An Implementation of Deep Convolution Segmentation for Crack Detection in Concrete

Faqih Ma'arif
Department of Civil Engineering, Universitas Negeri Yogyakarta, Sleman, Indonesia
faqih_maarif07@uny.ac.id (corresponding author)

Han Ay Lie
Department of Civil Engineering, Diponegoro University, Semarang, Indonesia
hanaylie@live.undip.ac.id

Slamet Widodo
Department of Civil Engineering, Universitas Negeri Yogyakarta, Sleman, Indonesia
swidodo@uny.ac.id

Zhengguo Gao
School of Transportation Science and Engineering, Beihang University, Beijing, China
gaozg@buaa.edu.cn

Fardiansyah Nur Aziz
Department of Electronics and Informatics Engineering, Universitas Negeri Yogyakarta, Sleman, Indonesia
fardiansaziz@gmail.com

Maris Setyo Nugroho
Department of Civil Engineering, Universitas Negeri Yogyakarta, Sleman, Indonesia
marissetyo@uny.ac.id

Received: 16 October 2025 | Revised: 17 November 2025 and 5 December 2025 | Accepted: 7 December 2025

 DOI: https://doi.org/10.48084/etasr.15569

ABSTRACT

Crack detection in concrete structures is crucial for maintaining safety and preventing structural damage. Traditional manual inspection methods have limitations in terms of efficiency, objectivity, and coverage on large-scale infrastructure. This study proposes an automated crack segmentation approach using a fully convolutional network with a modified architecture from the Visual Geometry Group. This method applies hierarchical feature learning over multiple deep-trained side outputs and addresses class imbalance with a balanced cross-entropy-based loss function. The training dataset consists of 300 real images representing road surfaces, walls, and concrete structures with real cracks, and 300 binary images as ground truth. Evaluations were conducted on various concrete block shapes, including square, T-shaped, and cylindrical forms. The third side output demonstrated the best performance with an F1-score of 83.3%, a precision of 77.0%, and a recall of 90.8%. The linear fusion strategy can effectively integrate multi-level features, resulting in an average Intersection over Union of 80.4%. The proposed model shows significant improvement over previous methods and can recognize crack patterns across various scales and structural shapes. These results confirm the potential of the proposed approach as a solid basis for automated infrastructure inspection systems.

Keywords-crack detection; deep learning; concrete; segmentation; convolutional neural network; monitoring

I. INTRODUCTION

Cracks in building structural elements due to excessive loads cannot be considered merely surface damage, but rather important indicators of the structure's integrity, safety, and long-term durability. In civil engineering, cracking is a precursor to structural degradation or failure if not promptly addressed [1]. Given the necessity for efficiency, accuracy, and speed in the inspection of large-scale structures, such as bridges, high-rise buildings, dams, and highways, automatic crack detection systems based on digital technology are required. Computer vision, through digital image analysis [2], is the primary solution to this issue and is generally divided into local and global feature-based methods. Authors in [3] applied advanced Convolutional Neural Network (CNN) architectures to recognize road surface anomalies, demonstrating that deep convolutional networks can effectively identify cracks and irregularities in pavement surfaces, and confirming the potential of deep learning for automated infrastructure inspection. However, variations in crack patterns and uneven lighting conditions still pose significant challenges in crack segmentation.

Various approaches have been developed to overcome these limitations. CNNs have demonstrated good performance in object detection and image segmentation [4], but still face challenges such as noise, shadows, and low-resolution images. Other methods, such as NDHM [5], crack tree [6], saliency [7], and Black Top Hat (BTH) [8], have been proposed, but they still fail to address crack intensity distribution, uneven lighting, or detect fine cracks. Several developments, such as ROA in BTH, are also ineffective for very small cracks. CNNs remain an option due to their ability to extract features and generate predictions through superpixels and labeling [9], but they are still limited in anticipating small-scale segmentation errors. Conditional Random Fields (CRFs) have been utilized to improve low-scale crack segmentation [10], and CNN–CRF integration has been applied to improve predictive performance on large datasets [11]. However, high-precision structural crack detection has yet to be optimally achieved. Several approaches, such as the HED architecture [12] and DeepCrack, have also been employed to detect cracks with varying pixel scales and depths.

