University of Nevada, Reno

Deep Learning Based Concrete Distress Detection System for Civil Infrastructure

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Computer Science and Engineering

by

Tamanna Yasmin

Dr. Hung (Jim) La - Thesis Advisor
December 2022

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

Tamanna Yasmin

entitled

Deep Learning Based Concrete Distress Detection System for Civil Infrastructure

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Hung (Jim) La, Ph.D.
Advisor

George Bebis, Ph.D.
Committee Member

Mohamed A. Moustafa, Ph.D.
Graduate School Representative

Markus Kemmelmeier, Ph.D., Dean
Graduate School

December, 2022

Abstract

In most civil concrete structures, the inspection of structural health is essential. A periodical inspection process ensures that the infrastructure will meet the functional requirements properly or not. To avoid hazardous situations in civil infrastructure, proper maintenance of concrete structures is necessary. The manual visual examination process might provide erroneous results while exploring critical parts of concrete surfaces. As a result, an accurate, safe, and dependable automated process is required for detecting concrete distress. Spalling is a critical distress that can damage concrete surfaces in civil infrastructure. Severe and harmful spalling needs to be taken care of to avoid life-threatening incidents by identifying the location of the distress. Aside from determining the location of the spalling, the severity level of the spalling must also be determined. These severity levels help determine how adverse the situation is and prioritize the process of fixing the spalling. Due to the impact of concrete distress, detecting surface defects like spallings caught the attention of researchers. In this thesis, we have presented approaches to detecting the location of spalling according to its severity level. The proposed methods use deep learning-based approaches and multi-class semantic segmentation. Our approaches have explored two major criteria to detect the spalling and categorize its severity level. Furthermore, we have conducted qualitative and quantitative analyses to demonstrate the performance achieved by the proposed methodologies.

Acknowledgments

The success of this thesis is heavily reliant on the encouragement and the guidance of many others. I want to take this opportunity to express my gratitude to the people who have been very supportive of the successful completion of this thesis.

I am very thankful for my advisor, Dr. Hung Manh La, for his tremendous support and help. I felt motivated and encouraged every time we had a conversation. Without his encouragement and guidance, this thesis would not have materialized. I would also like to thank my committee members, Dr. Mohamed A. Moustafa and Dr. George Bebis, for their advice and for taking the time to review the contents of my thesis.

Thanks to my husband, Mehedi Hasan, for his constant support and encouragement to complete this degree. I would like to thank my parents and my mother-in-law for their support and love.

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 (http://inspire-utc.mst.edu) at Missouri University of Science and Technology, 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 and any other entities.

Table of Contents

List of Figures

Chapter 1

Introduction

Distress in concrete surfaces happens when the surface is exposed to severe environmental phenomena and a lack of preventive measures to deal with the deterioration of structural health. Concrete distress can result in life-threatening incidents. Maintenance of these distress zones is crucial to avoiding hazardous situations [6]. Spalling is one kind of concrete abnormality that can destroy the structure's health, disrupt the appearance of the surface, and make it unhelpful. The heavy loads over the bridges and surface erosion on the walls gradually create the spallings. Spalling can make the pillars of bridges, the ceiling, and the walls of buildings prone to severe accidents. It is required to conduct regular inspections to ensure structural integrity. The inspection results are determined by the knowledge of the person in charge of the task [7]. There are critical positions, like the corners of the pillars of bridges, where manual detection of spalling can be very difficult. Therefore, an autonomous system to detect the exact location and condition of the spalling is very helpful. A regular, automatic inspection is crucial for civil infrastructure to ensure public safety. Keeping this in mind, several autonomous bridge inspection systems have been developed. [8–29]. Several approaches for crack detection have been proposed to avoid the manual inspection

process [30–34].

Because spalling is a type of concrete condition that can compromise the surface's structure, we need a careful and proper detection process to avoid problems with manual inspection. Depending on the condition of the spalling, the after-inspection process can be different. There are several conditions where we need precise information: the area of the spalling should be properly detected; the spalling can be large or deep enough that it needs to be fixed immediately; shallow or small spalling can get priority after the large or deep spalling; and we need a correct prediction of the priority to fix the spalling for a cost-effective solution. Therefore, detection of spalling is not the only solution to conducting an after-inspection process. We need to know the level of severity and condition of the spallings. The level of severity indicates how adverse the condition of the spalling is. As a result, depending on how the spalling is classified, the surface may have a large amount of spalling or deep spalling, where concrete materials are ejected from the surface [35], a small amount of spalling or shallow spalling, or no spalling at all. A deep spalling can be way more dangerous since the concrete surface keeps losing materials and makes it prone to severe accidents. Keeping that in mind, in this thesis we have presented the idea of detecting spalling and categorizing the level of severity.

1.1 Background Studies

In this section we have discussed several concepts related to our proposed approach. Several Deep Encoder-Decoder based architecture will be discussed. Along with these, some backbone networks will be described briefly in this section, which will be helpful to understand the motivation of proposed approaches.

Figure 1.1: Distress in concrete: (a) Spalling (b) Cracking

1.1.1 Distress in Concrete

Distress in concrete occurs due to exposure to extreme environmental hazards and a lack of preventive measures to protect the surface. This distress creates a surface abnormality and destroys the surface look. Aside from its appearance, this surface

abnormality can lead to dangerous situations. Distress in the concrete can make the civil infrastructure unusable.

