Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:20
https://doi.org/10.28919/cmbn/9738
ISSN: 2052-2541

CONCRETE CRACK SEGMENTATION USING U-NET BASED MODELS WITH VARIOUS PRE-TRAINED BACKBONES

MAHMUD ISNAN[1,2,*], DEDE FAUZI[3], ILFA STEPHANE[4], HERU SAPUTRA[5], HAKAS PRAYUDA[6]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

[2]Research Fellow at Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia 55183

[3]Department of Digital Business, Syedza Saintika University, Padang, Indonesia 25132

[4]Digital Business, Faculty of Information Technology and Creative Industries, Metamedia University, Padang, Indonesia 25173

[5]Information Systems, Faculty of Information Technology and Creative Industries, Metamedia University, Padang, Indonesia 25173

[6]Department of Civil Engineering, Faculty of Engineering Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia, 55183

Abstract: The use of deep learning to detect and analyze structural damage in concrete is gaining increasing attention, but exploration of U-Net with various pre-trained backbone models has not been conducted. This study aims to overcome these limitations by developing a U-Net-based concrete crack segmentation framework that utilizes eight modern and lightweight backbones, namely EfficientNetB0, VGG16, VGG19, ResNet34, ResNet50, DenseNet121, MobileNetV2, InceptionV3, and Xception. A total of 458 labeled images from the Concrete Crack Segmentation

*Corresponding author
E-mail address: mahmud.isnan@binus.ac.id
Received December 08, 2025

dataset were used as training and testing data, representing the variety of concrete surface conditions in real environments. Evaluation was conducted using Dice and Intersection over Union (IoU) metrics. The test results showed that the Xception backbone provided the best performance, with a Dice value of 0.8461 and an IoU of 0.7384, surpassing EfficientNetB0 and VGG16/VGG19 which were previously considered stable in segmentation tasks. This finding confirms that the depthwise separable convolution mechanism in Xception is capable of extracting thin crack features more representatively. This study provides an important contribution in selecting the optimal backbone for concrete crack segmentation models, while opening up opportunities for implementing more accurate and efficient structural condition monitoring technology on an industrial and public infrastructure scale.
Keywords: Concrete segmentation; U-Net; pre-trained models; CNN; instance segmentation.

2020 AMS Subject Classification: 68T07.

1. INTRODUCTION

Cracks in concrete surfaces are a crucial issue in the maintenance and health evaluation of structures because they can act as entry points for moisture and chloride ions, which can trigger reinforcement corrosion and accelerate material degradation [1,2]. This condition not only reduces the service life of infrastructure but also poses safety risks if not detected early and accurately [3,4]. With the increasing need for fast and reliable inspection processes, computer-based crack detection has become a key focus in the field of structural health monitoring [5,6]. However, it should be noted that computer-based approaches do not replace the role of engineers in concrete identification; they are simply tools to assist.

Furthermore, traditional inspection methods such as manual visual inspection are still widely used in practice. This approach is subjective, time-consuming, and not always effective when applied to large-scale structures or environments with high visual complexity. In response to these limitations, various studies have begun to utilize computer vision and deep learning techniques to automatically detect concrete cracks [7,8]. Semantic segmentation models have demonstrated superior performance due to their ability to extract spatial features and generate predictions in the form of pixel maps that represent the crack area more precisely [9]. Previous findings also indicate that CNN-based models can achieve high accuracy on both macro- and microscopic-scale crack images [2,10,11], with the U-Net architecture being one of the most widely used approaches as a baseline for material damage segmentation [12].

To address this gap, this study proposes a comprehensive performance analysis of eight feature extraction models (backbones), namely EfficientNetB0, VGG16, VGG19, ResNet34, ResNet50, DenseNet121, MobileNetV2, and Xception, integrated as encoders in a U-Net architecture for concrete crack segmentation. This approach allows a thorough evaluation of the segmentation performance and stability of each backbone on a dataset with limited size but high visual variability.

2. DATA AND METHODOLOGY

2.1. DATASET

The dataset used in this study comes from the Concrete Crack Segmentation Dataset, a collection of concrete crack images openly available through the Mendeley Data platform [13]. This dataset consists of 458 concrete surface images that have been annotated with two classes initially used for binary pixel classification. Then, this dataset is annotated to perform crack segmentation. In this dataset, there is only one segmentation class (crack segmentation) in JSON format. Of 458, there are 1010 crack labels. Figure 1 shows an example of a cementation crack in the dataset used [14].