However, there is still a research gap in accurate and reliable structural crack segmentation, especially for complex structural shapes and geometries. There has been no approach that systematically combines CNNs, CRFs, and deep supervision strategies to address pixel-based crack detection in concrete structures with diverse geometries.

The present study develops a CNN-based crack segmentation method enhanced by a Deeply Supervised Nets (DSN) strategy and CRF integration. The DeepCrack architecture is utilized to detect crack variations based on depth and scale, while CRF is deployed as a refinement for the final segmentation stage. This study uses open datasets and laboratory test images on rectangular beams, T-beams, and concrete cylinders to evaluate the method's performance. The main contribution of this study is the improvement of structural crack segmentation accuracy and the provision of a hierarchical training approach to support automated and more reliable computer vision-based infrastructure inspection.

II. METHODOLOGY

A. CNN Architecture

A deep learning architecture is used for pixel-based crack segmentation, where each pixel is classified as either a crack or a non-crack [13]. This architecture is built on the VGG-16 network, which acts as the primary feature extractor due to its proven capabilities in visual recognition tasks [14]. The side outputs are generated from five convolution blocks, following a DSN strategy to ensure that each level of the network is well-trained [15]. Each side output is then upsampled using a deconvolution layer [16] and combined to form the final feature representation called the fusion output.

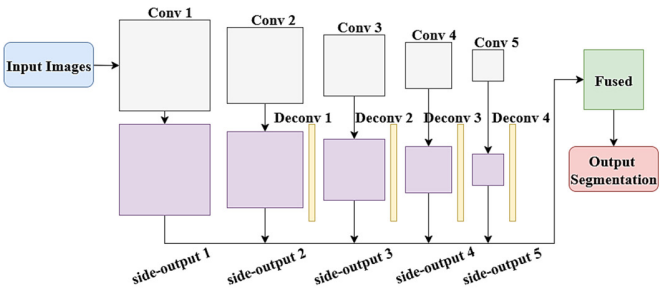

Fig. 1. Illustration of the research architecture.

B. Dataset and Data Preparation Process

The DeepCrack dataset was utilized for training and validation in this study because it provides a well-established benchmark containing manually annotated pixel-wise ground truth images. Manual annotation is the most reliable approach for generating accurate segmentation labels in crack detection research [17], although it is a time-consuming and resource-intensive process. The original images in the DeepCrack dataset [18] have non-uniform resolutions; therefore, all data were resized to 544 × 384 pixels according to the input requirements of the DeepCrack architecture.

In addition to the benchmark dataset, an experimental dataset was used in this study, which is publicly available in [19], consisting of 300 real images of concrete structures with visible cracks collected in the Civil Engineering Laboratory at Yogyakarta State University. These experimental images were used exclusively for testing purposes to evaluate the generalization capability of the trained model on unseen real-world crack patterns. During data collection, square beams, T-beams, and cylindrical concrete specimens were photographed under varying lighting and geometric conditions to simulate realistic inspection scenarios. The combination of both benchmark and experimental datasets allows the model to be trained under standardized conditions while being evaluated on real-world images, ensuring comprehensive performance validation across different environments.

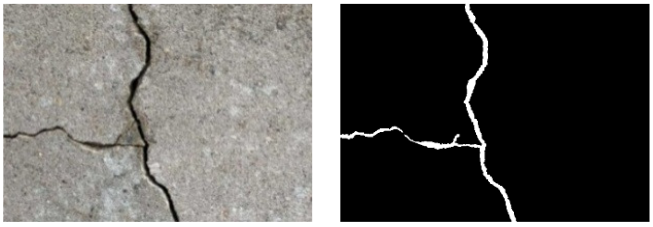

Fig. 2. Representative samples from the image dataset.

Pre-processing is performed to ensure data quality and consistency before use in training. Pixel values are normalized to the range (0, 1) by dividing each pixel value by 255 to accelerate model convergence and prevent the dominance of certain features due to scale differences [20]. Mask images are also ensured to have a proper binary threshold, so that each pixel has a pure binary value. The quality of the ground truth is strictly maintained, considering that deep learning models are very sensitive to inconsistent labels or erroneous annotations, which can lead to underfitting or overfitting to the wrong pattern [21]. Image augmentation techniques were applied to increase the diversity of the training samples. This strategy not only expands the visual diversity in the dataset but also improves the model's robustness to variations in lighting and crack orientation in real-world conditions [22]. Crack-specific augmentation is considered important because crack shape and direction are strongly influenced by material and environmental loads. With the right combination of pre-processing and augmentation, this dataset produces good generalization despite the limited number of images.

