Berger Scientfic Press

Innovation & Technology Advances, 2025, 3(2), 57-72
https://bergersci.com/index.php/ita/index
DOI: 10.61187/ita.v3i2.262

Small-Sample Object Detection of Surface Cracks in Concrete Structures of High-Rise Buildings via Multi-Level Transfer Learning

Junhong Wu[1], Ling Luo[2,*], Ni Liao[2]

[1]Southeast University Architectural Design and Research Institute Co., Ltd., Nanjing, Jiangsu, 210096, China
[2]School of Civil and Environmental Engineering, Chengdu Jincheng College, Chengdu, Sichuan, 610000, China

*Correspondence: wjhde163@126.com

Abstract: To ensure realistic crack effects on the complex surfaces of high-rise concrete structures and meet the demands of small-sample target detection, a small-sample target detection method for surface cracks in high-rise concrete structures is proposed under multi-level transfer learning. A two-dimensional maximum entropy threshold segmentation method is employed to segment images of high-rise concrete structures. After obtaining the target image, crack connectivity area filtering and crack linearity and rectangularity filtering are applied to remove isolated noise points. A multi-level transfer learning architecture is constructed by integrating multi-scale hybrid temporal convolutional networks, long short-term memory neural networks, and Attention mechanisms to generate distinct transfer learning hidden layers. Processed images are input as source domain data into this architecture, enabling knowledge transfer through the generated multi-level hidden layers. After small-sample hierarchical training, shared features between source and target domains are extracted. A cosine classifier outputs the crack category detection results for high-rise concrete structures. Test results demonstrate that this method accurately captures the irregular contours of mesh cracks and effectively distinguishes crack regions from backgrounds. It efficiently removes isolated point noise in images, maintaining smoothness metrics consistently between 0.008 and 0.015. The approach adapts to detecting cracks of diverse morphologies and categories.

Keywords: Multi-level transfer learning; High-rise buildings; Concrete structures; Surface cracks; Small samples; Object detection

How to cite this paper: Wu, J., Luo, L., Liao, N. Small-Sample Object Detection of Surface Cracks in Concrete Structures of High-Rise Buildings via Multi-Level Transfer Learning. *Innovation & Technology Advances*, 2025, *3*(2), 57–72. https://doi.org/10.61187/ita.v3i2.262

1. Introduction

Surface cracks in high-rise concrete structures serve as early warning signs of structural failure. If not promptly detected and repaired, these cracks may trigger chain reactions—such as reduced load-bearing capacity and reinforcement corrosion—under external forces like earthquakes [1], ultimately leading to catastrophic consequences like building collapse. However, cracks in high-rise concrete structures exhibit complex and diverse morphologies, including shrinkage cracks [2], temperature cracks, settlement cracks, and others. The causes, morphological characteristics, and severity of damage vary significantly among different crack types. Furthermore, the inspection environment is challenging: the building facades are subject to interference from factors like changing sunlight, surface stains, and material inhomogeneity, resulting in crack images with low contrast and high noise. Furthermore, data acquisition is costly [3]. Annotating crack samples in high-rise buildings requires involvement from professional structural engineers. Restrictions on building privacy protection limit the scale of publicly available datasets, resulting in a scarcity of annotated samples—a classic case of small-sample data. Traditional crack detection methods face significant limitations in small-sample object detection due

to insufficient diversity and class imbalance in the data. They exhibit low detection efficiency and susceptibility to subjective factors. Particularly in high-rise building facade inspections, the risks associated with high-altitude operations hinder large-scale implementation. Against this backdrop, achieving small-sample crack detection on high-rise concrete structures has become a critical research focus in structural safety monitoring and management.

Nguyen et al. [4] developed a method for crack detection in target structures. Starting with collected structural images, they employed image blending techniques to generate hybrid samples. Using a small amount of labeled data, they trained an initial CNN model to produce preliminary segmentation masks for cracks. This initial model was then applied to unlabeled domain data to generate high-confidence predictions as pseudo labels. Subsequently, an inter-learning mechanism was introduced to capture global contextual information, enhancing the model's understanding of the crack's overall structure and enabling precise localization of crack details. However, during application, this method may lose detection of crack diffusion edges when cracks exhibit significant spreading and connectivity. Shim et al. [5] aimed to detect cracks in concrete structures by acquiring stereo image pairs of concrete surfaces. These image pairs were fed into a Generative Adversarial Network (GAN). The generator produced accurate crack detection results, while the discriminator distinguished between generated results and real annotated data. To ensure the reliability of the discrimination results and avoid the limitations of a single discriminator, multiple discriminators were integrated. This approach combines the results from multiple discriminators to determine the detection outcome, ensuring the reliability of crack detection results in concrete structures. However, during application, this method exhibits low efficiency for detecting cracks in large concrete structures and has limited adaptability to different scenarios. If the labeled sample data is insufficient, the model may fail to accurately detect cracks. Alamdari et al. [6] focused on precise concrete structure detection. Using a robot-based platform equipped with multiple sensors, they collected surface data to construct a 3D point cloud model of the concrete surface. This approach identified macro-level elevation variations. Edge detection algorithms analyzed the captured images to capture subtle crack textures and morphological features, to preliminarily identify potential crack outlines. By integrating macro-level variations with detailed features, they produced final crack morphology detection results. However, the robotic platform faced significant limitations in application scenarios, struggling to adapt to complex environments or building inspection requirements. Kumar et al. [7] proposed an improved GrabCut detection method based on Sobel for automated concrete crack detection. This two-step model first segments crack regions using edge information detected by the enhanced Sobel algorithm, combined with edge detection results for image segmentation. The second step involves precise crack assessment. This evaluation integrates the segmentation results into a web-based tool, which runs the crack detection algorithm in the background and presents the results visually to the user, such as by marking the crack locations. However, during application, the web tool's crack analysis capabilities are relatively weak, potentially leading to poor detection of minute or blurred cracks.

Multi-level transfer learning represents an advanced form of transfer learning. Its core principle involves constructing a multi-tiered knowledge transfer structure. During the transfer process, knowledge transfer from the source domain/task to the target domain/task is divided into multiple levels (e.g., feature layer, parameter layer, semantic layer, etc.), with each level undertaking distinct transfer functions [8]. This creates a hierarchical knowledge transfer pathway, enabling models to achieve finer-grained and deeper knowledge transfer across different tasks, domains, or modalities. Therefore, this paper proposes a small-sample target detection method for surface cracks in high-rise building concrete structures under multi-level transfer learning. This approach effectively addresses the small-sample data bottleneck by leveraging multi-level transfer learning to

extract crack image features from different layers of the building concrete structure, thereby enhancing the reliability of crack target detection results.