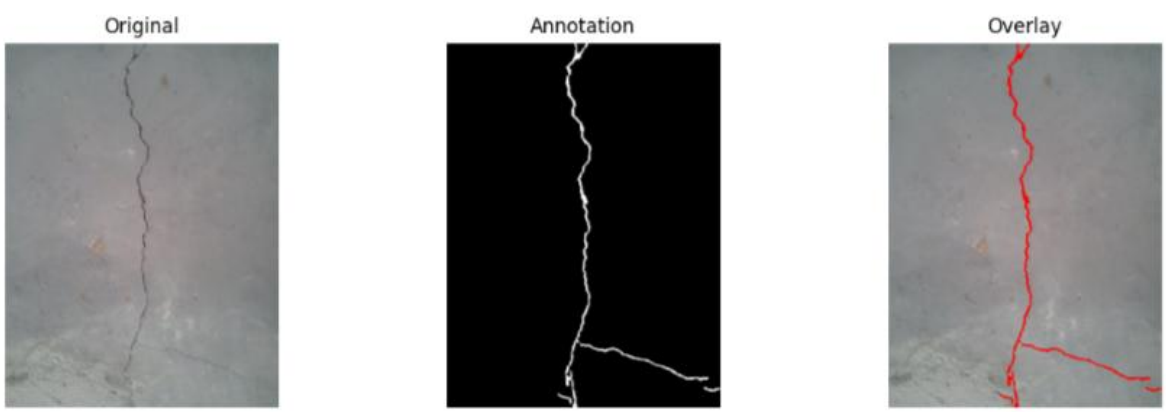

Figure 1. Sample of a crack segmentation dataset.

2.2. DATA PREPROCESSING

These images exhibit various characteristics of real-world conditions, including lighting variations, rough surface textures, irregular concrete geometry, and thin, branching, and irregular crack patterns. This variation presents unique challenges in the segmentation process and also ensures that model evaluation is conducted under visually representative conditions.

Before being used for training, all images and masks were resized to 512 × 512 pixels to ensure uniformity of input to the U-Net architecture. The dataset was then split into 367 training images and 91 validation images in a ratio of approximately 80:20, ensuring the model received a sufficient proportion of training data without compromising the need for generalization evaluation. The dataset was split randomly to ensure a diverse distribution of crack patterns across both subsets. During this preprocessing stage, an augmentation technique using the Albumentations library was used to increase data diversity. This technique also aimed to expand visual variation and reduce the risk of overfitting, given the relatively small dataset and the tendency for crack patterns to be thin and irregular. Augmentation included horizontal flip (p=0.5), vertical flip (p=0.2), random brightness–contrast adjustment (p=0.3), and Gaussian noise (p=0.2). After augmentation, the images were normalized using the ImageNet mean and standard deviation to stabilize the training process. This combination of preprocessing steps was designed to improve the model's robustness to variations in lighting, texture, and concrete surface structure, resulting in more consistent segmentation performance.

2.3. METHODOLOGY

2.3.1 U-NET ARCHITECTURE

Figure 2 illustrates the stages of this research methodology. The U-Net was used as the main architecture in this study due to its effectiveness in pixel-wise segmentation tasks for small, irregular objects, including concrete cracks. U-Net is a symmetric encoder–decoder model designed to capture local and global features equally through skip connections [15,16]. The encoder performs multi-level feature extraction using a series of convolution and downsampling operations, allowing for in-depth understanding of contextual information. At the bottleneck, the model processes features at the lowest resolution to obtain an abstract representation of the crack pattern. Meanwhile, the decoder performs incremental upsampling and recombines high-resolution features from the encoder through skip connections, enabling the recovery of spatial details essential for thin crack segmentation. Figure 3 illustrates the U-Net architecture.