C. Loss Function Formulation

The selection of the loss function is carried out to guide the network weight optimization process towards an optimal solution. One of the main challenges in crack segmentation datasets is the phenomenon of class imbalance, where the number of background pixels (majority class) far exceeds the number of cracked pixels (minority class). The use of standard loss functions in this condition causes the model to tend to be biased towards the majority class and results in low detection performance in the minority class. To overcome this problem, the class-balanced cross-entropy function has been applied as formulated in (1). This mechanism effectively reduces model bias by giving a higher penalty weight to misclassifications in the minority class, thus forcing the model to focus more on detecting cracked pixels [23]:

$$L_{side}(I,G,W,w) = \sum_{m=1}^{M} \alpha_m \ell_{side}\left(I,G,W,w^{(m)}\right) = \sum_{m=1}^{M} \alpha_m \Delta\left(P^{(m)},G,W,w^{(m)}\right) \quad (1)$$

To ensure an effective feature learning process across all network layers, a DSN strategy has been implemented. In this approach, the loss function is not only calculated on the final output, but also individually generated by the convolution block on each side output. Thus, the total loss function (L) optimized by the model is a combination of the combined loss of all side outputs (L_{side}) and the loss of the final fusion output (L_{fuse}), as formulated in (2) and (3). This approach has been shown to improve gradient flow and promote model convergence towards a more reliable solution [24]:

$$L_{fuse}(I,G,W) = -\sum_{j\in G_+} w_0 \, log \, Pr\left(P_j = 1|I,W\right) = -\sum_{j\in G_-} w_1 \, log \, Pr\left(P_j = 0|I,W\right) \quad (2)$$

$$L = L_{side} + L_{fuse} \quad (3)$$

D. Training and Validation Process

The training process was carried out end-to-end using the DeepCrack architecture according to the predetermined configuration. Model performance evaluation was carried out using the metrics Global Accuracy (G), Class Average Accuracy (C), Mean Intersection over Union (I/U), Precision (P), Recall (R), and F1-score (F). To evaluate the balance between sensitivity and specificity, Receiver Operating Characteristic (ROC) and Area Under Curve (AUC) analysis were also conducted. The G, C, and I/U values for all classes are calculated using (4-6). The variable n_{ij} is defined as the number of pixels originating from class i but classified as class j, with a total of n_{cls} classes. Meanwhile, $t_i = \sum_j n_{ij}$ represents the total number of pixels that are members of class i (including true positives and false positives):

$$G = \frac{\sum_i n_{ii}}{\sum_i t_i} \quad (4)$$

$$C = \left(\frac{1}{n_{cls}}\right) \frac{\sum_i n_{ii}}{t_i} \quad (5)$$

$$I/U = \left(\frac{1}{n_{cls}}\right) \frac{\sum_i n_{ii}}{\left(t_i + \sum_j n_{ij} - n_{ii}\right)} \quad (6)$$

In addition to the evaluation using semantic segmentation metrics, three common metrics in the field of crack detection are also calculated, namely Precision, Recall, and F1-score:

$$\text{Precision}(P) = \frac{\#True\ Positives}{\#True\ Positives + \#False\ Positives} \quad (7)$$

$$\text{Recall}(R) = \frac{\#True\ Positives}{\#True\ Positives + \#False\ Negatives} \quad (8)$$

$$\text{F1} - \text{Score}(F) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

Since the Precision-Recall (P-R) metric does not consider the number of True Negatives, the classic ROC metric is used as a complementary evaluation tool. Method performance is calculated based on individual scores. For the ROC curve, three metrics are evaluated: True Positive Rate (TPR), False Positive Rate (FPR), and AUC.

$$\text{TPR} = \frac{\#True\ Positives}{\#True\ Positives + \#FalseP\ ositives} \quad (10)$$

$$\text{FPR} = \frac{\#False\ Positives}{\#False\ Positives + \#True\ Negatives} \quad (11)$$

The best model is selected based on the highest I/U value on the validation data, and the model weights are saved for use in the final testing phase. This model checkpointing strategy is commonly utilized in deep neural network training to prevent performance degradation due to overtraining.