There are several types of concrete defects that can hamper the surface integrity of concrete: cracking, spalling, scaling, crazing, blistering, dusting, curling, and efflorescence. Cracking and spalling are more common and dangerous defects in buildings and bridges. Different types of concrete distress are shown in Figure 1.1.

Researchers are paying increasing attention to detect these distresses due to their impact on the surface. In this thesis, we present approaches to detect spalling and its severity levels. Spalling refers to the areas of concrete that have cracked and delaminated from the surface. The surface gets deteriorated, and flaking of concrete occurs due to spalling. Spalling can be very large, small, deep, or shallow. The severity of the spalling depends on its size and depth. Therefore, detecting and localizing spalling and its severity level are crucial to preventing any critical situation.

1.1.2 Deep Encoder-Decoder architecture

In our proposed work, deep encoder-decoder architecture refers deep learning-based pixel-wise segmentation model using encoder-decoder networks. These architectures contain two basic parts: encoder and decoder. The encoder-decoder structures are very popular in deep learning architecture for image segmentation [36]. These structures help to extract most useful features from images and make useful correlations among inputs within the network [36]. An overview of deep encoder-decoder architecture is shown in Figure 1.2.

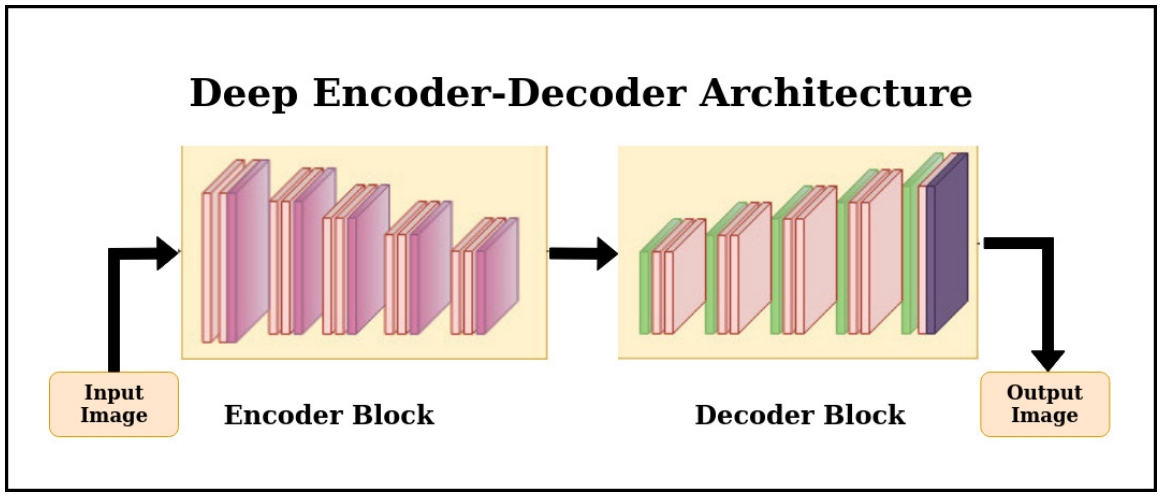

Figure 1.2: Overview of a deep architecture based on encoder-decoder network.

In recent years, several encoder-decoder based deep learning architectures for image segmentation have been proposed; SegNet [3], UNet [4], PSPNet [5], FCN [37], DeepLab [38], DeepCrack [39]. The encoder block usually a backbone network namely, ResNet-50 [40] [41], VGG-19 [42], Xception [43], and MobileNet [44]. The decoder blocks are commonly bilinear interpolation, deconvolution, or dense up sampling convolution [45].

A pre-trained classification network is usually used as the encoder network, followed by a decoder network. The encoder block consists of convolution, activation, and pooling layers. Instead of pooling layers, the decoder block contains upsampling layers. The encoder block creates low resolution discriminating features from the image data. To obtain the segmentation, the decoder blocks semantically project the discriminating features learned by the encoder on each pixel into high-resolution pixel space using upsampling layers [46].

1.1.3 Backbone Networks

The backbone network's task is to take image as input and extract the feature map to help the architecture with classification or segmentation. These network architectures may effectively extract an image's feature mapping, providing a strong base network for semantic segmentation [45]. Several backbone networks have been used in recent years in deep learning architectures as feature extractor: ResNet-50 [40] [41], VGG-19 [42], Xception [43], VGG-16 [47] and MobileNet [44].

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 [48] [49] [50] [51]. For affect detection in developing human computer interaction systems, ResNet-50 and VGG-16 have been used [47]. 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 facial expression, and to detect and localize rebar for bridge deck inspection and evaluation [52] [53] [54] [55] [56].

