Smart Structures and Systems, *Vol. 38, No. 1 (2026) 1-28*
https://doi.org/10.12989/sss.2026.38.1.001

Techno-Press

Label-efficient semi-supervised learning with hard negative samples for crack segmentation in concrete and asphalt structures

Muhammad Tanveer[1a], Soojin Cho*[1,2]

[1]*Department of Civil Engineering, University of Seoul, Dongdaemun-gu, Seoul 02504, South Korea*
[2]*Graduate School of Urban Bigdata Convergence, University of Seoul, Dongdaemun-gu, Seoul 02504, South Korea*

(*Received March 12, 2026, Revised June 29, 2026, Accepted July 9, 2026*)

Abstract. Crack segmentation is critical for maintaining structural safety, but fully supervised deep learning approaches require extensive pixel-level annotations, making large-scale deployment costly. This study adapts semi-supervised learning framework for crack segmentation that extends weak-to-strong consistency regularization with domain-aware hard negative sample (HNS) integration to minimize annotation needs while improving robustness. The framework utilizes a dual-stream architecture with weak-to-strong image- and feature-level perturbation-based consistency regularization, adapted from general semantic segmentation to the domain-specific challenges of structural crack inspection. HNS, which are visually deceptive non-crack patterns such as shadows, stains, and surface textures, were integrated into the unlabeled training stream, enabling the model to better distinguish cracks from background noise and substantially reduce false positives. The proposed method achieved performance comparable to that of a fully supervised model using only 20% of the labeled data, effectively reducing the labeling costs by 80%. Experiments on concrete and asphalt structures demonstrate that HNS integration improves precision, especially in visually complex conditions, while a mixed-domain model trained on both surfaces achieves strong generalization. Furthermore, the proposed method was compared with other state-of-the-art semi-supervised methods and consistently achieved improved performance in both concrete and asphalt datasets. These findings indicate that integrating SSL with HNS enhances crack segmentation, providing a scalable solution for real-world applications where background complexity remains challenging.

Keywords: concrete and asphalt structures; consistency regularization; crack segmentation; hard negative samples (HNS); semi-supervised learning (SSL)

1. Introduction

Infrastructure, including buildings, bridges, and asphalt roads, constitute essential societal assets. Among the various types of structural damage, cracks are particularly critical because their progression over time can significantly compromise structural integrity. Visual inspection has been one of the most common techniques used for structural health monitoring (SHM) for decades; however, it is time consuming and costly. With recent technological advancements, visual inspection

*Corresponding author, Ph.D., Professor, E-mail: soojin@uos.ac.kr
[a] Ph.D., E-mail: tanveer715@uos.ac.kr

http://www.techno-press.org/sss/5
ISSN: 1738-1584 (Print), 1738-1991 (Online)

has been increasingly replaced by image-processing techniques such as Gabor filtering, Otsu's method, edge detection, and threshold segmentation [1-4]. However, these algorithms face challenges in identifying cracks under changing environmental conditions and background noise in the images.

Deep learning models employ automated computer vision techniques and yield promising results in crack classification, detection, and pixel-wise segmentation. Flah et al. [5] combined image processing techniques with deep convolutional neural networks (CNN) for crack classification and quantification in concrete structures. Zhou and Song [6] optimized CNNs by studying hyperparameter selection for crack classification in laser-scanned asphalt images. Yang et al. [7] used the YOLOv3 model for structural crack detection, whereas Kim and Cho [8] applied AlexNet with probability thresholding to improve the detectability of concrete cracks. In addition, many studies have explored deep learning applications for crack detection and localization [9-11]. However, owing to the variability in crack characteristics, classification and detection methods often lack precision in capturing properties, such as crack length and shape. Consequently, deep learning-based pixel-wise segmentation, which learns distinctive features from crack data, has delivered promising results in recent years [12-14]. Alipour et al. [15] proposed CrackPix, a deep fully convolutional model that enables accurate pixel-level crack segmentation in concrete infrastructure, outperforming traditional methods and patch-based models. Sun et al. [16] enhanced the DeepLabV3+ model by adding a multi-scale attention module in the decoder known as DMA-Net for pavement crack segmentation, outperforming other traditional segmentation models. Chu et al. [17] proposed Tiny-Crack-Net for tiny crack segmentation and showed better performance on cracks as small as 0.05 mm. Tran et al. [18] introduced a novel crack segmentation method by optimizing the YOLOv7 and UNet models for high-resolution bridge deck images, achieving higher accuracy than other models. Yang et al. [19] developed a segmentation network for microcracks and complex background conditions in pavement images. Their approach used a multi-scale attention module and an additive fusion block to capture both local and global contextual information, outperforming other models under complex background conditions. Yang et al. [20] proposed a CNN-based crack detection approach that combines MobileNetV3 for initial classification with an improved DeepLabV3+ network for accurate semantic segmentation, achieving high detection accuracy and significantly reducing training time compared to traditional models. Recently, Mamba-based architectures have been introduced for crack segmentation tasks. Zhang et al. [21] proposed hybrid CNN-Mamba networks that integrate multi-shape convolution kernels, difference-based encoders, and cross-aware Mamba attention modules to enhance both spatial detail and global contextual understanding. These methods are particularly effective for capturing fine crack structures and complex edge patterns under diverse pavement conditions. However, all these crack segmentation models rely on highly pixel-wise labeled data, which are laborious and time-consuming.

To counteract the requirement for large amounts of labeled data, semi-supervised learning (SSL) techniques have been introduced to strike a balance between supervised and unsupervised learning approaches. Semantic segmentation based on SSL algorithms has been gaining increasing attention for addressing the need for large amounts of labeled data. Several SSL-based segmentation models that exhibit promising performance in image segmentation have been proposed. SSL algorithms primarily involve two methods: pseudo-labeling [22] and consistency regularization [23, 24]. In pseudo-labeling, labels are assigned to unlabeled data based on predictions generated from labeled images. In contrast, consistency regularization involves maintaining consistent predictions between two models or model branches when processing the same set of unlabeled images subjected to different data perturbation techniques. These perturbations typically include both weak and strong

transformations, such as horizontal and vertical flipping, color jittering, and CutOut [25], CutMix [26], MixMatch [27] and ClassMix [28] which help fully utilize unlabeled data and improve model robustness by consistently comparing predictions between the models. Recently, various methods have been proposed for semantic segmentation tasks that rely on pseudo-labeling and consistency regularization [29, 30]. Chen et al. [31] proposed a cross-pseudo supervision approach to enforce consistency between two networks with the same architecture but different initializations, using one-hot pseudo-segmentation maps from each network to supervise the other. Wang et al. [32] utilized unreliable pseudo-labels and anchor pixels to enhance model performance, whereas Adaptive Equalization Learning (AEL) proposed by Hu et al. [33] improved the performance of underperforming classes using strategies, such as adaptive CutMix, and addressed the class imbalance problem in SSL. These state-of-the-art SSL methods were applied to general semantic segmentation tasks and demonstrated good performances on the Cityscapes, Pascal VOC, and COCO benchmark datasets.

SSL algorithms have also been used in civil infrastructures for classification, detection, and pixel-wise segmentation. Guo et al. [34] used an SSL approach based on a CNN for façade defect classification on concrete surfaces. They developed a novel uncertainty filter to select reliable data and improve model performance, leveraging the mean teacher algorithm. Chun and Ryu [35] utilized a fully connected convolutional network (FCN) for asphalt damage detection using SSL by generating pseudo-labels for many unlabeled datasets. Generative Adversarial Networks (GANs) are widely used in SSL algorithms. Di et al. [36] proposed a Convolutional Autoencoder (CAE)-GAN model by combining a CAE and a GAN for defect classification in steel. Their performance improved by approximately 16% compared with that of traditional methods. Li et al. [37] implemented a semi-supervised framework using adversarial learning for pavement crack segmentation. They combined adversarial loss with cross-entropy loss, which improved the network's detectability and achieved better performance using a very small amount of labeled data. Similarly, Shim et al. [38, 39] proposed segmentation networks for concrete and asphalt damages using an adversarial learning approach. They also used super-resolution GANs to enhance data quality and integrated them with the adversarial learning framework, achieving improved crack segmentation performance with limited labeled data. Mohammed et al. [40] applied an SSL approach and customized a traditional U-Net to create a lightweight network, reducing computational costs for crack segmentation. Their proposed method outperformed other supervised learning models and reduced the need for labeled data by 40%. Wang and Su [41] implemented semi-supervised crack segmentation using the mean teacher algorithm and proposed EfficientUNet, which extracts multi-scale feature information of cracks. Their method performed well using only 60% of the labeled data but struggled to detect tiny cracks in complex background conditions. Tang et al. [42] proposed a semi-supervised learning framework based on consistency regularization and cross-supervision, which effectively utilized unlabeled post-earthquake images to enhance segmentation performance and accurately detect structural damages. Recently, Xiang et al. [43] developed SemiCrack, a semi-supervised crack segmentation model based on cross-pseudo supervision and contrastive learning. The segmentation network fused transformer and CNN architectures, enhancing both local feature extraction and global contextual understanding.

Nevertheless, these closest semi-supervised crack detection studies share a common limitation that motivates the present work. Adversarial-learning approaches reduce label dependence but do not explicitly model visually deceptive non-crack patterns, resulting in high false-positive rates in cluttered backgrounds. Mean-teacher methods such as EfficientUNet still require approximately 60% of the labels and, as the authors reported, show degraded performance on fine cracks under

complex conditions. SemiCrack improves feature learning through cross-pseudo supervision and contrastive learning but, like the other methods, is trained primarily on datasets containing relatively simple and well-defined crack patterns without incorporating hard negative samples (HNS). As a result, these methods have limited robustness when cracks are embedded in complex backgrounds with crack-like artifacts, leading to increased false-positive detections and reduced generalization in practical field applications. To address these limitations, the present study explicitly incorporates unlabeled HNS into the consistency-training framework to suppress false positives, achieves performance comparable to a fully supervised model using only 20% of the labeled data, and demonstrates strong cross-material generalization. Following were the contributions made.

1. Adapted and evaluated a weak-to-strong consistency regularization framework [44] for structural crack segmentation, demonstrating for the first time its effectiveness on domain-specific crack datasets characterized by severe class imbalance, subtle foreground patterns, and visually complex backgrounds, conditions absent from the general semantic segmentation benchmarks on which the original framework was validated. The adaptation leverages a dual-branch design with image-level and feature-level perturbations, enabling robust learning from unlabeled data under the unique challenges of crack inspection.
2. Curated and integrated domain-aware HNS into an unlabeled training stream for both concrete and asphalt datasets. These visually deceptive non-crack patterns, such as shadows, stains, and surface textures, were manually selected to reflect the complexities of real-world infrastructure surfaces. Their inclusion enables the model to better distinguish cracks from background noise, significantly reducing false positives and enhancing generalization across diverse material domains.
3. It achieved competitive or superior segmentation performance compared to fully supervised models, while using only 20% of the labeled data, reducing annotation requirements by 80%, and improving the scalability of crack segmentation systems.
4. A mixed-domain crack segmentation model was developed by combining datasets from concrete and asphalt surfaces, enabling the model to be generalized across diverse material types and adapted to various background textures.

