EA-YOLOv8: An Enhanced Crack Segmentation Algorithm for Infrastructure Maintenance

Feiyu Chen[1], Jinghui Qin[2,*], Jianyu Yin[3] and Xianfei Yin[4]

[1]Department of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou, China
[2]Department of Information Engineering, Guangdong University of Technology, Guangzhou, China
[3]Department of Civil and Environmental Engineering, University of Macau, Macau, China
[4]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China

flyversechen@gmail.com qinjinghui@gdut.edu.cn dc12714@um.edu.mo xianfyin@cityu.edu.hk

Abstract -
Automated crack detection plays a critical role in infrastructure maintenance and safety assessment. The detection and segmentation of concrete and pavement cracks present substantial technical challenges. These challenges stem from the inherent complexity of crack morphology, environmental lighting variations, and the textural similarities between cracks and their surrounding surfaces. In this paper, we introduce a novel segmentation algorithm named EA-YOLOv8 to address two critical gaps in current crack detection techniques. The first gap involves limited spatial information utilization. To solve this, we integrate the Efficient Local Attention (ELA) mechanism into the YOLOv8 architecture's Bottleneck structure. ELA enhances spatial feature extraction without compromising computational efficiency or channel information. This approach directly tackles the challenge of precise crack localization. The second gap relates to imprecise crack boundary detection. We develop an adaptive region-specific loss (ARS loss) that specifically targets the complexities of crack edge delineation. Traditional loss functions often fail to capture the nuanced boundaries of microscopic cracks. Experimental validation of EA-YOLOv8 was conducted on the Roboflow crack dataset. Through ablation studies and baseline comparisons with YOLOv5, YOLOv8, and the latest YOLOv11, the experimental results show the superior performance of EA-YOLOv8. In terms of results, EA-YOLOv8 improved precision by 10% compared to YOLOv5 and the latest YOLOv11, and increased precision by 5% over the original YOLOv8. These experimental findings validate the theoretical framework and practical applicability of our proposed EA-YOLOv8 in crack detection scenarios.

Keywords -

Crack segmentation; YOLOv8; Efficient local attention; Adaptive region-specific loss

1 Introduction

Infrastructure has experienced extensive environmental erosion and service load increases during recent decades. These factors significantly impact structural safety throughout the infrastructure lifecycle. Concrete and pavement cracks represent fundamental indicators of structural health conditions [1]. Precise and timely crack detection remains essential for infrastructure maintenance and public safety assurance.

The advancement of inspection technologies has transformed crack detection methodologies. Manual inspection methods maintain widespread application but exhibit limitations in labor requirements, time consumption, and subjective variations [2]. The integration of optical technologies with deep learning algorithms provides enhanced automated solutions for crack detection [1, 3]. While YOLO networks have shown potential in crack detection, existing methods still struggle with fundamental limitations [4, 5].

The primary challenges in current crack detection techniques include: Crack morphology presents extreme complexity. Cracks exhibit irregular patterns with dramatic variations in width and length, making feature extraction highly difficult [2, 6]. Most existing deep learning models fail to capture the nuanced characteristics of these structural irregularities.Background and crack region imbalance creates significant segmentation challenges. Current algorithms struggle to distinguish subtle crack boundaries from surrounding surface textures. This limitation leads to inaccurate detection, especially in regions with complex surface conditions. Boundary detection precision remains a critical weakness. Existing models typically cannot accurately delineate crack edges, particularly in areas with minimal contrast or complex surface textures.

To address these critical gaps, we developed EA-YOLOv8, an advanced segmentation algorithm with two key innovations. First, we integrated the Efficient Local Attention (ELA) mechanism to improve spatial feature recognition. ELA enables more precise crack localization by dynamically distributing attention weights across different image regions. This approach overcomes the traditional limitations in feature extraction. Second, we designed an Adaptive Region-Specific Loss (ARS loss) function specifically targeting crack boundary detection challenges. This innovative loss function enhances the model's ability to differentiate crack edges from background regions during training. By focusing on the most challenging detection areas, ARS loss significantly improves edge detection precision.

*Corresponding author

Experimental validation on the Roboflow crack dataset demonstrates the substantial improvements of EA-YOLOv8. Compared to existing methods, our approach increased precision by 10% over YOLOv5 and the latest YOLOv11, and improved performance by 5% compared to the original YOLOv8. These results validate the theoretical contributions and practical utility of our method in infrastructure crack detection.

