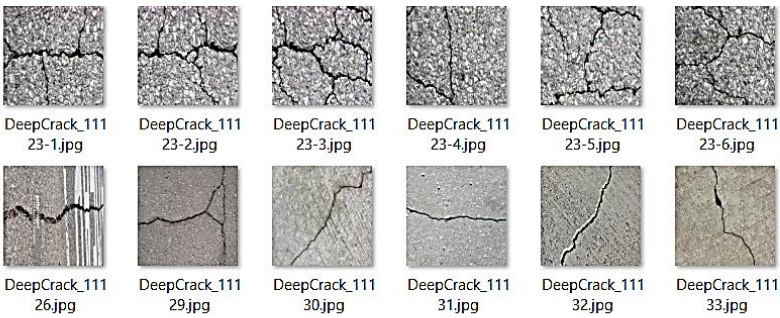

Figure 5. *Some sample images from the CrackVision12k dataset.*

These two datasets were chosen to ensure that the proposed method is evaluated under both diverse, large-scale conditions (CrackVision12k) and fine-grained, boundary-focused scenarios (DeepCrack), providing a balanced and comprehensive assessment of performance. This dual evaluation setting allows us to rigorously validate both the generalization and boundary-refinement capabilities of EB-UNet++.

4.2. Implementation

4.2.1. Baselines

To comprehensively evaluate our proposed pipeline, we compare it against five state-of-the-art crack segmentation models: FCN [18], Unet++ [24], DeepCrack [13], SegNet [1], and Hybrid-Segmentor [5]. All models were implemented using the PyTorch framework and trained on Kaggle using a TESLA T4 GPU. Pre-trained weights on ImageNet were used to initialize the model parameters. Each model was trained for 100 epochs with a batch size of 4, using the AdamW optimizer with a momentum of 0.9, a weight decay of 1e-4, and a learning rate of 0.001.

Alongside the proposed EB-UNet++, we also implement E-UNet++, which integrates EfficientNet-B2 into the UNet++ architecture without incorporating the BEM, for comparative experimentation. This ablation baseline enables us to directly isolate and quantify the contribution of the Boundary Extraction Module.

4.2.2. Evaluation metrics

To evaluate the performance of the proposed model, we used four common segmentation metrics as follows:

Precision measures the proportion of predicted crack pixels that are correct, where High precision indicates few false positives, reducing unnecessary pavement treatments:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Recall measures the proportion of actual crack pixels correctly detected. High recall ensures minimal missed cracks, which is critical for safety.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The **F1-score** is the harmonic mean of precision and recall that provides a balanced measure, especially for imbalanced datasets:

$$F - score = \frac{2 \times Precision \times Recall}{Precition + Recall} \tag{3}$$

IoU (Intersection over Union) measures the overlap between predicted and ground-truth crack regions. A higher IoU indicates more accurate segmentation and is a standard benchmark for model comparison.

$$IoU = \frac{\left|\sum_i \sum_j X(i,j) \cdot Y(i,j)\right|}{\left|\sum_i \sum_j [(X(i,j) + Y(i,j)) - X(i,j) \cdot Y(i,j)]\right|} \tag{4}$$

Together, these four metrics allow us to assess not only overall segmentation quality but also the trade-off between false positives, false negatives, and boundary-level localization accuracy.

4.3. Results

4.3.1. Results on CrackVision12k

Table 1 shows the performance comparison of EB-UNet++ against other methods on the

CrackVision12k dataset. The experimental results clearly demonstrate the superior performance of the proposed EB-UNet++ model. It achieves the highest scores across all evaluation metrics, with a Precision of 0.8081, Recall of 0.7966, F1-score of 0.8007, and IoU of 0.6698. Compared to traditional models such as FCN, UNet, and SegNet, EB-UNet++ shows significant improvements, particularly in terms of segmentation quality and boundary localization. Compared to the strongest baseline (Hybrid-Segmentor), EB-UNet++ improves IoU by nearly 4%, reflecting its enhanced ability to capture complex crack geometries in diverse real-world images.

Table 1. *Quantitative comparison on the CrackVision12k dataset.*

Models	Precision	Recall	F1 Score	IoU
FCN	0.8020	0.7020	0.7460	0.5980
Unet++	0.7890	0.7200	0.7500	0.6030
DeepCrack	0.7880	0.7040	0.7410	0.5920
SegNet	0.7500	0.7190	0.7300	0.5800
Hybrid-Segmentor	*0.8040*	0.7440	0.7700	0.6300
E-UNet++	0.7976	*0.7910*	*0.7890*	*0.6502*
EB-UNet++	**0.8081**	**0.7966**	**0.8007**	**0.6698**

Notably, the comparison between EB-UNet++ and its variant E-UNet++ (without BEM) highlights the effectiveness of the BEM. The consistent improvement in both F1-score and IoU confirms that boundary-aware learning plays a decisive role in crack detection tasks where irregular edges are critical.