III. RESULTS AND DISCUSSION

A. System Performance

Figure 3 shows the P-R curves of the nine crack segmentation methods. Side-output 3 achieves the highest F1-score of 0.833, followed by Side-output 2 (0.821) and Side-

output 1 (0.801). All methods maintain precision above 0.8 down to recalls of 0.4–0.5. DeepCrack-Basic, a combination of side-outputs 1 to 5, records an F1-score of 0.798 and ranks fourth. Although DeepCrack-Basic combines multiple side-outputs, its performance does not outperform the best side-output individually, indicating the need to optimize the fusion strategy. All side-output-based methods and DeepCrack-Basic exhibit a flatter curve in the low recall region, indicating high precision stability. In contrast, conventional methods, such as AutoCrack (F = 0.534) and RoadCNN (F = 0.357), exhibit lower performance with steeper curves. HED records the lowest F1-score among neural network-based methods (F = 0.625), while SegNet achieves moderate results (F = 0.756). The area under the PR curve indicates that side-outputs 2 and 3 provide the best balance between precision and recall. Evaluation on test data shows that DeepCrack significantly outperforms conventional methods and legacy CNN models.

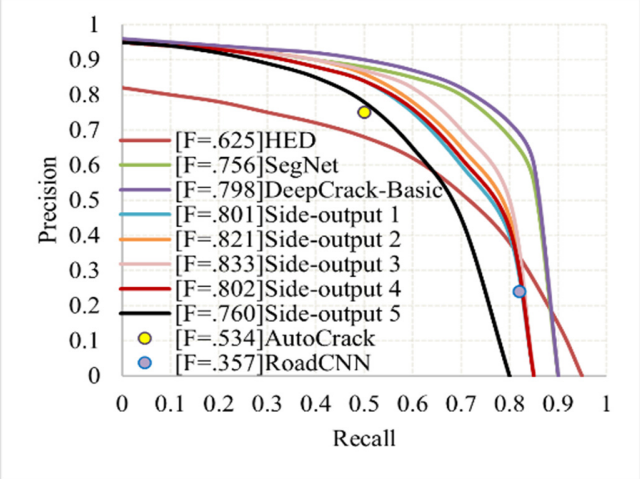

Fig. 3. PR curve for crack segmentation.

Figure 4 illustrates the ROC curves of the three crack segmentation methods. DeepCrack-Basic performed best with the highest AUC of 0.966, followed by HED (0.961) and SegNet (0.890). DeepCrack-Basic and HED have similar curves with a rapid increase in TPR at low FPRs, reaching a TPR of around 0.95 when the FPR is still below 0.1. SegNet exhibits a more gradual curve, with a more gradual increase in TPR at FPRs of 0-0.5. The most striking difference is seen at low FPRs (0-0.2), where DeepCrack-Basic and HED significantly outperform SegNet. All methods show convergence towards the (1, 1) point, reflecting the general characteristics of good classifiers. High AUC values of DeepCrack-Basic and HED (>0.96) indicate excellent discriminatory ability, while SegNet remains competitive, albeit by a smaller margin.

B. Side-Output Segmentation Analysis

Side-output segmentation evaluation was performed on three types of concrete structures: square, T-beam, and cylindrical, as shown in Figures 5-7. In square concrete, as depicted in Figure 5, side-output 1 captured fine crack details but was sensitive to noise, while side-outputs 2 and 3 provided the best balance between detail and clarity.

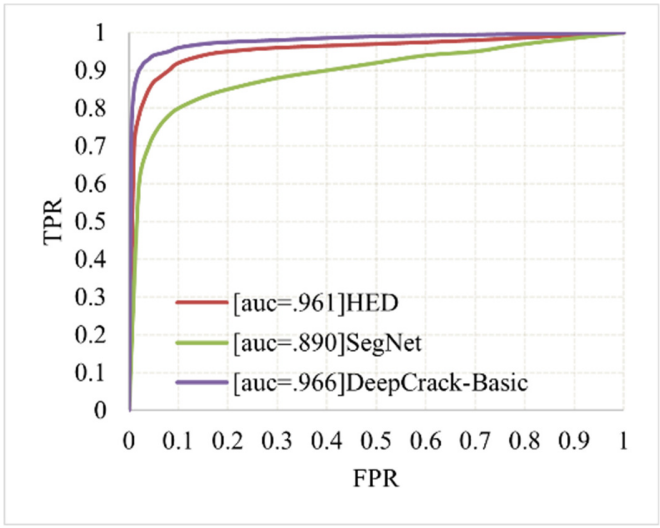

Fig. 4. ROC curve for crack segmentation.

