Computers and Structures 300 (2024) 107398

Contents lists available at ScienceDirect

Computers and Structures

journal homepage: www.elsevier.com/locate/compstruc

Concrete spalling detection system based on semantic segmentation using deep architectures

Tamanna Yasmin [a], Duc La [b], Kien La [c], Minh Tuan Nguyen [c], Hung Manh La [a,*]

[a] *Advanced Robotics and Automation Laboratory, Department of Computer Science and Engineering, University of Nevada, Reno, 1664 North Virginia Street, MS0171, Reno, 89557, NV, USA*

[b] *Kahlert School of Computing, University of Utah, 50 Central Campus Drive, Room 3190, Salt Lake, 84112, UT, USA*

[c] *Advanced Wireless Communication Network Laboratory, Department of Electrical Engineering, Thai Nguyen University of Technology, Thai Nguyen University, 666, 3-2 National road, Thai Nguyen, 240000, Viet Nam*

ARTICLE INFO

Dataset link: https://github.com/aralab-unr/Code-and-Dataset-for-Segmentation.git

Keywords:
Spalling detection
Spalling severity
Deep architecture
Encoder-decoder
Deep neural networks

ABSTRACT

This paper presents a method for detecting the location of spalling and assessing the severity level of the spalling in concrete surfaces. The proposed method is constructed based on deep learning architectures and multi-class semantic segmentation. The proposed method can detect each pixel as a non-spalling, a deep-spalling, or a shallow-spalling. The proposed method consists of three different deep learning architectures with several encoders as backbone networks. Both qualitative and quantitative analyses show that the deep learning architecture with a certain encoder network can detect spalling with different severity levels very well. Additionally, the paper proposes a method to analyze the deep spalling areas of concrete to show their severity levels. The performance analysis shows that this approach provides very convincing results with respect to the actual affected spalling areas. The results convey that this paper achieved a higher level of performance for detecting spalling and assessing the severity of the spalling.

1. Introduction

Concrete distress, such as spalling, poses life-threatening risks, necessitating regular maintenance to prevent hazardous incidents [1][2]. Spalling, a concrete abnormality caused by heavy loads and surface erosion, compromises structural integrity in bridges and buildings. It is required to conduct regular inspections to ensure structural integrity [3–5]. To address this issue, autonomous detection systems are now mandatory.

To manage spalling effectively, precise measurements and categorization are essential. Depending on spalling conditions, post-inspection actions vary. Priority is given to large or deep spalling, while smaller or shallower instances may follow. Detecting spalling alone is insufficient; assessing severity (based on size [6] or depth) is crucial to prioritize repairs. Spalling severity ranges from deep (high risk) to shallow or non-existent. Deep spalling significantly impacts structural health.

Therefore, this paper introduces spalling detection and categorization: deep, shallow, and non-spalling. We also propose a method for ranking the severity among deep spalling areas. The severity levels of spalling are shown in Fig. 1.

The rest of the paper is organized as follows: Section 1 also presents the related works and our contribution. Section 2 outlines the research methodology and our proposed work. Section 3 analyzes the results. Finally, the conclusion and future work of this paper are given in section 4.

1.1. Literature review

In recent years, there has been an increasing interest in spalling detection techniques for various infrastructure applications such as metro tunnels, subway networks, bridges, and railway surfaces. These methods often utilize machine vision, laser scanning, deep learning, and infrared thermography to detect and evaluate spalling.

One approach proposed for detecting spalling in subway networks is based on image processing [7]. The color image of the spalled area is processed to remove noise and extract surface features. A 3D visualization is created from the extracted features, and spalling severity and depth are detected using a projection of the spalling intensity curve and regression analysis.

* Corresponding author.
E-mail address: hla@unr.edu (H.M. La).

https://doi.org/10.1016/j.compstruc.2024.107398
Received 3 May 2023; Accepted 28 April 2024
Available online 20 May 2024

Fig. 1. Different level of spalling (a) non-spalling, (b) Deep Spalling, (c) Shallow Spalling.

Another approach suggested for metro tunnels utilizes surface roughness analysis based on a 3D mobile laser scanning system [8]. Point cloud data obtained from the scanning system is used to analyze the surface roughness, which is then used to detect concrete spalling.

A machine learning and vision-based approach has been developed for subways to detect and quantify spalling [9]. This approach involves extracting important features from images, removing noise, and detecting surface distresses in subways.

To detect concrete spalling automatically in multiple spots (within a single structural element or in multiple structural elements), a deep learning-based method has been developed [10]. In this work, an inexpensive depth camera is integrated with a faster region-based convolutional neural network (Faster R-CNN) to automatically detect, localize, and quantify the spalling damage.

Another deep learning-based real-time multi-drone approach has been proposed to detect surface defects in high-rise civil structures [11]. To detect five types of concrete surface defect images of two classes (crack and spalling): vertical crack, horizontal crack, diagonal crack, branch crack, and spalling, the authors have used the deep learning model YOLO-v3 (You Look Only Once-version3) and the edge computing principle.

A morphological attention ensemble learning method for surface defect detection at the bounding box level is proposed to detect three types of defects (crack, efflorescence, and spalling) [12,13]. The authors propose a specialized loss for each defect to demonstrate improved defect-recognition accuracy. Moreover, the deformable convolutional network (DCN) [14] and multi-task ensemble learning techniques have been exploited to adaptively extract defect features according to the feature shapes and to apply the loss of each defect, respectively.

Another approach based on Faster RCNN is proposed to detect four different damage types: surface cracks, spalling (including façade spalling and concrete spalling), and severe damage with exposed rebars and severely buckled rebars [15]. Since the purpose of their study is to present a timely assessment of defects and damages that occurred due to an earthquake to buildings, the authors manually evaluated the approach using annotated image data collected from damaged concrete buildings during several past earthquakes.

For rail surface spalling detection, a real-time visual inspection system has been proposed that utilizes image acquisition and image processing sub-systems [16]. Images captured by a camera are segmented, and spalling on the rail surface is detected using histogram curve information in the longitudinal direction of the track image.

An optical detection algorithm based on visual salience has been proposed for rail surface spalling detection [17]. This algorithm uses a threshold value to detect the difference between spalled and non-spalled areas after removing unnecessary noises from the neighborhood area of spalling.

A novel automated 3D spalling defects inspection system for railway tunnel linings has been proposed that uses laser intensity and depth information for accurate spalling detection [18]. A spalling intensity depurator network is also proposed for automatic feature extraction, and the system produces 3D inspection results with quantitative analysis of the spalled area.

Deep learning approaches have also been developed for automatic detection of cracks and spalling in buildings and bridges [19]. These approaches utilize Mask R-CNNs for continuous segmentation of damaged areas in bridges and buildings, and the deep CNN architectures can be extended for surface damage detection and evaluation.

To automatically detect concrete spalling, image texture and piecewise linear stochastic gradient descent logistic regression are used for pattern recognition [20]. Image textures are extracted from images, and statistical properties are used to categorize the condition of the concrete surface into non-spalled and spalled classes.

A terrestrial laser scanner was utilized in this study to simultaneously localize and quantify spalling defects on concrete surfaces, as reported by Kim et al. in 2015 [21]. The proposed method combines features with complementary properties to enhance the localization and quantification of spalling defects. To extract relevant information, such as the condition and size of the damaged portion of the concrete surface, a defect classifier was developed. The concrete structure was scanned using a terrestrial laser scanner, and a region of interest was selected for analysis. The scanner captured 3D coordinate information of the scanned points within the selected region. Once the raw scanned data was ready, the proposed method initiated the defect detection process.

Spalling in concrete structures can happen during fire conditions, as reported by Kodur et al. in 2021 [22]. The proposed approach considers factors like pore pressure, thermal gradients, and structural loading as contributors to spalling. Comparing the approach's predictions to experimental data from full-scale fire resistance tests on concrete beams of different strengths, the analysis reveals that pore pressure-induced stresses are the primary cause of spalling. However, thermal and mechanical stress levels also play a role in spalling. The extent of spalling significantly affects the fire resistance of concrete beams in severe fire scenarios.

The method proposed by Naser et al. in 2019 [23] is based on Machine Cognition (MC) to obtain expression in order to detect defects in concrete structures due to fire conditions. These expressions consider the geometric attributes, material composition, and distinct characteristics of reinforced concrete (RC) columns. Their purpose is to predict the occurrence and intensity of fire-induced spalling and assess the fire resistance of these structural components.

Another approach to identifying essential factors that influence the occurrence of fire-induced spalling in RC columns offers an exploration of how data science and machine learning techniques can be employed [24]. Nine distinct algorithms (naive Bayes, generalized linear model, logistic regression, fast large margin, deep learning, decision tree, random forest, gradient boosted trees, and support vector machine) have been used for this study to examine data collected from 185 fire experiments. These algorithms were similarly employed to pinpoint the

essential attributes influencing the likelihood of fire-induced spalling in RC columns and to create tools for immediate spalling prediction.

The study focused on investigating the effects of different types and sizes of specimens on concrete spalling when exposed to a hydrocarbon fire, as reported by Mohd et al. in 2018 [25]. Four different types of specimen sizes, including cylinders, columns, and panels, were analyzed to isolate the variables that affect concrete spalling. Additionally, three aggregate sizes were used in the concrete mixes to determine the impact of aggregate size on concrete spalling. The investigation also included analyzing the effect of aggregate type on concrete spalling.

Concrete spalling detection can be achieved using active infrared thermography, as reported by Tanaka et al. in 2006 [26]. Various irradiation devices, such as halogen lamps, xenon arc lamps, and far-infrared irradiation devices, can be used to heat the concrete surface and create a temperature gradient for detecting spalling. Active infrared thermography was chosen in this study due to its ability to provide measurements independent of meteorological conditions. Photogrammetry, laser scanning, and Light Detection and Ranging (LiDAR) are technologies commonly used for surface damage detection, including spalling, as reported by Zhang et al. in 2022 [27]. In the proposed method, point cloud data from the laser scanner was utilized to detect spalling and quantify its key properties in reinforced concrete columns. The first phase of the method involved removing noise points and calibrating the coordinate system of the captured point cloud data, which was then sliced into thin layers for analysis of the damaged areas. The second phase included detecting points corresponding to distressed and non-distressed areas. Finally, linear interpolation was used to calculate the spalling area and lost concrete volume.

A computer application has been developed to automatically evaluate spalling and detect spalling severity in concrete bridges [28]. The proposed approach utilizes a single-objective particle swarm optimization model based on the Tsallis entropy function to detect spalling. In the second phase, the severity of spalling is evaluated by generating a comprehensive analysis of the bridge deck image using the Daubechies discrete wavelet transform feature description algorithm. A hybrid artificial neural network-particle swarm optimization model is used in the second phase to accurately predict the spalling area and overcome the limitations of the gradient descent algorithm.

The timely and accurate detection of spalling and its severity is critical, and computer vision plays a vital role in this context by extracting numerical information from various sources such as depth images, digital images, videos, and 3D point clouds, processing the data, and taking appropriate actions [29]. A computer vision-based approach for classifying concrete spalling severity has been developed [30]. This method utilizes concrete images and categorizes the severity levels as shallow spall or deep spall. Features of the concrete surface, including statistical measurements of color channels, gray-level run length, and center-symmetric local binary pattern, are used to optimize the support vector machine classifier using the jellyfish search metaheuristic to divide the data into shallow spalling and deep spalling based on a decision boundary.

An entropy-based automated method has been proposed, which consists of three significant parts and utilizes computer vision technologies for spalling detection [31]. The spalling detection phase employs a segmentation model that integrates a multi-objective invasive weed optimization and information theory-based formalism of images. The feature extraction phase combines singular value decomposition and discrete wavelet transform to obtain efficient image information. The third phase involves developing a rating system for spalling severity based on its area and depth.