4.3.2. Results on DeepCrack

Table 2. *Quantitative comparison on the DeepCrack dataset.*

Models	Precision	Recall	F-Score	IoU
FCN	0.7358	0.8664	0.7801	0.6491
Unet++	0.7558	**0.8964**	0.8201	0.6891
DeepCrack	0.8016	0.8040	0.8028	0.6448
SegNet	0.6588	0.8797	0.7534	0.5871
Hybrid-Segmentor	0.8044	*0.8950*	0.8473	0.7356
E-UNet++	**0.8742**	0.8375	*0.8510*	*0.7400*
EB-UNet++	*0.8328*	0.8792	**0.8550**	**0.7474**

Table 2 reports the results on the DeepCrack dataset. As DeepCrack includes more challenging and fine-grained crack structures, the precision of boundary localization is critical, EB-UNet++ continues to demonstrate strong performance, achieving the highest F1-score (0.8550) and IoU (0.7474), outperforming even Hybrid-Segmentor, which was previously considered one of the strongest baselines.

While E-UNet++ attains the highest precision (0.8742), EB-UNet++ offers a better balance between precision and recall, resulting in improved overall segmentation quality. This balance demonstrates the robustness of EB-UNet++ in handling cracks with highly variable widths and fragmented boundaries. Compared to conventional models such as UNet, FCN, and SegNet, EB-UNet++ consistently outperforms them across all metrics. Furthermore, the improvement over E-UNet++ again confirms the positive impact of the BEM, particularly in handling fine and irregular crack boundaries present in the DeepCrack dataset. These results highlight the robustness and boundary sensitivity of EB-UNet++ in challenging real-world crack segmentation scenarios.

4.3.3. Ablation comparison

To assess the specific contribution of the Boundary Extraction Module, we conduct an ablation study by comparing the full model (EB-UNet++) with its variant (E-UNet++), which shares the same encoder–decoder structure but excludes the BEM.

On both CrackVision12k and DeepCrack datasets, EB-UNet++ consistently outperforms E-UNet++ in terms of F1-score and IoU, while maintaining competitive precision and recall. Specifically, on the CrackVision12k dataset, the addition of BEM improves the F1-score from 0.7890 to 0.8007 and IoU from 0.6502 to 0.6698. Similarly, on the DeepCrack dataset, the F1-score increases from 0.8510 to 0.8550, and IoU improves from 0.7400 to 0.7474.

Although the absolute improvements may appear modest, their consistency across two diverse benchmarks strongly validates the design rationale of BEM. In particular, these gains are concentrated at crack edges, where other models often fail.

These improvements, though modest in absolute values, are consistent across both datasets and indicate that the BEM enhances the network's ability to preserve and refine crack boundaries, particularly in scenarios with thin or fragmented structures.

Several qualitative examples, as shown in figure 6, visually demonstrate the proposed model's superior capability in accurately segmenting cracks and maintaining boundary continuity compared to other approaches.