2 Related Work

Advancements in deep learning has created substantial opportunities in complex detection tasks. Its application in crack detection demonstrates extensive implementation and profound impact. The field has witnessed numerous research achievements that establish the foundation for technological progression.

N.A.M. Yusof et al. [1] implemented convolutional neural network (CNN) methodologies for concrete crack classification. Their research validated the discriminative capabilities of deep learning in crack detection applications. Liu et al. [2] developed an enhanced deep convolutional neural network incorporating deep supervision mechanisms. This approach addressed segmentation accuracy limitations in complex environments and improved fine crack detail capture. Hassan Hosseinzadeh et al. [3, 7] utilized the YOLOv5 framework for pavement crack detection and segmentation. Their methodology achieved precise crack characterization while maintaining computational efficiency, addressing practical requirements in road maintenance applications. Similarly, Yin et al. [7, 8] implemented the YOLOv3 network for automated defect detection in municipal drainage systems. Their system detected six types of defects including cracks, achieving a mean average precision of 85.37%. This work demonstrated the practical application of deep learning in infrastructure maintenance, particularly for sewer pipe inspection where thousands of kilometers of CCTV footage require analysis. Some recent research has focused on minimizing the annotation burden for sewer defect detection [9, 10]. For example, a framework utilizing weakly supervised object localization for automated sewer defect detection was developed by Yin et al. [10], achieving high localization accuracy while requiring minimal annotation effort compared to traditional supervised approaches.

Current deep learning models exhibit specific limitations in crack detection applications. These limitations manifest in edge detection precision and fine-grained segmentation capabilities, particularly when processing irregular crack morphologies and complex background conditions. Crack characteristics, including width variations, morphological complexity, and environmental interactions, present significant challenges for accurate boundary delineation and feature extraction. Addressing annotation challenges in infrastructure inspection, Yin et al. [9] applied Weakly Supervised Object Localization (WSOL) for sewer defect detection using only image-level labels. Their approach achieved mean MaxBoxAccV2 scores of 64.33% with ResNet-50 backbone and established a benchmark for weakly supervised learning in infrastructure applications, eliminating the need for labor-intensive bounding box annotations while maintaining high classification accuracy.

Research efforts focus on architectural optimization strategies to address these technical limitations. Xiang et al. [7, 11] integrated Transformer modules into YOLOv5 architecture. The self-attention mechanism enhanced long-range dependency capture and contextual understanding within crack regions, expanding model detection capabilities. Dong et al. [4, 7] incorporated the Convolutional Block Attention Module (CBAM) into YOLOv5. This modification enabled adaptive attention distribution across feature dimensions, improving crack feature focus while reducing background interference. Liu [1] enhanced the DeepLabV3+ network through Dual Attention Module integration and HDD loss implementation. These modifications strengthened feature emphasis and optimized edge detection performance, advancing crack detection methodologies.

Contemporary research demonstrates that architectural enhancement through specialized modules and loss functions represents an effective approach for performance improvement. This study implements the Efficient Local Attention (ELA) mechanism [12] and Adaptive Region-Specific Loss (ARS loss) [11] within the YOLOv8 framework. Systematic ablation experiments validate the model's effectiveness. The research provides enhanced solutions for crack detection applications, contributing to technological advancement in this domain.

3 EA-YOLOv8

3.1 Overview

To improve the performance of crack segmentation, we introduce two architectural improvements to YOLOv8. The first modification integrates the Efficient Local Attention (ELA) mechanism into the C2F module. This integration creates the C2ELA module for crack feature extraction enhancement. The second modification implements Adaptive Region-Specific Loss (ARS loss) within the segmentation component. ARS loss operates in conjunction with BCE Loss for crack boundary localization. These improvements enhance the segmentation performance of the base architecture. Figure 1 depicts the architecture of the proposed EA-YOLOv8 framework.