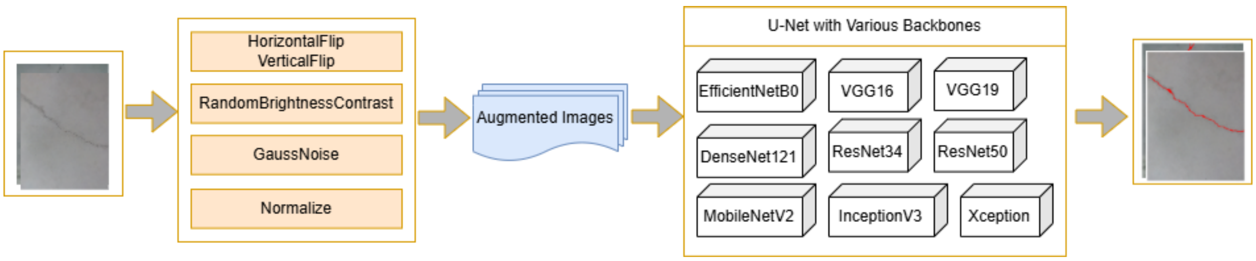

Figure 2. Research pipeline for crack segmentation.

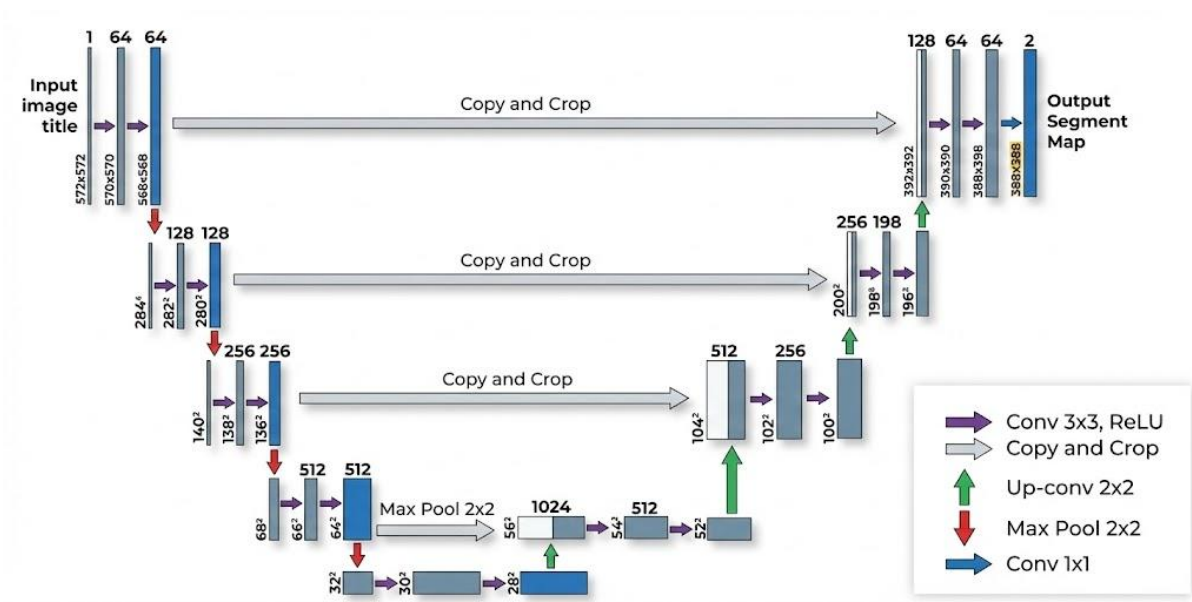

Figure 3. U-Net Architecture.

The U-Net architecture used in this study consists of four encoder blocks and four decoder blocks, each connected by skip connections to maintain information consistency between the extraction and reconstruction stages. All convolutions use ReLU activation, and upsampling is performed through transpose convolution. This structure was chosen for its ability to maintain smooth crack contours while providing precise segmentation results on complex concrete surfaces.

2.3.2 Backbone Encoder Variants

To evaluate the effect of feature extraction architectures on concrete crack segmentation performance, this study integrates eight modern backbones as encoders within the U-Net framework: EfficientNetB0 [17], VGG16, VGG19 [18], ResNet34, ResNet50 [19], DenseNet121 [20], MobileNetV2 [21], and InceptionV3 [22]. All backbones utilize pretrained weights from ImageNet to leverage transfer learning's ability to more effectively extract texture and visual pattern features on limited datasets. Each backbone has distinct architectural characteristics: VGG with a deep sequential convolutional structure, ResNet with residual connections to address gradient degradation, DenseNet with dense connectivity to maximize inter-layer information flow, MobileNetV2 with inverted residual blocks and lightweight depthwise separable convolutions, InceptionV3 with multi-scale processing, and EfficientNetB0 with compound scaling that balances network depth, width, and resolution.