The study introduces a method for identifying spalling damage using point cloud data and incorporating the damaged elements into a Building Information Model (BIM) by enhancing the as-built Industry Foundation Classes (IFC) model with semantic information [32]. The authors present a methodology for creating the as-built BIM, reconstructing the geometric properties of identified damage point clusters, and enhancing the associated IFC model with semantic information.

A deep neural network called MaDnet (material-and-damage-network), which is designed to perform dual tasks, it can identify both the material type (concrete, steel, asphalt) and distinguish between fine (cracks, exposed rebar) and coarse (spalling, corrosion) structural damage simultaneously [33]. The authors utilize semantic segmentation, which involves assigning material and damage labels to individual pixels in the image. The connection between material and damage is integrated by training shared filters using a multi-objective optimization approach.

Computer vision approaches offer solutions for detecting spalling severity on concrete surfaces [34]. The proposed work utilizes Extreme Gradient Boosting Machine and Deep Convolutional Neural Network (DCNN) to classify image data into shallow spall and deep spall. Feature extraction methods such as local binary pattern, center symmetric local binary pattern, local ternary pattern, and attractive repulsive center symmetric local binary pattern are used to extract properties of spalling from the concrete surface image. The prediction performance of the Extreme Gradient Boosting Machine is enhanced using the Aquila optimizer metaheuristic.

In summary, various approaches have been proposed for detecting spalling, or surface damage, in subway networks, metro tunnels, rail surfaces, buildings, and bridges. These approaches utilize image processing, laser scanning, machine learning, and other techniques for extracting features, removing noise, and detecting spalling severity and depth. Deep learning approaches, such as Mask R-CNNs, have been used for continuous segmentation of damaged areas in bridges and buildings. Terrestrial laser scanners have been utilized for localization and quantification of spalling defects on concrete surfaces. Active infrared thermography has been used for concrete spalling detection, and photogrammetry, laser scanning, and LiDAR have been used for surface damage detection. Computer applications and optimization models have also been developed for automated evaluation of spalling severity. These approaches aim to provide accurate and efficient detection of spalling in various structures, aiding in maintenance and repair efforts. However, there is a limited number of methods that focus on identifying and categorizing the severity of concrete spalling. We have provided a contrasting analysis between our proposed work and previous works in Table 1.

1.2. Contributions

Ensuring the structural health of concrete is crucial for maintaining the wellness of civil infrastructure. Therefore, detecting spalling and classifying its severity level has a significant impact on achieving this goal. Although there are several methods for detecting spalling, there are few approaches for classifying its severity level. This is important because it helps prioritize spalling maintenance, especially in critical areas.

Previous approaches for spalling detection have some limitations, such as that “non-spalling” areas were not categorized as a level of severity. To properly identify distressed surfaces and non-affected areas, these “non-spalling” areas should be included in the severity level. Additionally, spalling classes should be discretely segmented with proper visual mapping based on severity level. Thus, the classification of severity can be measured by how deep or shallow the spalling is, or whether there is no spalling at all. As a result of the benefits of image segmentation techniques in various fields, we considered this problem to be one of semantic segmentation.

To overcome these limitations, we propose a method for detecting and classifying spalling severity levels using deep architecture and encoder-decoder networks. Our approach uses pixel-by-pixel multiclass semantic segmentation to categorize spalling as non-spalling, shallow, or deep. We conducted a comparative analysis to determine the best combination of deep architecture and encoder-decoder networks. Ac-

Table 1
Comparison of spalling detection systems. Symbol: ✓Addressed, ☐ Not Addressed.

	Spalling Detection Surface	Spalling	Spalling Severity Level	Non-spalling
Dawood et al. [7]				
Dawood et al. [9]	Subway network	✓	☐	☐
Wu et al. [8]	Metro tunnel	✓	☐	☐
Beckman et al. [10]				
Hong et al. [12]				
Ghosh et al. [15]				
Kim et al. [21]				
Tanaka et al. [26]				
Zhang et al. [27]	Concrete surface	✓	☐	☐
Kumar et al. [11]				
Bai et al. [19]				
Abdelkader et al. [28]				
Isailovic et al. [32]	High-rise civil structure	✓	☐	☐
Zhou et al. [18]				
Pham et al. [16]	Rail Surface	✓	☐	☐
Hu et al. [17]	Rail Surface	✓	☐	✓
Hoang et al. [20]	Concrete Surface	✓	☐	✓
Kodur et al. [22]				
Naser et al. [23]				
Naser et al. [24]				
Mohd et al. [24]	Concrete surface under fire condition	✓	☐	☐
Hoang et al. [30]				
Nguyen et al. [34]	Concrete Surface	✓	Two severity levels	☐
Abdelkader et al. [31]	Concrete Surface	✓	A rating system	☐
Hoskere et al. [33]	Concrete, steel, asphalt	✓	☐	☐
Our proposed approach	Concrete surface	✓	Three severity levels, an additional ranking system	✓

cording to the severity level, deep spallings are more crucial since they affect the concrete surfaces more alarmingly. They are crucial and need further analysis. The affected area of a deep spalling may vary according to different sizes. We also proposed a pixel-wise severity ranking method to calculate the ranking of severity for deep spalling areas.

Our proposed approach is designed to identify distressed surfaces and non-affected areas accurately and to classify the severity of spalling more precisely. It offers several contributions, including a deep architecture-based method with different backbone networks, multi-class segmentation using pixel-by-pixel categorization, a pixel-wise severity ranking method, and qualitative and quantitative analysis to obtain the best results.

Overall, our proposed method provides an effective solution for detecting and classifying spalling severity levels. This will help engineers and maintenance personnel to prioritize and plan spalling maintenance activities more efficiently, resulting in improved infrastructure wellness and durability.

2. Research methodology

In this section, we have presented a comprehensive analysis of the different aspects of our proposed method for detecting spalling and severity levels. The approach for detecting spalling and spalling severity levels is based on Deep encoder-decoder networks. In recent years, several encoder-decoder-based deep convolutional networks have been proposed; SegNet [35], UNet [36], PSPNet [37], FCN [38], DeepLab [39], DeepCrack [40]. We have selected SegNet, PSPNet, and UNet for our proposed architecture. Our proposed approach delineated the use of SegNet, PSPNet, and UNet along with variations in the backbone networks to predict the best deep architecture-backbone network pair. A comparative analysis of the performance achieved from the three deep architectures in terms of different performance metrics has been discussed in the results and discussion section. We have employed different encoder modules leveraged within the context of the different deep architectures including ResNet-50 [41][42], VGG-19 [43], Xception [44], and MobileNet [45].

We have discussed several concepts related to our proposed approach. First, several deep encoder-decoder-based architectures will be discussed. The preparation of datasets and the data augmentation process will be included in this section. Along with these discussions, our proposed methodology for detecting spalling and severity levels using deep encoder-decoder networks will be outlined.

2.1. Deep encoder-decoder architecture

SegNet: The SegNet is an encoder-decoder network based architecture [35]. SegNet architecture-based image segmentation process has been used to extract abnormal skin lesions from dermoscopy image [46], for gland segmentation from colon cancer histology images [47], to detect dark spots in oil spill areas [48], for automated brain tumor segmentation on multi-modal MR image [49], to detect pixel level crack detection [50], and for the inspection and evaluation of bridge decks [51].

This architecture was proposed for pixel-wise semantic segmentation. The architecture for SegNet with encoder-decoder block is shown in Fig. 2. The encoder block of SegNet architecture contains 13 convolutional layers for feature maps which leads to object classification. The dense convolutions, ReLU non-linearity, and a non-overlapping max-pooling are performed by encoder block [52]. The max pooling is performed with a (2×2) window. The SegNet architecture avoids the fully connected layers to gain higher-resolution feature maps at the deepest encoder output. The down-sampling is the final step of the encoder. In the decoder block up-sampling and convolutions are performed [35]. The decoder conducts the up-sampling and calls the max pooling indices of the corresponding encoder layer. There is a K-class softmax classifier at the end to predict the class for each pixel.

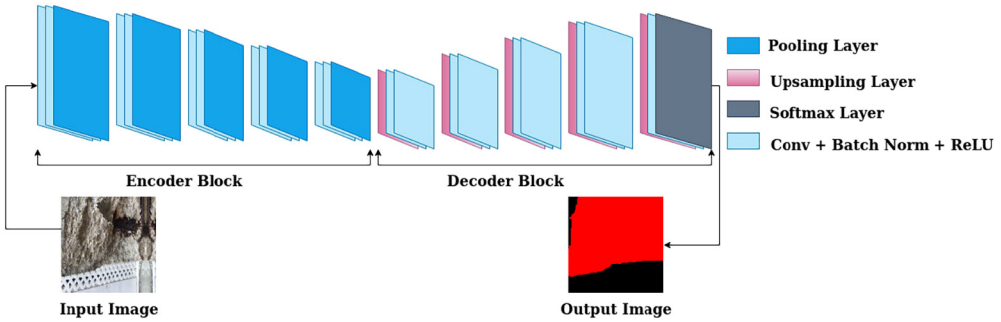

Fig. 2. Overview of SegNet Architecture [35].

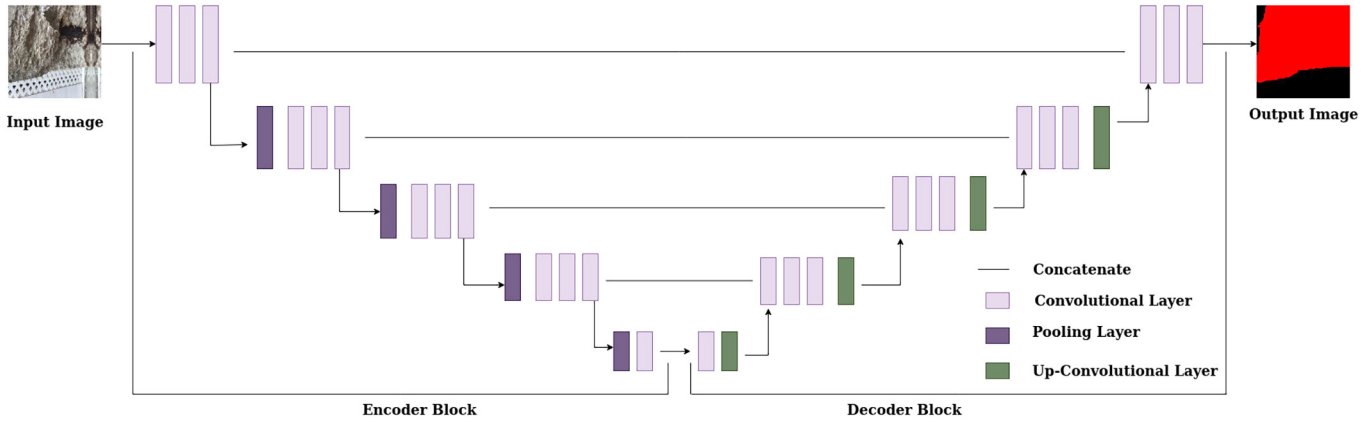

Fig. 3. Overview of UNet Architecture [36][58].

UNet: Several works used the UNet architecture for image segmentation; brain tumor image segmentation using UNet [53] and UNet-VGG16 [54], COVID-19 lung CT image segmentation [55], crack detection model [56], and dental panoramic image segmentation [57].

UNet is an encoder-decoder-based architecture consisting of four encoder and four decoder blocks. Fig. 3 shows the overview of UNet architecture. The encoder block contains two 3×3 convolutions [36].