Test images *Ground truth* *EB-UNet++* *Test images* *Ground truth* *EB-Unet++*

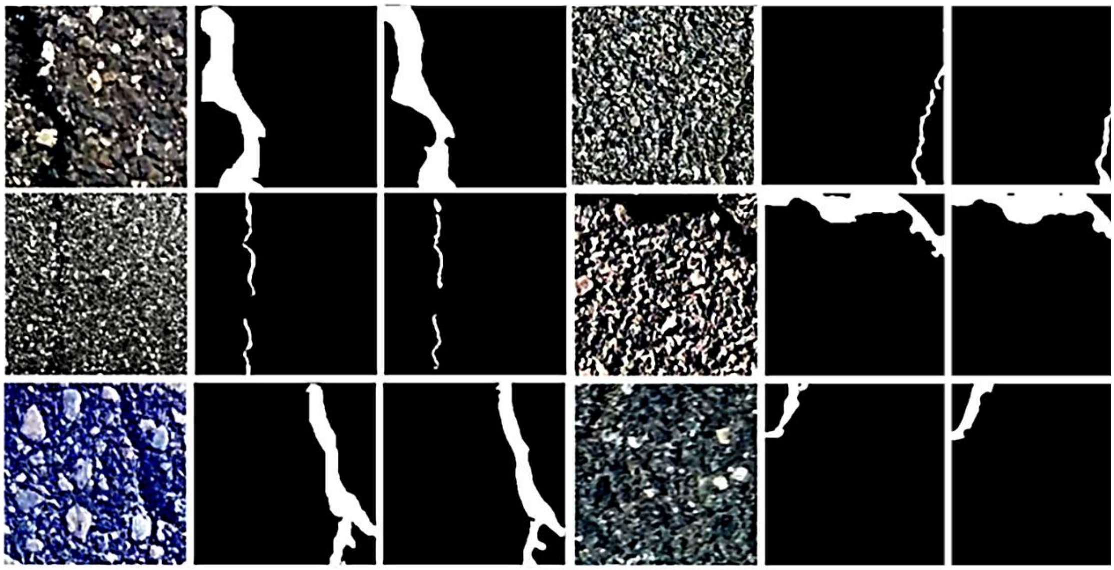

Figure 6. *Example crack segmentation results of EB-UNet++ compared with ground-truth annotations.*

5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose EB-UNet++, an enhanced crack segmentation network that integrates EfficientNet-B2 as the encoder and introduces a Boundary Extraction Module (BEM) to explicitly refine crack boundaries. By combining efficient feature encoding, dense multi-scale fusion, and boundary-aware representation, EB-UNet++ addresses the long-standing challenge of accurately delineating narrow and irregular cracks, which remain prone to misclassification in existing models. Experimental results on two widely used public benchmarks demonstrate that our method consistently outperforms state-of-the-art architectures in terms of both segmentation accuracy and boundary localization, confirming the effectiveness and robustness of the proposed design.

Future work: In future works, we aim to extend EB-UNet++ to support multi-class crack segmentation across diverse pavement materials, develop lightweight variants suitable for deployment on mobile and edge devices, and explore the integration of temporal information from video streams to enable real-time, continuous pavement monitoring in practical scenarios.

Acknowledgment: *This research was supported by the National-level Scientific Research Project, code: KC-4.0-51/19-25, and by the National Foundation for Science and Technology Development (NAFOSTED).*

REFERENCES

[1]. Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. *"SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation."* IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 2481–2495 (2015).

[2]. Chen *et al*. *"A multiscale enhanced pavement crack segmentation network coupling spectral and spatial information of UAV hyperspectral imagery."* Int. J. Appl. Earth Obs. Geoinformation, 128, 103772 (2024).

[3]. Dai *et al*. *"Deformable Convolutional Networks."* 2017 IEEE International Conference on Computer Vision (ICCV), 764–773 (2017).

[4]. Goo, June Moh, Xenios Milidonis, Alessandro Artusi, Jan Boehm, and Carlo Ciliberto. *"Hybrid-Segmentor: A Hybrid Approach to Automated Fine-Grained Crack Segmentation in Civil Infrastructure."* ArXiv abs/2409.02866 (2024).

[5]. Goo, June Moh *et al*. *"CrackVision12K."* University College London. Dataset. (2024). https://doi.org/10.5522/04/26946472.v1

[6]. Han, Chengjia, Handuo Yang, and Yaowen Yang. *"Enhancing pixel-level crack segmentation with visual mamba and convolutional networks."* Automation in Construction (2024).

[7]. Hu, Xiaopeng et al. *"BBCNet: Boundary-body coherence network with adaptive self-attention distillation for enhanced crack segmentation."* Digit. Signal Process., 162, 105148 (2025).

[8]. Huyan, Ju, Wei Li, Susan L. Tighe, Zhengchao Xu, and Junzhi Zhai. *"CrackU-net: A novel deep convolutional neural network for pixelwise pavement crack detection."* Structural Control and Health Monitoring, 27 (2020).

[9]. Ji, Honggeun, Jina Kim, Syjung Hwang, and Eunil Park. *"Automated Crack Detection via Semantic Segmentation Approaches Using Advanced U-Net Architecture."* Intelligent Automation & Soft Computing (2022).

[10]. Liu, Gaoyang, Wei Ding, Jian-Hua Shu, Alfred Strauss, and Yuan-yu Duan. *"Two-Stream Boundary-Aware Neural Network for Concrete Crack Segmentation and Quantification."* Structural Control and Health Monitoring (2023).

[11]. Liu, Huajun, Xiangyu Miao, Christoph Mertz, Chengzhong Xu, and Hui Kong. *"CrackFormer: Transformer Network for Fine-Grained Crack Detection."* 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 3763–3772 (2021).

[12]. Lin, Qing Huo, Wei Li, Xiangpan Zheng, Haoyi Fan, and Zuoyong Li. *"DeepCrackAT: An effective crack segmentation framework based on learning multi-scale crack features."* Eng. Appl. Artif. Intell., 126, 106876 (2023).