1.2 Literature Review

In recent years, a number of spalling detection techniques have been proposed. The proposed methods detects spalling in metro tunnels, subway networks, Bridges, and railway surfaces. Methods for spalling localization, evaluation, and detection based on machine vision, laser scanning, Deep learning, infrared thermography have enticed the attraction of researchers. However, there aren't many methods for identifying and categorizing the severity of concrete spalling. An image processing based approach has been proposed for detecting spalling in subway networks [57]. The color image of spalling area has been processed to remove the noise and to find out the surface

features of spalling area. A 3D visualization has been created from the extracted features of spalling. To detect the severity and spalling depth the authors used projection of the spalling intensity curve and regression analysis. A concrete automatic spalling detection technique has been suggested for metro tunnels [58]. A spalling detection system for rail surface has been proposed based on real-time visual inspection system [59]. Two sub-systems have been used in this proposed method; one image acquisition system and one image processing system. The image acquisition sub-process captures data or images using a camera. The image processing sub-process removes any abnormal condition on the images like bright or light shadow on the rail track. The rail track images captured by the camera are segmented first using a trail track extractor. Then the spalling on the rail surface are detected by using the information based on histogram curves in the longitudinal direction of the track image. This method uses a 3D cloud point that contains data on the inner wall and outlier points to detect spalling damage on the tunnel surface. A machine learning and vision-based approach for subways has been developed to detect and quantify the spalling [60]. This approach combines two processes; process of extracts important features about spalling from images by removing noise, and detects surface distresses in subways. To detect spalling on the rail surface an algorithm has been proposed [61]. This is an optical detection algorithm based on visual salience. There is a difference between the normal surface and the distressed surface. if the unnecessary noises can be removed form the neighborhood area of spalling then the difference between spalling and non spalling area become more discrete. Based on this concept the authors detects the spalling area by using a threshold value. An automated 3D spalling defects inspection system in railway tunnel linings has been proposed [7]. The method used laser intensity and depth information for spalling detection. A mobile laser scanning system has been used for preparing a database

containing intensity images and depth images of the railway tunnel linings. These information are used to detect the volume of spalling accurately. For extracting the concrete spalling features automatically a spalling intensity depurator network is also proposed. The proposed work also produces 3D inspection results containing a quantitative analysis of the spalling defected area. To detect two crucial structural failures, cracks and spallings automatically in buildings and bridges deep learning approaches are developed [62]. The work has proposed three different types of Mask R-CNNs. The damaged area in the bridges and buildings are checked continuously by using the segmentation process. The deep CNN architectures can be extended to detect and evaluate surface damages. A detection technique based on Faster RCNN is proposed for several damage types [63]. This work taking into account surface crack, facade spalling and concrete spalling, and severely buckled rebars and spalling with exposed rebars. In Faster RCNN, a total of nine anchor boxes were used with three different scales and three different aspect ratios [64]. The authors of the proposed work have used seven different scales and eight different aspect ratios, a total to 56 anchor boxes to improve detection accuracy. Since, Automatic and early detection of concrete damages are very crucial, image texture and a piece-wise linear stochastic gradient descent logistic regression are used to detect automatically detect spalling in concrete [35]. Image textures are used to extract features from images namely, statistical properties of color channels, gray-level co occurrence matrix, and gray-level run lengths. These extracted features from image texture are used to categorize condition of concrete surface. The pattern recognition process is conducted by piece-wise linear stochastic gradient descent logistic regression to detect two classes of spalling as non-spall and spall using image textures. A terrestrial laser scanner has been used to simultaneously localize and quantify spalling defects on concrete surfaces [65]. The features which have complementary properties to each other are helpful for improved

localization and quantification of spalling defects have been developed and combined in the proposed method. To extract the information, for example, condition of the damage portion of investigated surface region, accurate localization, and size a defect classifier is developed. The concrete structure is scanned by the terrestrial laser scanner and a region of interest is selected. The scanner captures the 3D coordinate information of the scanned points inside the selected region. The proposed method then start the process of defect detection one the raw scanned data is ready. Different types and sizes of the specimen crucially effect spalling in concrete under hydrocarbon fire exposure are crucial [66]. This proposed work isolated the variables that effect concrete spalling when exposed to a hydrocarbon fire. Four different types of specimen size were analyzed to conduct the study consisting of cylinders, columns, and panels. In concrete mixes to determine the effect of aggregate size on concrete spalling three aggregate sizes were used. The investigation on aggregate type has been done as well to determine the effect on concrete spalling. Concrete spalling can be detected by Active infrared thermography [67]. The irradiation devices namely, halogen lamps, xenon arc lamps and far-infrared irradiation devices are used to detect concrete spalling by heating the surface of the concrete. Active Infrared thermography uses artificial heat to produce a temperature gradient and the passive infrared method uses the natural conditions caused by solar radiation. The proposed work used active infrared method since it is capable of taking measurements unhampered by meteorological conditions. Technologies that generates point cloud such as photogrammetry, laser scanning and Light Detection and Ranging (LiDAR) has been used for surface damage detection (spalling) [68]. The proposed work used point cloud data to detect the spalling and quantify its key properties in Reinforced concrete columns. In first phase of the method the noise points are removed and the coordinate system of the captured point cloud data are calibrated and sliced into thin layers for the damaged

areas. The second phase includes detecting the point for distressed and non distressed area. Finally, by using linear interpolation the spalling area and lost concrete volume are calculated.