A ReLU activation function comes after each convolution. The encoder component of the UNet architecture functions as an image feature extractor and gathers the image's features. Each encoder block's number of feature channels is doubled and its spatial dimensions are cut in half by the encoder network. A link connects the encoder blocks and decoder blocks together. The resulting output of the encoder blocks' ReLU activation function connects to the matching decoder blocks. Two (3×3) convolutions are used in the connection between the encoder and decoder blocks, and each convolution is followed by a ReLU activation function. By providing supplementary information, this connection enables the decoder to build stronger semantic features. The decoder network has half the number of feature channels and doubles the spatial dimensions. A (2×2) transpose convolution is present in the decoder's initial stage. Using a concatenation process of convolution and connection, the feature maps are transferred through the connection between the encoder and decoder. A segmentation mask is created in the decoder section. A (1×1) convolution with sigmoid activation is applied to the output generated by the final decoder. Using an activation function, the segmentation mask is transformed into pixel-wise categorization.

PSPNet: The architectural overview of PSPNet is shown in Fig. 4. PSPNet is one of the most well-recognized image segmentation models. PSPNet-based semantic segmentation process used in image semantic segmentation [59], pavement distress detection [60] and crack detection [61][62], arms and hands segmentation for egocentric perspective using image segmentation [63], and image segmentation for coronary angiography [64].

This architecture has two blocks like most semantic segmentation models: PSPNet encoder and PSPNet decoder. The PSPNet encoder consists of the CNN backbone with dilated convolutions [65] and the pyramid pooling module. Dilated convolution layers are used in place of the typical convolutional layers in the backbone's last layers, which helps to increase the receptive field. The last two blocks of the backbone contain these dilated convolution layers. As a result, the feature that is added at the end of the backbone has more features. During convolution, the value of dilation indicates the sparsity. In comparison to standard convolution, dilated convolution has a broader receptive field. The size of the used context information is found from the size of the receptive field. The pyramid pooling module is the primary component of this model since it enables the model to recognize the global context in the image and classify the pixels according to that context. The backbone's feature map is pooled at different sizes, passed through a convolution layer, and then upsampled to bring the pooled features up to the size of the original feature map. The original feature map and the upsampled maps are finally concatenated before being sent to the decoder. This method aggregates the overall context by fusing the information at different scales. The decoder will then take those features and turn them into predictions by feeding them into its layers once the encoder has extracted the image's features. The decoder is another network that processes inputted characteristics to provide predictions.

2.2. Backbone network

Several CNN's (convolutional neural network) backbone networks have made significant advancements with the highest quality performances over the past few years. These network architectures may effectively extract an image's feature mapping, providing a strong base network for semantic segmentation [66]. We have used ResNet-50, VGG-19, MobileNet, and Xception as feature extractors in our deep architectures.

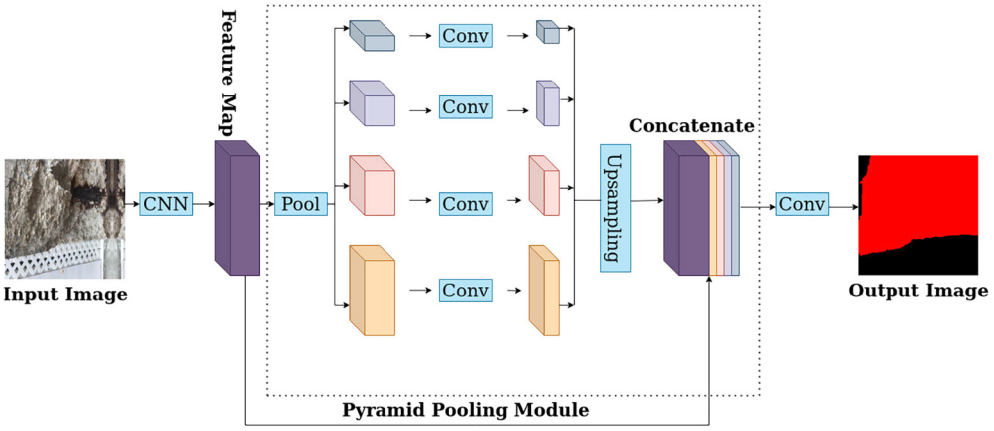

Fig. 4. Overview of PSPNet Architecture [37].

ResNet-50, VGG-19, and Xception are CNN with 50 layers, 19 layers and 71 layers deep, respectively [42][43][44]. MobileNet is one kind of CNN designed for mobile and embedded vision applications [45]. MobileNet, VGG-19, Xception have been used in the field of medical imaging, eye's region classification, namely apple leaf diseases identification, skin lesion classification, and diagnosis of pneumonia from chest X-Ray images [67][68][69][70]. In the area of image recognition or image classification, VGG-19, ResNet-50, and Xception are used to classify images and malware data, to recognize the facial expression, and to detect and localize rebar for bridge deck inspection and evaluation [71][72][73][74][75].

2.3. Preparation of dataset

Large volumes of data are required to train, validate, and test the models in deep network architectures [76]. As a result, managing a well-balanced dataset is a crucial step. For our proposed method, we have collected images of different buildings and bridges. For bridge data collection, we mainly used our developed robots integrated with non-destructive evaluation sensors and cameras to collect the bridge deck surface images [77–81]. We collected images at different times of the day to maintain the non-uniformity of the environment. We have employed a data augmentation procedure for our proposed architecture to help with the data management issue. We assigned the labels of non-spalling, deep spalling, and shallow spalling, along with the labels of severity, to each image in our collection. As a result, during the training process, pixel mapping is automatically generated from the labeling of the image.

The non-spalling area and the severity levels of spalling are annotated with RGB combinations because the labeling of images follows the RGB range. The process of annotating images is an arduous and time-consuming task [82]. We have annotated each image according to the spalling area; Deep spalling area, shallow spalling area, and non-spalling area. The example of the annotation process of an image is shown in Fig. 5.

We have prepared a collection of images for each original image and respective annotated image using the data augmentation procedure. The augmentation procedure chooses a random picture for each image as well as a random pixel point for the tagged image. A selected image and pixel map of the relevant original image is made in accordance with that. The augmentation method generates a number of sub-images at random from the pixel locations by flipping or rotating the pixel map. The augmentation process of a sample image is shown in Fig. 6.

2.4. Proposed architecture

We have proposed the method using three different types of deep architectures with different backbone networks for detecting the spalling

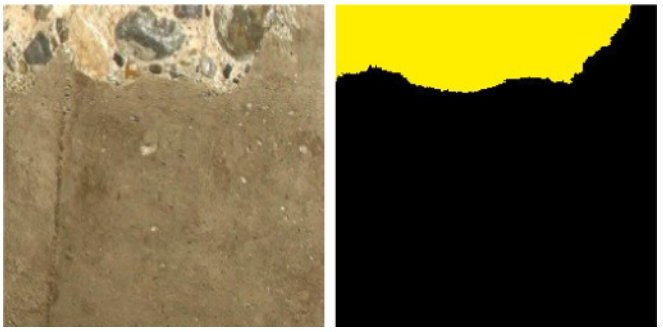

Fig. 5. Data Annotation (a) Image of shallow spalling, (b) Annotation of shallow spalling.

and level of severity. The Encoder-Decoders model used in deep learning-based image segmentation technology is trained from start to end [83]. A pre-trained CNNs model, such as the ResNet pre-trained model, MobileNet pre-trained model, or VGG pre-trained model, is the encoder. We have implemented this deep architecture with PSPNet (Fig. 4), UNet (Fig. 3), and SegNet (Fig. 2).

For the encoder part, we have employed ResNet-50, VGG-19, Xception, and MobileNet. The encoder block has convolution and pooling layers. A set of down-sampled feature maps are produced by each part of the encoder using an input picture or feature map. The pooling layers help the encoder to form integrated feature points after the feature is extracted from the convolution layers.

The decoder is essentially a mirrored encoder. The difference between the decoder block and encoder block is the up-sampling layer instead of the pooling layers. It gradually upsamples the encoder's output and semantically projects into high-resolution pixel space from the low-resolution identifiable feature maps.

The advantage of employing deep learning-based image segmentation architecture is, the segmentation model differentiate each pixel at the pixel level as well as projects the features with the different category at various stages into the pixel space learned by the encoder to fully segment the target region [83]. Moreover, using the concatenation process the decoder connects to the corresponding encoder and helps to reduce the loss that happened during the down-sampling process. Therefore, due to the advantage and performance of deep learning-based image segmentation architecture in several fields [84][85][86], we have proposed the use of deep architectures with encoder-decoder networks to detect spalling and severity level. We have considered the spalling and severity detection process as multi-class image segmentation. Therefore, as the output or semantic segmentation of the given data from the encoder-decoder network, we get the segmented area of spalling; non-

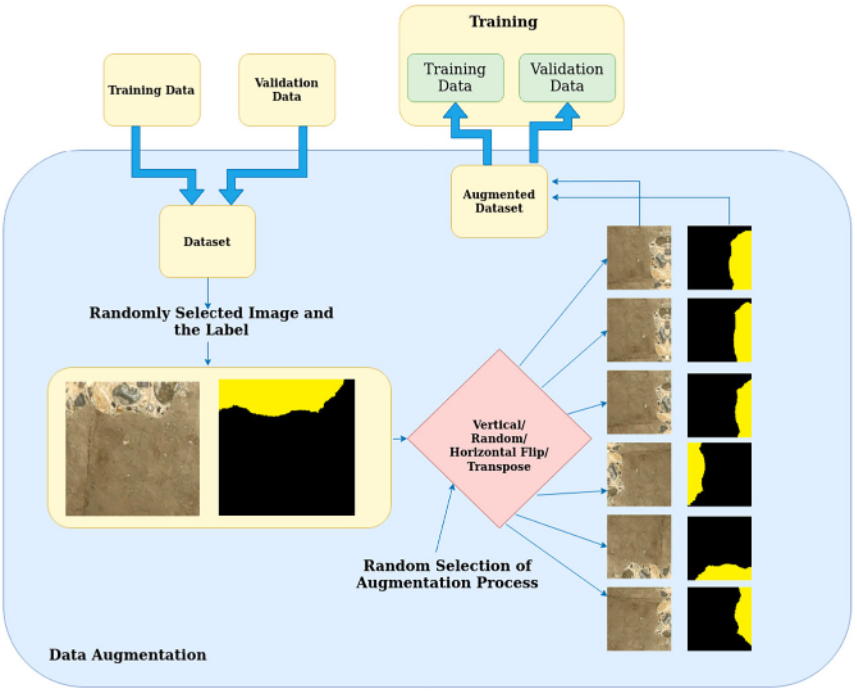

Fig. 6. Data Augmentation Process.

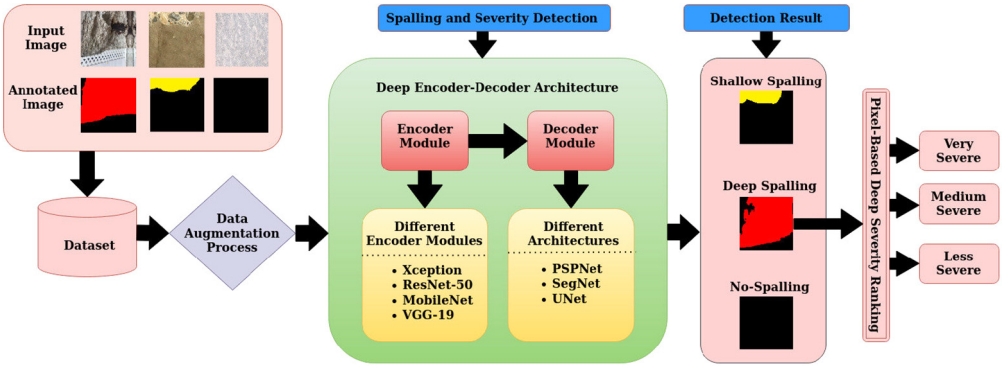

Fig. 7. Overview of the proposed methodology.

spalling, deep spalling, or shallow spalling. Fig. 7 shows the overview of the proposed methodology to detect the spalling severity levels.