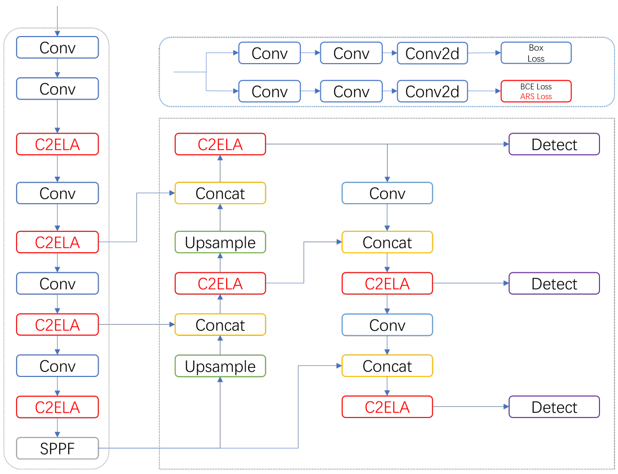


Figure 1. The network architecture of EA-YOLOv8

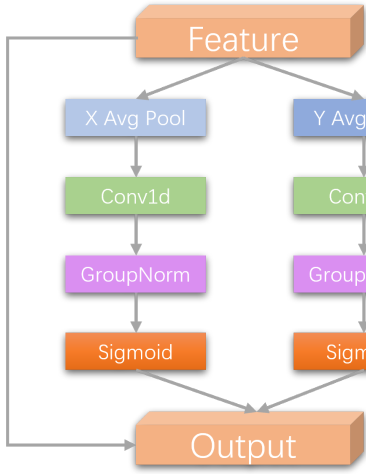


Figure 2. The structure of Efficient Local Attention.

3.2 Efficient Local Attention (ELA)

We introduce the operational mechanisms of ELA (Efficient Local Attention) for crack recognition in this section. Figure 2 illustrates the structural architecture of ELA. ELA utilizes strip pooling operations for coordinate information extraction. The pooling process extracts horizontal and vertical features from input feature maps. This directional processing method enables orientation-specific pattern recognition. Crack features demonstrate distinct directional properties. The coordinate extraction mechanism captures feature variations along crack extension directions. This process establishes detailed morphological analysis capabilities.

The architecture deploys 1D convolution for bidirectional positional enhancement. Unlike standard 2D convolutions that process pixels in all directions simultaneously, 1D convolution works like a signal filter that processes information along a single dimension (either horizontal or vertical). This approach is similar to how radar systems scan in specific directions to detect objects. 1D convolution has demonstrated computational efficiency advantages over 2D convolution operations, requiring fewer calculations while still capturing essential directional information. This convolution method integrates directional features and maintains spatial continuity information, helping the model "connect the dots" when detecting elongated crack structures.

ELA incorporates Group Normalization (GN) [13] for feature distribution stabilization. In simple terms, GN works like adjusting the contrast of different parts of an image separately to ensure all important details remain visible regardless of lighting conditions. Environmental factors in crack images introduce feature value variations. These factors include illumination conditions and background complexities. GN [13] executes group-based feature normalization procedures. This normalization method reduces environmental interference effects. The stabilized feature distributions support consistent pattern recognition during model training.

The integration of ELA into YOLOv8's architecture is implemented by replacing the C2f modules with our proposed C2ELA modules. The bottleneck structure in YOLOv8's C2f module performs feature extraction operations through convolutional processing and residual connections. Our implementation embeds the ELA mechanism within this Bottleneck structure according to the following steps:

- We initialize the model with pretrained YOLOv8 weights from Ultralytics.
- The bottleneck components within C2f are identified and modified to include ELA.
- Each bottleneck module processes the input through a 1×1 convolution to reduce channel dimensions.
- The ELA placement occurs after these convolutional layers but before the residual connection.
- The output from ELA is then passed through another 1×1 convolution to restore channel dimensions.
- Finally, a residual connection adds the original input to the processed features.

This architectural arrangement optimizes feature processing capabilities for crack detection. Figure3 presents the C2ELA module configuration. The structural integration combines ELA and Bottleneck functionalities, enhancing the model's ability to capture the distinctive direc-

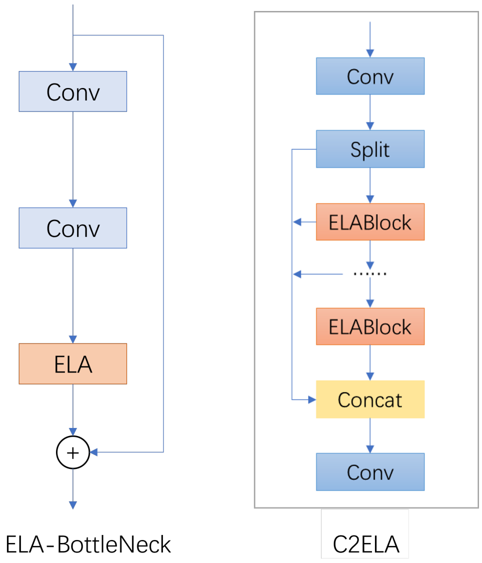


Figure 3. The structure of C2ELA.

tional patterns of cracks while maintaining computational efficiency.