A computer application for automatically evaluating spalling and detecting spalling severity in concrete bridges has been developed [69]. To detect spalling this proposed approach developed a single-objective particle swarm optimization model based on Tsallis entropy function. Then in the second phase the severity of spalling is evaluated by generating a compendious analysis of the bridge deck image using Daubechies discrete wavelet transform feature description algorithm. The second phase conducts a hybrid artificial neural network-particle swarm optimization model as well for the prediction of spalling area correctly. This hybrid model is used to overcome the drawbacks of the gradient descent algorithm. It is crucial to detect spalling and severity correctly and in a timely manner. The concept of computer vision includes extract numeric information from depth images, digital images, videos, and 3D point clouds, process the information, and take action [70]. A computer vision-based approach to classify concrete spalling severity has been developed [71]. To detect the severity the proposed method uses concrete images. The level of severity here be categorized as deep spall or shallow spall. The features of concrete surface including statistical measurement of color channels, gray-level run length, and center-symmetric local binary pattern helps the support vector machine classifier optimized by the jellyfish search metaheuristic to divide the data into shallow spalling and deep spalling using a decision boundary. An entropy-based automated method has been developed including three significant parts using computer vision technologies [72]. The spalling detection phase proposed a segmentation model that adopts a multi-objective invasive weed optimization and information theory-based formalism of images to detect spall. In the feature extraction phase, to get the efficient image information an integration of sin-

gular value decomposition and discrete wavelet transform are integrated . The third phase includes the process of developing a rating system of spalling severity based on its area and depth. Computer vision approaches provide solution to detect spalling severity in concrete [73]. The proposed work used Extreme Gradient Boosting Machine and Deep Convolutional Neural Network (DCNN) to classify image data into shallow spall and deep spall. The features extracted from image of concrete surface to get the properties of spalling. The feature extraction methods include, local binary pattern, center symmetric local binary pattern, local ternary pattern, and attractive repulsive center symmetric local binary pattern. The Aquila optimizer metaheuristic enhanced the prediction performance of Extreme Gradient Boosting Machine.

1.3 Motivation and Contribution

The maintenance of the structural health of concrete is crucial. Therefore, several automated systems have been developed for the inspection of concrete bridges. The detection of spalling and severity level classification have a significant impact on ensuring the wellness of civil infrastructure. The literature review section describes several methods for spalling detection. The severity level of spalling is important as well, since it helps to prioritize the process of spalling maintenance. For the detection of spalling severity classification, there are very few approaches. Our proposed approaches detect the spalling and the level of severity of the spalling. Moreover, we have analyzed the importance of including unaffected areas in the spalling severity level. The severity of the spalling can be determined by how deep or shallow it is, as well as how large or small it is. The most important part is the discrete segmentation of spalling classes based on severity level in order to categorize spalling severity with proper visual pixel mapping. As a result of the benefits of image segmentation

techniques in various fields, we considered this problem to be one of semantic segmentation. As a consequence, we have proposed two approaches to detect spalling and classify the severity level. We have designed our proposed method around the use of deep architecture using different encoder-decoder networks. Our proposed approaches use pixel-by-pixel multiclass semantic segmentation to classify the severity levels. To determine the best result and combination of the deep architecture with encoder-decoder networks, we have conducted comparative analyses.

Our main contribution of this thesis are mentioned below:

- deep architecture based method with different encoder-decoder networks to detect the spalling and the severity level.
- Multi-class segmentation using pixel-wise categorization to determine the severity level of spalling.
- Qualitative and quantitative analysis of the deep architectures to get the best result and identify the best combination with encoder-decoder networks.
- Chapter 2 delineates the deep architecture based proposed approach to detect spalling and severity level of spalling as Large spalling, small spalling, and no-spalling.
- Chapter 3 presents an extended approach of the method described in Chapter 2. In this proposed approach the spalling severity level have been categorized as deep spalling, shallow spalling, and no-spalling.
- Chapter 3 further proposes a pixel-wise severity ranking method to calculate the ranking of severity for deep spalling areas.

1.4 Thesis Organization

We have organized this thesis as follows: Chapter 1 discusses the background concepts related to our proposed work and the literature review that presents the related works. Chapter 2 presents the proposed method for detecting spalling and the severity level, including the research methodology, results, and discussion. Chapter 3 covers another proposed method for detecting and localizing spalling and severity level, along with the methodology of the proposed work, results, and discussion. In chapter 4, we provided concluding remarks that summarized our research work as well as some potential future research directions.

Chapter 2

Spalling Severity Detection System-1

2.1 Introduction

Structural health inspection is vital to civil infrastructure, and concrete is essential to that [33]. Monitoring any structural distress periodically on roads, bridge decks, highways, and buildings is crucial. Since proper inspection and maintenance in these areas are necessary to avoid severe, life-threatening disasters, any spalling in the concrete can lead to serious accidents [71]. Proper inspection and timely maintenance should be done to avoid unwanted events [74]. Another crucial part is detecting the severity of spalling and ensuring proper maintenance based on the detection result. Different types of concrete spalling and their severity are shown in Figure 2.1. Traditional methods have been employed to detect and inspect structural defects. Manually detecting spalling is time-consuming and prone to human errors while detecting these anomalies on the concrete, especially if they happen at a crucial point like under the breeze or underwater beams [34]. We need an autonomous system with little or no human intervention that can solve the issues with traditional methods.