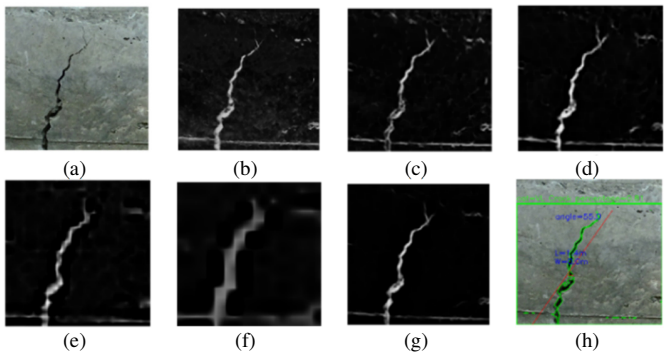

Fig. 5. Crack segmentation results in square concrete: (a) original image, (b)-(f) side-output 1-5, respectively, (g) linear fusion, (h) crack measurement.

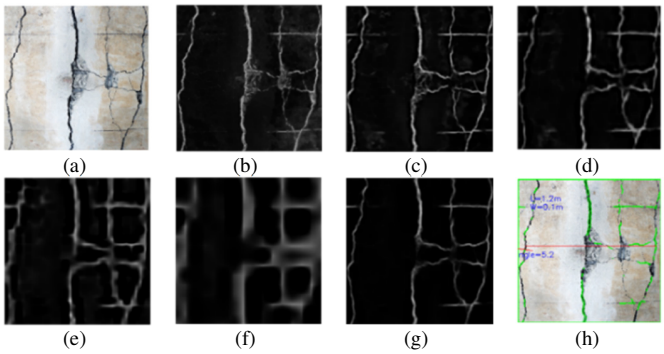

Fig. 6. Crack segmentation results on T-beam concrete: (a) original image; (b)-(f) side-output 1-5, respectively; (g) linear fusion; (h) crack measurement.

T-beam, as displayed in Figure 6, had higher geometric complexity. Side-outputs 2 and 3 maintained detection consistency under varying illumination, while deeper layers (4 and 5) exhibited shape resilience with lower spatial precision. In cylindrical-shaped concrete, presented in Figure 7, curvature caused intensity distortion, but side-outputs 2 and 3 were still able to distinguish cracks from geometric artifacts. Overall,

linear fusion results produced the most accurate and stable segmentation by combining details from shallow layers and context from deep layers, demonstrating its effectiveness across a wide range of structural shapes.

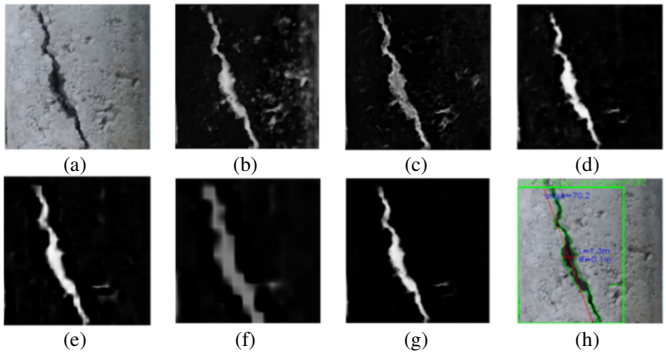

Fig. 7. Crack segmentation results in cylindrical concrete: (a) original image, (b)-(f) side-output 1-5, respectively, (g) linear fusion, (h) crack measurement.

These results demonstrate that the geometry of concrete structures significantly influences the characteristics and accuracy of crack segmentation. For square concrete, the model performed best because the flat surface facilitates visual feature extraction. T-beam concrete introduced additional challenges such as uneven lighting and crack concentration at corners, but the model still demonstrated good adaptability, particularly in the intermediate layers. For cylindrical concrete, successful segmentation demonstrated that abstract features from the inner layers are robust to geometric distortion, which is important for real-world applications. The hierarchical features of learning and linear fusion strategies proved effective in handling varying structural complexities. Side-outputs 2 and 3 consistently provided the best detection across all concrete types, supporting the notion that the intermediate layers of the network play a crucial role in combining local sensitivity and robustness to noise. These results provide new insights, according to which, side-output-based segmentation architectures can not only handle variations in surface texture but also adapt to complex geometric shapes, making them suitable for various types of infrastructure inspection.