3.3 Adaptive Region-Specific (ARS) Loss

We propose the Adaptive Region-Sensitive Loss (ARS-Loss) mechanism, which employs Tversky Loss [14] for sub-region computation. This approach surpasses traditional Dice Loss limitations through parameters α and β, which regulate false positive (FP) and false negative (FN) penalties. The formulation is expressed as:

$$L_{Tversky} = 1 - \frac{\sum_{i \in V} p_i \cdot g_i + \varepsilon}{y} \quad (1)$$

$$y = \sum_{i \in V} p_i \cdot g_i + \alpha \sum_{i \in V} p_i (1 - g_i) + \beta \sum_{i \in V} (1 - p_i) \cdot g_i + \varepsilon \quad (2)$$

The equation incorporates specific variables: p_i represents pixel-wise crack probability prediction, g_i denotes corresponding ground truth labels, V specifies the sub-region pixel set, and ε maintains numerical stability.

The core innovation of ARS-Loss lies in its adaptive parameters that evolve according to regional error distributions through:

$$\alpha_{\text{Adaptive}} = A + B \cdot \frac{FP}{FP + FN} \quad (3)$$

$$\beta_{\text{Adaptive}} = A + B \cdot \frac{FN}{FP + FN} \quad (4)$$

The ARS-Loss implementation follows a systematic procedure integrated with the YOLOv8 framework:

1) Forward Pass and Feature Extraction: Input images are processed through the YOLOv8 backbone and neck, generating feature maps at different scales.

2) Segmentation Head Processing: The segmentation head transforms the feature maps to produce prediction probability maps $P \in \mathbb{R}^{H \times W}$, where each pixel value $p_i \in [0, 1]$ represents the probability of belonging to the crack class.

3) Region Subdivision: Both prediction maps and ground truth labels are subdivided into a regular grid structure of 16×16×16 (width × height × channels) sub-regions. This subdivision is implemented as follows:

a. For an input of size $H \times W$, we compute region dimensions $r_h = H/16$ and $r_w = W/16$

b. Each sub-region $R_{m,n}$ covers pixels from indices $(m \cdot r_w, n \cdot r_h)$ to $((m + 1) \cdot r_w - 1, (n + 1) \cdot r_h - 1)$

c. The grid structure creates 256 sub-regions, enabling localized error analysis

4) Sub-region Error Analysis: For each sub-region $R_{m,n}$, we compute:

a. False Positives (FP): $\sum_{i \in R_{m,n}} p_i \cdot (1 - g_i)$

b. False Negatives (FN): $\sum_{i \in R_{m,n}} (1 - p_i) \cdot g_i$

c. True Positives (TP): $\sum_{i \in R_{m,n}} p_i \cdot g_i$

5) Adaptive Parameter Calculation: For each sub-region $R_{m,n}$, the adaptive parameters are calculated as:

a. $\alpha_{m,n} = 0.3 + 0.7 \cdot \frac{FP_{m,n}}{FP_{m,n} + FN_{m,n}}$

b. $\beta_{m,n} = 0.3 + 0.7 \cdot \frac{FN_{m,n}}{FP_{m,n} + FN_{m,n}}$

6) Sub-region Loss Computation: The Tversky loss for each sub-region is calculated using the region-specific adaptive parameters:

$L_{Tversky} = 1 - \frac{\sum_{i \in V} p_i \cdot g_i + \varepsilon}{y}$

$y = \sum_{i \in V} p_i \cdot g_i + \alpha \sum_{i \in V} p_i (1 - g_i) + \beta \sum_{i \in V} (1 - p_i) \cdot g_i + \varepsilon$

7) Total Loss Aggregation: The overall ARS-Loss is computed as the mean of all sub-region losses:

$L_{ARS} = \frac{1}{256} \sum_{m=0}^{15} \sum_{n=0}^{15} L_{m,n}$

8) Loss Integration with YOLOv8: The ARS-Loss is integrated with the standard YOLOv8 loss components:

$L_{total} = \lambda_{box} \cdot L_{box} + \lambda_{cls} \cdot L_{cls} + \lambda_{ARS} \cdot L_{ARS}$

where $\lambda_{box} = 0.05$, $\lambda_{cls} = 0.5$, and $\lambda_{ARS} = 1.0$ are weighting coefficients determined through ablation studies to balance the different loss components.