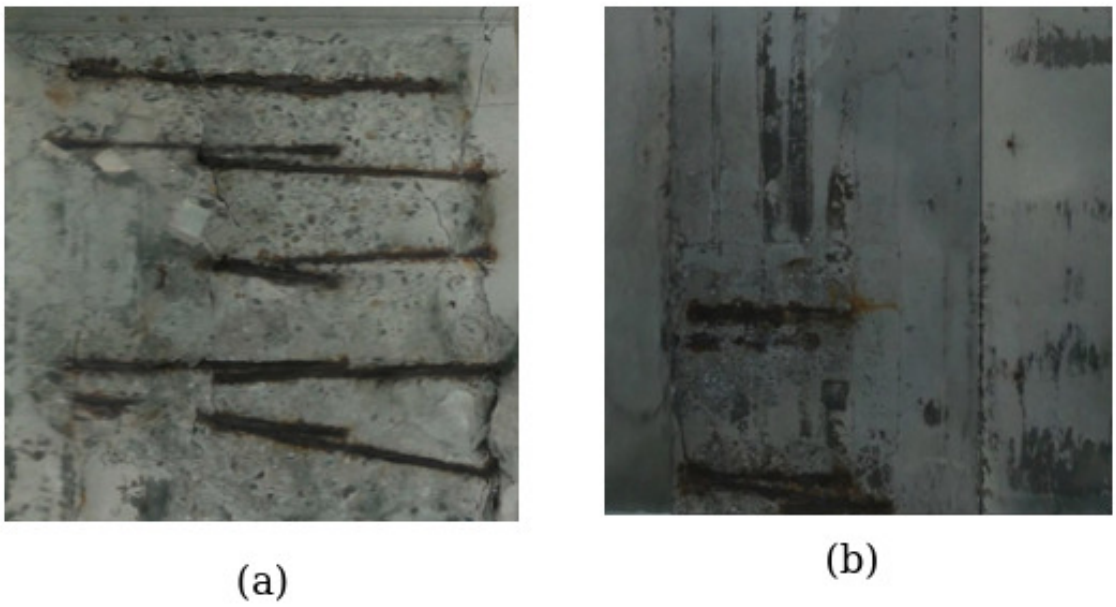

Figure 2.1: Different level of spalling: (a) Large Spalling (b) Small Spalling.

Deep learning methods are instrumental in detecting and inspecting concrete spalling. Several image processing methods have emerged. However, these approaches produce unnecessary image features. Even though these methods are simple and computationally inexpensive [75], they need a filtration process to remove unwanted features. Sometimes, the filtration process can remove useful features and keep the ones that are not necessary. Approaches combining machine learning and image processing are computationally expensive and require preprocessing of images. Convolutional neural networks (CNNs) are experts in this case for classifying and detecting spalling in concrete structures. They can extract spatial-visual features from images that are very useful to increase performance to detect structural defects [34]. The challenges that lie in detecting and classifying the severity level of spalling are extracting features and implementing appropriate methods in real-life applications. Furthermore, one of the major challenges is managing large amounts of data to extract the feature from the concrete environment. Our proposed work leads to overcoming these

challenges. While previous works have focused on surface defects like cracks and spallings, few have addressed spalling severity detection. Spalling severity can be measured using depth (deep or shallow), size (large or small), or no spalling at all. The most crucial part of determining the severity level is segmenting the spalling area in a properly identifiable manner. Therefore, we have proposed the use of deep architecture using different encoder-decoder networks with pixel-by-pixel multiclass semantic segmentation to classify the severity levels of spalling as no-spalling, small spalling, or large spalling. To get the optimal deep architecture, we tested several encoder-decoder networks to compare and analyze the performance of the detection processes. Moreover, we have provided a comparative analysis of the deep architectures with different encoder-decoder networks to predict the best result among the proposed architectures.

The main contributions of this proposed work include:

1. An encoder-decoder-based deep architecture to detect the spalling in concrete.

2. Detection of the level of spalling severity using multi-class pixel-wise segmentation.

3. Comparative analysis between deep architectures with different encoder-decoder networks for spalling detection and severity level.

2.2 Research Methodology

Our proposed method for detecting and classifying spalling severity level is based on Deep encoder-decoder networks. Several encoder-decoder based deep convolutional networks have been proposed [76] [4] [5]. In this work, we have selected SegNet [76] and UNet [4] for our proposed architecture. Moreover, in results and discussion sec-

tion, we discuss the comparison between these two architectures for different encoder-decoder networks based on spalling detection and severity level classification.