For the deep architecture-based proposed method, we have annotated the images of deep spalling based on the exposed reinforcing steel bars. The annotated images are used to detect the spalling severity level. These deep spalling areas based on the reinforcing steel bars can be large, very large, or small. Along with the depth, the ratio of the affected deep spall areas helps provide more insight into the severity. For that reason, we have proposed a method to calculate the ranking of severity for the deep spalling area. This proposed method determines the affected deep spalling area using pixel-wise calculation. Afterwards, the ranking of deep spalling areas is determined according to the ratio of the affected area (number of affected pixels) with respect to the overall area (number of total pixels). We have determined the ratio using Equation (1), where D_pixel provides the number of total pixels of deep spalling area, and T_pixel counts the total number of pixels for the entire image. The ranking of deep spalling areas is categorized as "very severe," "medium severe," and "less severe" based on the value of the ratio using a predefined threshold.

$$Ratio = \frac{D\_pixel}{T\_pixel} \tag{1}$$

Hence, we first detected spalling and its severity. Moreover, we discussed the comparative analysis of the performances achieved by the deep architectures. Using pixel-wise calculation, we determined the severity ranking for deep spalling areas. We have provided a comparative analysis of severity ranking for three selected image categories in the Result and Discussion section.

3. Result and discussion

In this section, we are going to present the performance analysis of the proposed deep architecture with different encoder-decoder networks. The results and experimental analysis part includes dataset preparations, experimental setup, and qualitative and quantitative analysis of proposed architectures.

3.1. Experimental setup

The dataset contains images of spalling in buildings and bridges. These images have different types of noises, namely, oil spills, faded colors, and stones. It is difficult to detect any abnormality in the crucial corners of bridges, for example, the intersection of pillars, due to

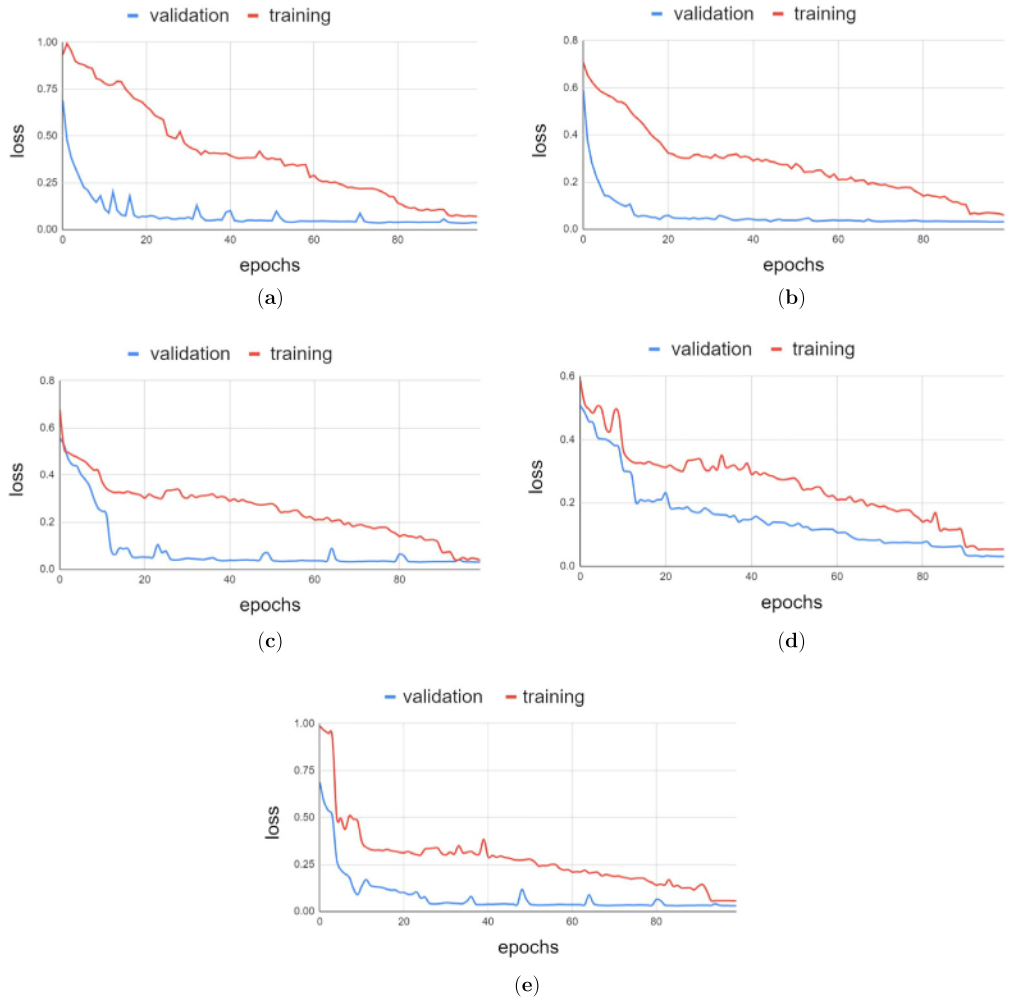

Fig. 8. Categorical cross-entropy (CCE) loss curves for training and validation (a) PSPNet, (b) PSPNet with ResNet-50, (c) PSPNet with Xception, (d) PSPNet with MobileNet, (e) PSPNet with VGG-19.

the difference in light. The images were taken at different times of the day to avoid any impact of light and shadow on the result of spalling detection.

We have used GIMP (GNU Image Manipulation Program) to annotate our images in the pixel-by-pixel map. GIMP is one of the most popular illustration and image editing programs available [87]. We have annotated each image according to the spall class. The reason behind the image size (256×256) is to focus on the specific spall class with any noises or light differences. Our dataset contains different categories of images: only deep spalling, deep spalling with non-spalling area, only shallow spalling, shallow spalling with non-spalling area, and non-spalling. The spalling area is categorized as deep spalling when the reinforcing steel bars are exposed. The shallow spalling areas are the ones whose condition lies between deep spalling and non-spalling. Moreover, the spalling without exposed steel bars was considered shallow spalling in this paper.

The method was trained and tested on a system with a GTX 1080 GPU. The size of the image was (256×256). For the multi-class classification problem, we used categorical cross-entropy (CCE) loss, which is also known as Softmax loss. The Adam optimizer was used to optimize the architecture with a learning rate of 0.001.

We have used an augmentation process (described in Fig. 6) during the training and validation phases. The use of the augmentation process during the training and validation phases has the advantage of avoiding overfitting problems [82]. The dataset contains 10000 images for training and another 2000 images for validation. The CCE loss curve of the training and validation for PSPNet, SegNet, and UNet are shown in Fig. 8, 9, and 10, respectively. We recorded the loss for all the deep encoder-decoder combinations. In the loss curve, the training loss and validation loss show how the model fits the training data and the new data, respectively. The loss is measured by the error between its predicted output and the true output. Our goal is to get the loss value as close as 0. During the starting phase, Fig. 8, 9, and 10 show gaps between the training and validation curves. The use of the augmentation process during the training and validation phases helped to reduce the gaps gradually. Since the gaps were reducing and both of the loss curves were getting close to 0, the model was fitting well. The training and validation curves started converging approximately after 90 epochs which indicates a desirable characteristic. We maintained 100 epochs for all the architectures to avoid overfitting.

The testing phase was conducted on 300 images. First, we tested the proposed approach with 100 images to determine the difference in performance achieved based on the number of test images. The difference in performance for these two datasets is negligible (approximately 0.01% for all metrics), which justifies a consistent performance. Hence, we have presented the performance analysis only for the dataset with 300 images. The proposed approach trained for 100 epochs. Therefore, on each epoch, it was trained on a different dataset because of the augmentation process.

Because spalling is detected based on its severity levels, we have presented non-statistical qualitative analysis as well as quantitative analysis with statistical measurements. In the quantitative analysis sub-

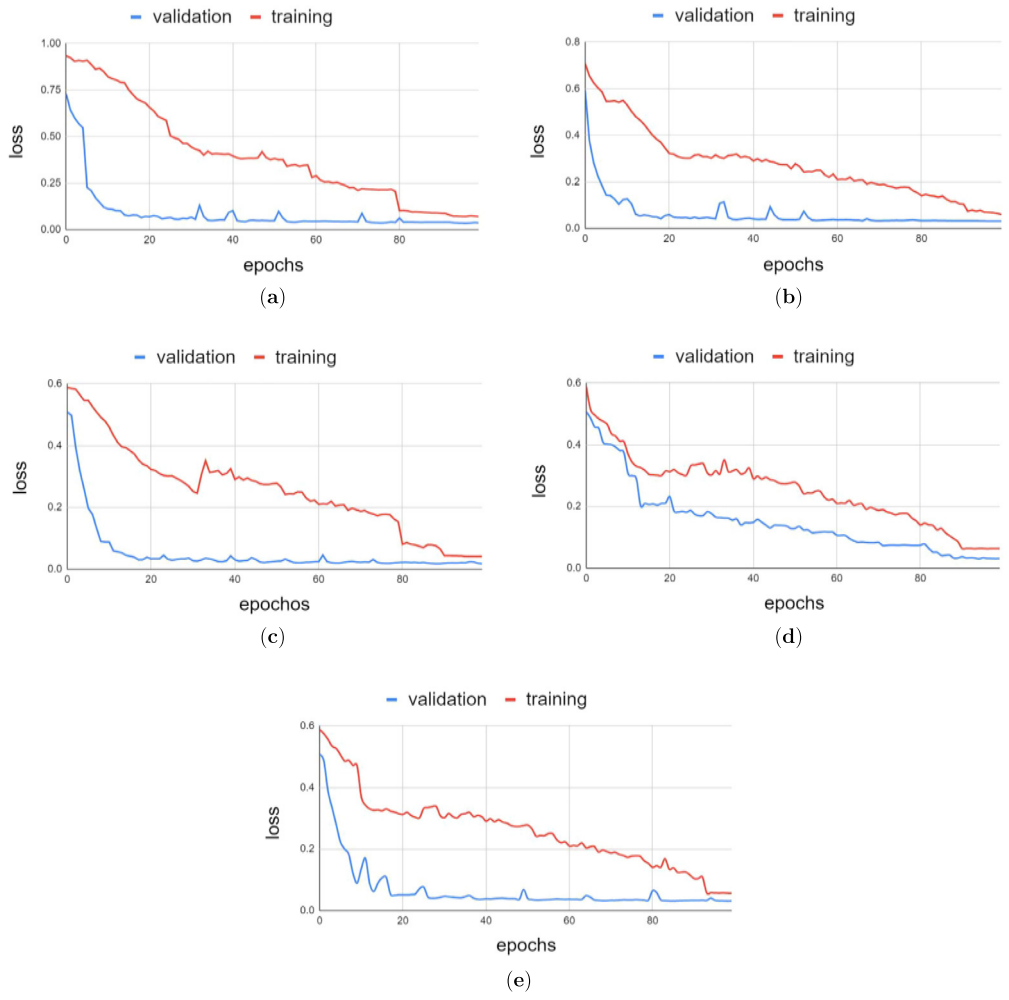

Fig. 9. CCE loss curves for training and validation (a) SegNet, (b) SegNet with ResNet-50, (c) SegNet with Xception, (d) SegNet with MobileNet, (e) SegNet with VGG-19.

section, we evaluated the performance of three deep architectures with different encoder-decoder networks using different metrics. The qualitative analysis subsection describes the performance comparison based on the results of spalling detection and severity level.