9) Backpropagation and Weight Update: During the backward pass, gradients are computed with respect to the total loss. The adaptive nature of ARS-Loss ensures that regions with higher FP receive stronger penalties for false positives (higher α), while regions with higher FN receive stronger penalties for false negatives (higher β). This region-sensitive adaptation guides the model to focus on challenging areas with imbalanced error distributions.

The implementation of ARS-Loss starts with the pre-trained YOLOv8 model (weights from Ultralytics) and modifies the loss computation module while maintaining the original network architecture. This approach requires minimal changes to the YOLOv8 framework, making it easily applicable as a drop-in enhancement for crack segmentation tasks.

The region-sensitive methodology enhances feature discrimination capacity and crack boundary detection precision by dynamically adjusting penalties based on local error characteristics. This is particularly effective for thin crack structures where traditional global loss functions often struggle to provide sufficient gradient signals.

4 Experiments

4.1 Dataset and Implementation details

The experimental dataset consists of the Roboflow Crack Segmentation Dataset. Roboflow developed this public dataset for traffic and infrastructure safety research applications. The dataset incorporates 4,029 static images from road and wall environments. Each image contains pixel-level annotations. The image collection represents multiple infrastructure conditions and crack characteristics. The dataset supports comprehensive safety analysis tasks.

The research implementation adopts Roboflow's original data organization. The dataset division follows the provided training, testing, and validation partitions. Table 1 specifies the image distribution across these subsets.

Table 1. The dataset splitting of Roboflow Crack Segmentation dataset.

Dataset	Roboflow Crack Seg
Train	3717
Validation	200
Test	112

All experiments are conducted with an NVIDIA GeForce RTX 4070 GPU (12GB VRAM). The model is implemented with PyTorch[1] 2.4.1 framework running on CUDA 12.6. Additional libraries include NumPy 1.26.3 for numerical operations and Ultralytics for YOLOv8 implementation. All experiments are conducted with Python 3.10.14. The model initialization incorporates YOLOv8s-seg pre-trained weights from Ultralytics. The training configuration is concluded in Table 2.

Table 2. Hyperparameters for model training.

Parameter		Parameter	
Epoch	150	Momentum	0.937
Optimizer	SGD	Warmup Epoch	3
lr0	0.01	Warmup Momentum	0.8
lrf	0.01	Warmup Bias lr	0.1

[1] https://pytorch.org/

Table 3. Performances of various models on the Roboflow dataset.

	Precision	Recall	mAP50	F1
YOLOv5	0.75592	0.68675	0.68545	0.71968
YOLOv8	0.80325	0.6506	0.67648	0.718911
YOLOv11	0.75672	**0.69478**	0.65088	0.72442
ARS-YOLOv8	0.80307	0.66265	0.69921	0.726134
ELA-YOLOv8	0.83489	0.6506	0.69087	0.731313
EA-YOLOv8	**0.85669**	0.65462	**0.70297**	**0.742146**

5 Results and Discussion

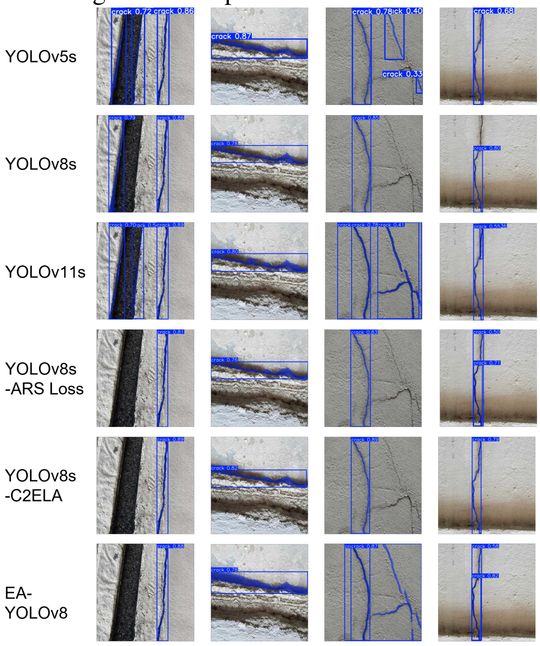
This investigation was executed utilizing the Roboflow dataset corpus. Our methodology encompasses a systematic evaluation between EA-YOLOv8 and preceding YOLO variants: YOLOv5, YOLOv8, YOLOv11, ELA-YOLOv8, and ARS-YOLOv8. The architectural design of ELA-YOLOv8 incorporates the ELA attention mechanism explicitly for feature extraction augmentation. The ARS-YOLOv8 framework implements the ARS loss function specifically for optimization procedures. Statistical metrics were acquired through methodological validation protocols. The empirical results, as documented in Table 3, elucidate the comparative analysis. The novel EA-YOLOv8 architecture, synthesizing the ELA attention mechanism and ARS loss function, manifests enhanced detection efficacy compared to singular-component implementations.