C. Statistical Analysis of Side-Output Performance in Crack Segmentation

Table I presents the quantitative evaluation results of each side-output in the hierarchical feature learning architecture. In the semantic segmentation metric, side-output 3 achieved the highest performance with a global accuracy of 97.5%, a class average accuracy of 96.1%, and an *I/U* of 80.2%. Side-output 1 achieved the lowest global accuracy (95.3%) and a *I/U* of 72.8%. Performance consistently improved from side-outputs 1-3 but declined in side-outputs 4 and 5, with *I/U* of 77.6% and 76.2%, respectively. In the crack-specific detection metric, side-output 3 demonstrated the highest recall (90.8%), with a precision of 77.0%, resulting in the best F1-score of 83.3%. Side-outputs 1 and 2 demonstrated higher precision (76.8% and 77.8%), but lower recall (83.6% and 86.9%). Side output 5 showed the lowest precision (69.3%) but maintained a recall of 84.2%. The F1-score decreased for side-outputs 4 (80.2%) and 5 (76.0%). The linear fusion results recorded the highest *I/U* (80.4%), slightly exceeding side-output 3, but with a lower-class average accuracy (91.5%). Precision and recall were 78.5% and 79.8%, respectively, resulting in an F1-score of 79.8%.

TABLE I. RESULT OF SIDE OUTPUTS

Outputs	Metrics (%)					
	G	***C***	***I/U***	***P***	***R***	***F***
Side-output 1	95.3	95.5	72.8	76.8	83.6	80.1
Side-output 2	97	95.6	77.5	77.8	86.9	82.1
Side-output 3	97.5	96.1	80.2	77	90.8	83.3
Side-output 4	97	96.1	77.6	74.6	86.6	80.2
Side-output 5	96.6	95.2	76.2	69.3	84.2	76.0
Fused results	97.5	91.5	80.4	78.5	81.2	79.8

These results support the hypothesis that the intermediate layers of a convolutional network provide an optimal balance between spatial resolution and context understanding. The third-order side-output consistently performed best on both semantic and detection metrics, confirming the effectiveness of features from intermediate depths. The performance increases from shallow to intermediate layers, followed by a decrease in deeper layers, indicating a trade-off between feature abstraction and spatial precision. The high recall value of the third-order side-output indicates its ability to capture more crack pixels, which is important in structural monitoring because false detections (missed cracks) are more critical than false positives. The linear fusion strategy also proved to balance the contributions of each feature level, resulting in more stable and reliable predictions for field applications.

D. Performance on Benchmark and Experimental Datasets

To ensure an objective evaluation, the proposed model is compared with a baseline variant of DeepCrack–Basic [18] that uses the core architecture without additional CRF modules or feature fusion. The DeepCrack–Basic model achieves an F1-score of 79.8% and an *IoU* of 80.1% on the benchmark DeepCrack dataset. The proposed model demonstrates comparable performance on this dataset, confirming that the training and implementation processes have been carried out correctly and are consistent with the benchmark study. Table II presents the comparative performance of the proposed model.

TABLE II. COMPARATIVE PERFORMANCE OF PROPOSED MODEL

Dataset	Method	*F* (%)	*I/U* (%)	Source
DeepCrack (Benchmark)	DeepCrack–Basic	79.8	80.1	[18]
DeepCrack (Benchmark)	Proposed model	Comparable	Comparable	This study
Laboratory (Experiment)	Proposed model	83.3	80.4	This study

a. "Comparable" indicates that the performance of the proposed model on the DeepCrack benchmark is similar to the baseline DeepCrack–Basic variant, as no separate benchmark test split was performed in this study.

Next, the model is tested using a laboratory dataset consisting of 300 concrete images with various lighting conditions and geometries. On this dataset, the model achieves an F1-score of 83.3% and an *IoU* of 80.4%, demonstrating

strong generalization capabilities to previously unseen crack patterns. These results demonstrate that, despite its simpler configuration compared to advanced DeepCrack variants (e.g., those using CRF or additional augmentations), the proposed model maintains competitive and stable performance on both the benchmark and laboratory datasets.