2. Small-Sample Object Detection of Surface Cracks in High-Rise Concrete Structures

2.1. Image Segmentation of Surface Cracks in High-Rise Building Concrete Structures

Due to the limited number of crack samples available for training and recognition in small-sample object detection, this paper employs image segmentation on surface cracks of high-rise concrete structures. This approach ensures the reliability of subsequent crack detection results by enabling the detection model to focus more on the intrinsic features of cracks (such as shape, texture, and grayscale) while reducing interference from irrelevant background information. Consequently, the model can more effectively learn key crack features from a small sample size. Surface cracks in high-rise concrete structures exhibit minimal contrast with their backgrounds, while building surfaces often feature color variations and other stains [9]. Addressing these characteristics, this paper employs a two-dimensional maximum entropy threshold segmentation method to process the images of surface cracks in high-rise concrete structures [10]. This method accentuates the contrast between cracks and background, identifies the optimal segmentation threshold based on this contrast, and clearly separates cracks from complex backgrounds [11]. By integrating two-dimensional information—pixel grayscale and neighborhood average grayscale—it ensures that the segmentation process does not disrupt the crack propagation edges or connected regions.

If the surface crack image of a high-rise concrete structure is represented as $g(x,y)$, processing $g(x,y)$ using both pixel grayscale values and neighborhood average grayscale yields its two-dimensional histogram $q(x,y)$. The calculation formula is:

$$q(x,y)=\frac{1}{9}\sum_{m=-3/2}^{3/2}\sum_{n=-3/2}^{3/2}g(x+v,y+d) \tag{1}$$

Where: v and d are the corresponding gray values of $g(x,y)$ and $q(x,y)$, respectively; $v,d=0,...,L-1$; 9 is the total number of pixels within the 3×3 neighborhood of the image.

Combining the calculation result of $q(x,y)$, the two-dimensional thresholding function f_e is determined, with its calculation formula as follows:

$$f_e=\begin{cases}z_0, g(x,y)\le e, q(x,y)\le r\\ z_1, g(x,y)>e, q(x,y)>r\end{cases} \tag{2}$$

Where: (e,r) represents the two-dimensional threshold vector; z_0 and z_1 denote the grayscale values of the crack region and background region, respectively.

By combining (e,r) to segment $q(x,y)$, it forms the target, background, and boundary regions. The target and background regions occupy the largest proportions, with significant differences in pixel grayscale values between their adjacent neighborhoods. Therefore, to ensure effective image segmentation [12] and distinguish targets from backgrounds, we design the probability distributions of pixel grayscale values for targets and backgrounds. By integrating their posterior probabilities, we perform normalization processing on other image regions to obtain the discrete two-dimensional entropy O, calculated as follows:

$$O=-\sum_{v=1}^{e}\sum_{d=1}^{r}p_{vd}\lg p_{vd} \tag{3}$$

Where: p_{vd} denotes the joint probability density.

Based on the calculated O, the total entropy function $\varphi(e,r)$ is computed using the following formula:

$$\varphi(e,r)=\lg\left[p_1(1-p_1)\right]+\frac{O_1}{p_1}+\frac{O_2-O_1}{1-p_1} \tag{4}$$

Where: $p_1=\sum_{v=1}^{e}\sum_{d=1}^{r}p_{vd}$; $O_1=-\sum_{v=1}^{e}\sum_{d=1}^{r}p_{vd}\lg p_{vd}$; $O_2=-\sum_{v=1}^{L-1}\sum_{d=1}^{L-1}p_{vd}\lg p_{vd}$.

Obtain the maximum value of $\varphi(e,r)$ to derive the optimal threshold vector $(\tilde{e},\tilde{r})$ for the image, calculated as:

$$(\tilde{e},\tilde{r})=\max_{0\le v,d\le L-1}\varphi(e,r) \tag{5}$$

Combine the optimal threshold vector to complete target and background segmentation [13], thereby obtaining the final concrete surface crack target image $I(x,y)$. Use this segmentation result as input for subsequent processing.

2.2. Image Filtering of Surface Cracks in High-Rise Concrete Structures

Cracks in high-rise concrete structures exhibit complex and diverse characteristics [13], including shrinkage cracks, temperature cracks, settlement cracks, and other types. The causes, morphological features, crack propagation, and connectivity of different cracks all exhibit certain differences. Therefore, to ensure effective detection of cracks with different morphologies and propagation states prior to small-sample target detection of surface cracks in high-rise concrete structures, this paper applies filtering to segmented target images. This processing primarily involves two aspects: crack connected area filtering and crack linearity and rectangularity filtering. Details of both filtering methods are described below:

(1) Crack Connected Area Filtering:

Images captured from the surface of high-rise concrete structures are influenced by shooting conditions (such as lighting and dust) or the inherent performance of the imaging equipment. This can affect the illumination and color of the captured images, introducing isolated noise points. Although crack regions typically possess a certain scale, if the cracks are minute, these isolated noise points can compromise the accuracy of subsequent crack detection. Therefore, this paper applies area filtering to images based on the premise that cracks on high-rise concrete structures possess a certain area scale. This processing removes isolated point noise within the crack area.

If the connected regions of cracks in the segmented image $I(x,y)$ of the high-rise concrete structure surface are denoted as $L(x,y)$, the number of crack connected regions is N, and the connected crack area is represented as A_i, the calculation formula are:

$$A_i=\sum_x\sum_y L_i(x,y) \tag{6}$$

$$L_i(x,y)=\begin{cases}0, A_i<A_0\\ 1, A_i\ge A_0\end{cases} \tag{7}$$

Where: A_0 represents the threshold used to identify isolated point noise, $i=1,2,...,N$.

(2) Crack Linearity and Rectangularity Filtering:

Surface cracks in concrete structures exhibit distinct linear characteristics, such as horizontal and vertical cracks, which propagate linearly. Consequently, the aspect ratio of cracks exceeds that of isolated point noise clusters, and these clusters display pronounced convexity. Based on this, the paper defines the aspect ratio of the crack's bounding rectangle H_i to characterize this convexity. By employing an aspect ratio threshold $\widehat{H}_0$ and a rectangularity threshold $\widehat{B}_0$, it precisely removes blocky noise. This approach avoids the

issue of incorrectly filtering out some diagonal cracks and network cracks (which have smaller aspect ratios) when using aspect ratio filtering alone.