In our implementation, the backbone is used solely as an encoder to extract feature representations at various depth levels, while the U-Net decoder structure is kept uniform throughout the experiment. This approach ensures that performance differences are entirely due to variations in the feature extraction capabilities of each backbone, rather than architectural changes during the reconstruction phase. Thus, the resulting comparative analysis is more objective in assessing the effectiveness of each backbone for the concrete crack segmentation task.

2.3.3 Training Process

The hyperparameters used in the model training process are summarized in Table 1. The Adam optimizer was used to adjust the initial learning rate to 1×10^{-3}, chosen for its stability in accelerating convergence on CNN-based segmentation tasks. The loss function used was a combination of Dice Loss and Binary Cross-Entropy (BCE) Loss. This combination was designed to balance the model's sensitivity to relatively small crack areas (via Dice Loss) while maintaining optimization stability in cases of pixel imbalance (via BCE).

Training was conducted for 30 epochs with a batch size of 4, adjusting for GPU memory limitations while preserving spatial detail in high-resolution data. To improve training stability, the *ReduceLROnPlateau* learning rate scheduler was used with a decrease factor of 0.5 and a patience of three epochs. This scheduler automatically reduces the learning rate when the validation loss shows no improvement, preventing the model from getting stuck in local minima. Furthermore, an early stopping mechanism was implemented by monitoring the validation loss value to stop training if performance no longer improves, thereby reducing the risk of overfitting. Training also leverages *Automatic Mixed Precision* (AMP) to improve computational efficiency and speed up execution on GPUs without sacrificing numerical accuracy. The best model is then selected based on the lowest validation loss achieved during the training process, so the reported performance reflects the most optimal configuration of each tested backbone.

Table 1. Hyperparameter settings.

Training Hyperparameters	**Values**
Optimizer	Adam
Loss function	Dice Loss and Binary Cross-Entropy (BCE)
Learning Rate Scheduler	ReduceLROnPlateau
Training Epochs	30
learning rate	1×10^{-3}

2.3.4 Model Performance Evaluation

The model performance of each backbone in the U-Net architecture is evaluated using two main metrics, namely the Dice Coefficient and Intersection over Union (IoU). These two metrics are widely used in semantic segmentation tasks because they assess the suitability of the predicted area to the ground truth at the pixel level, making them very relevant for detecting thin, irregular concrete cracks that have a small pixel proportion compared to the background. Mathematically, both metrics are defined in Table 2.

Table 2. Definitions and formulas for metric evaluation.

Metric	**Calculation Formula**
Dice Coefficient	$Dice = \frac{2\,x\,\|P \cap G\|}{\|P\| + \|G\|}$
Intersection over Union (IoU)	$IoU = \frac{\|P \cap G\|}{\|P \cup G\|}$

Where P represents the predicted pixels, while G represents the ground truth pixels. Dice places greater emphasis on the overlap between predictions and actual labels, making it more sensitive to thin objects like cracks. Meanwhile, IoU provides a more stringent assessment because it compares the intersection area to the entire combined area. Using both metrics together provides a comprehensive picture of segmentation quality.

3. Experiment Results and Discussion

Table 3 summarizes the performance of U-Net with several backbones based on evaluations of crack segmentation experiments. Overall, the results show that Xception is the highest-performing backbone, with a Dice value of 0.8461, an IoU of 0.7384, and the lowest validation loss of 0.0928. This indicates that Xception's depthwise separable convolution mechanism is able to capture thin, irregular crack patterns more effectively than other backbones.

Table 3. Comparison of backbone performance on the U-Net model.