[13]. Liu, Yahui, Jian Yao, Xiaohu Lu, Renping Xie, and Li Li. *"DeepCrack: A deep hierarchical feature learning architecture for crack segmentation."* Neurocomputing, 338, 139–153 (2019).

[14]. Mehta, Vedant. *"Towards Real-Time Polyp Segmentation During Colonoscopy Using an EfficientNet-Based UNet Architecture."* 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE), 447–451 (2023).

[15]. Pham, Thang Quoc, Thai Hoang Le, Khai Dinh Lai, Dat Quoc Ngo, Tan Van Pham, Quang Hua, Khang Le, Huyen Le, and Tuyen Nguyen. *"CAAF-ResUNet: Adaptive Attention Fusion with Boundary-Aware Loss for Lung Nodule Segmentation."* Medicina, 61 (2025).

[16]. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. *"U-Net: Convolutional Networks for Biomedical Image Segmentation."* ArXiv abs/1505.04597 (2015).

[17]. Ren, Yupeng, Jisheng Huang, Zhiyou Hong, Wei Lu, J. Yin, Lejun Zou, and Xiaohua Shen. *"Image-based concrete crack detection in tunnels using deep fully convolutional networks."* Construction and Building Materials, 234, 117367 (2020).

[18]. Shelhamer, Evan, Jonathan Long, and Trevor Darrell. *"Fully convolutional networks for semantic segmentation."* 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431–3440 (2014).
[19]. Shen, Aoran, Ruxin Chen, Yueze Zhu, and Ruohan Hu. *"Segmentation of Multi-Organ Functional Tissue Units Using UNet-EfficientNet-B8."* Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science (2023).
[20]. Tan, Mingxing, and Quoc V. Le. *"EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks."* ArXiv abs/1905.11946 (2019).
[21]. Tao, Huaqi, Bing Liu, Jinqiang Cui, and Hong Zhang. *"A Convolutional-Transformer Network for Crack Segmentation with Boundary Awareness."* 2023 IEEE International Conference on Image Processing (ICIP), 86–90 (2023).
[22]. Zhao, Fan, Yu Chao, and Linyun Li. *"A Crack Segmentation Model Combining Morphological Network and Multiple Loss Mechanism."* Sensors (Basel, Switzerland), 23 (2023).
[23]. Zhou, Yang, Raza Ali, Norrima Binti Mokhtar, Sulaiman Wadi Harun, and Masahiro Iwahashi. *"MixSegNet: A Novel Crack Segmentation Network Combining CNN and Transformer."* IEEE Access, 12, 111535–111545 (2024).
[24]. Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. *"UNet++: A Nested U-Net Architecture for Medical Image Segmentation."* Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 11045, 3–11 (2018).

TÓM TẮT

EB-UNet++: Mạng phân đoạn vết nứt nâng cao kết hợp EfficientNet-B2 và UNet++ với khối trích xuất biên

Nhận dạng vết nứt mặt đường là một bài toán quan trọng trong hệ thống giao thông thông minh và công tác bảo trì hạ tầng. Tuy nhiên, việc phân đoạn chính xác các vết nứt vẫn gặp nhiều thách thức do hình dạng không đều, độ tương phản thấp so với nền và điều kiện ánh sáng hoặc bề mặt thay đổi. Trong nghiên cứu này, chúng tôi đề xuất kiến trúc EB-UNet++, một mô hình học sâu mới nhằm nâng cao hiệu quả phân đoạn vết nứt. Mô hình tích hợp khả năng trích xuất đặc trưng mạnh mẽ của EfficientNet-B2 vào bộ mã hoá UNet++, giúp cải thiện khả năng trích xuất đặc trưng đa tỉ lệ một cách hiệu quả và ổn định. Để tăng cường khả năng nhận diện biên và giảm nhiễu, chúng tôi bổ sung thêm khối trích xuất biên vào mạng. Kết quả thực nghiệm trên các tập dữ liệu vết nứt chuẩn cho thấy EB-UNet++ vượt trội hơn nhiều mô hình hiện đại khác về độ chính xác phân đoạn và nhận diện biên, đạt được các chỉ số IoU và F1-score cao hơn. Kiến trúc đề xuất hứa hẹn khả năng ứng dụng thực tế trong các hệ thống giám sát cơ sở hạ tầng và kiểm tra đường bộ tự động.

Từ khóa: Nhận dạng vết nứt mặt đường; Phân đoạn vết nứt; Mô-đun trích xuất biên (BEM); Kiểm tra bề mặt đường; Trích xuất đặc trưng đa tỉ lệ.