If the rectangularity of the fracture-connected region is denoted as B_i, the fracture line and rectangularity filtering formulas are:

$$B_i = \frac{A_i}{\overleftrightarrow{A}_i} \tag{8}$$

$$H_i = \frac{S_i}{E_i} \tag{9}$$

$$L_i(x,y) = \begin{cases} 0, H_i < \widehat{H}_0 and H_i > \widehat{H}_0 \\ 1, \text{other} \end{cases} \tag{10}$$

Where: $\overleftrightarrow{A}_i$ denotes the area of the crack's bounding rectangle; S_i denotes the length of the crack's bounding rectangle; E_i denotes the width of the crack's bounding rectangle.

Combining the above completes the image filtering process for surface cracks in high-rise concrete structures. Area filtering, crack line shape filtering, and rectangularity filtering effectively remove isolated noise points from the images. The processed images retain minute cracks, diagonal cracks, and networked cracks $\tilde{I}(x,y)$, providing reliable basis for subsequent crack detection.

2.3. Concrete Surface Crack Detection Based on Multi-Level Transfer Learning

2.3.1. Overall Structure of the Multi-Level Transfer Learning Detection Method

Each hidden layer in the transfer learning architecture corresponds to a feature. Due to the limited sample size of surface images for high-rise concrete structures, features from a single hidden layer are insufficient [14,15] to comprehensively capture crack characteristics. This limitation prevents detection results from fully representing the geometric morphology of concrete cracks, such as continuity, tip morphology, and overall orientation. Therefore, this paper integrates multi-scale hybrid temporal convolutional networks (TCN), BiLSTM, and Attention mechanisms to generate distinct hidden layers. Knowledge transfer occurs through these layers, enhancing the transfer learning method's ability to extract target deep features [16]. The overall structure of the concrete structure surface crack detection method based on multi-level transfer learning is shown in **Figure 1**.

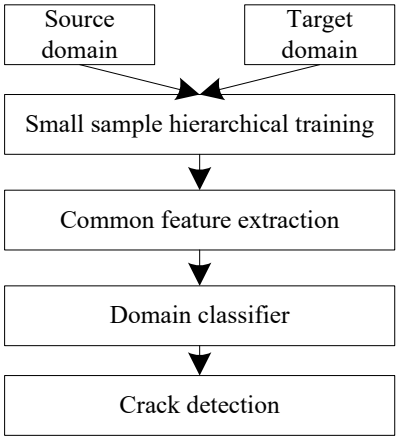


Figure 1. Overall structure of concrete structure surface crack detection method based on multi-level transfer learning.

The task of small-sample object detection for surface cracks in concrete structures involves learning relevant knowledge from the source domain dataset and transferring

this knowledge to the target domain task using a limited target domain dataset. If the target task training set is denoted as X_1 and the target task test set as X_2, the primary function of the test set is to evaluate the algorithm's generalization capability on small samples. Based on this, the dataset for the high-rise concrete structural surface crack small-sample object detection task $\widehat{X}$ comprises three parts: X_1, X_2 and X_e. Where, X_e represents the source domain target categories, while X_1 and X_2 collectively form the target domain dataset $\tilde{X}$ for the concrete structural surface crack small-sample detection task. The formulas for each dataset are:

$$\begin{cases} \tilde{X} = X_1 \cup X_2 \\ X_1 = X_e \cup \tilde{X} \\ \tilde{X} = X_e \cup M_0 \\ \tilde{X} = \left\{ \left[\tilde{I}(x,y), c \right], \tilde{g} \in G, c \in C \right\} \\ c = \left\{ \left(c_j, l_j \right), j = 1,2,...,M \right\} \end{cases} \tag{11}$$

Where: $\tilde{I}(x,y)$ denotes the input target image; c_j denotes the candidate region in the input image; the corresponding crack category is denoted by l_j; M denotes the total number of categories.

2.3.2. Feature Extraction Based on Multi-Level Transfer Learning

Multi-level transfer learning for small-sample target detection of surface cracks in high-rise concrete structures primarily employs TCN, BiLSTM, and Attention mechanisms to generate distinct hidden layers for multi-level knowledge transfer. This process acquires knowledge from the source domain X_e and the learning task (concrete surface crack detection), thereby providing transferable knowledge for subsequent crack detection tasks in the target domain. The feature extraction workflow based on multi-level transfer learning is illustrated in **Figure 2**.

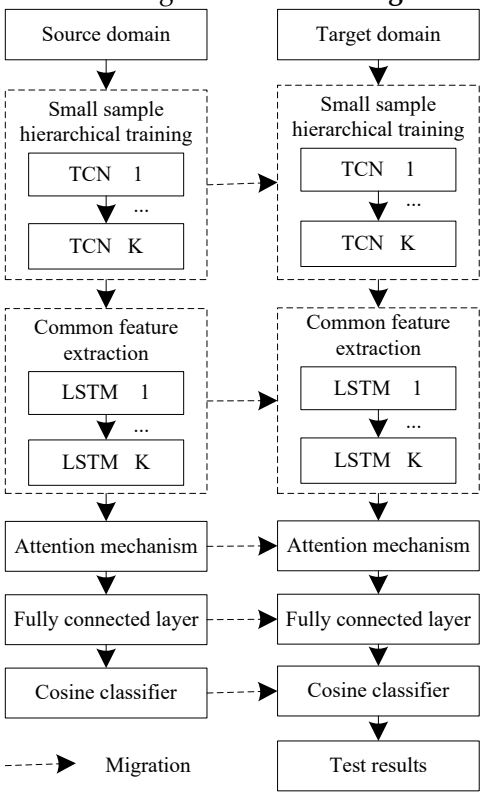


Figure 2. Feature extraction process based on multi-level transfer learning.