3.2. Quantitative analysis

This section presents a performance-based statistical analysis and deep learning-based image segmentation architectures with different encoder-decoder networks. The overall performance for spalling detection with severity level is shown in Table 2.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$IoU = \frac{TP}{TP + FN + FP} \tag{5}$$

We have used Dice loss, mIoU, Precision, Recall, and Accuracy metrics for the performance analysis. The dice loss referred to the loss level for the combination architecture with different encoder-decoder networks. We have performed the calculation on Equation (2), (3), and (4) to find out the Accuracy, Precision, and Recall respectively. From Equation (5), we get the calculation for IoU for each class which helps to calculate the mean value of IoU for all classes.

Table 3 describes the quantitative measures used for evaluating the performance of deep network architectures. The lower Dice Loss values are more appropriate since they reflect the degree of loss incurred by the different combinations of network frameworks employed in the proposed system for spalling and severity detection. The higher values for all other performance measures reflect the proposed spalling and severity detection system's improved performance. The suggested system performs well for spalling and severity detection as the mIoU, precision, recall, and accuracy values increase.

The statistical performance for PSPNet is shown in Table 2; the result for PSPNet architecture with encoders namely, Xception, ResNet-50, MobileNet, and VGG-19, respectively. According to the metrics discussed above, PSPNet architecture with Xception gives the best result among all the combinations (**e.g., Dice Loss: 5.94%, mIoU: 88.78%, Precision: 94.67%, Recall: 98.43%, Accuracy: 96.06%**). The result for ResNet-50 is pretty close to Xception. PSPNet with default encoder-decoder network gives comparatively poor results than with the other encoder-decoder networks (**e.g., Dice Loss: 8.33%, mIoU: 84.62%, Precision: 92.53%, Recall: 90.82%, Accuracy: 92.40%**). The performance of VGG-19 with the PSPNet architecture shows that it closely follows the performance of PSPNet with the default encoder-decoder network (**e.g., Dice Loss: 7.82%, mIoU: 85.49%, Precision: 94.44%, Recall: 92.95%, Accuracy: 92.97%**). For PSPNet, the decreasing CNN layers (during employing Xception, ResNet-50, MobileNet, VGG-19) have a negative impact on the performance for spalling and severity level detection.

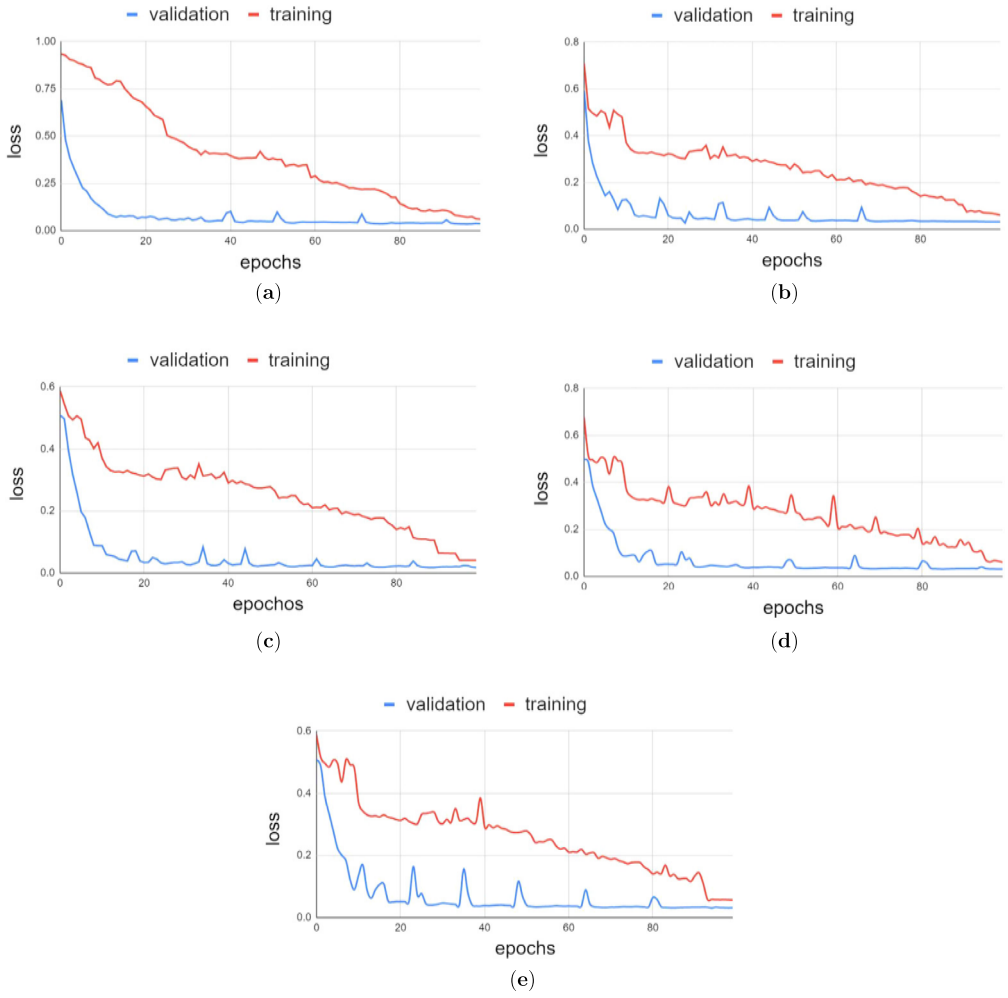

Fig. 10. CCE loss curves for training and validation (a) UNet, (b) UNet with ResNet-50, (c) UNet with Xception, (d) UNet with MobileNet, (e) UNet with VGG-19.

Table 2
Performance comparison among Deep architectures with different backbone networks.

Base Model	Encoder	Dice Loss (%)	mIoU (%)	Precision (%)	Recall (%)	Accuracy (%)
SegNet	-	9.58	82.49	93.99	95.37	92.40
”	**Xception**	**6.80**	**87.97**	**90.86**	**98.43**	**95.19**
”	ResNet-50	6.96	86.97	93.77	96.39	94.57
”	MobileNet	8.49	84.35	95.36	95.79	94.23
”	VGG-19	16.30	71.96	93.83	92.34	92.0
UNet	-	11.78	78.90	92.73	91.42	89.65
”	**Xception**	**7.28**	**86.43**	**92.75**	**97.23**	**93.79**
”	ResNet-50	11.58	79.24	90.68	92.89	92.04
”	MobileNet	7.36	86.29	93.91	96.13	93.52
”	VGG-19	18.16	69.26	89.49	86.53	88.59
PSPNet	-	8.33	84.62	92.53	90.82	92.40
”	**Xception**	**5.94**	**88.78**	**94.67**	**98.43**	**96.06**
”	ResNet-50	6.90	87.16	92.08	97.20	95.58
”	MobileNet	7.44	86.13	91.19	93.46	93.58
”	VGG-19	7.82	85.49	94.44	92.95	92.97

In Table 2 the results are shown for UNet framework with different encoder-decoder networks with metrics Dice loss, mIoU, Precision, Recall, and Accuracy. The results are provided for UNet architecture with encoders namely, Xception, ResNet-50, MobileNet, and VGG-19. Similarly to the performance of PSPNet, UNet architecture provides the best result with Xception among all UNet architecture combinations (**e.g., Dice Loss: 7.28%, mIoU: 86.43%, Precision: 92.75%, Recall: 97.23%, Accuracy: 93.79%**). The comparative analysis shows that MobileNet follows the performance of Xception quite closely. Unlike PSPNet, ResNet-50 provides a pretty low performance than Xception and MobileNet. The performance of VGG-19 encoder is comparatively poor than the other encoder-decoder networks for UNet architecture (**e.g., Dice Loss: 18.16%, mIoU: 69.26%, Precision: 89.49%, Recall: 86.53%, Accuracy: 88.59%**).

Table 3
Quantitative measures used for evaluating the performance of deep network architectures. Spalling pixels belong to the positive class and non-spalling pixels belong to the negative class.

Measure	Definition	Description
TP	True Positive	Number of accurately identified spalling pixels
FP	False Positive	Number of pixels erroneously labeled as spalling pixels
TN	True Negative	Number of accurately identified non-spalling pixels
FN	False Negative	the number of pixels detected as non-spalling erroneously

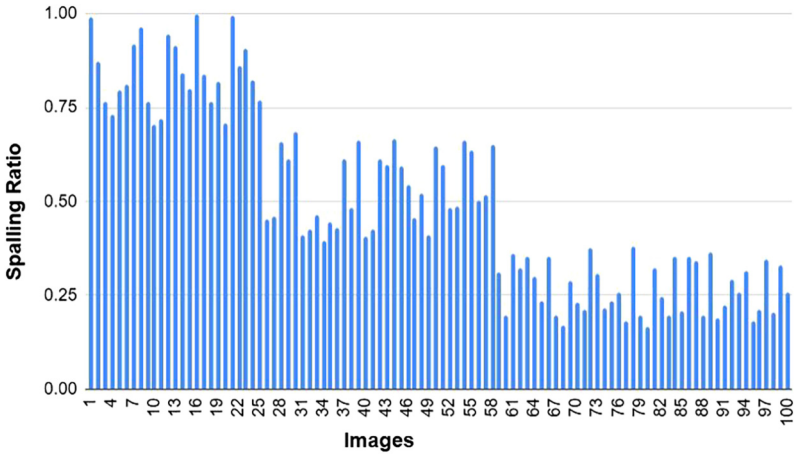

Fig. 11. Spalling ratio of 100 test images for deep spalling areas.

The comparative analysis of the statistical performance of SegNet architecture with default encoder-decoder and Xception, ResNet-50, MobileNet, and VGG-19 are shown in Table 2. The SegNet architecture with VGG-19 encoder performs similarly to the UNet architecture with VGG-19 encoder; comparatively poor results than the other encoder-decoder networks (SegNet with VGG-19: **Dice Loss: 16.30%, mIoU: 71.96%, Precision: 93.83%, Recall: 92.34%, Accuracy: 92.0%**). The comparative analysis shows that the SegNet architecture with Xception provides the best result among all the combinations (**e.g., Dice Loss: 6.80%, mIoU: 87.97%, Precision: 90.86%, Recall: 98.43%, Accuracy: 95.19%**). For SegNet architecture, the performance achieved with ResNet-50 matches the performance of Xception pretty closely. When using Xception, ResNet-50, MobileNet, and VGG-19 with SegNet architecture, the performance of spalling and severity level detection suffers as the number of CNN layers decreases.

The above discussion and performance evaluation shown in Table 2 infer that PSPNet gives comparatively good performance for detecting spalling and severity levels among the three deep architectures. For all three deep architectures, Xception gives the best result. The VGG-19 provides comparatively poor performance compared to other encoder-decoder networks for detecting spalling and severity levels with SegNet and UNet architectures. For PSPNet architecture, the VGG-19 encoder and PSPNet with the default encoder-decoder network both provide poorer performance than other encoder-decoder networks.

Table 4 shows the results for the severity ranking of deep spalling for three image categories.

Several ranking methods are proposed for civil infrastructure [88] [89]. Moreover, we have analyzed our dataset and observed that we should categorize the severity ranking for deep spalling areas. We have 300 images for testing the deep architectures for spalling severity detection. Among the 300 images, there are around 100 images of deep spalling areas. Based on our observation, we have defined the ranking for deep spalling areas as very severe, medium severe, and less severe. Fig. 11 shows the spalling ratio of 100 test images for deep spalling areas. According to the spalling ratio and from our observations of the dataset, we have defined thresholds for the severity ranking. The predefined thresholds for the severity ranking are defined as: less severe when Ratio ≤ 0.39, medium severe when, $0.4 \leq$ Ratio ≤ 0.69, and very severe when Ratio ≥ 0.7. Table 4 displays the ratio, which is expressed in terms of 100%. We have selected three different categories of images from the 100 deep spalling images. In Table 4, we presented the results of the severity ranking for each deep architecture with different backbone networks based on each selected image category. The **Input Size** displays the total number of pixels of the image, which is the same for all the images, **Spalling Size** refers to the number of pixels affected by deep spalling, **Ratio** is calculated using Equation (1) and shown in 100%, and **Severity Ranking** determines the ranking of severity based on the ratio and the predefined threshold value. We compared the severity ranking to ground truth for each image category. For most of the models, the severity ranking follows the ranking of ground truth very closely.