Backbone	Best Epoch	Dice	IoU	Val Loss
Xception	30	**0.8461**	**0.7384**	**0.0928**
EfficientNetB0	27	0.8432	0.7339	0.0959
VGG16	23	0.8427	0.7335	0.0961
VGG19	16	0.8425	0.7331	0.0966

ResNet34	24	0.8387	0.7278	0.0971
DenseNet121	28	0.8373	0.7255	0.0983
ResNet50	30	0.8347	0.7209	0.0994
MobileNetV2	29	0.8293	0.7137	0.1053
InceptionV3	23	0.8129	0.6902	0.1140

EfficientNetB0, VGG16, and VGG19 follow by a relatively small margin, each with a Dice value above 0.842 and an IoU above 0.733. These backbones demonstrate strong performance stability, although they are unable to surpass the feature representation generated by Xception. The ResNet34, DenseNet121, and ResNet50 architectures exhibit intermediate performance, with Dice values ranging from 0.8347 to 0.8387. Meanwhile, MobileNetV2 and InceptionV3 demonstrated the lowest performance, particularly InceptionV3, which only achieved a Dice of 0.8129 and an IoU of 0.6902. The performance decline for these two backbones is likely due to suboptimal architectural designs for extracting linear and thin structures such as cracks.

This comparison confirms that backbone encoder selection significantly influences the success of concrete crack segmentation. The results also demonstrate that lightweight yet information-rich architectures, such as Xception and EfficientNetB0, can deliver superior performance on limited-scale datasets. Therefore, this table serves as an objective basis for determining the most effective backbone for use in automated segmentation-based structural condition monitoring systems.

Figure 4 illustrates the training and validation dice scores of all backbones. Similar to the model performance results that indicate Xception U-Net is the best model, the training and validation curves demonstrate rapid convergence. The stable and minimally fluctuating curve pattern illustrates the model's ability to efficiently learn crack features from the initial epoch. All other backbones also show consistent performance improvements, but with varying degrees of stability. Overall, this trend confirms that Xception U-Net not only excels in the final metrics but also has more optimal learning dynamics than other architectures.