By focusing on labeling efficiency and robust segmentation in complex environments, this study advances crack segmentation technology toward scalable and practical structural health monitoring applications.

2. Proposed methodology

This section presents an adapted semi-supervised framework for crack segmentation and describes its core components. Building upon the established weak-to-strong consistency regularization paradigm [44], this work adapts the dual-branch perturbation strategy to the specific challenges of structural crack inspection, including severe foreground-background class imbalance, visually deceptive background textures, and limited annotated data. The key domain-specific contribution introduced in this study is the systematic curation and integration of hard negative samples (HNS) into the unlabeled training stream, which directly targets the false positive problem endemic to crack segmentation in real-world infrastructure. For the base segmentation model, the approach employs DeepLabV3+ [45] with a ResNet-101 [46] backbone, which provides strong multi-scale feature extraction capabilities suited to cracks of varying widths and textures. The datasets used in this study encompass diverse crack types and background conditions across both

concrete and asphalt surfaces, carefully constructed to support domain-specific training and to improve generalization in practical applications.

2.1 Semi-supervised consistency framework

In semi-supervised semantic segmentation algorithms, both labeled and unlabeled data are used during training. The labeled dataset, denoted as $D_l = \{x_l, y_l\}$, comprises input images x_l paired with the corresponding ground-truth labels y_l. By contrast, the unlabeled dataset, denoted as $D_u = \{x_u\}$, consists of input images $\{x_u\}$ without labels. Because labeled data are often limited due to the high cost of annotation, semi-supervised learning aims to maximize the utility of unlabeled data by enforcing consistency between perturbed versions of the same image. In this study, a weak-to-strong perturbation-based consistency regularization technique is applied at both the image and feature levels to enhance the model's ability to learn from unlabeled data, thereby ensuring the model produces consistent predictions across different perturbations of the same input, thereby improving robustness and generalization. Fig. 1 provides an overview of the proposed semi-supervised method, illustrating the augmentation strategies and consistency learning process.

2.1.1 Image-level perturbation strategy

For each unlabeled image x_u, a weak augmentation technique denoted as A_u^w, is first applied. This includes simple geometric transformations, such as cropping and horizontal flipping, resulting in a weakly augmented image represented as x_u^w. Weak perturbations serve as a stable reference for generating pseudo-labels that guide consistency training under stronger transformations. Additionally, two separate streams of strong augmentation, denoted as ${A_u}^{s_1}$, ${A_u}^{s_2}$, were applied on top of x_u^w to create two strongly augmented variants, $x_u^{s_1}$ and $x_u^{s_2}$, respectively. These augmentations involve more aggressive transformations, including color jittering, blurring, and cutmix. It is important to note that $x_u^{s_1}$ and $x_u^{s_2}$ are not identical because they undergo different combinations of strong augmentation and distinct regions of CutMix patches. This diversity in perturbations helps the model become more invariant to different input variations. The predictions for weakly and strongly perturbed images are expressed as

$$ {p_u}^w = \hat{F}(A_u^w(x_u^w)), \tag{1} $$

$$ {p_u}^{s_1} = F({A_u}^{s_1}(A_u^w(x_u^{s_1}))), \tag{2} $$

$$ {p_u}^{s_2} = F({A_u}^{s_2}(A_u^w(x_u^{s_2}))), \tag{3} $$

In the above Eqs. (1)-(3), ${p_u}^w$ corresponds to the prediction of the weakly perturbed image, and ${p_u}^{s_1}$, ${p_u}^{s_2}$ represent the predictions for the two strongly perturbed images. A teacher-student model framework is adopted, where teacher model $\hat{F}$ generates pseudo-labels based on weakly perturbed images, while student model F learns from strongly perturbed images to refine its predictions. The teacher and student models were kept identical in structure to maintain simplicity and efficiency. Weak augmentation preserves the semantic content of the image, allowing the teacher branch to generate more reliable pseudo-labels. The student branch is then trained using stronger perturbations to learn augmentation-invariant feature representations. Applying strong augmentations to both branches may reduce pseudo-label reliability because excessive image distortions can alter the semantic structure of subtle crack patterns.

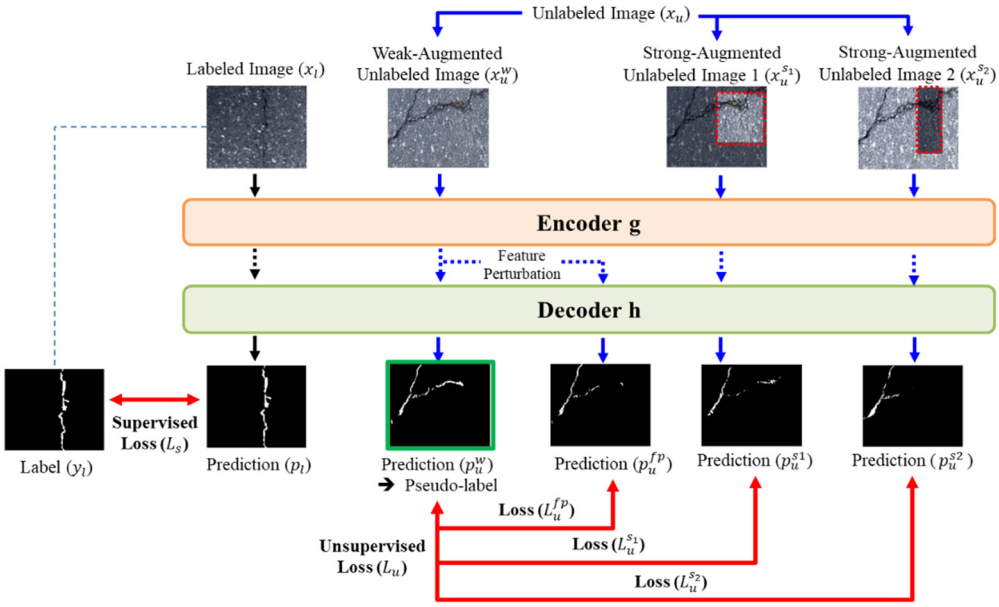


Figure 1. Overview of semi-supervised crack segmentation method

2.1.2 Feature-level perturbation strategy

In addition to image-level perturbations, feature-level perturbations are introduced to enhance the robustness of the model. These perturbations were applied to the feature representations extracted from the weakly perturbed images x_u^w, as shown in Fig. 1. The feature-level perturbation is applied to the same weakly augmented image used by the teacher branch for pseudo-label generation. Consequently, any differences between the teacher prediction and the feature-perturbed prediction arise solely from feature-space perturbations rather than from different image augmentations. The motivation behind feature-level perturbation is to expand the perturbation space, thereby enhancing the model's adaptability while reducing its dependence on hyperparameter tuning for strong augmentations. Unlike image-level perturbations, which require careful optimization of augmentation parameters for different datasets, feature-level perturbations operate directly in the feature space, offering greater flexibility and improved computational efficiency. The prediction of the feature perturbation ${p_u}^{fp}$ obtained from the segmentation model consists of encoder g and decoder h as

$$ {e_u}^w = g(x_u^w), \tag{4} $$

$$ {p_u}^{fp} = h(\rho({e_u}^w)), \tag{5} $$

In the above Eqs. (4)-(5), ${e_u}^w$ is the features extracted from the weakly perturbed image x_u^w, and ρ denotes feature perturbation technique. In this study, a channel dropout layer was utilized for the feature perturbation mechanism. This technique randomly deactivates feature channels during training, encouraging the model to learn from incomplete representations and improve its generalization. Channel dropout is a simple yet effective perturbation strategy that introduces feature-level diversity with a minimal computational overhead. Its effectiveness in consistency-based semi-supervised segmentation was also demonstrated in a recent study [44], where it

outperformed alternatives, such as noise injection and virtual adversarial training, in terms of stability and accuracy.

Each minibatch of unlabeled data maintains a total of four feedforward streams during training, including the simplest stream ($p_u{}^w$) corresponding to the weak perturbation in Eq. (1), two image-level strong perturbation streams ($p_u{}^{s_1}$, $p_u{}^{s_2}$) as defined in Eqs. (2)-(3), and the feature-level perturbation stream ($p_u{}^{fp}$) as illustrated in Eq. (5). By enforcing consistency across these streams, the model learns to generate stable predictions under varying perturbations, leading to an improved segmentation performance.

2.1.3 Combined supervised and consistency loss

Because both labeled and unlabeled data were utilized, the training objective consisted of a combination of supervised and unsupervised loss terms. For the labeled data, the model used supervised loss as a cross-entropy loss function that optimized the model performance by comparing the model predictions of the labeled dataset and ground-truth labels as follows

$$L_{sup} = H(y_l, p_l), \tag{6}$$

In the above Eq. (6) $H(y_l, p_l) = -\sum y_l \log p_l$ represents the cross-entropy loss function that minimizes the discrepancy between model's predictions and the ground-truth labels:y_l represents ground-truth labels and p_l represents the predicted probability distribution for the labeled dataset. This loss function ensures that the model correctly classifies the labeled samples by reducing the difference between the predicted and actual class probabilities.

For the unlabeled dataset, the unsupervised loss is derived from the predictions obtained through the four feedforward streams, which include weak-to-strong image-level and feature-level perturbations. The prediction obtained from the weakly augmented unlabeled image, denoted as $p_u{}^w$, was converted into pseudo-labels. These pseudo-labels act as a reference to enforce consistency between the model predictions under strong and feature perturbations. The unsupervised loss function encourages the alignment between these predictions and is formulated as follows

$$L_u = \frac{1}{B_u}\sum \mathbb{1}\left(max(p_u{}^w) \geq \tau\right).\left(\lambda H\left(p_u{}^w, p_u{}^{fp}\right) + \frac{\mu}{2}\left(H(p_u{}^w, p_u{}^{s_1}) + H(p_u{}^w, p_u{}^{s_2})\right)\right), \tag{7}$$

In the above Eq. (7), B_u represents the batch size of the unlabeled dataset, and τ denotes the confidence threshold used to filter out unreliable pseudo-labels. The selection of τ plays a crucial role in determining which pseudo-labels contribute to the training process, ensuring that only confident predictions are retained. In this study, τ is set to zero to retain all pseudo-labels during training. This choice reflects the imbalanced nature of the crack datasets, where discarding low-confidence predictions could result in the loss of many valid but subtle crack regions. Additional experiments evaluating other threshold values are presented in Appendix A (Fig. A1), demonstrating that the model's performance remains robust at a zero threshold but degrades gradually as the threshold increases. The loss weights λ and μ control the contributions of feature-level and image-level perturbations to the unsupervised loss, and are set to 0.5 and 0.25, respectively, to maintain a balance between the two perturbation types. The total loss function for the semi-supervised method is formulated by combining both the supervised and unsupervised losses, ensuring effective optimization of the model using both labeled and unlabeled data. The overall loss function is expressed as

$$L = L_{sup} + \lambda L_u^{fp} + \mu L_u^{s_1} + \mu L_u^{s_2}. \quad (8)$$

Where L_{sup} corresponds to the supervised cross-entropy loss and L_u^{fp}, $L_u^{s_1}$, and $L_u^{s_2}$ represent the unsupervised consistency losses for feature-level and image-level perturbations, respectively. This combined loss formulation enables the model to effectively learn from a limited amount of labeled data while leveraging the additional information provided by unlabeled samples, ultimately enhancing segmentation accuracy.