3.3. Qualitative analysis

This section presents the qualitative analysis of the proposed approach to show the non-statistical performance of the deep architectures with different encoder-decoder networks. The performance evaluation of different deep architectures for detecting spalling and severity level segmentation has been shown in Fig. 12, Fig. 13, and Fig. 14. The results highlight the overall performance of spalling and severity detection based on deep spalling, shallow spalling, and non-spalling images.

In Fig. 12, the results are shown for the PSPNet framework with different encoder-decoder networks. We have mentioned earlier that the images in our dataset are categorized as only deep spalling, deep spalling with non-spalling area, only shallow spalling, shallow spalling with non-spalling area, and non-spalling. For the PSPNet framework, only deep spalling, shallow spalling with non-spalling, and non-spalling areas were chosen as input to present the performance evaluation. In Fig. 12, we have original image, ground truth which is pixel-by-pixel

Table 4
Results for severity ranking of deep spalling for three different image categories.

Input Image	Model Name	Input Size (No. of Pixels)	Spalling Size (No. of Pixels)	Ratio (%)	Severity Ranking
Category 1	Ground Truth	65536	65536	100	Very Severe
”	PSPNet	”	65003	99.19	Very Severe
”	PSPNet(Xception)	”	65536	100	Very Severe
”	PSPNet(ResNet-50)	”	65401	99.79	Very Severe
”	PSPNet(MobileNet)	”	60221	91.89	Very Severe
”	PSPNet(VGG-19)	”	60006	91.56	Very Severe
Category 2	Ground Truth	”	20481	31.25	Less Severe
”	SegNet	”	20222	30.86	Less Severe
”	SegNet(Xception)	”	20500	31.28	Less Severe
”	SegNet(ResNet-50)	”	20377	31.09	Less Severe
”	SegNet(MobileNet)	”	20147	30.74	Less Severe
”	SegNet(VGG-19)	”	20110	30.69	Less Severe
Category 3	Ground Truth	”	29721	45.35	Medium Severe
”	UNet	”	35314	53.88	Medium Severe
”	UNet(Xception)	”	31027	47.34	Medium Severe
”	UNet(ResNet-50)	”	25889	39.50	Less Severe
”	UNet(MobileNet)	”	32276	49.25	Medium Severe
”	UNet(VGG-19)	”	35080	53.52	Medium Severe

mapping of original image for each spalling class, result for PSPNet architecture, result for PSPNet architecture with encoders namely, Xception, ResNet-50, MobileNet, and VGG-19, respectively. The comparative analysis in Fig. 12 shows that, PSPNet architecture with Xception gives the best result among all the combinations. ResNet-50 provides pretty similar performance to Xception. In comparison to ground truth, MobileNet gives some inaccurate predictions for deep spalling, shallow spalling, and non-spalling images.

The PSPNet architecture with the default encoder and VGG-19 network gives comparatively poor results compared to the other encoder-decoder networks. Among the three severity classes of spalling, non-spalling areas are predicted to be more accurate for all the architecture combinations.

The results are shown for the UNet framework with different encoder-decoder networks in Fig. 13. For the UNet framework, we have chosen here deep spalling with a non-spalling area, shallow spalling with a non-spalling area, and non-spalling area as input for performance evaluation. A sticker serves as noise in the shallow spalling image. We have the original image, ground truth, which is a pixel-by-pixel mapping of the original image for each spalling class, the result for UNet architecture, and the result for UNet architecture with encoders namely, Xception, ResNet-50, MobileNet, and VGG-19, as shown in Fig. 13, respectively. Fig. 13 for Xception shows that the UNet architecture gives the best result among all the combinations. The comparative analysis shows that MobileNet follows the results of Xception. ResNet-50 provides pretty low performance compared to Xception and MobileNet, unlike PSPNet. In comparison to ground truth, the results provided by Unet architecture, UNet architecture with ResNet-50, and VGG-19 MobileNet have some inaccurate predictions for deep spalling, shallow spalling, and non-spalling images. Fig. 13 shows that the VGG-19 encoder performs poorly in comparison to the other encoder-decoder networks.

The comparative analysis of SegNet architecture with the default encoder-decoder and with Xception, ResNet-50, MobileNet, and VGG-19 is shown in Fig. 14. The categorization of images for performance evaluation of the SegNet framework was chosen here as deep spalling with the non-spalling area, shallow spalling with non-spalling, and non-spalling. In Fig. 14, we have the original image of deep spalling, shallow spalling, and non-spalling area, ground truth which is pixel-by-pixel mapping of the original image for each spalling class, result for SegNet architecture, result for SegNet architecture with encoders namely, Xception, ResNet-50, MobileNet, and VGG-19, respectively. The SegNet architecture with the VGG-19 encoder gives comparatively poor results compared to other encoder-decoder networks like the UNet architecture. VGG-19 shows poor performance, especially for non-spalling and shallow spalling classes. The non-spalling is predicted pretty accurately by most of the architecture combinations, except for the SegNet architecture with VGG-19. In Fig. 14, the comparative analysis presents that the SegNet architecture with Xception gives the best result among all the combinations for all the spalling severity classes. The results for ResNet-50 show that it matches the result for Xception pretty closely. SegNet architecture with the default encoder-decoder network gives poor results compared to ground truth, especially for shallow spalling. MobileNet gives some incorrect predictions for deep and shallow spalling areas.

Based on the discussion above and the performance shown in Fig. 12, 13, and 14, it can be concluded that most deep architectures with encoder-decoder networks provide comparatively good results for non-spalling areas. The performance evaluation for predicting deep spalling and shallow spalling closely follows the performance evaluation for predicting non-spalling areas. The performance evaluation shows that, among the three deep architectures, PSPNet shows the best performance for detecting spalling and severity classification. The Xception gives the best results for detecting deep spalling, shallow spalling, and non-spalling with SegNet, UNet, and PSPNet deep architectures. Comparatively, VGG-19 shows poor performance in detecting spalling and severity levels with UNet and SegNet architectures. For PSPNet, the VGG-19 encoder closely follows the performance of PSPNet with the default encoder-decoder network.

4. Conclusions and future work

This paper presents an innovative deep learning-based approach to detect spalling and its severity levels in civil infrastructure using encoder-decoder networks. The proposed method fills the gap in the literature, where very few methods exist for detecting the severity level of spalling accurately. Our study shows that deep learning-based architectures with encoder-decoder networks offer high performance in detecting spalling severity levels in different fields, including civil infrastructure.

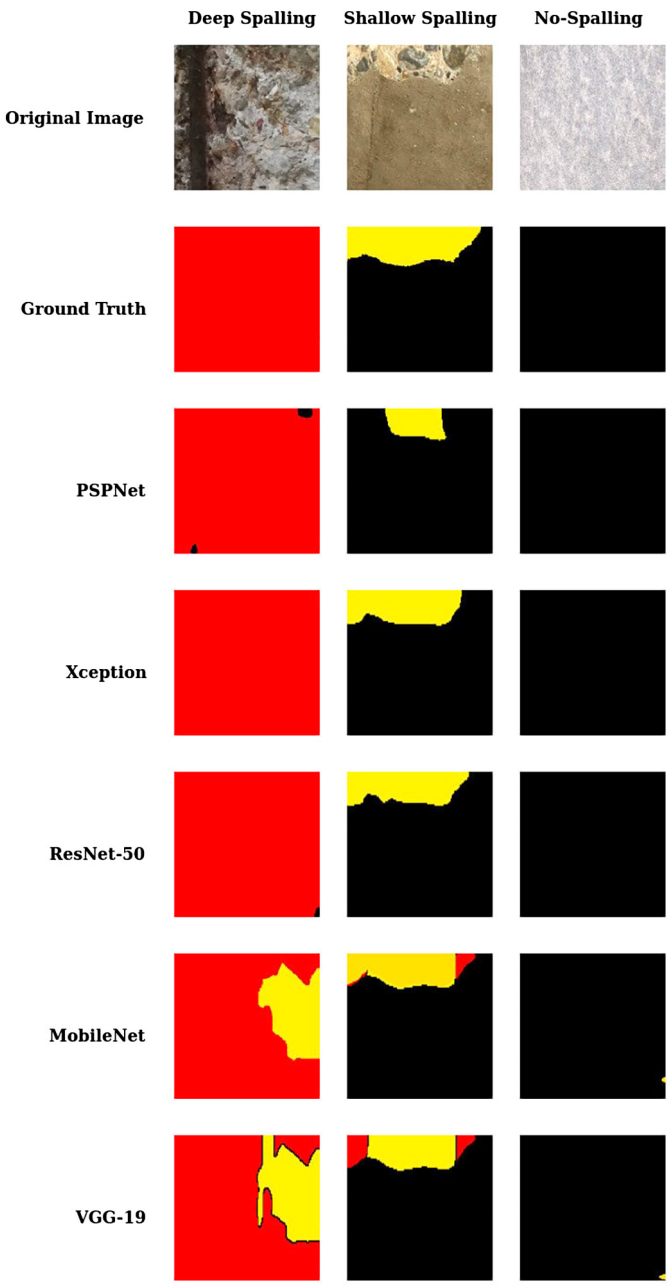

Fig. 12. Results are shown for the PSPNet framework with different encoder-decoder networks.

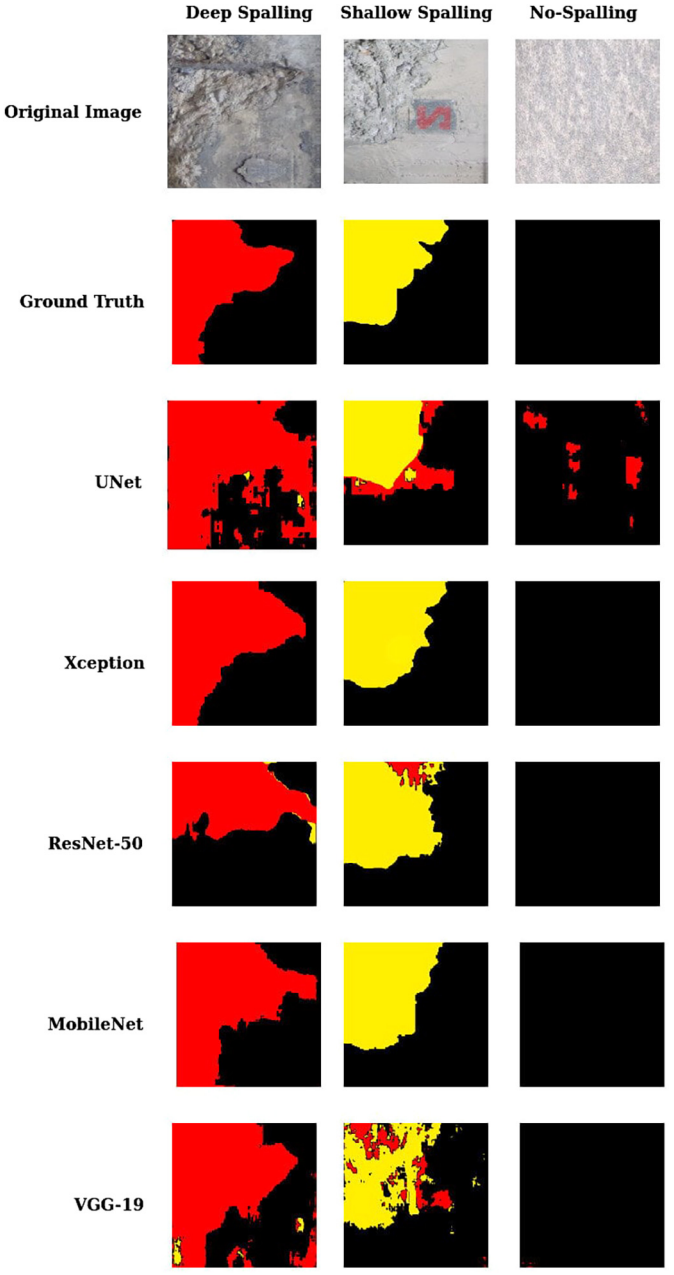

Fig. 13. Results are shown for the UNet framework with different encoder-decoder networks.