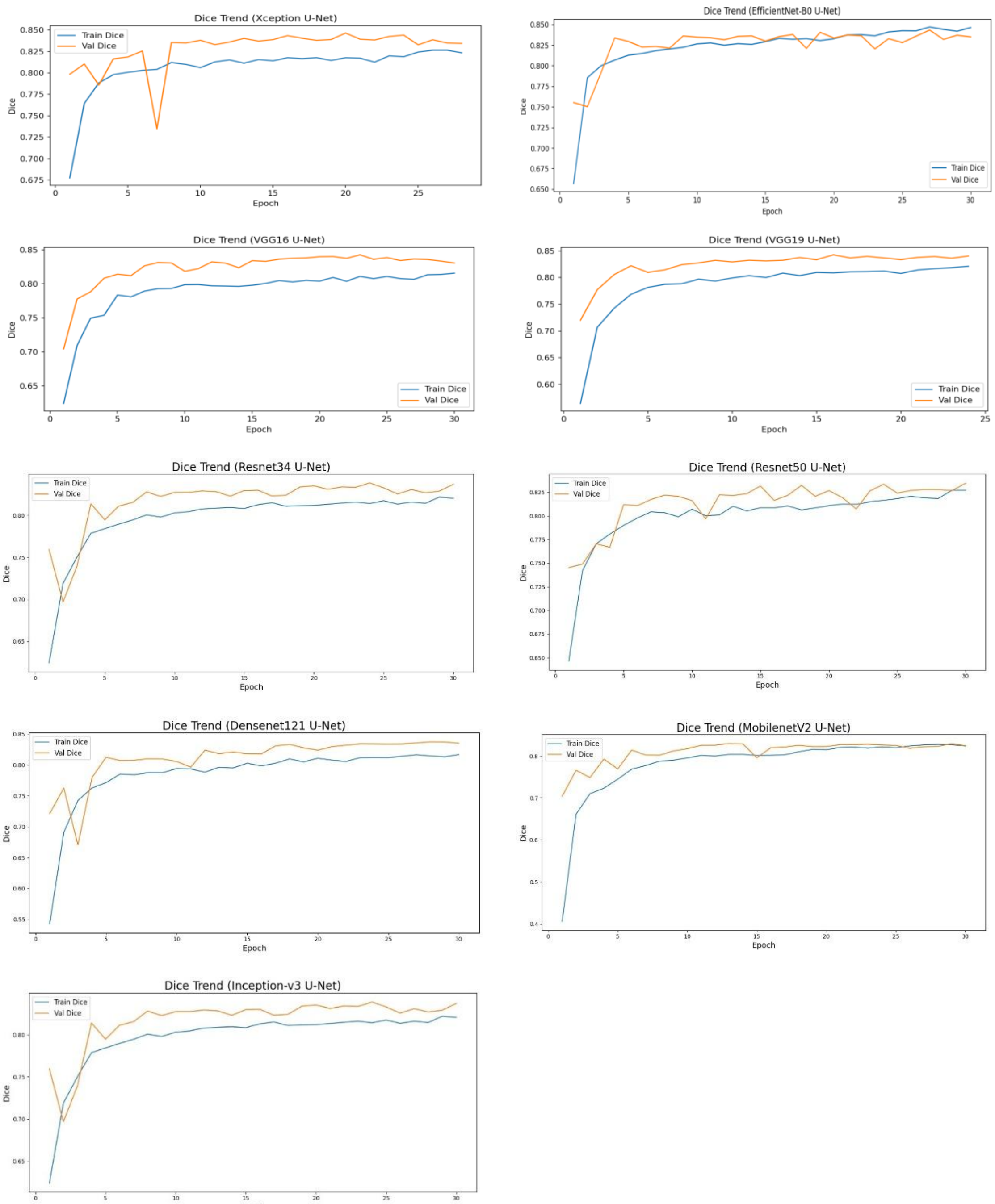

Figure 4. Training and validation dice score curves.

In figure 5, segmentation results are depicted for concrete segmentation using U-Net with Xception backbone. The model yields near-perfect accuracy in delineating thin and irregular crack structures across various validation images. The predicted mask (Overlay Pred, shown in green) shows a high degree of spatial alignment with the ground-truth mask (Overlay GT, shown in red), indicating that the model is capable of capturing fine-grained structural details even when the cracks exhibit low contrast or complex branching patterns.

The results of this study are in line with the findings of [23], who reported that lightweight architectures such as Xception are able to provide superior performance on relatively small dataset infrastructures due to their ability to balance feature depth and complexity. This finding also supports the study of [24], which showed that crack patterns are more easily separated using a model that is sensitive to fine spatial details. On the other hand, research by [25] noted that models with simple structures often perform well in concrete microstructure segmentation. This is consistent with the performance of VGG16 and VGG19 in this study, which, although not the best backbones, are still able to compete with modern architecture. Thus, the results of this study reinforce the understanding that backbone effectiveness is strongly influenced by the structure of the segmented objects and the scale of the dataset used.

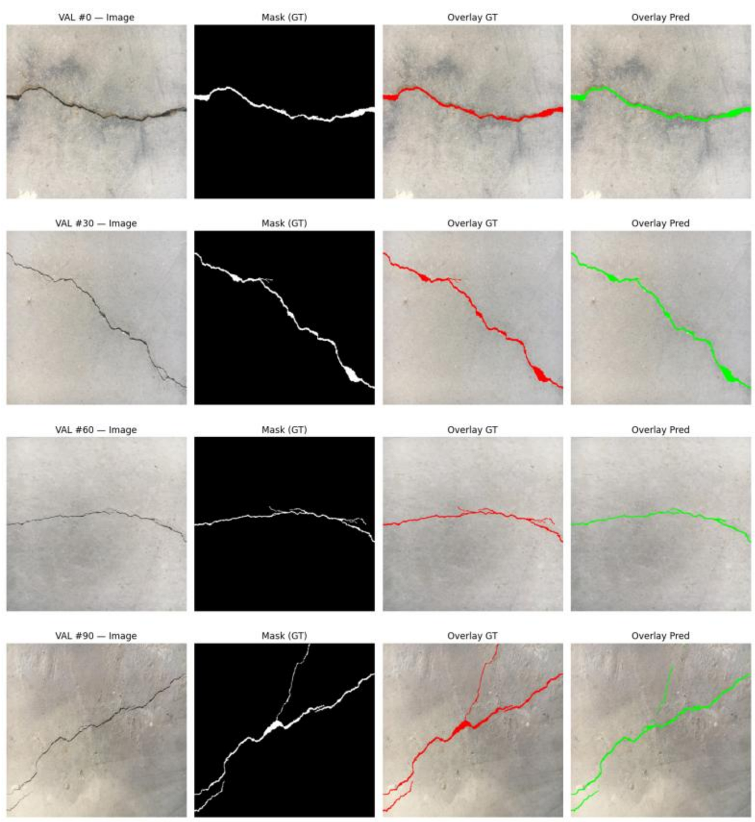

Figure 5. Concrete segmentation result using U-Net with Xception backbone.