2.2 Backbone Network: DeepLabV3+

In this study, the DeepLabV3+ model with ResNet-101 as the backbone network was used for crack segmentation. DeepLabV3+ is a state-of-the-art encoder-decoder model that enhances the spatial pyramid pooling module introduced in its predecessor DeepLabV3. The overall architecture of DeepLabV3+ is illustrated in Fig. 2.

The encoder module of DeepLabV3+ incorporates ResNet-101 as its deep convolutional neural

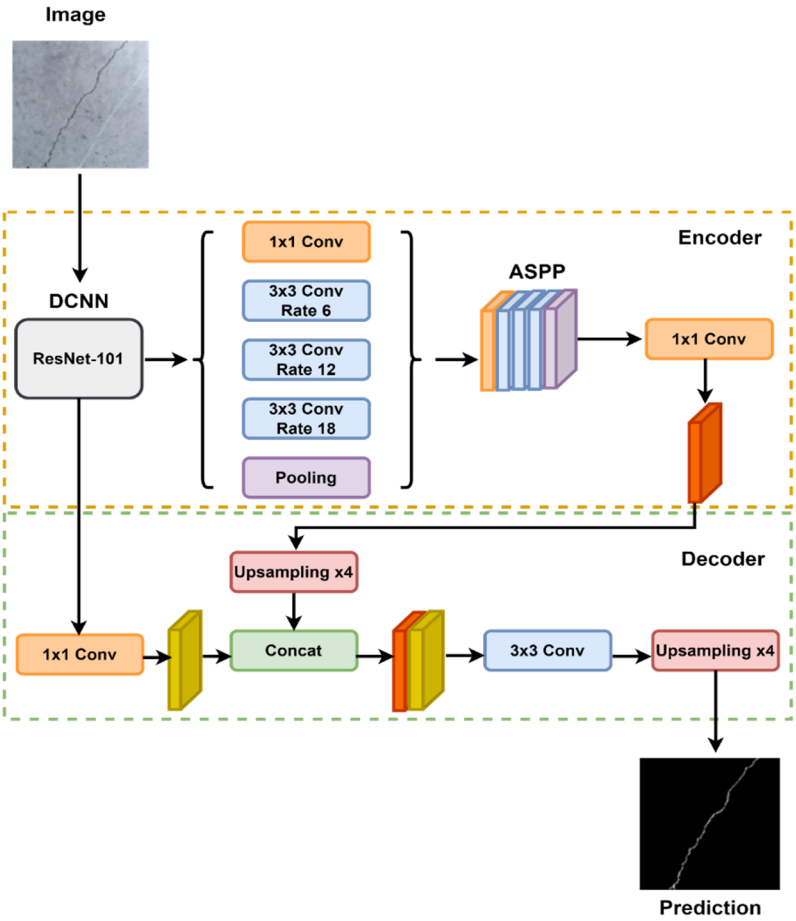


Figure 2. Overview of DeepLabV3+ model

network (DCNN) backbone, responsible for extracting both low- and high-level features from the input image. The extracted low-level features were passed into the decoder through a 1×1 convolution layer to ensure that fine details were preserved. By contrast, high-level or contextual features are forwarded to the Atrous Spatial Pyramid Pooling (ASPP) module within the encoder. The ASPP module plays a critical role in capturing multi-scale contextual information. It consists of five branches: a 1×1 convolution, three 3×3 convolutions with dilation rates of 6, 12, and 18, and an image pooling layer. The inclusion of convolutions with varying dilation rates allows the model to capture contextual information at multiple scales, making it highly effective for segmenting cracks of different sizes. After feature extraction, the outputs of all five branches in the ASPP module were concatenated and refined using a 1×1 convolution to ensure a comprehensive representation of the spatial information.

In the decoder module, the high-level features obtained from the ASPP underwent 4× upsampling before being concatenated with the low-level features extracted earlier in the encoder. This fusion of high- and low-level features helps restore fine details in the segmented output. The concatenated features are further processed through a 3×3 convolution layer to refine spatial information. Finally, a second 4× upsampling operation is applied to generate the final pixel-wise segmentation predictions. By integrating the ASPP module for multi-scale feature extraction with an effective encoder-decoder structure, DeepLabV3+ significantly enhances segmentation accuracy, making it well-suited for detecting cracks with varying textures and scales.

2.3 Hard Negative Samples (HNS) integration

This study introduces the use of HNS along with positive crack instances within a semi-supervised framework to enhance segmentation accuracy in visually complex environments. HNS are non-crack patterns that visually resemble actual cracks, including shadows, stains, surface grooves, and linear textures. These patterns can easily confuse the model, resulting in false positives during prediction. In this study, HNS were manually identified through visual inspection of the unlabeled images. Regions exhibiting a crack-like appearance were selected as challenging negative examples. This manual curation enabled the model to learn from difficult background conditions without the need for additional labeling, thereby reducing the misclassification of non-cracked areas.

HNS were incorporated exclusively into the unlabeled branch of the semi-supervised framework, allowing the model to leverage both labeled crack annotations and diverse unlabeled background patterns during training. Through repeated exposure to HNS, the model gradually learns to distinguish between the actual crack features and visually similar non-crack artifacts, resulting in improved detection precision. The inclusion of the HNS helps the model develop finer discriminative features and enhances its robustness in complex scenes. This is particularly important in real-world conditions, where background textures and lighting variations often resemble crack patterns. To ensure domain-specific effectiveness, separate HNS datasets were curated for concrete and asphalt surfaces. Each set reflected the visual characteristics and noise profiles typical of the respective material types. Overall, integrating unlabeled HNS into the training process improved model generalization and reduced reliance on large-scale labeled data. This approach strengthened the ability of the model to detect cracks accurately across a wide range of real-world conditions.

2.3.1 Concrete crack dataset

The concrete crack dataset used in this study originated from real concrete structures in South Korea and was supplemented with additional data from online sources containing images with

visible cracks. Various types of concrete cracks with different widths were included to construct a comprehensive dataset. Manual annotation was performed using the Adobe Photoshop labeling tool to create the corresponding masks. To enhance the dataset diversity and improve the model performance, additional HNS were added sourced from the AI Hub, featuring images of real concrete structures such as stains, concrete joints, tree branches, and surface irregularities with color variations that visually resemble actual cracks, as illustrated in Fig. 3. The inclusion of such challenging HNS alongside true crack samples helps the model distinguish between real cracks and visually similar non-crack textures. This training method enables the model to identify subtle features that define cracks while minimizing false positives. As a result, the model achieves a more reliable level of accuracy in distinguishing real cracks on concrete surfaces. The dataset contains images with varying resolutions, ranging from a high resolution of 1600 × 1200 pixels to a lower resolution of 796 × 716 pixels. For training and evaluation purposes, the dataset has been divided into distinct training and testing sets; specific details are provided in Table 1.

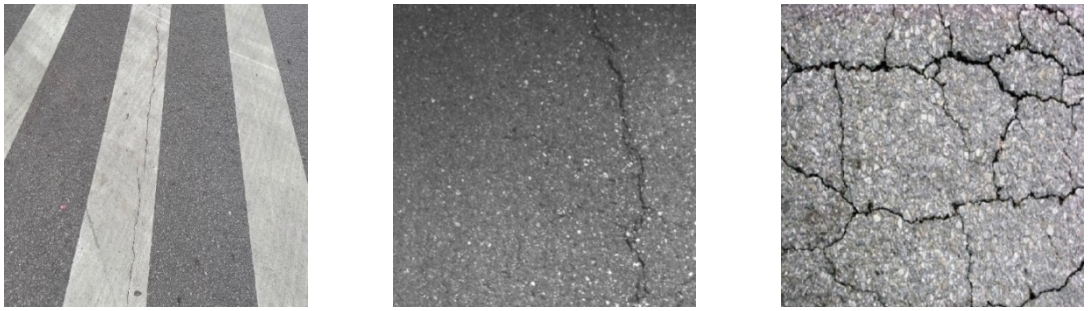
(a) Concrete cracks

(b) HNS

Figure 3. Example images from concrete cracks dataset with HNS

Table 1. Quantitative details of concrete and asphalt datasets

Dataset	Image type	Training images	Validation images	Testing images	Total images
Concrete cracks	Normal cracks	968	196	232	5296
	HNS	3900			
Asphalt cracks	Normal cracks	1076	229	262	2790
	HNS	1223			



(a) Asphalt cracks

(b) HNS

Figure 4. Example images from concrete cracks dataset with HNS

2.3.2 Asphalt crack dataset

The asphalt crack dataset used in this study is a composite of several publicly available datasets, including the Crack Forest Dataset (CFD) [47], Crack-500 [48], DeepCracks [49], and GAPs-384 [50] along with additional images sourced from real asphalt roads. These datasets feature various crack types, including longitudinal, transverse, and alligator patterns. To enhance the model performance, training incorporated unlabeled HNS specific to the asphalt dataset. The challenging samples were curated from the RoboFlow website and sourced from actual asphalt roads in South Korea. The HNS includes elements such as lane markings, grooves, pavement joints, construction seams, repaired interfaces, and other surface discontinuities that visually resemble cracks but were not annotated as crack defects in the reference datasets, as illustrated in Fig. 4. Although some of these structures may represent physical material discontinuities, they are not considered crack labels according to the annotation protocol adopted by the source datasets and were therefore treated as hard negative samples. These HNS are critical for refining the model's ability to differentiate between real cracks and other visual elements. This dual exposure to both real and non-crack patterns improve the accuracy of the model and reduces the number of false positives. Additional details regarding the asphalt dataset, including the numbers of training and testing images, are listed in Table 1.

3. Training configuration and dataset structuring for evaluation

This section presents the training configuration and scenario-based experimental design used to evaluate the proposed semi-supervised framework. All experiments followed a consistent training

setup to ensure comparability, and the evaluation scenarios were structured around different dataset compositions. These include domain-specific configurations for concrete and asphalt cracks, inclusion of HNS in the unlabeled stream, and a mixed-domain setting combining both concrete and asphalt data. The goal of this design was to analyze the effectiveness and generalization capability of the framework across different material types and background complexities.