We incorporated three different deep architectures and four backbone networks in our proposed methodology to achieve the best performance. Our results indicate that the PSPNet-based deep architecture with the Xception encoder offers the best performance. We have also conducted statistical and non-statistical analyses to demonstrate the proposed method's high performance.

Our study has several potential future directions, including improving the proposed deep architecture's efficiency by reducing power consumption and memory requirements while achieving better performance in detecting spalling and severity levels. Additionally, this approach's adaptability to detect various concrete distresses using a deep architecture-based combined detection process is worth exploring. Overall, our proposed method provides a promising solution for detecting and classifying spalling severity levels in civil infrastructure, which is crucial for ensuring the structural health of concrete.

CRediT authorship contribution statement

Tamanna Yasmin: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Duc La:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Kien La:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Conceptualization. **Minh Tuan Nguyen:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Data curation. **Hung Manh La:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

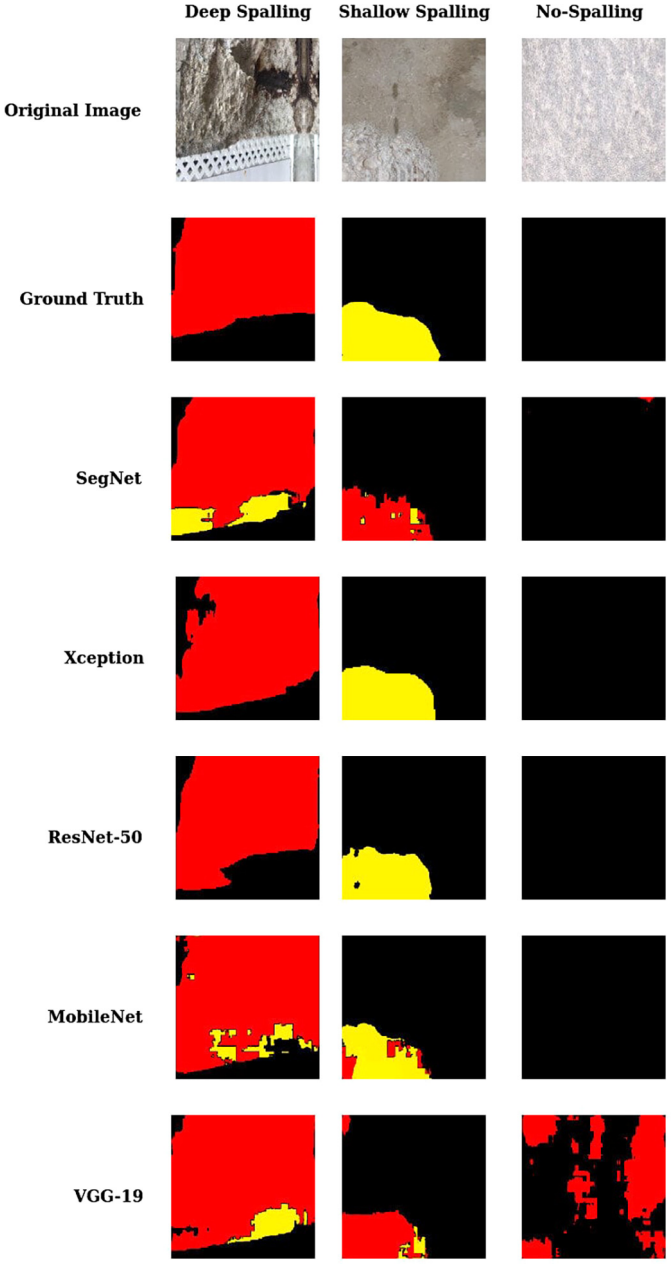

Fig. 14. Results are shown for the SegNet framework with different encoder-decoder networks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and code generated or used during the study will be available at https://github.com/aralab-unr/Code-and-Dataset-for-Segmentation.git.

Acknowledgement

This work is supported by the U.S. National Science Foundation (NSF) under grants NSF-CAREER: 1846513 and NSF-PFI-TT: 1919127, and the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (USDOT/OST-R) under Grant No. 69A3551747126 through INSPIRE University Transportation Center, and the Japan NineSigma through the Penta-Ocean Construction Ltd. Co. under Agreement No. SP-1800087. The views, opinions, findings, and conclusions reflected in this publication are solely those of the authors and do not represent the official policy or position of the NSF, the USDOT/OST-R, or any other entities.

References

[1] Ahmed H, La HM, Gucunski N. Review of non-destructive civil infrastructure evaluation for bridges: state-of-the-art robotic platforms, sensors and algorithms. Sensors 2020;20:3954. https://doi.org/10.3390/s20143954.
[2] Ahmed H, La HM. Steel defect detection in bridges using deep encoder-decoder networks. In: Structural health monitoring 2021; 2021.
[3] Ahmed H, La HM, Pekcan G. Rebar detection and localization for non-destructive infrastructure evaluation of bridges using deep residual networks. In: International symposium on visual computing. Springer; 2019. p. 631–43.
[4] Prasanna P, Dana KJ, Gucunski N, Basily BB, La HM, Lim RS, et al. Automated crack detection on concrete bridges. IEEE Trans Autom Sci Eng 2016;13:591–9. https://doi.org/10.1109/TASE.2014.2354314.
[5] Dinh TH, Ha QP, La HM. Computer vision-based method for concrete crack detection. In: 2016 14th international conference on control, automation, robotics and vision (ICARCV); 2016. p. 1–6.
[6] Yasmin T, Le C, La HM. Deep architecture based spalling severity detection system using encoder-decoder networks. In: Advances in visual computing: 17th international symposium, ISVC 2022, proceedings, Part II. Springer; 2022. p. 332–43.
[7] Dawood T, Zhu Z, Zayed T. Detection and quantification of spalling distress in subway networks. In: Proceedings of the 21st international symposium on advancement of construction management and real estate. Springer; 2018. p. 607–15.
[8] Wu H, Ao X, Chen Z, Liu C, Xu Z, Yu P. Concrete spalling detection for metro tunnel from point cloud based on roughness descriptor. J Sens 2019;2019. https://doi.org/10.1155/2019/8574750.
[9] Dawood T, Zhu Z, Zayed T. Machine vision-based model for spalling detection and quantification in subway networks. Autom Constr 2017;81:149–60. https://doi.org/10.1016/j.autcon.2017.06.008.
[10] Beckman GH, Polyzois D, Cha Y-J. Deep learning-based automatic volumetric damage quantification using depth camera. Autom Constr 2019;99:114–24. https://doi.org/10.1016/j.autcon.2018.12.006.
[11] Kumar P, Batchu S, Kota SR, et al. Real-time concrete damage detection using deep learning for high rise structures. IEEE Access 2021;9:112312–31. https://doi.org/10.1109/access.2021.3102647.
[12] Hong Y, Yoo SB. Oasis-net: morphological attention ensemble learning for surface defect detection. Mathematics 2022;10:4114. https://doi.org/10.3390/math10214114.
[13] Gibb S, La HM, Louis S. A genetic algorithm for convolutional network structure optimization for concrete crack detection. In: 2018 IEEE congress on evolutionary computation (CEC); 2018. p. 1–8.
[14] Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, et al. Deformable convolutional networks. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 764–73.
[15] Ghosh Mondal T, Jahanshahi MR, Wu R-T, Wu ZY. Deep learning-based multi-class damage detection for autonomous post-disaster reconnaissance. Struct Control Health Monit 2020;27:e2507. https://doi.org/10.1002/stc.2507.
[16] Pham D, Ha M, Xiao C. A novel visual inspection system for rail surface spalling detection. IOP conference series: materials science and engineering, vol. 1048. IOP Publishing; 2021. p. 012015.
[17] Hu Z, Zhu H, Hu M, Ma Y. Rail surface spalling detection based on visual saliency. IEEJ Trans Electr Electron Eng 2018;13:505–9. https://doi.org/10.1002/tee.22594.
[18] Zhou M, Cheng W, Huang H, Chen J. A novel approach to automated 3d spalling defects inspection in railway tunnel linings using laser intensity and depth information. Sensors 2021;21:5725. https://doi.org/10.3390/s21175725.
[19] Bai M, Sezen H. Detecting cracks and spalling automatically in extreme events by end-to-end deep learning frameworks. In: ISPRS annals of photogrammetry and remote sensing spatial information science, XXIV ISPRS congress. International Society for Photogrammetry and Remote Sensing; 2021. p. 161–8.
[20] Hoang N-D, Nguyen Q-L, Tran X-L. Automatic detection of concrete spalling using piecewise linear stochastic gradient descent logistic regression and image texture analysis. In: Complexity 2019; 2019.
[21] Kim M-K, Sohn H, Chang C-C. Localization and quantification of concrete spalling defects using terrestrial laser scanning. J Comput Civ Eng 2015;29:04014086. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000415.
[22] Kodur V, Banerji S. Modeling the fire-induced spalling in concrete structures incorporating hydro-thermo-mechanical stresses. Cem Concr Compos 2021;117:103902. https://doi.org/10.1016/j.cemconcomp.2020.103902.
[23] Naser M. Heuristic machine cognition to predict fire-induced spalling and fire resistance of concrete structures. Autom Constr 2019;106:102916. https://doi.org/10.1016/j.autcon.2019.102916.

[24] Naser M. Observational analysis of fire-induced spalling of concrete through ensemble machine learning and surrogate modeling. J Mater Civ Eng 2021;33:04020428. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003525.

[25] Mohd Ali A, Sanjayan J, Guerrieri M. Specimens size, aggregate size, and aggregate type effect on spalling of concrete in fire. Fire Mater 2018;42:59–68. https://doi.org/10.1002/fam.2457.

[26] Tanaka H, Tottori S, Nihei T. Detection of concrete spalling using active infrared thermography. Q Rep RTRI 2006;47:138–44. https://doi.org/10.2219/rtriqr.47.138.

[27] Zhang H, Zou Y, del Rey Castillo E, Yang X. Detection of rc spalling damage and quantification of its key properties from 3d point cloud. KSCE J Civ Eng 2022;26:2023–35. https://doi.org/10.1007/s12205-022-0890-y.

[28] Abdelkader EM, Moselhi O, Marzouk M, Zayed T. Evaluation of spalling in bridges using machine vision method. In: ISARC. Proceedings of the international symposium on automation and robotics in construction, vol. 37. IAARC Publications; 2020. p. 1136–43.

[29] Reja VK, Varghese K, Ha QP. Computer vision-based construction progress monitoring. Autom Constr 2022;138:104245. https://doi.org/10.1016/j.autcon.2022.104245.

[30] Hoang N-D, Huynh T-C, Tran V-D. Concrete spalling severity classification using image texture analysis and a novel jellyfish search optimized machine learning approach. Adv Civ Eng 2021;2021. https://doi.org/10.1155/2021/5551555.

[31] Mohammed Abdelkader E, Moselhi O, Marzouk M, Zayed T. Entropy-based automated method for detection and assessment of spalling severities in reinforced concrete bridges. J Perform Constr Facil 2021;35:04020132. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001544.