In contrast, backbones like InceptionV3 and MobileNetV2 performed the lowest in this study. InceptionV3's multi-scale design, while effective for complex and diverse objects, appears to be suboptimal for learning small linear structures like cracks. Its aggressive feature abstraction approach can lead to the loss of fine details that are key to crack segmentation. MobileNetV2 also exhibits limitations in extracting crack patterns, likely due to its inverted residual block architecture, which focuses more on model efficiency than representational capabilities for thin structures.

Furthermore, this study has several limitations that should be noted. First, the dataset used is relatively limited, with only 458 images, all of which are close-up images of concrete surfaces. This condition limits the variety of backgrounds, extreme lighting conditions, and confounding objects (e.g., shadows, concrete chips, or surface blemishes). Second, the model has not been tested on large-scale field images, such as bridges, building walls, or other structural elements, so its generalizability in real-world environments still needs further evaluation. Third, this study only compares CNN backbones and does not evaluate recent architectures such as Vision Transformers, Swin-UNet, or SegFormer, which have shown promising results in thin object segmentation.

4. CONCLUSION

This study evaluates eight CNN backbones combined with a U-Net architecture for concrete crack segmentation and shows that Xception is the most superior configuration, with a Dice score of 0.8461 and an IoU of 0.7384, outperforming EfficientNetB0, VGG16, and VGG19, which still exhibit competitive performance. The success of Xception is primarily driven by the use of depthwise separable convolution, which more efficiently captures thin and branching crack patterns, while other models such as EfficientNetB0 offer a good balance between accuracy and computational efficiency, and VGG exhibits stability due to its simple yet effective layered convolution structure. In contrast, backbones such as InceptionV3 perform less well, indicating that a multi-scale approach is not always optimal for detecting small linear structures. Overall, the results of this study confirm that the combination of U-Net and Xception is a highly effective choice for computer vision-based concrete crack detection systems, with contributions including a comparative analysis between the models, the development of a robust segmentation pipeline, and empirical recommendations for the development of structural health monitoring systems. Future research could be directed at exploring Transformer-based backbones, expanding the dataset with more varied field imagery, and integrating the model into drone-based automated inspection platforms or edge devices to improve readiness for real-world implementation.

ACKNOWLEDGMENT

This research was supported by the Universitas Muhammadiyah Yogyakarta (UMY) Research Fellowship Program.

AUTHOR CONTRIBUTIONS

Mahmud Isnan: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization, Dede Fauzi: Formal analysis, Writing – original draft, Visualization, Ilfa Stephane: Writing – original draft, Visualization, Heru Saputra: Writing – original draft, Visualization, Hakas Prayuda: Writing - Review & Editing, Supervision.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] L.F. Rincon, Y.M. Moscoso, A.E.A. Hamami, J.C. Matos, E. Bastidas-Arteaga, Degradation Models and Maintenance Strategies for Reinforced Concrete Structures in Coastal Environments Under Climate Change: A Review, Buildings 14 (2024), 562. https://doi.org/10.3390/BUILDINGS14030562.

[2] Y. Jusman, A. Zaki, A. Priambada, A. Firdaus, Classification of Concrete Crack Through Texture Analysis and K-Nearest Neighbors Models, in: 2025 International Conference on Computer, Control, Informatics and its Applications (IC3INA), IEEE, 2025: pp. 376-380. https://doi.org/10.1109/IC3INA68387.2025.11325269.

[3] W. Wang, C. Su, Automatic Concrete Crack Segmentation Model Based on Transformer, Autom. Constr.

139 (2022), 104275. https://doi.org/10.1016/J.AUTCON.2022.104275.

[4] Y. Jusman, A. Zaki, A. Firdaus, A. Priambada, Concrete Crack Identification Through Hu Moment Shape Analysis and K-Nearest Neighbors Models, in: 2025 5th International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), IEEE, 2025: pp. 378-382. https://doi.org/10.1109/ICE3IS66769.2025.11281148.

[5] P.S. Chakurkar, D. Vora, S. Patil, S. Mishra, K. Kotecha, Data-Driven Approach for Ai-Based Crack Detection: Techniques, Challenges, and Future Scope, Front. Sustain. Cities 5 (2023), 1253627. https://doi.org/10.3389/FRSC.2023.1253627.