3.1 Training setup and augmentation strategy

The proposed semi-supervised learning (SSL) method is trained on a workstation equipped with four NVIDIA RTX Titan GPUs, each with 24 GB of memory. Training and testing were conducted in parallel across the GPUs using a distributed training strategy. Stochastic Gradient Descent (SGD) was used as the optimizer, with a momentum of 0.9 and a weight decay of 1e-4 to update the model weights during training. A multistep learning rate schedule was employed starting at 0.001 and adjusted after each iteration using a polynomial learning rate strategy. The model was trained for 240 epochs, and its performance was evaluated after each epoch to monitor the progress. The training resolution was set to 800 × 800 pixels, and an overall batch size of eight was used, in which each minibatch contained eight labeled and eight unlabeled images.

For a computational overhead comparison between the supervised and semi-supervised models, the supervised model required approximately 2 h to complete 240 epochs. In contrast, the semi-supervised method required 12–15 h to complete the same number of epochs, primarily because of the additional processing of unlabeled data and the computation of the unsupervised loss across multiple perturbation branches. Despite the extended training time, the semi-supervised model typically reached its peak performance within the first 50 epochs, corresponding to approximately 4 h of training. An early stopping mechanism was used to terminate the training process when no further performance improvements were observed. In terms of GPU memory consumption, the semi-

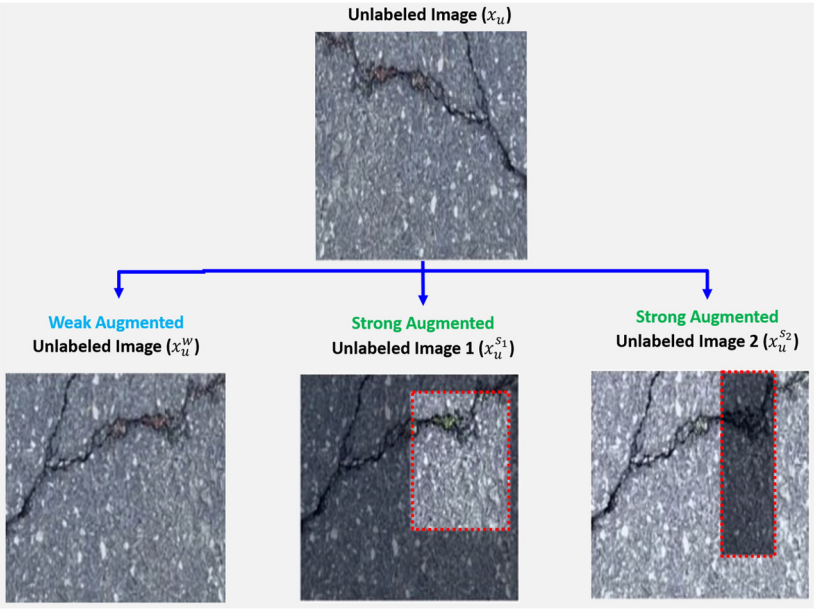


Figure 5. Example of unlabeled data augmentation technique

supervised framework utilized approximately 24 GB of memory per GPU, which was nearly double the memory requirement of the supervised baseline (12 GB per GPU). Although the semi-supervised method introduces additional computational costs, the improvement in segmentation performance, particularly in data-scarce scenarios, demonstrates the practical value of the proposed approach.

Various image-augmentation techniques were applied to the unlabeled dataset to generate weakly and strongly perturbed versions, as illustrated in Fig. 5. The weak augmentation stage involved geometric transformations like random cropping and horizontal flipping, each of which was applied independently with a probability of 0.5. For strong augmentations, a series of photometric transformations were applied following the weak augmentations. First, color jittering was applied with a probability of 0.8, affecting brightness, contrast, saturation, and hue, with the respective parameters set to 0.5, 0.5, 0.5, and 0.25. Subsequently, a random grayscale transformation was applied with a probability of 0.2. To improve generalization and reduce the sensitivity to noise, a blur operation was introduced with a probability of 0.5. Finally, the CutMix technique was applied, in which a random region from one image was selected and pasted onto another. This technique encourages the model to focus on different image regions, thereby enhancing the dataset diversity.

3.2 Scenario-based dataset composition for model evaluation

The model was trained on multiple datasets, including concrete cracks, asphalt cracks, and their combinations, incorporating the HNS to optimize the performance for each dataset. Different experiments were conducted as shown in Table 2. Scenario 1 focused on the detection of concrete cracks and included two scenarios. In Scenario 1-1 (the baseline experiment), both the labeled and unlabeled datasets contained only positive concrete cracks. To assess the impact of HNS, Scenario 1-2 introduced concrete HNS into the unlabeled dataset while keeping the labeled dataset unchanged. This allowed the model to learn from challenging negative samples, thereby improving its ability to distinguish between crack and non-crack instances. Similarly, Scenario 2 focuses on asphalt crack segmentation. In Scenario 2-1 (baseline experiment), the labeled and unlabeled datasets contained only positive asphalt cracks. In Scenario 2-2, asphalt HNS was added to the unlabeled dataset to enhance the performance of the semi-supervised model. Importantly, the paired scenario design (Scenarios 1-1 vs. 1-2, and Scenarios 2-1 vs. 2-2) constitutes an ablation study for the HNS component: by holding the SSL framework, architecture, training configuration, and labeled data constant while varying only the inclusion of HNS in the unlabeled stream, the isolated contribution of HNS can be directly quantified. Finally, Scenario 3 explored the generalization of the model across mixed datasets comprising concrete and asphalt cracks. In Scenario 3-1, both the labeled and unlabeled datasets contained a combination of concrete and asphalt cracks. This scenario aimed to evaluate the performance of a generalized model trained on mixed-domain data, in comparison to models trained exclusively on individual datasets in Scenario 1-1 (concrete) and Scenario 2-1 (asphalt). These experiments examined the effect of unlabeled HNS on the performance of the semi-supervised model and aimed to develop a generalized model for both concrete and asphalt crack datasets.

4. Results and discussion

This section presents the results and discussion for all scenarios, as listed in Table 2, for both the concrete and asphalt crack datasets. The analysis included both quantitative and qualitative evaluations of the model performance across the datasets, with an overall comparison. Although the

Table 2. Analyzing data combinations: An overview of scenarios

Types of cracks	Scenarios	Sub-scenarios	Labeled data	Unlabeled data
Concrete cracks	Scenario 1	Scenario 1-1	Concrete	Concrete
		Scenario 1-2	Concrete	Concrete & HNS
Asphalt cracks	Scenario 2	Scenario 2-1	Asphalt	Asphalt
		Scenario 2-2	Asphalt	Asphalt & HNS
Mixed cracks (Concrete + Asphalt)	Scenario 3	Scenario 3-1	Concrete & Asphalt	Concrete & Asphalt

datasets used in this study contained various crack types, including longitudinal, transverse, and microcracks, they were not annotated with specific subtype labels. Therefore, a per-class analysis could not be performed. Instead, an evaluation was conducted based on the overall segmentation performance across all crack types combined. To evaluate the performance of the proposed method, the mean Intersection over Union (mIoU) metric was employed in all the experiments. The mIoU can be defined as

$$mIoU = \frac{TPs}{TPs + FPs + FNs}. \tag{9}$$

where true positives (TPs) represent the pixels in the prediction that overlap with the true labels, false positives (FPs) refer to the pixels in the prediction that do not overlap with the true labels, and FNs (false negatives) refer to the pixels in the true labels that do not overlap with the prediction, also known as missing crack pixels.

4.1 Concrete cracks

Fig. 6 depicts the performance of both the supervised and semi-supervised models for segmenting concrete cracks across different scenarios. In the semi-supervised model training, 1%, 5%, 10%, and 20% of the data were used as labeled data, whereas the remaining data were used as unlabeled data. In contrast, supervised model training uses only the same labeled data ratios (1%, 5%, 10%, and 20%) without incorporating unlabeled data. The fully supervised model trained with 100% labeled data served as a benchmark for comparison with the semi-supervised models. The transition from 20% labeled data in Scenario 1-1 of the semi-supervised model to 100% labeled data in the fully supervised model represents the model's maximum performance when utilizing all labeled data. In Scenario 1-1, both the supervised and semi-supervised models were trained without incorporating the HNS into the dataset. Starting with only 1% labeled data, the semi-supervised model achieved an mIoU of 63.07%, which is 9.88% higher than that achieved by the supervised model (53.19%). As expected, the performances of both models improved as the amount of labeled data increased. At a 20% labeled data ratio, the semi-supervised model achieved an mIoU of 70.09%, which is only 1.92% lower than that of the fully supervised model (72.01%). Across all data splits, the semi-supervised model consistently outperformed the supervised model, demonstrating the advantages of leveraging unlabeled data during training. This contrast underscores the limitations of supervised learning in scenarios where labeled data are scarce.

In Scenario 1-2, concrete HNS were exclusively added to the unlabeled dataset, whereas the labeled data remained unchanged, as in Scenario 1-1. To assess the impact of HNS on the model

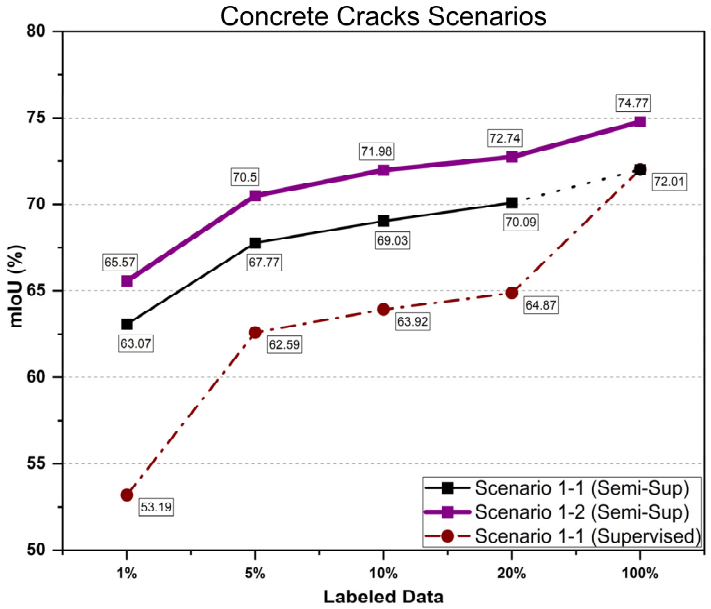


Figure 6. Quantitative comparison of all scenarios of concrete cracks dataset

performance, a consistent number of HNS (3,900 unlabeled HNS images) was incorporated at each data ratio. The inclusion of the HNS improved the performance of the semi-supervised model. At the 20% labeled data ratio, the semi-supervised model in Scenario 1-2 achieved an mIoU of 72.74%, surpassing the fully supervised model's 72.01% mIoU. Furthermore, when 100% labeled data (positive cracks) were used alongside unlabeled HNS, the performance of the semi-supervised model further increased from 72.01% to 74.77% mIoU. The addition of HNS enhanced the ability of the model to distinguish between positive and negative instances, leading to improved segmentation accuracy, indicating that incorporating HNS into unlabeled data not only boosts the performance of semi-supervised models but also reduces reliance on large labeled datasets.