Feature extraction based on multi-level transfer learning involves learning across multiple hidden layers. The overall process can be divided into three stages: small-sample hierarchical partitioning, common feature extraction, and crack target detection. The primary purpose of small-sample hierarchical partitioning is to partition the source domain image data into an array of data, where the number of groups matches the number of layers in the method. After partitioning, each layer corresponds one-to-one and is trained sequentially. Common Feature Extraction primarily aims to map source and target domain data from their original feature spaces into a shared latent space [17]. This enables extraction of a common representation across all domains, thereby obtaining low-level domain-invariant features to ensure effective detection of minute cracks. Crack target detection employs a cosine classifier to identify cracks, determine their categories, and output detection results. Details of each component are described below:

(1) Small-Sample Hierarchical Training:

The target domain dataset $\tilde{X}$ for small-sample crack detection on concrete structures—comprising X_1 and X_2 — is partitioned. This partitioning follows a hierarchical structure from lower to higher levels based on multi-level transfer learning. The partitioned data is denoted as $\left\{\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_K\right\}$, where K represents the number of groups corresponding to the number of levels in the multi-level transfer learning.

After partitioning, attention training begins with the lower-level data $\tilde{X}_1$ using TCN. Specifically, the $\tilde{X}_1$ data is used to train the K_1 layer of the transfer learning model. Upon completion, the weights of this layer are retained. After loading these weights, the K_2 layer of the transfer learning model is trained using the data $\tilde{X}_2$ to obtain its weights. The crack detection error is then compared between the two layers. If the crack detection error of the lower layer is smaller than that of the upper layer, the upper layer's weights are used as the initial training weights for the lower layer. Otherwise, the lower layer's weights are used for training the next layer.

Repeat the above steps to complete training for all small-sample data. Use the training results as input $\hat{X}$ for subsequent layers for further processing.

(2) Common Feature Extraction:

Feature extraction is a critical step in multi-layer transfer learning, primarily aiming to extract common feature representations between the source domain and target domain through spatial mapping. To ensure the effectiveness of common feature extraction [18], this paper employs LSTM for this feature extraction. In the detection of small-sample cracks on concrete structures, data from the source domain and target domain may exhibit differences in feature representations. LSTMs can map raw data into a common feature space and, based on the distinct characteristics of concrete structures and crack data, extract more generalizable patterns from both domains. These patterns are not confined to the current small-sample data, enabling the model to adaptively learn and extract common features [19]. This enhances the model's ability to detect cracks across diverse concrete structures and environmental conditions, improving its generalization capabilities. Consequently, the model achieves more accurate crack detection even when confronted with limited target domain data.

The results of the aforementioned small-sample hierarchical training $\hat{X}$ is fed into the LSTM. By integrating the LSTM's forget gate, input gate, and output gate, common features are extracted. The calculation formulas for each gate are as follows:

$$\psi_1 = \vec{f}\left(\tilde{w}_1^2 \hat{X}_t + \tilde{w}_1 \hat{y}_{t-1} + b_1\right) \tag{12}$$

$$\psi_2 = \vec{f}\left(\vec{w}_3^2 \hat{X}_t + \vec{w}_2 \hat{y}_{t-1} + b_2\right) \tag{13}$$

$$\vec{\psi}_3 = \vec{\vec{f}}\left(\hat{w}_2^3 \hat{X}_t + \hat{w}\hat{y}_{h-1} + \vec{b}_3\right) \tag{14}$$

Where: ψ_1, ψ_2 and $\vec{\psi}_3$ represent the output of the forget gate, the input gate, and the feature extraction result of the output gate, respectively; $\vec{\vec{f}}$ represents the activation function; $\tilde{w}_1^2$ and $\tilde{w}_1$ denote the weights between the forget gate and the input gate, and between the forget gate and the historical output, respectively; $\vec{\vec{w}}_3^2$ and $\vec{\vec{w}}_2$ denote the weights between the input gate and the previous output, and between the input gate and the output gate, respectively; $\hat{w}_2^3$ and $\hat{w}$ denote the weights between the output gate and the input gate output, and between the output gate and the historical output, respectively; b_1, b_2 and $\vec{b}_3$ denote the bias terms for the forget gate, input gate, and output gate, respectively; $\hat{y}_{t-1}$ denotes the crack state information from the previous time step.

When extracting common features between the source domain and target domain through the aforementioned steps, to enhance its ability to capture features across different concrete structures and environmental conditions [20], this paper introduces an adaptive penalty factor adjustment mechanism based on weight dispersion. The magnitude of the penalty factor is dynamically adjusted according to the accuracy of crack detection results and the degree of weight dispersion. If the weight dispersion of the LSTM is denoted as χ_w, a higher value indicates greater weight dispersion, which may cause the extracted common features to overlook fine-grained details. Conversely, a lower value indicates more concentrated weights; excessive concentration may lead to extracted common features favoring fine-grained details while neglecting the overall feature distribution. The calculation formula for χ_w is:

$$\chi_w = \sqrt{\frac{1}{\kappa - 1}\sum_{\kappa=1}^{\kappa}\left(w_\kappa - \bar{w}\right)^2} \tag{15}$$

Where: κ denotes the number of weights; w_κ represents the weight at position κ; $\bar{w}$ denotes the mean of the weights.

To reflect the crack detection accuracy of multi-level transfer learning, the detection error variation rate $\Delta\varepsilon$ is defined, calculated as follows:

$$\Delta\varepsilon = \frac{\varepsilon_{\tau+1} - \varepsilon_\tau}{\varepsilon_\tau} \tag{16}$$

Where: τ denotes the iteration count.

Combining χ_w and $\Delta\varepsilon$ determines the adaptive regularization factor λ. This factor automatically adjusts weight magnitudes based on the LSTM's learning state to better extract shared features. The calculation formula for λ is:

$$\lambda = \Delta\varepsilon \cdot \chi_w \tag{17}$$

We weights of each LSTM layer are adaptively adjusted based on the above formula to ensure effective extraction of shared features between the target domain $\tilde{X}$ and source domain X_e for the small-sample detection task of surface cracks in concrete structures.

(3) Crack Object Detection:

This section integrates an Attention mechanism, a fully connected layer, and a cosine classifier. The primary function of the Attention mechanism is to filter the extracted common features—specifically, to screen source domain features based on target domain task requirements. It identifies key features closely related to the target domain task from a large volume of source domain features, enabling better focus on critical characteristics such as geometry and texture associated with cracks while ignoring irrelevant features, thereby enhancing feature utilization efficiency. Traditional classifiers relying on fully connected layers require sufficient training samples to find the optimal classification hyperplane between categories, which fails to meet the demands of small-sample detection

tasks. Therefore, the paper adopts a cosine classifier. This classifier performs classification by calculating the cosine similarity between candidate region features and category weight vectors, making it more suitable for small-sample crack detection tasks. Because cosine similarity focuses on the directional consistency of feature vectors, it can effectively distinguish different crack categories by leveraging semantic directional information from features, even with limited samples, thereby improving classification accuracy for crack categories under small-sample conditions.