[6] J.M.G. Payawal, D. Kim, Image-Based Structural Health Monitoring: A Systematic Review, Appl. Sci. 13 (2023), 968. https://doi.org/10.3390/APP13020968.

[7] Y. Jusman, A. Zaki, A. Firdaus, A. Priambada, Concrete Crack Classification Using Gray-Level Co-Occurrence Matrix Features Based on Support Vector Machine, in: 2025 4th International Conference on Electronics Representation and Algorithm (ICERA), IEEE, 2025: pp. 533-538. https://doi.org/10.1109/ICERA66156.2025.11087368.

[8] Ihksan M, Isnan M, Fauzi D, Prayuda H. Performance Comparison of ResNet50 and CNN-LSTM in Concrete Classification. 2025 5th International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS) 2025:226–31. https://doi.org/10.1109/ICE3IS66769.2025.11281310.

[9] Y. Song, Z. Huang, C. Shen, H. Shi, D.A. Lange, Deep Learning-Based Automated Image Segmentation for Concrete Petrographic Analysis, Cem. Concr. Res. 135 (2020), 106118. https://doi.org/10.1016/J.CEMCONRES.2020.106118.

[10] B. Hilloulin, I. Bekrine, E. Schmitt, A. Loukili, Modular Deep Learning Segmentation Algorithm for Concrete Microscopic Images, Constr. Build. Mater. 349 (2022), 128736. https://doi.org/10.1016/J.CONBUILDMAT.2022.128736.

[11] I.F.A. Abdalla, A. Zaki, S.A.P. Rosyidi, S. Riyadi, Automated Identification of Concrete Cracks Utilizing State-Of-The-Art Deep Learning Models, in: 2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), IEEE, 2025: pp. 360-365. https://doi.org/10.1109/ICIMCIS68501.2025.11327287.

[12] J. Shi, J. Dang, M. Cui, R. Zuo, K. Shimizu, et al., Improvement of Damage Segmentation Based on Pixel-Level Data Balance Using Vgg-Unet, Appl. Sci. 11 (2021), 518. https://doi.org/10.3390/APP11020518.

[13] Ç.F. Özgenel, Concrete Crack Segmentation Dataset, (2019). https://doi.org/10.17632/jwsn7tfbrp.1.

[14] Dataset Ninja, Concrete Crack Segmentation Dataset, 2019. https://datasetninja.com/concrete-crack-segmentation-dataset (accessed December 6, 2025).

[15] M. AnbuDevi, K. Suganthi, Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET, Diagnostics 12 (2022), 3064. https://doi.org/10.3390/DIAGNOSTICS12123064.

[16] W. Weng, X. Zhu, INet: Convolutional Networks for Biomedical Image Segmentation, IEEE Access 9 (2021), 16591-16603. https://doi.org/10.1109/ACCESS.2021.3053408.

[17] M. Tan, Q. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, in: Proceedings of the 36th International Conference on Machine Learning, PMLR 97:6105-6114, 2019.

[18] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2014.

[19] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016: pp. 770-778. https://doi.org/10.1109/CVPR.2016.90.

[20] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017: pp. 2261-2269. https://doi.org/10.1109/CVPR.2017.243.

[21] A.G. Howard, M. Zhu, B. Chen, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv:1704.04861 (2017), https://doi.org/10.48550/ARXIV.1704.04861.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the Inception Architecture for Computer Vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016: pp. 2818-2826. https://doi.org/10.1109/CVPR.2016.308.

[23] E.S. Chahal, A. Patel, at al. Unet Based Xception Model for Prostate Cancer Segmentation from MRI Images, Multimed. Tools Appl. 81 (2021), 37333-37349. https://doi.org/10.1007/S11042-021-11334-9.

[24] X. Zhang, D. Rajan, B. Story, Concrete Crack Detection Using Context-aware Deep Semantic Segmentation Network, Comput. Civ. Infrastruct. Eng. 34 (2019), 951-971. https://doi.org/10.1111/MICE.12477.

[25] A.N. Beskopylny, E.M. Shcherban', S.A. Stel'makh, L.R. Mailyan, B. Meskhi, et al., Method for Concrete Structure Analysis by Microscopy of Hardened Cement Paste and Crack Segmentation Using a Convolutional Neural Network, J. Compos. Sci. 7 (2023), 327. https://doi.org/10.3390/JCS7080327.