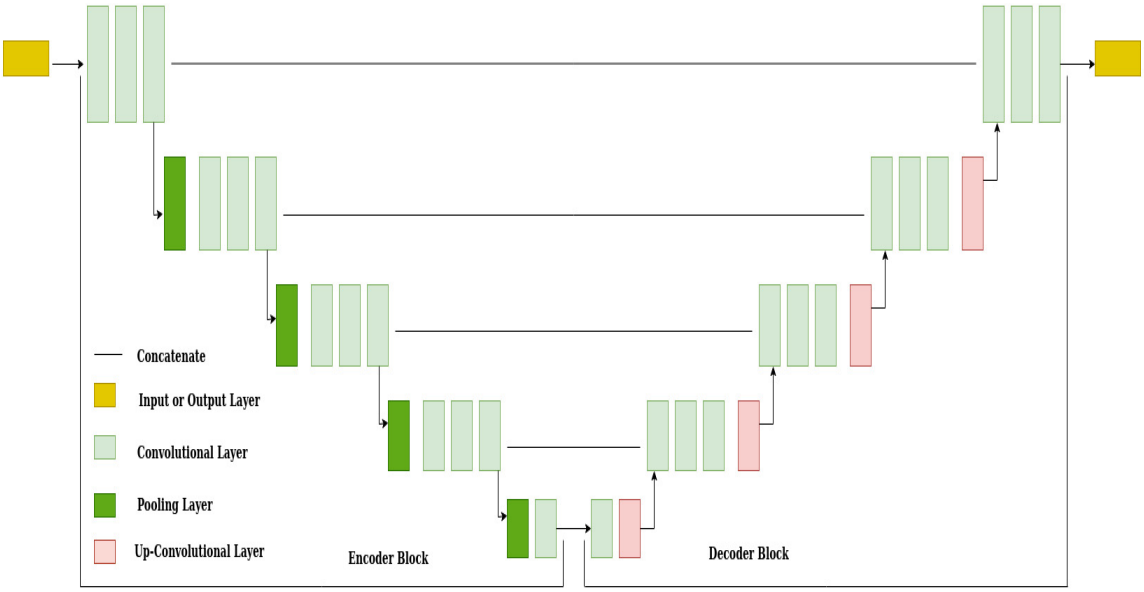

Figure 2.2: Architecture for UNet with Encoder and Decoder [1]

2.2.1 UNet

The UNet architecture consists of encoder and decoder blocks. The architecture for UNet is shown in Figure 2.2. Each encoder block contains two 3×3 convolutions [4]. Each of the convolutions is followed by a ReLU activation function. The encoder part of the UNet architecture works as a feature extractor and acquires the features of the image. The encoder network has half the spatial dimensions and doubles the number of feature channels of each encoder block. The encoder blocks and decoder blocks are connected via a link. The resulted output of the ReLU activation function from the encoder blocks makes a connection to the corresponding decoder blocks. The connection between the encoder-decoder block contains two 3×3 convolutions, and each of the convolutions is followed by a ReLU activation function. This connection helps the decoder to produce better semantic features by providing supplementary information.

The decoder network has half the number of feature channels and doubles the spatial dimensions. The starting phase of the decoder contains a 2×2 transpose convolution. The feature maps are passed through the connection between the encoder and decoder using a concatenation process of convolution and the connection. In the decoder part, a segmentation mask is generated. The resulting output produced from the last decoder is passed through a 1×1 convolution with sigmoid activation. The segmentation mask is converted into pixel-wise classification using an activation function.

2.2.2 SegNet

The SegNet architecture consists of encoder and decoder networks and is proposed for pixel-wise semantic segmentation. Figure 2.3 shows the architecture for SegNet with encoder-decoder block.

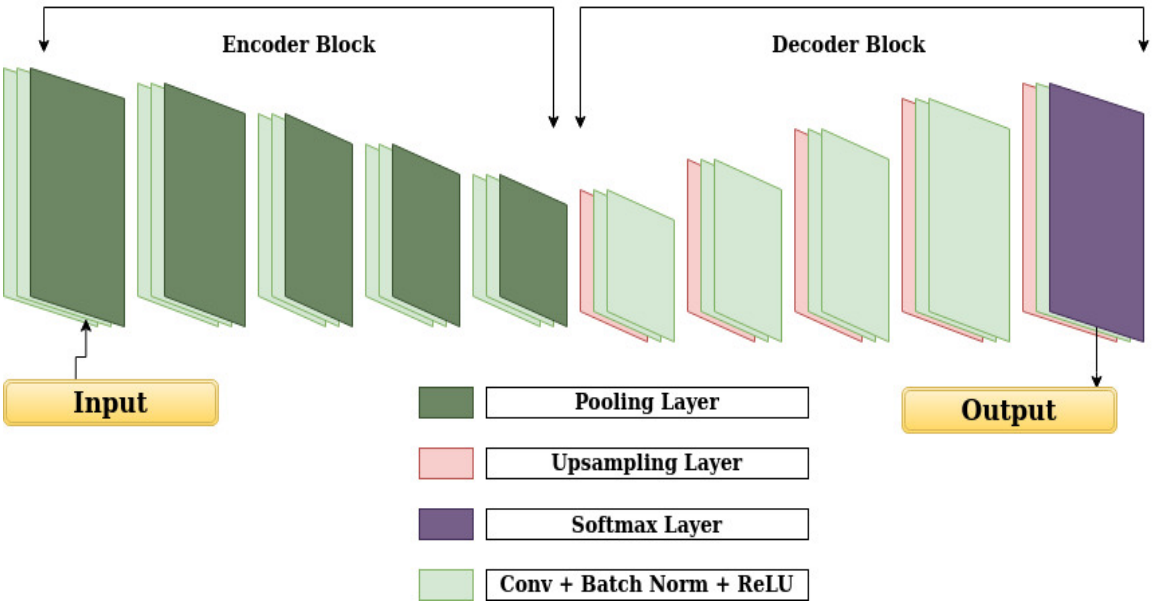

Figure 2.3: Architecture for SegNet with Encoder and Decoder [2]

The encoder network has 13 convolutional layers for feature maps, leading to object classification. The encoder network performs dense convolutions, ReLU non-

linearity, a non-overlapping max pooling [76]. The max pooling is done with a 2 × 2 window. The final step of the encoder is down-sampling. The decoder network performs up-sampling and convolutions [3]. In the end, there is a softmax classifier for each pixel. The max pooling indices of the corresponding encoder layer are called when the decoder conducts the up-sampling. In the end, there is a K-class softmax classifier to identify the class for each pixel. Figure 2.3 shows the architecture for SegNet with encoder-decoder block.