Following this dataset-based evaluation, the results of the present study were compared with several previous deep learning-based approaches used for crack segmentation in concrete. The proposed model demonstrates improved performance compared to the HED method, SegNet, and the baseline DeepCrack–Basic architecture. Specifically, the former achieved an F1-score of 83.3%, which represents an improvement over HED (62.5%), SegNet (75.6%), and DeepCrack–Basic (79.8%). This improvement highlights the effectiveness of the class-balanced cross-entropy loss function and the linear fusion strategy in enhancing the balance between precision and sensitivity. Furthermore, the developed model excels in segmenting complex concrete geometries, such as T-beams and cylindrical specimens, which have not been extensively addressed in previous studies. These results demonstrate that the introduced approach provides better generalization to geometric distortions and lighting variations, making it more applicable to real-world infrastructure inspection systems. Overall, this research not only contributes to improving detection accuracy but also expands the applicability of CNN-based models for more complex structural health monitoring scenarios.

IV. CONCLUSIONS

This study proposes a novel approach for crack segmentation in concrete structures by integrating hierarchical feature learning, a class-balanced cross-entropy loss function, and linear feature fusion strategies to improve detection accuracy on complex structural shapes. The developed model successfully maintains high precision and good sensitivity on non-planar surfaces, such as T-beams and cylinders, with the best results recorded on the 3rd side output (F1-score 83.3%, precision 77.0%, and recall 90.8%). Comparative evaluations demonstrate superior performance compared to HED, SegNet, and DeepCrack–Basic variants, with a Mean Intersection over Union (*I/U*) of 80.4% and strong generalization capability across various beam types. The main contribution of this study lies in the combination of linear fusion strategies and class-balanced loss functions that effectively address class cohesion and improve segmentation stability. Although the limited dataset size may affect generalization, further research is proposed to adopt advanced data augmentation, attention mechanisms, and edge computing-based real-time re-monitoring systems. This study makes an important contribution to the development of a reliable and efficient computer vision-based automatic detection system for monitoring health structures.

ACKNOWLEDGMENT

The authors would like to thank the Directorate of Research and Community Service, Universitas Negeri Yogyakarta, for providing financial support for this research under contract number T/697/UN34.9/PT.01.03/2025.

REFERENCES

[1] T. Hong, M. J. Chae, D. Kim, C. Koo, K. S. Lee, and K. H. Chin, "Infrastructure Asset Management System for Bridge Projects in South Korea," *KSCE Journal of Civil Engineering*, vol. 17, no. 7, pp. 1551–1561, Nov. 2013, https://doi.org/10.1007/s12205-013-0408-8.

[2] D. Ai, G. Jiang, S.-K. Lam, P. He, and C. Li, "Computer Vision Framework for Crack Detection of Civil Infrastructure—a Review," *Engineering Applications of Artificial Intelligence*, vol. 117, Jan. 2023, Art. no. 105478, https://doi.org/10.1016/j.engappai.2022.105478.

[3] D. Doan Van, "Application of Advanced Deep Convolutional Neural Networks for the Recognition of Road Surface Anomalies," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10765–10768, Jun. 2023, https://doi.org/10.48084/etasr.5890.

[4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018, https://doi.org/10.1109/TPAMI.2017.2699184.

[5] Q. Li and X. Liu, "Novel Approach to Pavement Image Segmentation Based on Neighboring Difference Histogram Method," in *2008 Congress on Image and Signal Processing*, Sanya, China, May. 2008, pp. 792–796, https://doi.org/10.1109/CISP.2008.13.

[6] Q. Zou, Y. Cao, Q. Li, Q. Mao, and S. Wang, "CrackTree: Automatic Crack Detection From Pavement Images," *Pattern Recognition Letters*, vol. 33, no. 3, pp. 227–238, Feb. 2012, https://doi.org/10.1016/j.patrec.2011.11.004.

[7] W. Xu, Z. Tang, J. Zhou, and J. Ding, "Pavement Crack Detection Based on Saliency and Statistical Features," in *2013 IEEE International Conference on Image Processing*, Melbourne, VIC, Australia, Sept. 2013, pp. 4093–4097, https://doi.org/10.1109/ICIP.2013.6738843.

[8] W. Zhang, Z. Zhang, D. Qi, and Y. Liu, "Automatic Crack Detection and Classification Method for Subway Tunnel Safety Monitoring," *Sensors*, vol. 14, no. 10, pp. 19307–19328, Oct. 2014, https://doi.org/10.3390/s141019307.