Overall, the results confirm that the adapted semi-supervised framework with HNS is an effective strategy for improving crack segmentation accuracy while minimizing the need for extensive labeled data. Since all other components, architecture, SSL framework, training configuration, and labeled data remain identical between Scenarios 1-1 and 1-2, the observed gains are attributable solely to the inclusion of HNS in the unlabeled stream, confirming its isolated contribution. Additional performance metrics, such as mean Precision, Recall, and F1 score for each data ratio, are provided in Appendix A and Table A1. The precision improvements are particularly notable: HNS primarily suppresses false positives (non-crack regions misclassified as cracks) rather than improving recall, consistent with the role of negative samples in refining decision boundaries. The inclusion of HNS in Scenario 1-2 resulted in higher precision values, indicating reduced false positives and improved discrimination between cracks and the background.

Fig. 7 presents a qualitative analysis of the semi-supervised model's performance on concrete cracks in Scenario 1-1 and Scenario 1-2, using 20% labeled data. Scenario 1-2, which incorporated HNS, demonstrated significant improvements in crack segmentation. The first two rows of Fig. 7 compare performance on simple concrete cracks (positive cracks), where both models from Scenario 1-1 and Scenario 1-2 successfully predicted the cracks, closely matching the ground truth labels. To

evaluate the impact of HNS, the next two rows depict images with more challenging background conditions, such as shadows and texture variations. In the third-row image, where the crack is partially obscured by a shadow, the model trained under Scenario 1-1 failed to detect most of the crack pixels. In contrast, the model trained under Scenario 1-2 successfully identified the cracks, with predictions closely aligned with the true labels. The final row illustrates model performance on a highly complex background, where spalling with rebar exposure and structural elements (e.g., a door joint) resemble actual cracks. In this case, the model in Scenario 1-1 produced numerous false positives. However, the model in Scenario 1-2 was able to identify the true cracks more accurately, with predictions better aligned with the ground truth. Although a small number of false-positive pixels remained, the Scenario 1-2 model demonstrated significantly improved performance in isolating crack features from background noise. Overall, the qualitative comparison between Scenario 1-1 and Scenario 1-2 shows that incorporating HNS into the unlabeled dataset enhances

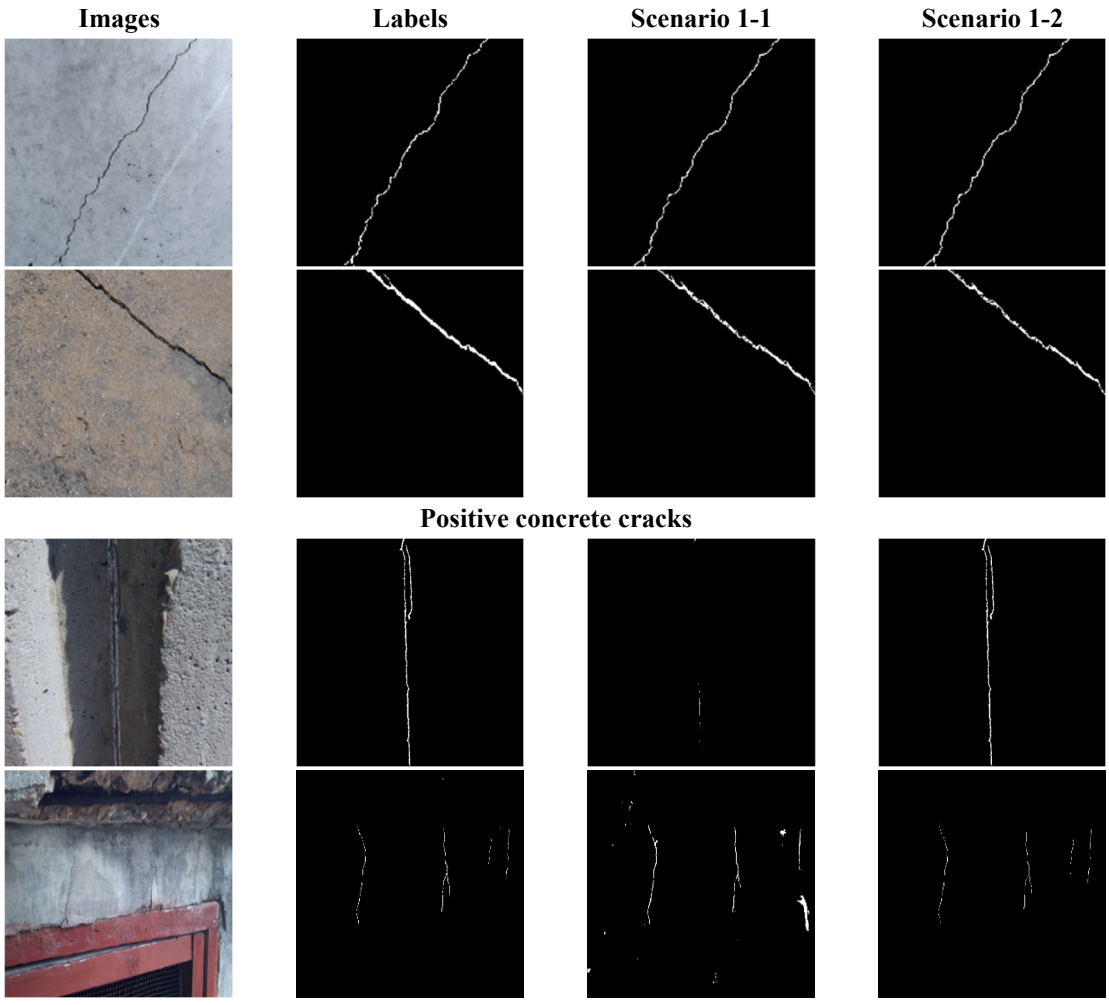


Figure 7. Comparison of prediction results of concrete crack scenarios in both positive and HNS cracks

the model's robustness and its ability to handle complex crack segmentation scenarios. While both models performed well on simple cracks, Scenario 1-2 consistently outperformed Scenario 1-1 in visually complex environments, reducing false positives and improving prediction accuracy.

4.2 Asphalt cracks

Fig. 8 compares the supervised and semi-supervised models of asphalt cracks. Similar to the concrete crack dataset, labeled data ratios of 1%, 5%, 10%, and 20% were used, and the remaining data were utilized as unlabeled data to train the semi-supervised model. In Scenario 2-1, both the supervised and semi-supervised models were trained without incorporating the HNS into the dataset. With 1% labeled data, the semi-supervised model achieved an mIoU of 65.15%, which is 4.11% higher than the supervised model's mIoU of 61.04%. As expected, increasing the amount of labeled data improved the performance of both models. However, the semi-supervised model consistently outperformed the supervised model across all labeled data ratios, particularly between 1% and 10%, demonstrating the effectiveness of leveraging unlabeled data in low-annotation settings. At a 20% labeled data ratio, the semi-supervised model achieved an mIoU of 70.59%, which was only 2.31% lower than that of the fully supervised model (72.90%).

In Scenario 2-2, HNS were exclusively included in the unlabeled dataset, whereas the labeled data ratios remained the same as in Scenario 2-1. A consistent set of 1,223 HNS was added to the unlabeled dataset at each data split to evaluate its impact on model performance. The inclusion of HNS significantly improved the model's performance. For instance, at the 1% labeled data split, the semi-supervised model with HNS achieved a mIoU of 68.64%, representing a 3.49% absolute gain over Scenario 2-1 (65.15%) and a 7.60% improvement over the supervised model (61.04%). At the 20% labeled data ratio, the semi-supervised model in Scenario 2-2 achieved an mIoU of 73.11%, outperforming both the fully supervised model (72.90%) and the semi-supervised model in Scenario

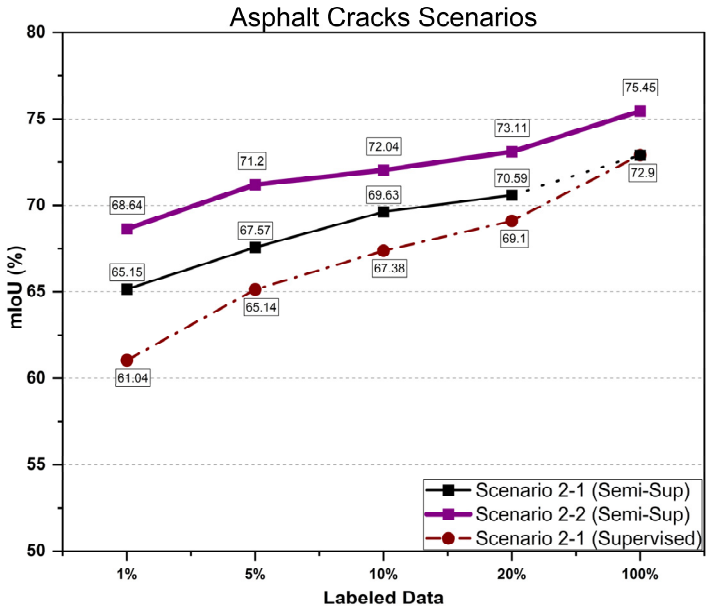


Figure 8. Quantitative comparison of all scenarios of asphalt cracks dataset

2-1 (70.59%). Furthermore, when 100% of the labeled data were combined with unlabeled HNS in the semi-supervised model, the mIoU increased from 72.90% (fully supervised model) to 75.45%, thereby demonstrating the effectiveness of the HNS to help the model differentiate between positive and negative crack instances, leading to enhanced segmentation performance. The corresponding values of the mean Precision, Recall, and F1 score across all label ratios are presented in Appendix A and Table A2. In particular, the improved precision of the HNS-enhanced models highlights the ability of the framework to suppress false positives in challenging asphalt backgrounds. Since all training components, architecture, SSL framework, training configuration, and labeled data remain identical between Scenarios 2-1 and 2-2, the observed performance gains are attributable solely to the inclusion of HNS in the unlabeled stream, confirming their isolated contribution in the asphalt domain. Consistent with the concrete results in Section 4.1, the gains manifest more strongly in precision than in recall, reflecting the role of negative samples in tightening the decision boundary against false positives rather than improving sensitivity to crack pixels.