The experimental evaluation compares EA-YOLOv8 with the YOLOv5 baseline model. EA-YOLOv8 exhibits performance enhancements across evaluation metrics. YOLOv5 achieves baseline performance with 0.75592 precision and 0.68675 recall. YOLOv11, despite being the most recent architecture in the YOLO family, achieves 0.75672 precision and 0.69478 recall, showing only marginal improvement over YOLOv5 in crack detection tasks. YOLOv8 and its architectural variants demonstrate precision improvements. These results indicate YOLO architecture progression.

Precision defines the ratio of correctly predicted positive samples (True Positives, TP) among total positive predictions. EA-YOLOv8 achieves 0.85669 precision in crack identification. This result surpasses both YOLOv5 (0.75592), YOLOv11 (0.75672) and YOLOv8 (0.80325) baselines. The visualized results in Figure 4 column one show detection errors in baseline models. EA-YOLOv8 reduces false detections through two mechanisms. The ELA mechanism enhances directional feature extraction capabilities. The ARS mechanism enables regional attention optimization. These components reduce False Positive (FP) predictions.

Recall represents the proportion of actual positive samples correctly predicted as positive (TP). EA-YOLOv8 achieves 0.65462 recall. While this is lower than

Figure 4. Comparison of Each Model



YOLOv11's recall (0.69478), which demonstrates the highest recall among the baseline models, the third column in Figure 4 shows detection limitations in YOLOv8, ELA-YOLOv8, and ARS-YOLOv8. EA-YOLOv8 combines ARS loss regional weight adjustment with ELA feature extraction. This integration improves positive sample detection.

Mean Average Precision (mAP50) evaluates model performance in object detection tasks. EA-YOLOv8 achieves 0.70297 mAP50 ,which significantly outperforms YOLOv11's 0.65088 mAP50. The ELA mechanism extracts discriminative features. The ARS loss focuses on challenging regions. These components enhance detection accuracy across scales.

F1 score provides the harmonic mean of precision and recall. EA-YOLOv8 achieves 0.742146 F1 score , surpassing YOLOv11's 0.724428 F1 score. This metric demonstrates the effectiveness of ELA and ARS loss integration.

This research presents EA-YOLOv8 for crack detection applications. The model integrates ELA attention mechanism and ARS loss function. Ablation experiments on the Roboflow dataset validate model effectiveness. EA-YOLOv8 demonstrates performance improvements over YOLOv5, YOLOv11 and YOLOv8 baselines. The model surpasses ELA-YOLOv8 and ARS-YOLOv8 variants in precision, recall, mAP50, and F1 metrics.

6 Conclusion and Limitation

In this paper, we introduce a novel segmentation algorithm named EA-YOLOv8 to incorporate two technical innovations into the YOLOv8 architecture. First, we deploy the Efficient Local Attention (ELA) mechanism following the final convolutional layer within the Bottleneck structure of the C2f module. ELA can optimize spatial information utilization while preserving channel dimensions and computational efficiency, resulting in improved crack localization capabilities. Second, we develop an adaptive region-specific loss (ARS loss) as an auxiliary loss to specifically target the difficulties in crack boundary delineation, leading to refined edge detection results. Experimental validation of EA-YOLOv8 was conducted on the Roboflow crack dataset. Through comprehensive ablation studies and baseline comparisons with YOLOv5, YOLOv8, and the latest YOLOv11, the experimental results demonstrate the EA-YOLOv8's superior performance. Quantitatively, EA-YOLOv8 exhibits a 10% precision improvement over YOLOv5, a 10% improvement over YOLOv11 and a 5% performance gain compared to the original YOLOv8. While YOLOv11 demonstrates slightly higher recall than EA-YOLOv8, our model achieves superior precision, mAP50, and F1 scores, indicating better overall performance for crack detection tasks. These experimental results validate both the theoretical framework and practical applicability of our proposed EA-YOLOv8 in crack detection scenarios.