2.2.3 Preparation of Dataset

In Deep network architectures, we need large amounts of data [33] to train, validate, and test the model. Therefore, collecting the dataset is a crucial part. For our proposed architecture, we have used a data augmentation process that alleviates the problem of managing the data. We have collected images of different bridges for our proposed work. We also collected the image at different times of the day to maintain the non-uniformity of the environment. The severity level of spalling is defined based on the spalling area and the overall area of the input images. The spalling severity level is large if the overall affected area is more than 50%. Small spalling areas are defined as those with an overall affected area of 1 to 49%.

For each image in our dataset, we annotated the image as non-spalling or spalling with the levels of severity. Therefore, pixel mapping is automatically created during the training process by labeling the image. Since the label of images follows the RGB range, the non-spalling area and the severity levels of spalling are annotated with RGB combinations.

After labeling every image, we have used the data augmentation process to prepare a dataset of sub-images for every original image. For each image, the augmentation process selects a random image and a random pixel point for the labeled image.

According to that point, a sub-image and pixel map of the corresponding original image are created. From the pixel point, the augmentation method generates several sub-images randomly by flipping or rotating the pixel map.

2.2.4 Proposed Deep Architecture

We have proposed the method using two different types of deep architectures with different encoder-decoder networks as backbone for detecting the spalling and the level of severity. Due to the advantage and performance of deep learning-based image segmentation architecture in several fields [77] [78] [79], we have proposed the use of deep architectures with encoder-decoder networks to detect spalling and severity level. First, we detect spalling and the severity level using SegNet and UNet architectures. Finally, we discuss the comparative analysis of their performances with different encoder-decoder networks for detecting the spalling and the level of severity. We have considered spalling detection and severity level classification as multi-class semantic segmentation. We have implemented the proposed deep architecture with UNet (Figure 2.2), and SegNet (Figure 2.3). For both of the architectures, we have used three different encoders as the backbone network: a pre-trained CNNs model, such as the ResNet pre-trained model, Xception, or the VGG pre-trained model.

The overview of proposed deep architecture is shown in Figure 2.4. The proposed architecture has an encoder block and a decoder block, making it an encoder-decoder network. For the encoder part, we have employed ResNet-50, VGG-19, and Xception. 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 encoder to form integrated feature points after the feature extracted from the convolution layers.

The difference between the decoder block and encoder block is the up-sampling

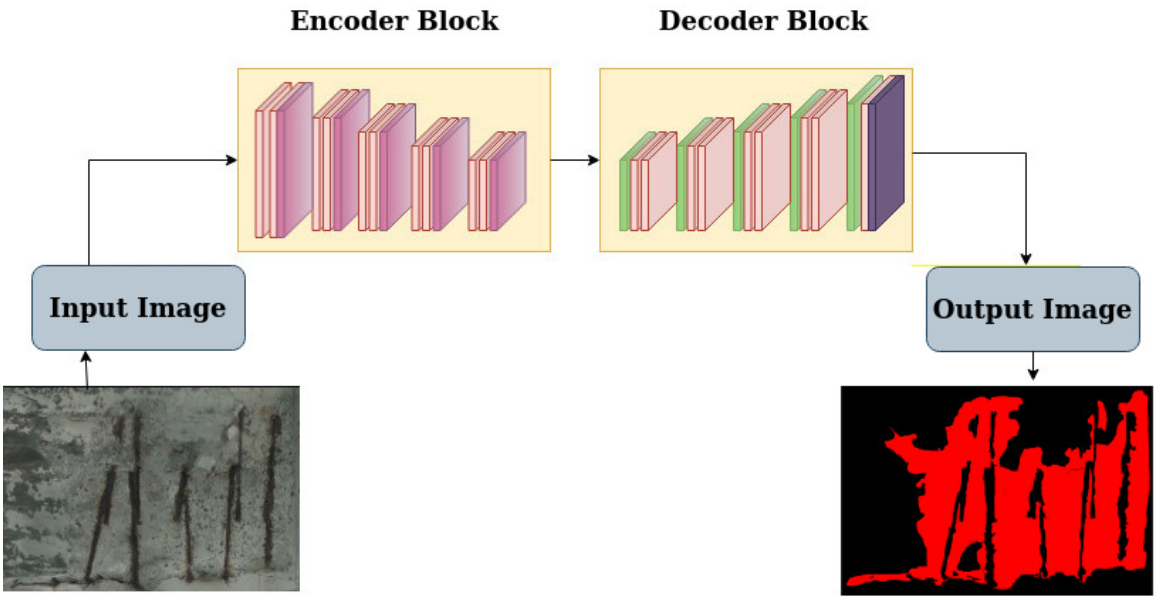

Figure 2.4: Proposed Encoder-Decoder Based Deep Architecture

layer instead of the pooling layers. The decoder block is essentially a mirrored image of encoder. It gradually upsamples the encoder's output and semantically projects into high resolution pixel space from the low-resolution identifiable feature maps. For our proposed method, we have used the default decoder network in the decoder block as the UNet and SegNet architectures shown in Figure 2.2, and 2.3 respectively.