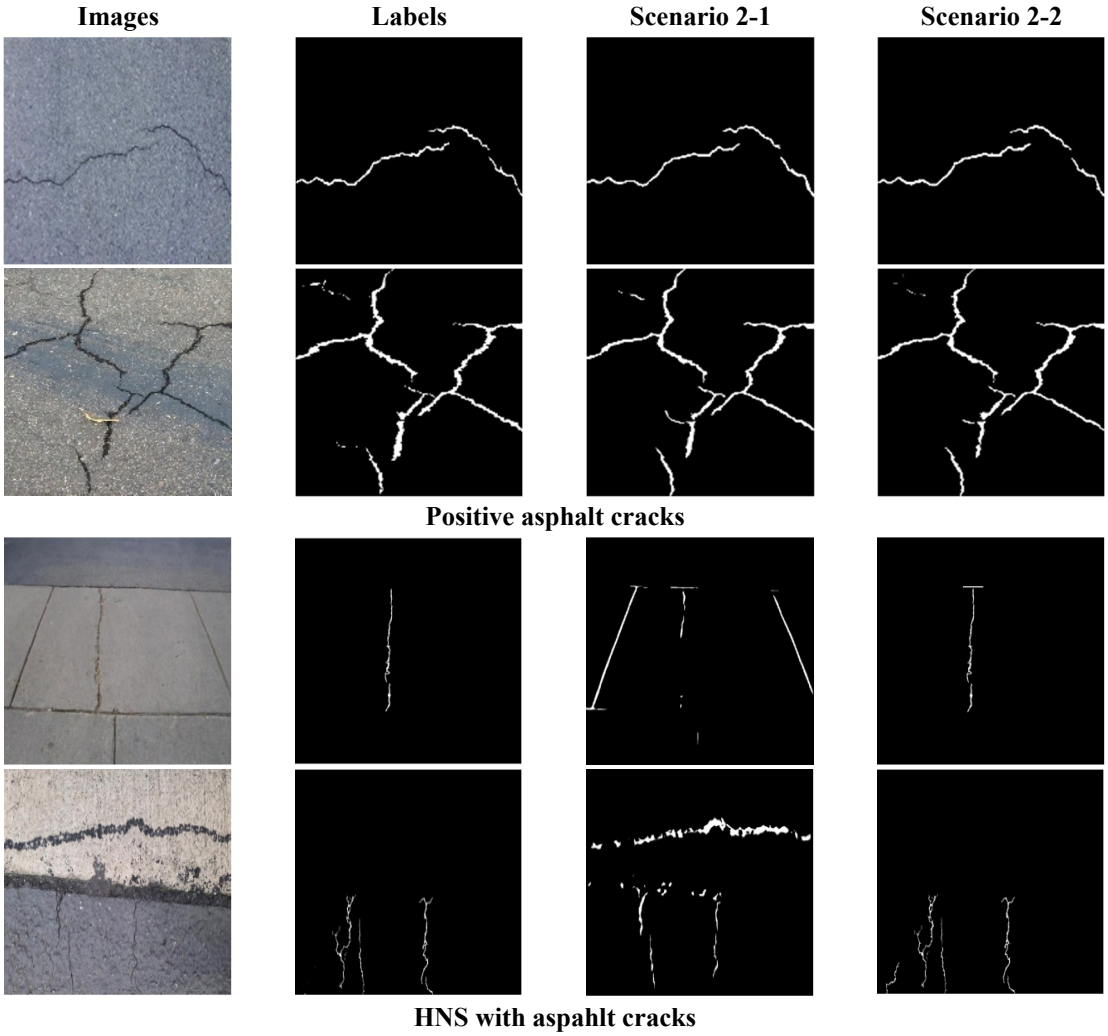


Figure 9. Comparison of prediction results of asphalt crack scenarios in both positive and HNS cracks

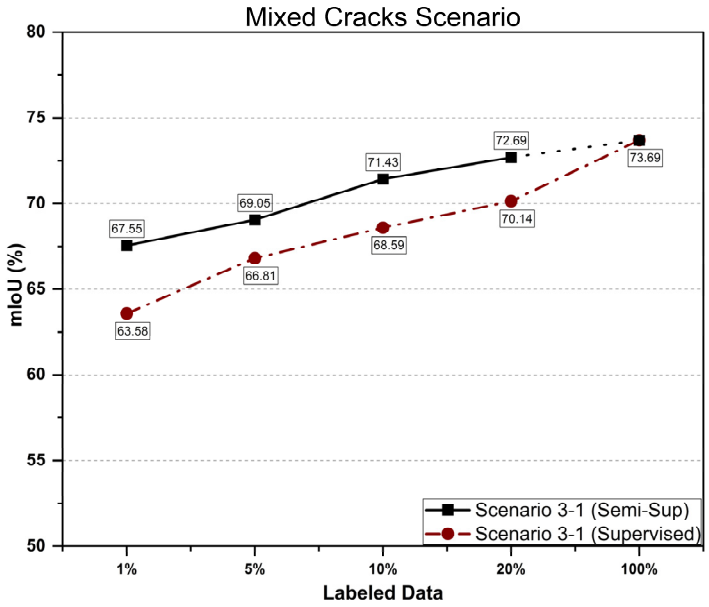
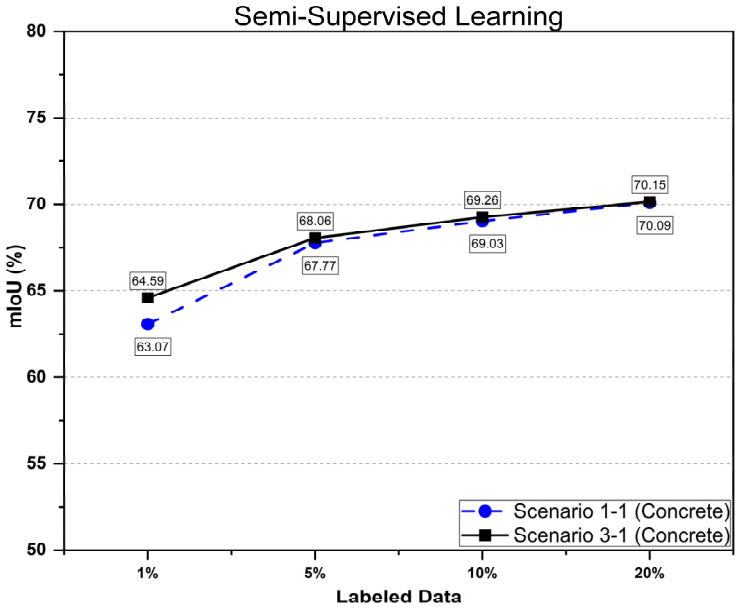


Figure 10. Quantitative comparison of supervised and semi-supervised models of mixed cracks dataset

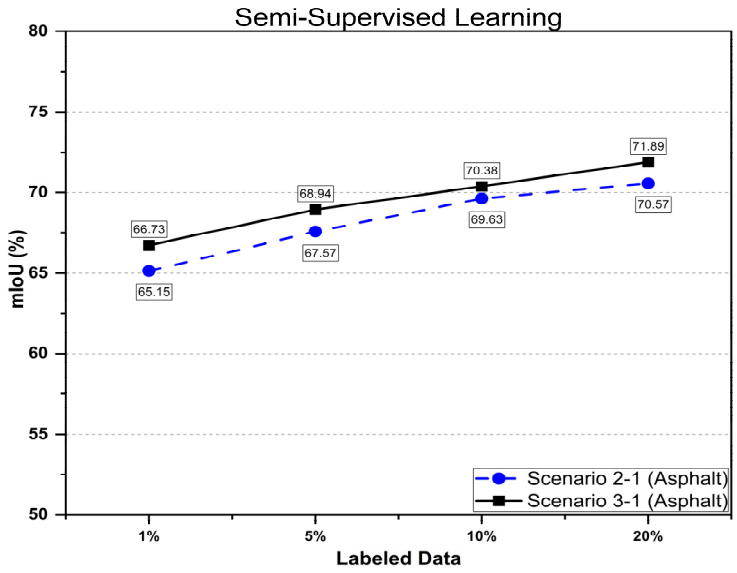
Fig. 9 illustrates a qualitative comparison of the performance of the semi-supervised model in Scenario 2-1 and Scenario 2-2, using 20% labeled data, with Scenario 2-2 incorporating HNS. The results demonstrate noticeable improvements in crack segmentation when HNS were included in the unlabeled dataset. The first two rows of Fig. 9 compare model performance on simple asphalt cracks (positive cracks), where both the Scenario 2-1 and Scenario 2-2 models produced predictions that closely matched the ground truth. The next two rows evaluate model behavior on more complex crack images with challenging background conditions. In the third-row image, the scene includes an asphalt road adjacent to a concrete pavement with crack separations and joints. The Scenario 2-1 model misclassified the pavement joints as cracks and failed to detect the actual crack pixels. In contrast, the Scenario 2-2 model more accurately predicted the crack pixels, with only a small number of false positives. The final row presents a highly complex scene involving cracks on an asphalt surface, a concrete pavement, a separation between the two surfaces, and charcoal stains that visually resemble cracks. While the Scenario 2-1 model detected some crack pixels, it misclassified the surface separation and charcoal stains as cracks, leading to a high false positive rate. Conversely, the Scenario 2-2 model successfully identified the actual crack pixels and significantly reduced false positives. It demonstrated a better ability to distinguish true cracks from visual elements that mimic crack patterns, such as stains and material separations. Overall, the qualitative comparison between Scenario 2-1 and Scenario 2-2 confirms that incorporating HNS into the unlabeled dataset significantly enhances the model's robustness. While both models performed well on simple cracks, the Scenario 2-2 model consistently outperformed Scenario 2-1 in complex environments by reducing false positives and improving segmentation accuracy.

4.3 Mixed cracks (Concrete and asphalt)

Fig. 10 illustrates the performance of the supervised and semi-supervised models in Scenario 3-1, which utilized both concrete and asphalt crack datasets. The goal of this scenario was to develop



(a)



(b)

Figure 11. Overall comparison of scenario 3 with scenarios 1 and 2 (a) Concrete cracks, (b) Asphalt cracks

a generalized and robust semi-supervised model capable of segmenting cracks across both surface types. In Scenario 3-1, the model was trained using the combined labeled and unlabeled data from Scenarios 1-1 and 2-1, without including any HNS. The test sets used for evaluation were the same as those in Scenarios 1-1 (concrete) and 2-1 (asphalt), allowing for consistent and direct comparison. As expected, the performance of both the supervised and semi-supervised models improved with increasing amounts of labeled data. However, the mixed-crack semi-supervised model consistently

achieved higher mIoU scores than the corresponding supervised model across all labeled data ratios. By leveraging a larger pool of unlabeled data, the semi-supervised model was able to gain additional insights during training. For example, with only 1% labeled data, the mixed-crack semi-supervised model achieved an mIoU of 67.55%, outperforming the supervised model, which reached 63.58%. At the 20% labeled data ratio, the semi-supervised model achieved an mIoU of 72.69%, just 1.00% below the fully supervised model trained with 100% labeled data (73.69%). These results demonstrate the effectiveness of the semi-supervised learning approach in utilizing unlabeled data, even when combining different crack types such as concrete and asphalt. A detailed analysis of the Precision, Recall, and F1 scores for Scenario 3-1 is provided in Appendix A and Table A3. These metrics further confirm that the model trained on mixed-domain data maintains strong generalization capabilities while achieving balanced precision and recall across surface types.