The extracted common features $\vec{\psi}_3$ are fed into a fully connected layer. The cosine classifier then calculates the cosine similarity $\mu_{i,j}$ between the feature vector $\vec{\psi}_3$ and the category weight vector. The calculation formula is:

$$\mu_{i,j} = \beta \frac{\breve{w}_j \vec{\psi}_3^i}{\left\| \breve{w}_j \right\| \times \left\| \vec{\psi}_3^i \right\|} \tag{18}$$

Where: $\breve{w}_j$ denotes the weight vector for crack categories; β represents the scaling factor.

Given a candidate region c_j in the input image, where $\mu = (\mu_1, \mu_1, ..., \mu_c)$ denotes the cosine similarity between its feature and the category weight vector, the probability P_j of the small-sample target category for surface cracks in high-rise concrete structures is calculated as:

$$P_j = \frac{\exp(\mu_j)}{\sum_{j=1}^{C} \exp(\mu_j)} \tag{19}$$

Where: C denotes the number of crack categories.

Based on the calculation results from the above formula (19), the crack category can be determined. The category with the highest probability value corresponds to the crack, thereby completing the small-sample target category detection for surface cracks in high-rise concrete structures.

3. Test Analysis

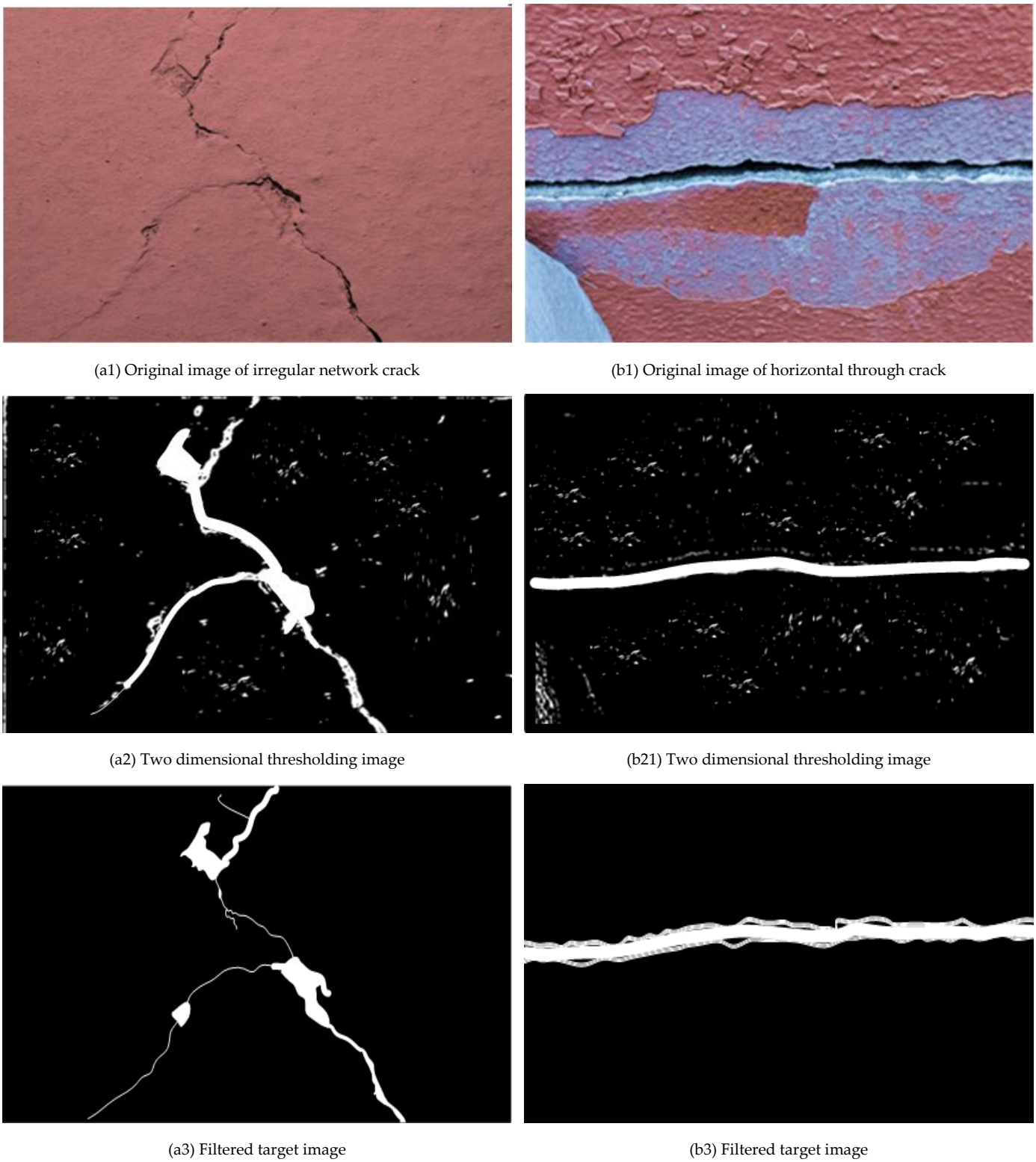
To evaluate the effectiveness of the proposed method for small-sample target detection of surface cracks in high-rise concrete structures, this study conducted relevant tests using multiple 38-story residential buildings in a specific region. Analysis was based on surface image data collected by the property management using drones. The collected images primarily encompass multiple crack categories with diverse morphologies. Each category features a limited number of annotated images—no more than 15—representing typical small-sample image data. The actual field environment of the high-rise building under investigation is shown in **Figure 3**. Details of the collected small-sample data are presented in **Table 1**.

Figure 3. Actual collection environment of high-rise buildings on-site.

Table 1. Details of Small Sample Data Collected.

Serial number	Crack category	Sample quantity/ sheet	Details
1	Horizontal cracks	12	The crack direction is parallel or nearly parallel to the horizontal plane of the concrete surface, ranging from micro cracks (<0.1mm) to through cracks (>1mm).
2	Longitudinal crack	15	Distributed in a straight line along the mid span of the structure, with large spans.
3	Oblique crack	8	Tilting downwards at a 45 degree angle to the edge of the slab, it often occurs at the corner of the exterior wall or in the span of the two-way shear wall.
4	Irregular network cracks	10	Irregular diffusion state.
5	Shrinkage cracks	9	Wide in the middle and narrow at both ends, often caused by concrete settlement and shrinkage.