[32] Isailović D, Stojanovic V, Trapp M, Richter R, Hajdin R, Döllner J. Bridge damage: detection, ifc-based semantic enrichment and visualization. Autom Constr 2020;112:103088. https://doi.org/10.1016/j.autcon.2020.103088.

[33] Hoskere V, Narazaki Y, Hoang TA, Spencer Jr B. Madnet: multi-task semantic segmentation of multiple types of structural materials and damage in images of civil infrastructure. J Civ Struct Health Monit 2020;10:757–73. https://doi.org/10.1007/s13349-020-00409-0.

[34] Nguyen H, Hoang N-D. Computer vision-based classification of concrete spall severity using metaheuristic-optimized extreme gradient boosting machine and deep convolutional neural network. Autom Constr 2022;140:104371. https://doi.org/10.1016/j.autcon.2022.104371.

[35] Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–95. https://doi.org/10.1109/tpami.2016.2644615.

[36] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.

[37] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 2881–90.

[38] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3431–40.

[39] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 2017;40:834–48. https://doi.org/10.1109/tpami.2017.2699184.

[40] Zou Q, Zhang Z, Li Q, Qi X, Wang Q, Wang S. Deepcrack: learning hierarchical convolutional features for crack detection. IEEE Trans Image Process 2018;28:1498–512. https://doi.org/10.1109/tip.2018.2878966.

[41] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.

[42] Koonce B. Resnet 50. In: Convolutional neural networks with swift for tensorflow. Springer; 2021. p. 63–72.

[43] Gómez-Flores W, de Albuquerque Pereira WC. A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound. Comput Biol Med 2020;126:104036. https://doi.org/10.1016/j.compbiomed.2020.104036.

[44] Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–8.

[45] Sinha D, El-Sharkawy M. Thin mobilenet: an enhanced mobilenet architecture. In: 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON). IEEE; 2019. p. 0280–5.

[46] Kadry S, Taniar D, Damaševičius R, Rajinikanth V, Lawal IA. Extraction of abnormal skin lesion from dermoscopy image using vgg-segnet. In: 2021 seventh international conference on bio signals, images, and instrumentation (ICBSII). IEEE; 2021. p. 1–5.

[47] Tang J, Li J, Xu X. Segnet-based gland segmentation from colon cancer histology images. In: 2018 33rd youth academic annual conference of Chinese association of automation (YAC). IEEE; 2018. p. 1078–82.

[48] Guo H, Wei G, An J. Dark spot detection in sar images of oil spill using segnet. Appl Sci 2018;8:2670. https://doi.org/10.3390/app8122670.

[49] Alqazzaz S, Sun X, Yang X, Nokes L. Automated brain tumor segmentation on multi-modal mr image using segnet. Comput Vis Media 2019;5:209–19. https://doi.org/10.1007/s41095-019-0139-y.

[50] Song C, Wu L, Chen Z, Zhou H, Lin P, Cheng S, et al. Pixel-level crack detection in images using segnet. In: International conference on multi-disciplinary trends in artificial intelligence. Springer; 2019. p. 247–54.

[51] Ahmed H, La HM, Tavakolli A. Use of deep encoder-decoder network for sub-surface inspection and evaluation of bridge decks. In: Structural health monitoring 2021; 2021.

[52] Badrinarayanan V, Handa A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. ArXiv preprint arXiv:1505.07293, 2015. https://doi.org/10.48550/arXiv.1505.07293.

[53] Aghalari M, Aghagolzadeh A, Ezoji M. Brain tumor image segmentation via asymmetric/symmetric unet based on two-pathway-residual blocks. Biomed Signal Process Control 2021;69:102841. https://doi.org/10.1016/j.bspc.2021.102841.

[54] Pravitasari AA, Iriawan N, Almuhayar M, Azmi T, Irhamah I, Fithriasari K, et al. Unet-vgg16 with transfer learning for mri-based brain tumor segmentation. TELKOMNIKA (Telecommu Comput Electron Control) 2020;18:1310–8. https://doi.org/10.12928/telkomnika.v18i3.14753.

[55] Saood A, Hatem I. Covid-19 lung ct image segmentation using deep learning methods: unet vs. segnet. BMC Med Imaging 2021;21:2–10. https://doi.org/10.1186/s12880-020-00529-5.

[56] Liu F, Wang L. Unet-based model for crack detection integrating visual explanations. Constr Build Mater 2022;322:126265. https://doi.org/10.1016/j.conbuildmat.2021.126265.

[57] Sivagami S, Chitra P, Kailash GSR, Muralidharan S. Unet architecture based dental panoramic image segmentation. In: 2020 international conference on wireless communications signal processing and networking (WiSPNET). IEEE; 2020. p. 187–91.

[58] Li J, Li W, Jin C, Yang L, He H. One view per city for buildings segmentation in remote-sensing images via fully convolutional networks: a proof-of-concept study. Sensors 2019;20:141. https://doi.org/10.3390/s20010141.

[59] Yang C, Guo H. A method of image semantic segmentation based on pspnet. In: Mathematical problems in engineering 2022; 2022.

[60] Zhong J, Zhu J, Huyan J, Ma T, Zhang W. Multi-scale feature fusion network for pixel-level pavement distress detection. Autom Constr 2022;141:104436. https://doi.org/10.1016/j.autcon.2022.104436.

[61] Shu J, Li J, Zhang J, Zhao W, Duan Y, Zhang Z. An active learning method with difficulty learning mechanism for crack detection. Smart Struct Syst 2022;29:195–206. https://doi.org/10.12989/sss.2022.29.1.195.

[62] Wang J-J, Liu Y-F, Nie X, Mo Y. Deep convolutional neural networks for semantic segmentation of cracks. Struct Control Health Monit 2022;29:e2850. https://doi.org/10.1002/stc.2850.

[63] Sarah H, Clua E, Vasconcelos CN. Arms and hands segmentation for egocentric perspective based on pspnet and deeplab. In: International conference on human-computer interaction. Springer; 2020. p. 152–70.

[64] Zhu X, Cheng Z, Wang S, Chen X, Lu G. Coronary angiography image segmentation based on pspnet. Comput Methods Programs Biomed 2021;200:105897. https://doi.org/10.1016/j.cmpb.2020.105897.

[65] Yu F, Koltun V, Funkhouser T. Dilated residual networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 472–80.

[66] Xing Y, Zhong L, Zhong X. An encoder-decoder network based fcn architecture for semantic segmentation. Wirel Commun Mob Comput 2020;2020. https://doi.org/10.1155/2020/8861886.

[67] Carvalho T, De Rezende ER, Alves MT, Balieiro FK, Sovat RB. Exposing computer generated images by eye's region classification via transfer learning of vgg19 cnn. In: 2017 16th IEEE international conference on machine learning and applications (ICMLA). IEEE; 2017. p. 866–70.

[68] Bi C, Wang J, Duan Y, Fu B, Kang J-R, Shi Y. Mobilenet based apple leaf diseases identification. Mob Netw Appl 2020:1–9. https://doi.org/10.1007/s11036-020-01640-1.

[69] Chhabra M, Kumar R. A smart healthcare system based on concatenation of resnet50v2 and xception model for detecting pneumonia from medical images. In: 2022 international conference on machine learning, big data, cloud and parallel computing (COM-IT-CON), vol. 1. IEEE; 2022. p. 161–7.

[70] Ayan E, Ünver HM. Diagnosis of pneumonia from chest x-ray images using deep learning. In: 2019 scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT). IEEE; 2019. p. 1–5.

[71] Bansal M, Kumar M, Sachdeva M, Mittal A. Transfer learning for image classification using vgg19: caltech-101 image data set. J Ambient Intell Humaniz Comput 2021:1–12. https://doi.org/10.1007/s12652-021-03488-z.

[72] Rezende E, Ruppert G, Carvalho T, Ramos F, De Geus P. Malicious software classification using transfer learning of resnet-50 deep neural network. In: 2017 16th IEEE international conference on machine learning and applications (ICMLA). IEEE; 2017. p. 1011–4.

[73] Lo WW, Yang X, Wang Y. An xception convolutional neural network for malware classification with transfer learning. In: 2019 10th IFIP international conference on new technologies, mobility and security (NTMS). IEEE; 2019. p. 1–5.

[74] Li B, Lima D. Facial expression recognition via resnet-50. Int J Cogn Comput Eng 2021;2:57–64. https://doi.org/10.1016/j.ijcce.2021.02.002.

[75] Ahmed H, La HM, Tran K. Rebar detection and localization for bridge deck inspection and evaluation using deep residual networks. Autom Constr 2020;120:103393. https://doi.org/10.1016/j.autcon.2020.103393.
[76] Billah UH, Tavakkoli A, La HM. Concrete crack pixel classification using an encoder decoder based deep learning architecture. In: International symposium on visual computing. Springer; 2019. p. 593–604.
[77] Van Nguyen L, Gibb S, Pham HX, La HM. A mobile robot for automated civil infrastructure inspection and evaluation. In: 2018 IEEE international symposium on safety, security, and rescue robotics (SSRR); 2018. p. 1–6.
[78] Gibb S, Le T, La HM, Schmid R, Berendsen T. A multi-functional inspection robot for civil infrastructure evaluation and maintenance. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS); 2017. p. 2672–7.
[79] Gibb S, La HM, Le T, Nguyen L, Schmid R, Pham H. Nondestructive evaluation sensor fusion with autonomous robotic system for civil infrastructure inspection. J Field Robot 2018;35:988–1004. https://doi.org/10.1002/rob.21791.
[80] La HM, Gucunski N, Dana K, Kee S-H. Development of an autonomous bridge deck inspection robotic system. J Field Robot 2017;34:1489–504. https://doi.org/10.1002/rob.21725.
[81] Le T, Gibb S, Pham N, La HM, Falk L, Berendsen T. Autonomous robotic system using non-destructive evaluation methods for bridge deck inspection. In: 2017 IEEE international conference on robotics and automation (ICRA); 2017. p. 3672–7.
[82] Billah UH, La HM, Tavakkoli A. Deep learning-based feature silencing for accurate concrete crack detection. Sensors 2020;20:4403. https://doi.org/10.3390/s20164403.
[83] Liu Y, Zhang Z, Liu X, Wang L, Xia X. Efficient image segmentation based on deep learning for mineral image classification. Adv Powder Technol 2021;32:3885–903. https://doi.org/10.1016/j.apt.2021.08.038.
[84] Zhang Z, Wu C, Coleman S, Kerr D. Dense-inception u-net for medical image segmentation. Comput Methods Programs Biomed 2020;192:105395. https://doi.org/10.1016/j.cmpb.2020.105395.
[85] Peng H, Xue C, Shao Y, Chen K, Xiong J, Xie Z, et al. Semantic segmentation of litchi branches using deeplabv3+ model. IEEE Access 2020;8:164546–55. https://doi.org/10.1109/access.2020.3021739.
[86] Liu S, Li M, Li M, Xu Q. Research of animals image semantic segmentation based on deep learning. Concurr Comput, Pract Exp 2020;32:e4892. https://doi.org/10.1002/cpe.4892.
[87] Sparavigna AC. A method for the segmentation of images based on thresholding and applied to vesicular textures. ArXiv preprint arXiv:1612.01131, 2016. https://doi.org/10.48550/arXiv.1612.01131.
[88] Mokhtari S, Wu L, Yun H-B. Statistical selection and interpretation of imagery features for computer vision-based pavement crack–detection systems. J Perform Constr Facil 2017;31:04017054. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001006.
[89] Shah YU, Jain S, Tiwari D, Jain M. Development of overall pavement condition index for urban road network. Proc, Soc Behav Sci 2013;104:332–41. https://doi.org/10.1016/j.sbspro.2013.11.126.