[9] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning Hierarchical Features for Scene Labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013, https://doi.org/10.1109/TPAMI.2012.231.

[10] L. Zhou, K. Fu, Z. Liu, F. Zhang, Z. Yin, and J. Zheng, "Superpixel Based Continuous Conditional Random Field Neural Network for Semantic Segmentation," *Neurocomputing*, vol. 340, pp. 196–210, May 2019, https://doi.org/10.1016/j.neucom.2019.01.016.

[11] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road Crack Detection Using Deep Convolutional Neural Network," in *2016 IEEE International Conference on Image Processing*, Phoenix, AZ, USA, Sept. 2016, pp. 3708–3712, https://doi.org/10.1109/ICIP.2016.7533052.

[12] S. Xie and Z. Tu, "Holistically-Nested Edge Detection," in *2015 IEEE International Conference on Computer Vision*, Santiago, Chile, Dec. 2015, pp. 1395–1403, https://doi.org/10.1109/ICCV.2015.164.

[13] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, https://doi.org/10.1109/TPAMI.2016.2644615.

[14] Z. Xu, Q. Zhang, F. Hao, Z. Ren, Y. Kang, and J. Cheng, "VGG-CAE: Unsupervised Visual Place Recognition Using VGG16-Based Convolutional Autoencoder," in *Pattern Recognition and Computer Vision*, vol. 13020, H. Ma, L. Wang, C. Zhang, F. Wu, T. Tan, Y. Wang, J. Lai, and Y. Zhao, Eds. Cham, Switzerland: Springer International Publishing, 2021, pp. 91–102.

[15] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-Supervised Nets," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, San Diego, CA, USA, May 2014, vol. 38, pp. 562–570, https://doi.org/10.48550/arXiv.1409.5185.

[16] H. Noh, S. Hong, and B. Han, "Learning Deconvolution Network for Semantic Segmentation," in *2015 IEEE International Conference on*

Computer Vision, Santiago, Chile, Dec. 2015, pp. 1520–1528, https://doi.org/10.1109/ICCV.2015.178.

[17] A. Mohan and S. Poobal, "Crack Detection Using Image Processing: A Critical Review and Analysis," *Alexandria Engineering Journal*, vol. 57, no. 2, pp. 787–798, Jun. 2018, https://doi.org/10.1016/j.aej.2017.01.020.

[18] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li, "DeepCrack: a Deep Hierarchical Feature Learning Architecture for Crack Segmentation," *Neurocomputing*, vol. 338, pp. 139–153, Apr. 2019, https://doi.org/10.1016/j.neucom.2019.01.036.

[19] Faqih Ma'arif, Han Ay Lie, Slamet Widodo, Zhengguo Gao, Fardiansyah Nur Aziz, and Maris Setyo Nugroho, "YSU Concrete Crack Image Dataset (Version 1.0)." Zenodo, Dec. 03, 2025, https://doi.org/10.5281/ZENODO.17810477.

[20] H. Sugimori, K. Shimizu, H. Makita, M. Suzuki, and S. Konno, "A Comparative Evaluation of Computed Tomography Images for the Classification of Spirometric Severity of the Chronic Obstructive Pulmonary Disease with Deep Learning," *Diagnostics*, vol. 11, no. 6, May 2021, Art. no. 929, https://doi.org/10.3390/diagnostics11060929.

[21] C. Ding, T. Pereira, R. Xiao, R. J. Lee, and X. Hu, "Impact of Label Noise on the Learning Based Models for a Binary Classification of Physiological Signal," *Sensors*, vol. 22, no. 19, Sept. 2022, Art. no. 7166, https://doi.org/10.3390/s22197166.

[22] Y. Cha, W. Choi, and O. Büyüköztürk, "Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, May 2017, https://doi.org/10.1111/mice.12263.

[23] T. Kong, A. Yao, Y. Chen, and F. Sun, "HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, Jun. 2016, pp. 845–853, https://doi.org/10.1109/CVPR.2016.98.

[24] D. Li and Q. Chen, "Dynamic Hierarchical Mimicking Towards Consistent Optimization Objectives," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, Jun. 2020, pp. 7639–7648, https://doi.org/10.1109/CVPR42600.2020.00766.