The impact of mixing crack datasets from different domains (concrete and asphalt) on model performance is further analyzed in Fig. 11, which evaluates the generalization ability of the mixed-crack model by comparing its performance on individual crack types against models trained specifically for each material. Fig. 11(a) compares the semi-supervised mixed-crack model with the concrete-specific model from Scenario 1-1 when evaluated on concrete cracks. The results indicate that the mixed-crack model slightly outperformed the concrete-specific model at each labeled data ratio, suggesting that combining the two crack types during training did not negatively affect performance. For example, with 1% labeled data, the mixed model achieved an mIoU of 64.59%, compared to 63.07% for the concrete-only model. Fig. 11(b) compares the performance of the mixed-crack model on asphalt cracks against the asphalt-specific model from Scenario 2-1. Here too, the mixed-crack model consistently outperformed the asphalt-specific model across all labeled data ratios. For instance, with 1% labeled data, the mixed model achieved an mIoU of 66.73%, compared to 65.15% for the asphalt-only model. These modest yet consistent gains suggest that exposure to a broader range of crack patterns enhanced the model's generalization capability without signs of overfitting or performance inflation. Since no HNS were included and there was no overlap between training and testing images, the improvements are more likely attributable to the increased diversity of training data rather than data leakage. In summary, training with a mixed crack dataset proves beneficial in applications requiring crack monitoring across both concrete and asphalt surfaces, such as in bridges or multi-surface infrastructure, enabling comprehensive and adaptable segmentation across varied real-world conditions.

5. Comparison with other semi-supervised methods

To further evaluate the effectiveness of the adapted semi-supervised framework, we conducted comparative experiments against three state-of-the-art semi-supervised segmentation methods: FixMatch [29], U2PL [32], and AugSeg [30]. All methods were implemented using the same encoder–decoder architecture, DeepLabV3+ with a ResNet-101 backbone, to ensure a fair comparison. The experiments were conducted under four training scenarios involving both concrete and asphalt crack datasets, using only 20% of the labeled data, with the remaining 80% treated as unlabeled data.

Table 3 presents the mIoU values for each method under Scenarios 1-1 and 1-2 (concrete cracks) and Scenarios 2-1 and 2-2 (asphalt cracks). The proposed method consistently outperformed the baseline semi-supervised approaches across all scenarios, achieving the highest mIoU in every case. Notably, while competing methods such as FixMatch, U2PL, and AugSeg exhibited variable

Table 3. Comparison with other semi-supervised methods

Methods	Concrete Cracks		Asphalt Cracks	
	Scenario 1-1 mIoU (%)	Scenario 1-2 mIoU (%)	Scenario 2-1 mIoU (%)	Scenario 2-2 mIoU (%)
FixMatch	67.29	69.27	66.90	68.49
U2PL	66.78	68.15	68.23	69.47
AugSeg	68.00	70.05	67.44	69.45
Ours	**70.09**	**72.74**	**70.59**	**73.11**

performance depending on the scenario, the proposed method maintained robust and stable results across both datasets, with and without the inclusion of HNS. For example, in Scenario 1-2 (concrete with HNS), the proposed method achieved an mIoU of 72.74%, compared to 70.05% for AugSeg, 69.27% for FixMatch, and 68.15% for U2PL. Similarly, in Scenario 2-2 (asphalt with HNS), the proposed method achieved an mIoU of 73.11%, clearly outperforming the other methods, highlighting the generalization capability and consistency of the adapted framework in addressing real-world crack segmentation challenges under limited supervision, particularly in the presence of complex backgrounds and visually deceptive patterns.

6. Conclusions

This study presented an adapted semi-supervised learning framework for crack segmentation that extends an established consistency regularization paradigm with domain-aware HNS integration, designed to reduce annotation costs while maintaining high accuracy in challenging environments. The method effectively employed a weak-to-strong dual-stream perturbation strategy for consistency regularization, enabling the model to extract meaningful information from unlabeled data. A key contribution was the incorporation of HNS exclusively into the unlabeled dataset, which significantly enhanced the robustness of the model in segmenting complex crack patterns across both concrete and asphalt surfaces. Extensive experiments demonstrated the critical role of HNS in improving model performance under challenging segmentation conditions. To develop a generalized model applicable to both surface types, additional experiments were conducted by combining the concrete and asphalt crack datasets. The performance of the mixed-domain model was compared with models trained separately on each dataset, highlighting the potential benefits of dataset integration. Furthermore, the performance of the method was compared with other state-of-the-art semi-supervised methods. Based on the experimental results, the following conclusions can be drawn:

Concrete cracks:

- The semi-supervised model with HNS achieved an mIoU of 72.74% using only 20% labeled data, slightly outperforming the fully supervised model that achieved an mIoU of 72.01% with 100% labeled data.
- Further performance improvements were observed when the semi-supervised model was trained using 100% of the labeled data while incorporating HNS exclusively into the unlabeled dataset, achieving an mIoU of 74.77%.

Asphalt cracks:
- Similarly, the semi-supervised model with HNS achieved an mIoU of 73.11% using only 20% of the labeled data, surpassing the fully supervised model's mIoU of 72.90%.
- Additionally, when the semi-supervised model was trained with 100% labeled data and HNS in the unlabeled dataset, the model's performance improved further, achieving an mIoU of 75.45%.

Semi-supervised vs supervised models:
- Across both the concrete and asphalt crack datasets, the semi-supervised model consistently outperformed the supervised model for each labeled data ratio.
- These results highlight the effectiveness of the proposed SSL approach in leveraging unlabeled data, significantly reducing the reliance on labeled data and lowering annotation costs by approximately 80%.

Mixed cracks:
- The semi-supervised model trained on a mixed crack dataset (concrete and asphalt cracks) exhibited a similar performance trend, outperforming the supervised model for each data ratio.
- When compared with models trained separately on individual datasets, the mixed dataset model showed a slight improvement, suggesting that combining datasets is an effective approach, particularly when a generalized model is needed to detect cracks across both concrete and asphalt surfaces.

Comparison with other semi-supervised methods:
- The adapted SSL method consistently outperformed state-of-the-art approaches, including FixMatch, U2PL, and AugSeg, across all scenarios.
- Particularly in Scenario 1-2 and Scenario 2-2, which used HNS in both concrete and asphalt datasets, the proposed method achieved the highest mIoU values, suggesting improved robustness and generalization capability under challenging conditions, attributable primarily to the HNS training strategy rather than to a fundamentally different SSL formulation.

In summary, the adapted SSL framework demonstrated strong potential for advancing crack segmentation in structural inspection tasks, particularly in data-scarce environments. The integration of domain-aware HNS into the unlabeled dataset significantly improved model performance, especially under complex background conditions, and constitutes the primary practical contribution of this study. These findings emphasize the potential of adapting established SSL frameworks with domain-specific training strategies to enhance segmentation accuracy, reduce reliance on large labeled datasets, and ultimately contribute to more efficient infrastructure maintenance and monitoring.

Despite the encouraging results, several limitations remain. The hard negative samples used in this study were manually curated based on domain knowledge, which requires additional effort and may introduce subjective selection bias. Furthermore, the proposed framework was evaluated using concrete and asphalt datasets, and further validation on other infrastructure types and environmental conditions would strengthen its generalizability. Future work will investigate automated hard negative sample mining and evaluate the framework across broader structural inspection datasets.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (RS-2025-00522517).

Appendix A

Fig. A1 illustrates the proposed method mIoU performance across different confidence thresholds (τ) ranging from 0.0 to 0.5. The evaluation covered both concrete and asphalt crack scenarios, with and without HNS. The results indicate that the model maintains stable performance at lower thresholds, but mIoU decreases slightly as threshold values increase. This suggests that a threshold of zero provides a favorable balance between retaining pseudo-labels and maintaining segmentation accuracy, particularly in visually complex and imbalanced datasets.

To complement the primary evaluation metric, mIoU, presented in the main text, additional

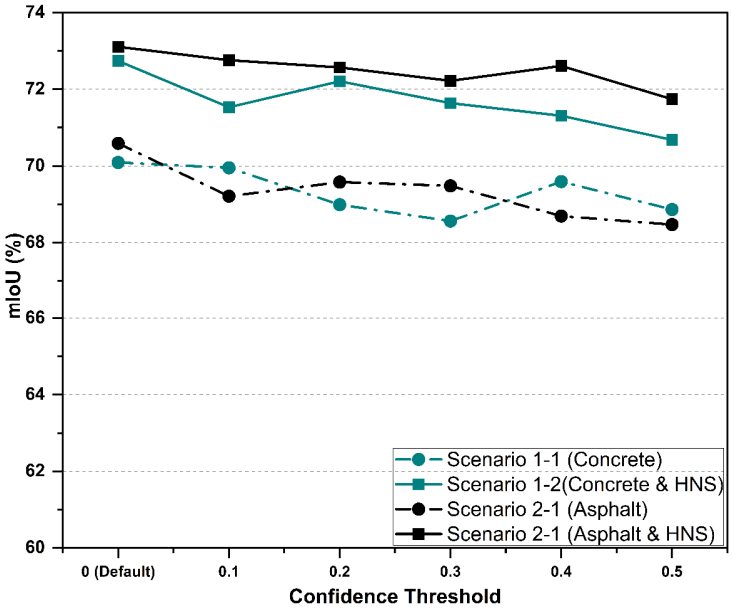


Figure A1. mIoU vs. confidence threshold τ across concrete and asphalt crack scenarios

Table A1. Evaluation results of concrete crack scenarios

Labeled data	Scenario 1-1				Scenario 1-2			
	mPrecision (%)	mRecall (%)	mF1 (%)	mIoU (%)	mPrecision (%)	mRecall (%)	mF1 (%)	mIoU (%)
1%	63.78	76.19	69.43	63.07	69.38	71.62	70.48	65.57
5%	67.98	77.54	72.45	67.77	75.78	80.38	78.01	70.50
10%	75.51	76.19	75.85	69.03	80.52	78.19	79.34	71.98
20%	78.36	79.86	79.10	70.09	83.74	79.46	81.54	72.74

Table A2. Evaluation results of asphalt crack scenarios

Labeled data	Scenario 2-1				Scenario 2-2			
	mPrecision (%)	mRecall (%)	mF1 (%)	mIoU (%)	mPrecision (%)	mRecall (%)	mF1 (%)	mIoU (%)
1%	67.12	74.16	70.46	65.15	74.87	73.54	74.20	68.64
5%	67.26	77.13	71.86	67.57	80.32	77.36	78.81	71.20
10%	71.53	79.16	75.15	69.63	83.19	78.44	80.75	72.04
20%	75.27	81.86	78.43	70.59	87.66	77.91	82.50	73.11

Table A3. Evaluation results of mixed crack scenario

Labeled data	Scenario 3-1			
	mPrecision (%)	mRecall (%)	mF1 (%)	mIoU (%)
1%	69.35	75.11	72.12	67.55
5%	73.64	76.28	74.94	69.05
10%	78.39	80.36	79.36	71.43
20%	82.80	78.93	80.82	72.69

performance metrics, including the mean precision, mean Recall, and F1 score, across all experimental scenarios and labeled data ratios were explored. These metrics offer deeper insight into the behavior of the model in distinguishing cracks from background noise. In particular, high precision in hard-negative sample (HNS) scenarios reflects improved suppression of false positives, whereas high recall indicates the effective detection of fine or subtle cracks. The results in Tables A1, A2, and A3 demonstrate consistent improvements across all metrics, reinforcing the robustness and generalization ability of the proposed semi-supervised framework under limited supervision.