Research limitations require further investigation. EA-YOLOv8 recall performance remains below ARS-YOLOv8 levels. This indicates potential feature extraction misalignment between ELA and ARS loss components. ARS loss weight adjustment mechanisms may experience constraints from ELA-generated features. These interactions affect regional learning processes. Future research directions include attention mechanism compatibility analysis with ARS loss. Additional investigation of YOLOv8 detection head, feature pyramid, and SPPF module modifications may enhance crack segmentation performance. Furthermore, this study utilized only a single dataset for training and testing. Future work should incorporate multiple diverse datasets for cross-validation to better verify the model's reliability and generalizability across different crack detection scenarios.

References

[1] NAM Yusof, MK Osman, MHM Noor, A Ibrahim, NM Tahir, and NM Yusof. "Crack detection and classification in asphalt pavement images using deep convolution neural network". In: *2018 8th IEEE international conference on control system, computing and engineering (ICCSCE)*. IEEE. 2018, pp. 227–232.

[2] Yahui Liu, Jian Yao, Xiaohu Lu, Renping Xie, and Li Li. "DeepCrack: A deep hierarchical feature

learning architecture for crack segmentation". In: *Neurocomputing* 338 (2019), pp. 139–153.

[3] Hassan Hosseinzadeh, Ali Ghiami Bajgirani, and Mohadeseh Delavarian. "Detection and classification of asphalt pavement cracks using YOLOv5". In: *Journal of Modeling in Engineering* 21.72 (2023), pp. 175–193.

[4] Zhaohui Dong, Guijie Zhu, Zhun Fan, Jiacheng Liu, Huanlin Li, Yuwei Cai, Huaxing Huang, Ze Shi, Weibo Ning, and Liu Wang. "Automatic Pavement Crack Detection Based on YOLOv5-AH". In: *2022 12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE. 2022, pp. 426–431.

[5] Xuezhi Xiang, Zhiyuan Wang, and Yulong Qiao. "An improved YOLOv5 crack detection method combined with transformer". In: *IEEE Sensors Journal* 22.14 (2022), pp. 14328–14335.

[6] Jie Liu. "Road Crack Detection Using HDD LOSS and Dual Attention Module with DeepLabv3+". In: *2023 3rd International Conference on Digital Society and Intelligent Systems (DSInS)*. IEEE. 2023, pp. 148–152.

[7] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. "YOLO by Ultralytics, January 2023". In: *URL https://github. com/ultralytics/ultralytics* 2 ().

[8] Xianfei Yin, Yuan Chen, Ahmed Bouferguene, Hamid Zaman, Mohamed Al-Hussein, and Luke Kurach. "A deep learning-based framework for an automated defect detection system for sewer pipes". In: *Automation in construction* 109 (2020), p. 102967.

[9] Jianyu Yin, Xianfei Yin, Yifeng Sun, and Mi Pan. "Bridging the Annotation Gap: Innovating Sewer Defects Detection with Weakly Supervised Object Localization". In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. Vol. 41. IAARC Publications. 2024, pp. 669–674.

[10] Jianyu Yin, Xianfei Yin, Mi Pan, and Long Li. "Scalable and transparent automated sewer defect detection using weakly supervised object localization". In: *Automation in Construction* 174 (2025), p. 106152.

[11] Yizheng Chen, Lequan Yu, Jen-Yeu Wang, Neil Panjwani, Jean-Pierre Obeid, Wu Liu, Lianli Liu, Nataliya Kovalchuk, Michael Francis Gensheimer, Lucas Kas Vitzthum, et al. "Adaptive region-specific loss for improved medical image segmentation". In: *IEEE transactions on pattern analysis and machine intelligence* 45.11 (2023), pp. 13408–13421.

[12] Wei Xu and Yi Wan. "ELA: Efficient local attention for deep convolutional neural networks". In: *arXiv preprint arXiv:2403.01123* (2024).

[13] Yuxin Wu and Kaiming He. "Group normalization (2018)". In: *URL https://arxiv. org/abs* (1803).

[14] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. "Tversky loss function for image segmentation using 3D fully convolutional deep networks". In: *International workshop on machine learning in medical imaging*. Springer. 2017, pp. 379–387.