Before applying the proposed method for small-sample target detection of surface cracks in high-rise concrete structures, the collected small-sample images undergo two-dimensional histogram segmentation and filtering. To evaluate the segmentation and filtering performance of this method, two images of high-rise concrete structure surfaces were randomly selected: **Figure 4**(a1) depicts an image with irregular network-like cracks, while **Figure 4**(b1) shows an image with horizontal through-cracks. These two images represent cracks of varying severity. Two-dimensional histogram segmentation was performed using the method described in this paper. Filtering was then applied to the segmented results to obtain the final processed images, as shown in **Figure 4**.



(a1) Original image of irregular network crack

(b1) Original image of horizontal through crack

(a2) Two dimensional thresholding image

(b21) Two dimensional thresholding image

(a3) Filtered target image

(b3) Filtered target image

Figure 4. Image segmentation and filtering processing results.

Based on the test results in **Figure 4**:

(1) Analysis from the perspective of image segmentation effectiveness: The paper employs a two-dimensional maximum entropy threshold segmentation method, which fully considers pixel grayscale values and their neighborhood average grayscale infor-

mation. This approach accurately captures the irregular contours of mesh cracks, effectively distinguishing crack regions from the background. This successfully extracts cracks from complex backgrounds while preserving crack continuity and structural integrity. The segmentation results clearly reveal the linear orientation and width variations of cracks, avoiding the discontinuities caused by uneven grayscale in traditional methods (e.g., no breakpoints at crack intersections). This demonstrates the method's adaptability to cracks of various morphologies.

(2) From the perspective of image filtering effects: Combining crack connected component filtering with linearity and rectangularity filtering effectively removes isolated noise points in the image. It handles the extended edges of fine cracks and reticular cracks well, without causing breakage or excessive smoothing. Furthermore, it preserves the multi-directional extension characteristics and linear features of reticular cracks, maintaining the integrity of the crack diffusion connected components, thereby enhancing the accuracy of subsequent detection.

(3) Based on the above analysis, it is concluded that the proposed two-dimensional maximum entropy threshold segmentation method combined with the two-stage filtering strategy (connectivity area filtering + linearity and squareness filtering) effectively improves the preprocessing quality of concrete surface crack images in high-rise buildings. This provides a reliable data foundation for subsequent small-sample object detection tasks based on multi-level transfer learning. This method effectively preserves crack morphological integrity while significantly suppressing noise interference, demonstrating strong practicality and robustness.

To further validate the filtering effectiveness of the proposed method, the smoothness metric η_o is adopted as the evaluation criterion. This metric assesses the method's ability to remove isolated noise points in images and the smoothness of object edges after filtering. The metric ranges from 0 to 1, with lower values indicating superior performance. The calculation formula for η_o is:

$$\eta_o = \frac{\left\{\sum\left[\tilde{I}_{i+1} - \tilde{I}_i\right]^2\right\}}{\left\{\sum\left[I_{i+1} - I_i\right]^2\right\}} \tag{20}$$

Where: $\tilde{I}_{i+1}$ and $\tilde{I}_i$ represent the filtered results of adjacent pixels $i+1$ and i, respectively; I_{i+1} and I_i denote the pre-filtered pixel values.

By applying the proposed method to filter target images of different crack types, the smoothness index η_o results obtained after filtering for various crack sizes are presented in **Table 2**.

Table 2. Smoothness Index Test Results.

Crack size /mm	Horizontal cracks	Longitudinal crack	Oblique crack	Irregular network cracks	Shrinkage cracks
0.1	0.012	0.009	0.008	0.013	0.009
0.2	0.011	0.013	0.010	0.011	0.011
0.3	0.009	0.011	0.011	0.014	0.013
0.4	0.010	0.012	0.009	0.012	0.010
0.5	0.008	0.010	0.01	0.013	0.008
0.6	0.011	0.008	0.012	0.015	0.008
0.7	0.013	0.009	0.011	0.011	0.012
0.8	0.012	0.012	0.013	0.012	0.013
0.9	0.011	0.011	0.009	0.014	0.011
1.0	0.009	0.013	0.010	0.013	0.009

Based on the test results in **Table 2**:

(1) Overall filtering effectiveness analysis: The proposed method maintains smoothness index values between 0.008 and 0.015 across all crack types and widths. This demonstrates that the combined filtering approach (connected area filtering + linearity and rectangularity filtering) effectively suppresses noise while preserving edge smoothness for diverse crack morphologies and dimensions.

(2) Smoothness analysis for different crack types: Horizontal cracks exhibit smoothness index values ranging from 0.008 to 0.013, reaching their lowest values (0.008 and 0.009) at 0.5 mm and 1.0 mm widths. Smoothness metrics for longitudinal cracks range from 0.008 to 0.013, with optimal performance at 0.6mm and 0.9mm widths (0.008 and 0.011), exhibiting minimal overall fluctuation; The smoothness index for diagonal cracks ranges from 0.008 to 0.013, with the lowest values (0.008 and 0.009) observed at widths of 0.1mm and 0.9mm; The smoothness index for irregular network cracks was slightly higher, ranging from 0.011 to 0.015, particularly reaching 0.015 and 0.014 at widths of 0.6mm and 0.9mm; The smoothness index for shrinkage cracks ranged from 0.008 to 0.013, with the lowest values (0.008) observed at 0.5mm and 0.6mm widths. Among these, irregular network cracks exhibited the highest smoothness index. This indicates that network cracks, due to their complex structure and numerous branches, retain more details after filtering, resulting in slightly lower smoothness values that remain within an acceptable range. Therefore, the proposed method demonstrates good filtering performance and adaptability for all crack types.

To evaluate the crack detection effectiveness of the proposed method for different crack categories, 10 small sample images were randomly selected. Crack detection was performed using the proposed method, and the crack category for each image was determined based on the small sample target category probability of surface cracks in high-rise concrete structures, calculated using formula (19). The detection results are shown in **Table 3**.

Table 3. Probability Results of Target Categories for Small Sample Cracks.