References

1. Zhao, H., Qin, G., Wang, X. (2010). Improvement of canny algorithm based on pavement edge detection. Proceedings of the 2010 3rd International Congress on Image and Signal Processing, Yantai, China, October, 2, 964-967. https://doi.org/10.1109/CISP.2010.5646923.
2. Zhu, S., Xia, X., Zhang, Q., Belloulata, K. (2007). An image segmentation algorithm in image processing based on threshold segmentation. Proceedings of the 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, 673-678.
3. Kabir, S., Rivard, P., He, D.C., Thivierge, P. (2009). Damage assessment for concrete structure using image processing techniques on acoustic borehole imagery. Construction and Building Materials, 23(10), 3166-3174. https://doi.org/10.1016/j.conbuildmat.2009.06.013.
4. Abdel-Qader, I., Abudayyeh, O., Kelly, M.E. (2003). Analysis of edge-detection techniques for crack identification in bridges. Journal of Computing in Civil Engineering, 17(4), 255-263. https://doi.org/10.1061/(asce)0887-3801(2003)17:4(255).
5. Flah, M., Suleiman, A.R., Nehdi, M.L. (2020). Classification and quantification of cracks in concrete structures using deep learning image-based techniques. Cement and Concrete Composites, 114, 103781. https://doi.org/10.1016/j.cemconcomp.2020.103781.
6. Zhou, S., Song, W. (2020). Deep learning-based roadway crack classification using laser-scanned range images: A comparative study on hyperparameter selection. Automation in Construction, 114, 103171.

https://doi.org/10.1016/j.autcon.2020.103171.
7. Yang, C., Chen, J., Li, Z., Huang, Y. (2021). Structural crack detection and recognition based on deep learning. Applied Sciences, 11(6), 2868. https://doi.org/10.3390/app11062868.
8. Kim, B., Cho, S. (2018). Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. Sensors, 18(10), 3452. https://doi.org/10.3390/s18103452.
9. Wang, L., Kawaguchi, K.I., Wang, P. (2020). Damaged ceiling detection and localization in large-span structures using convolutional neural networks. Automation in Construction, 116, 103230. https://doi.org/10.1016/j.autcon.2020.103230.
10. Ali, L., Alnajjar, F., Jassmi, H.A., Gocho, M., Khan, W., Serhani, M.A. (2021). Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. Sensors, 21(5), 1688. https://doi.org/10.3390/s21051688.
11. Wang, S., Nishio, M. (2024). Review for vision-based structural damage evaluation in disasters focusing on nonlinearity. Smart Structures and Systems, 33(4), 263-279. https://doi.org/10.12989/sss.2024.33.4.263.
12. Dung, C.V. (2019). Autonomous concrete crack detection using deep fully convolutional neural network. Automation in Construction, 99, 52-58. https://doi.org/10.1016/j.autcon.2018.11.028.
13. Zhou, S., Song, W. (2020). Concrete roadway crack segmentation using encoder-decoder networks with range images. Automation in Construction, 120, 103403. https://doi.org/10.1016/j.autcon.2020.103403.
14. Hsieh, Y.A., Tsai, Y.J. (2020). Machine learning for crack detection: Review and model performance comparison. Journal of Computing in Civil Engineering, 34(5), 04020038. https://doi.org/10.1061/(asce)cp.1943-5487.0000918.
15. Alipour, M., Harris, D.K., Miller, G.R. (2019). Robust pixel-level crack detection using deep fully convolutional neural networks. Journal of Computing in Civil Engineering, 33(6), 04019040. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000854.
16. Sun, X., Xie, Y., Jiang, L., Cao, Y., Liu, B. (2022). DMA-Net: DeepLab with multi-scale attention for pavement crack segmentation. IEEE Transactions on Intelligent Transportation Systems, 23(10), 18392-18403. https://doi.org/10.1109/TITS.2022.3158670.
17. Chu, H., Wang, W., Deng, L. (2022). Tiny-Crack-Net: A multiscale feature fusion network with attention mechanisms for segmentation of tiny cracks. Computer-Aided Civil and Infrastructure Engineering, 37(14), 1914-1931. https://doi.org/10.1111/mice.12881.
18. Tran, T.S., Nguyen, S.D., Lee, H.J., Tran, V.P. (2023). Advanced crack detection and segmentation on bridge decks using deep learning. Construction and Building Materials, 400, 132839. https://doi.org/10.1016/j.conbuildmat.2023.132839.
19. Yang, L., Bai, S., Liu, Y., Yu, H. (2023). Multi-scale triple-attention network for pixelwise crack segmentation. Automation in Construction, 150, 104853. https://doi.org/10.1016/j.autcon.2023.104853.
20. Zhang, J., Yang, D., Cai, Y.Y., Xu, E.D., Yuan, Y., Wang, Y.J. (2022). Lightweight deep learning-based automated concrete surface inspection: Crack classification and pixel-level segmentation. Smart Structures and Systems, 35(3), 163-175. https://doi.org/10.12989/sss.2025.35.3.163.
21. Zhang, J., Li, D., Zeng, Z., Zhang, R., Wang, J. (2025). Dual-branch crack segmentation network with multi-shape kernel based on convolutional neural network and Mamba. Engineering Applications of Artificial Intelligence, 150, 110536. https://doi.org/10.1016/j.engappai.2025.110536.
22. Lee, D.H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: ICML 2013 Workshop: Challenges in Representation Learning, 3(2), 896.
23. Jeong, J., Lee, S., Kim, J., Kwak, N. (2019). Consistency-based semi-supervised learning for object detection. Advances in Neural Information Processing Systems, 32.
24. Xie, Q., Dai, Z., Hovy, E., Luong, T., Le, Q. (2020). Unsupervised data augmentation for consistency training. Advances in Neural Information Processing Systems, 33, 6256-6268.
25. DeVries, T., Taylor, G.W. (2017). Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552.
26. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. Proceedings of the 2019 IEEE/CVF International Conference

on Computer Vision (ICCV), 6023-6032.
27. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C. (2019). MixMatch: A holistic approach to semi-supervised learning. Advances in Neural Information Processing Systems, 32, 1-14.
28. Olsson, V., Tranheden, W., Pinto, J., Svensson, L. (2021). Classmix: Segmentation-based data augmentation for semi-supervised learning. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 1369-1378.
29. Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems, 33, 596-608.
30. Zhao, Z., Yang, L., Long, S., Pi, J., Zhou, L., Wang, J. (2023). Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11350-11359. https://doi.org/10.1109/cvpr52729.2023.01092.
31. Chen, X., Yuan, Y., Zeng, G., Wang, J. (2021). Semi-supervised semantic segmentation with cross pseudo supervision. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2613-2622. https://doi.org/10.1109/CVPR46437.2021.00264.
32. Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X. (2022). Semi-supervised semantic segmentation using unreliable pseudo-labels. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4248-4257. https://doi.org/10.1109/CVPR52688.2022.00421.
33. Hu, H., Wei, F., Hu, H., Ye, Q., Cui, J., Wang, L. (2021). Semi-supervised semantic segmentation via adaptive equalization learning. Advances in Neural Information Processing Systems, 34, 22106-22118.
34. Guo, J., Wang, Q., Li, Y. (2021). Semi-supervised learning based on convolutional neural network and uncertainty filter for façade defects classification. Computer-Aided Civil and Infrastructure Engineering, 36(3), 302-317. https://doi.org/10.1111/mice.12632.
35. Chun, C., Ryu, S.K.(2019). Road surface damage detection using fully convolutional neural networks and semi-supervised learning. Sensors, 19(24), 5501. https://doi.org/10.3390/s19245501.
36. He, D., Xu, K., Zhou, P., Zhou, D. (2019). Surface defect classification of steels with a new semi-supervised learning method. Optics and Lasers in Engineering, 117, 40-48. https://doi.org/10.1016/j.optlaseng.2019.01.011.
37. Li, G., Wan, J., He, S., Liu, Q., Ma, B. (2020). Semi-supervised semantic segmentation using adversarial learning for pavement crack detection. IEEE Access, 8, 51446-51459. https://doi.org/10.1109/ACCESS.2020.2980086.
38. Shim, S., Kim, J., Cho, G.C., Lee, S.W. (2020). Multiscale and adversarial learning-based semi-supervised semantic segmentation approach for crack detection in concrete structures. IEEE Access, 8, 170939-170950. https://doi.org/10.1109/ACCESS.2020.3022786.
39. Shim, S., Kim, J., Lee, S.W., Cho, G.C. (2022). Road damage detection using super-resolution and semi-supervised learning with generative adversarial network. Automation in Construction, 135, 104139. https://doi.org/10.1016/j.autcon.2022.104139.
40. Mohammed, M.A., Han, Z., Li, Y., Al-Huda, Z., Li, C., Wang, W. (2022). End-to-end semi-supervised deep learning model for surface crack detection of infrastructures. Frontiers in Materials, 9, 1058407. https://doi.org/10.3389/fmats.2022.1058407.
41. Wang, W., Su, C. (2021). Semi-supervised semantic segmentation network for surface crack detection. Automation in Construction, 128, 103786. https://doi.org/10.1016/j.autcon.2021.103786.
42. Tang, W., Mondal, T.G., Wu, R.T., Subedi, A., Jahanshahi, M.R. (2023). Deep learning-based post-disaster building inspection with channel-wise attention and semi-supervised learning. Smart Structures and Systems, 31(4), 365-381. https://doi.org/10.12989/sss.2023.31.4.365.
43. Xiang, C., Gan, V.J., Guo, J., Deng, L. (2023). Semi-supervised learning framework for crack segmentation based on contrastive learning and cross pseudo supervision. Measurement, 217, 113091. https://doi.org/10.1016/j.measurement.2023.113091.

44. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y. (2023). Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7236-7246. https://doi.org/10.1109/cvpr52729.2023.00699.
45. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49.
46. Bello, I., Fedus, W., Du, X., Cubuk, E.D., Srinivas, A., Lin, T.Y., Shlens, J., Zoph, B. (2021). Revisiting resnets: Improved training and scaling strategies. Advances in Neural Information Processing Systems, 34, 22614-22627.
47. Shi, Y., Cui, L., Qi, Z., Meng, F., Chen, Z. (2016). Automatic road crack detection using random structured forests. IEEE Transactions on Intelligent Transportation Systems, 17(12), 3434-3445. https://doi.org/10.1109/TITS.2016.2552248.
48. Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., Ling, H. (2019). Feature pyramid and hierarchical boosting network for pavement crack detection. IEEE Transactions on Intelligent Transportation Systems, 21(4), 1525-1535. https://doi.org/10.1109/TITS.2019.2910595.
49. Liu, Y., Yao, J., Lu, X., Xie, R., Li, L. (2019). DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing, 338, 139-153. https://doi.org/10.1016/j.neucom.2019.01.036.
50. Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., Ebersbach, D., Stoeckert, U., Gross, H.M. (2017). How to get pavement distress detection ready for deep learning? A systematic approach. 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, May, 2039-2047. https://doi.org/10.1109/IJCNN.2017.7966101.