Image	Horizontal cracks	Longitudinal crack	Oblique crack	Irregular network cracks	Shrinkage cracks
1	0.16	0.11	0.39	0.21	0.82
2	0.23	0.75	0.14	0.19	0.19
3	0.77	0.13	0.22	0.24	0.31
4	0.52	0.51	0.71	0.13	0.24
5	0.33	0.34	0.13	0.76	0.16
6	0.19	0.79	0.22	0.33	0.24
7	0.25	0.44	0.86	0.25	0.13
8	0.14	0.23	0.11	0.81	0.21
9	0.83	0.16	0.25	0.12	0.16
10	0.11	0.28	0.37	0.19	0.85

Based on the test results in **Table 3**: The method described in this paper can detect qualified crack categories by calculating the target category probability of small-sample cracks on the surface of high-rise concrete structures. The maximum probability value corresponds to the crack category:

The maximum category probabilities for Image 1 and Image 10 are 0.82 and 0.85, respectively, corresponding to shrinkage cracks;

The maximum category probabilities for Image 2 and Image 6 are 0.75 and 0.79, respectively, corresponding to longitudinal cracks;

(3) The maximum category probabilities for Image 3 and Image 9 are 0.77 and 0.83, respectively, corresponding to horizontal cracks;

(4) The maximum category probabilities for Image 4 and Image 7 are 0.71 and 0.86, respectively, corresponding to diagonal cracks;

(5) The maximum class probability results for Image 5 and Image 8 are 0.76 and 0.81, respectively, corresponding to oblique cracks.

Therefore, the method described in this paper demonstrates effective detection performance for surface cracks of various categories in high-rise concrete structures. It adapts to the detection needs of different crack morphologies and effectively handles structurally complex, multi-branched crack patterns.

Among the detection results, Image 4 yielded a maximum class probability of 0.71, yet horizontal and vertical crack probabilities were also relatively high (0.52 and 0.51). This discrepancy stems from the small sample size in the study and limitations in the annotation of image capture angles, leading to some overlap in the method's learning of crack orientations. Although these two probabilities are relatively high, the paper employs multi-level hierarchical processing to extract multi-level crack features, which does not compromise the final classification accuracy.

To further analyze the effectiveness of the proposed method in small-sample target detection, the methods from [4] and [6] were simultaneously employed as comparison methods. The correlation coefficient ϑ_j was adopted as the evaluation metric to measure the detection performance of the three methods on small-sample crack targets. This metric quantifies the similarity between detected results and actual results, with values ranging from -1 to 1. A smaller value indicates poorer detection performance, while a larger value signifies better detection effectiveness. while a larger value indicates superior detection performance. The formula for calculating ϑ_j is:

$$f_e = \begin{cases} z_0, g(x,y) \le e, q(x,y) \le r \\ z_1, g(x,y) > e, q(x,y) > r \end{cases} \tag{21}$$

Where: ς_j represents the difference between the detected result and the actual result; n denotes the number of detection samples.

Crack detection was performed on cracks of varying sizes using the methods from Reference [4], Reference [6], and the method proposed in this paper. The detection effectiveness of the three methods was evaluated based on the ϑ_j calculated using formula (21), as shown in **Figure 5**.

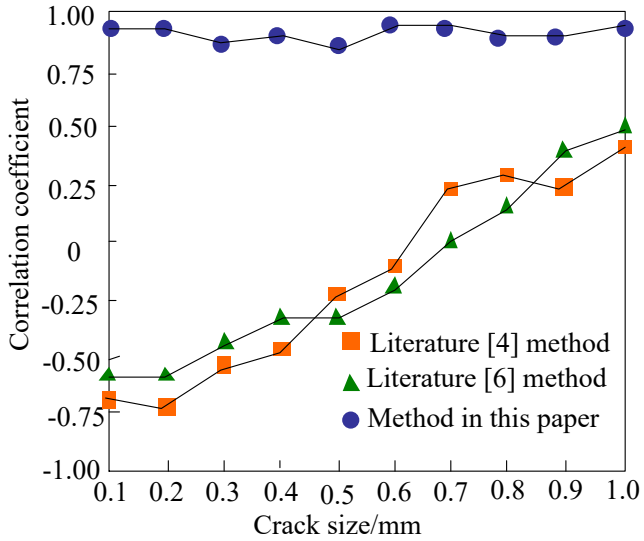


Figure 5. Correlation coefficient test results of three methods for crack detection.

Based on the test results in **Figure 5**: After detecting cracks of different sizes using the methods from Reference [4], Reference [6], and the proposed method in this paper, the method from Reference [4] showed poor detection performance for extremely small cracks

(0.1–0.6 mm), with the correlation coefficient ϑ_j below 0. For larger cracks (>0.7 mm), the correlation coefficient ϑ_j showed a significant increase, reaching above 0.25 in all cases. Furthermore, the correlation coefficient fluctuated between increases and decreases during this detection process. Consequently, this method not only performed poorly on micro-cracks but also exhibited instability. The method in Reference [6] similarly yielded lower correlation coefficient ϑ_j across all crack widths compared to the proposed method, with the worst performance observed for small cracks (0.1–0.6 mm). The proposed method maintains a high correlation coefficient ϑ_j close to 1 across all crack widths (0.1 mm–1.0 mm), demonstrating strong generalization capability and stability. This is achieved by: Utilizing a multi-level transfer learning architecture to extract multi-level, multi-scale crack features; Effectively preserving crack structure and suppressing noise during preprocessing (segmentation + filtering); Maintaining high classification accuracy with the cosine classifier even under small sample conditions. Therefore, the proposed multi-level transfer learning method significantly outperforms the comparison methods in crack detection tasks, demonstrating superior robustness and accuracy under conditions of small sample size, multi-scale data, and complex environmental conditions.

4. Conclusion

This paper addresses challenges in detecting surface cracks on high-rise concrete structures—including complex crack morphology, scarce samples, and high environmental interference—by proposing a few-shot object detection method based on multi-level transfer learning. The approach innovatively integrates TCN, BiLSTM, and Attention mechanisms to construct a multi-level transfer learning architecture, effectively enhancing crack detection accuracy and robustness under few-shot conditions. Key findings are as follows:

(1) Innovative Preprocessing Method: A two-stage filtering strategy (connectivity area filtering + linearity and rectangularity filtering) removes isolated noise points by considering both scale and morphological characteristics. Connectivity area filtering eliminates minute noise points, while linearity and rectangularity filtering preserve complex crack patterns like diagonal and network cracks, addressing the misclassification issues of traditional filtering methods on intricate crack morphologies.

(2) Classification Mechanism Innovation: Replaces traditional classifiers with a cosine classifier. Classification is performed by calculating the cosine similarity between feature vectors and category weight vectors, emphasizing semantic consistency in feature direction. This approach effectively utilizes limited sample feature information in small-sample scenarios, enhancing crack category classification accuracy and overcoming the limitation of traditional classifiers requiring large sample sizes.

Combining these innovative advantages, the proposed method maintains excellent edge smoothness and structural integrity during application—whether for 0.1mm micro-cracks or complex network cracks—while demonstrating strong preprocessing robustness. Under small-sample conditions (no more than 15 annotated images per category), it accurately identifies detection probabilities for various crack types including horizontal, vertical, diagonal, irregular network, and shrinkage cracks. This resolves the inefficiency and susceptibility to subjective factors inherent in traditional methods under small-sample data, providing an effective solution for small-sample crack detection on the surfaces of high-rise concrete structures.

References

1. Opabola, E. A., Elwood, K. J. Seismic Performance of Reinforced Concrete Beams Susceptible to Single-Crack Plastic Hinge Behavior. Journal of Structural Engineering, 2023, 149(4), 1-15. https://doi.org/10.1061/JSENDH.STENG-11424
2. Maio, U. D., Gaetano, D., Greco, F., et al. The damage effect on the dynamic characteristics of FRP-strengthened reinforced concrete structures. Composite Structures, 2023, 309, 116731. https://doi.org/10.1016/j.compstruct.2023.116731

3. Aldao, E., L. Fernández-Pardo, F. Veiga-López, et al. Synthetic Data Generation Techniques for Enhancing Crack Detection in Railway Concrete Sleepers. Journal of Computing in Civil Engineering, 2025, 39(3), 1.1-1.13. https://doi.org/10.1061/JCCEE5.CPENG-6158
4. Nguyen, Q. D., Thai, H. T., Nguyen, S. D. Self-training method for structural crack detection using image blending-based domain mixing and mutual learning. Automation in Construction, 2025, 170, 105892. https://doi.org/10.1016/j.autcon.2024.105892
5. Shim, S., Kim, J., Cho, G. C., et al. Stereo-vision-based 3D concrete crack detection using adversarial learning with balanced ensemble discriminator networks. Structural Health Monitoring, 2023, 22(2), 1353-1375. https://doi.org/10.1177/14759217221097868
6. Alamdari, A. G., Ebrahimkhanlou, A. (2024). A multi-scale robotic approach for precise crack measurement in concrete structures. Automation in Construction, 2024, 158, 105215. https://doi.org/10.1016/j.autcon.2023.105215
7. Kumar, C., Sinha, A. K. Automated Crack Detection and a Web Tool Using Image Processing Techniques in Concrete Structures. Russian Journal of Nondestructive Testing, 2023, 59(11), 1119-1135. https://doi.org/10.1134/S1061830923600569
8. Zhang, Z.X., Li L.J., Xie, G. A Model Robustness Optimization Method Integrating Transfer Learning and Adversarial Training. Computer Simulation, 2024, 41(5), 383-389.
9. Loverdos, D., Sarhosis, V. Pixel-level block classification and crack detection from 3D reconstruction models of masonry structures using convolutional neural networks. Engineering Structures, 2024, 310, 118113. https://doi.org/10.1016/j.engstruct.2024.118113
10. Vincens, B., Corres, E., Muttoni, A. (2024). Image-based techniques for initial and long-term characterization of crack kinematics in reinforced concrete structures. Engineering Structures, 2024, 317, 118492. https://doi.org/10.1016/j.engstruct.2024.118492
11. Shamsabadi, E. A., Erfani, S. M. H., Xu, C., et al. Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling. Automation in construction, 2024, 158, 105181. https://doi.org/10.1016/j.autcon.2023.105181
12. Jirakitpuwapat, W. A Regret Bound for the AdaMax Algorithm with Image Segmentation Application. Mathematical Methods in the Applied Sciences, 2025, 48(9):10208-10214. https://doi.org/10.1002/mma.10879
13. Thapliyal, S., Kumar, N. ASCAEO: accelerated sine cosine algorithm hybridized with equilibrium optimizer with application in image segmentation using multilevel thresholding. Evolving Systems, 2024, 15(4), 1297-1358. https://doi.org/10.1007/s12530-023-09552-7
14. Fernandez, I., Berrocal, C. G., Almfeldt, S., et al. Monitoring of new and existing stainless-steel reinforced concrete structures by clad distributed optical fibre sensing. Structural Health Monitoring, 2023, 22(1), 257-275. https://doi.org/10.1177/14759217221081149
15. Sharma, G., Singh, M., Berwal, K. Video Salient Object Detection Via Multi-level Spatiotemporal Bidirectional Network Using Multi-scale Transfer Learning. IETE Journal of Research, 2024, 70(11), 8077-8088. https://doi.org/10.1080/03772063.2024.2370952
16. Mukherjee, S., Peng, L., Udpa, L., et al. Dynamic Defect Detection in Fast, Robust NDE Methods by Transfer Learning Based Optimally Binned Hypothesis Tests. Research in Nondestructive Evaluation: A Journal of the American Society for Nondestructive Testing, 2024, 35(2), 70-101. https://doi.org/10.1080/09349847.2024.2316916
17. Singh, S. A., Choudhari, S. J., Desai, K. A. Augmenting human-guided progressive learning with machine vision systems for robust surface defect detection. Advanced engineering informatics, 2024, 62, 102906. https://doi.org/10.1016/j.aei.2024.102906
18. Russel, N. S., Selvaraj, A. MultiScaleCrackNet: A parallel multiscale deep CNN architecture for concrete crack classification. Expert Systems with Application, 2024, 249, 123658. https://doi.org/10.1016/j.eswa.2024.123658
19. Arafin, P., Billah, A. M., Issa, A. Deep learning-based concrete defects classification and detection using semantic segmentation. Structural health monitoring, 2023, 23(1), 383-409. https://doi.org/10.1177/14759217231168212
20. Raushan, R., Singhal, V., Jha, R. K. Damage detection in concrete structures with multi-feature backgrounds using the YOLO network family. Automation in construction, 2025, 170, 105887. https://doi.org/10.1016/j.autcon.2024.105887