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 different category at various stages into the pixel space learned by the encoder to fully segment the target region [80]. Moreover, using the concatenation process the decoder connects to the corresponding encoder and help to reduce the loss happened during down-sampling process. 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 as no-spalling, large spalling, or small spalling.

2.3 Result and Discussion

This section describes the comparative analysis of the proposed architecture with different encoder-decoder networks. Moreover, we will discuss the data processing and system configuration used for training and testing the models.

2.3.1 Experimental setup

In this study, we have collected images of different types of spalling from bridges to train the model for classifying the severity level. These images contain various noises like faded colors, stones, and oil spills. As a result, the training data contains noises similar to those found in the real world. It is difficult to detect any abnormality in the crucial corners of bridges, for example, the intersection of pillars, due to 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. For the multi-class classification problem, we used categorical cross-entropy. The Adam optimizer was used to optimize the architecture with a learning rate of 0.001. The training and testing ran on a system with a GTX 1080 GPU.

The GIMP software was used for the dataset to generate a pixel map. We have augmented the images and prepared a dataset of 10000 images for training and 2000 images for validation with a resolution height $\times$ width of 1024×1024. For each test, 100 images were used. For each encoder-decoder network, we have used a sub-sample size of each image with height $\times$ width: 416×608.

To evaluate the performance, we have used several metrics, which will be described in the next. Since we have used an encoder-decoder-based architecture to detect the spalling and multi-class segmentation to classify the severity level, we have compared the performances of two deep encoder-decoder-based architectures for different

encoder-decoder networks.

2.3.2 Quantitative Analysis

This section has prepared a performance-based quantitative analysis for two deep architectures with different encoder-decoder networks. Table 2.1 shows the overall performance for spalling detection with severity level classification. We have used Dice loss, 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 used Equation 2.1, 2.2, and 2.3 for Accuracy, Precision, and Recall respectively. Here, true positive (TP) means the number of spalling pixels correctly predicted, and true negative (TN) means the number of spaling detected as non-spalling, which are non-spalling areas by pixel mapping. False positive (FP) means the number of pixels detected as spalling incorrectly, and false negative (FN) means the number of pixels detected as non-spalling erroneously.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.3}$$

With two deep architectures UNet [4] and SegNet [76], we used VGG19, Xception and ResNet50 as encoder-decoder networks. The results from table 2.1 show that when we used the UNet framework with the ResNet-50 combination performed better than any combination. Therefore, we can say that the complexity of the number of layers does not negatively impact the performance of the UNet framework.

While using ResNet-50 with SegNet architecture provided a different set of perfor-

Table 2.1: Quantitative performance comparison between two Deep architectures with different encoder-decoder networks.

Method	Dice Loss(%)	Precision(%)	Recall(%)	Accuracy(%)
UNet (VGG19)	8.5	85.4	91.3	98.3
UNet (Xception)	8.6	85.5	90.3	98.2
UNet (ResNet-50)	7.9	92.3	91.9	98.5
SegNet (VGG19)	8.5	85.6	91.2	98.3
SegNet (Xception)	8.8	85.2	90.0	98.0
SegNet (ResNet-50)	8.7	85.4	90.2	98.1

mances (Dice Loss: 8.7%, precision: 85.4%, Recall: 90.2%, and Accuracy: 98.1%). With SegNet architecture, the most promising results are given by the SegNet-VGG19 combination. The performances, including Dice Loss, Precision, Recall, and Accuracy for all the other architecture-encoder combinations, can be seen in Table 2.1. We do not have any comparative study with previous works. We have used multi-class segmentation for classifying the severity level as no-spalling, small, and large spalling. Previous studies classifies severity level as shallow or deep spalling [73] [71] and predicted severity rating according to area and depth [72]. In our proposed approach, the comparative analysis provides the best result between the deep architectures with different encoder-decoder networks for detecting the spalling severity.

In the experimental setup section, we mentioned the validation dataset. The validation and training loss for three deep encoder-decoder combinations are shown in Figure 2.5, 2.6, and 2.7 (UNet with ResNet-50, UNet with VGG19, and SegNet with VGG19, respectively). We have calculated the loss using categorical cross-entropy (for multi-class classification). For UNet with ResNet-50 (Figure 2.5), around epoch 80, the validation loss stopped improving enough. After 85 epochs, approximately, the training loss started converging to the validation loss.

The UNet with VGG19 in Figure 2.6, shows that, the validation loss stopped improving enough around epochs 80. The training loss started converging to the validation loss around 90 epochs.

For SegNet with VGG19 (Figure 2.7), the validation loss stopped improving enough around epochs 80. The training loss started converging to the validation loss around 90 epochs.

For all three deep encoder-decoder combinations shown in Figure 2.5, 2.6, and 2.7 (UNet with ResNet-50, UNet with VGG19, and SegNet with VGG19, respectively), the models fit nicely around 100 epochs. More training may result in overfitting to the models. Therefore, we selected 100 epochs to train the models.

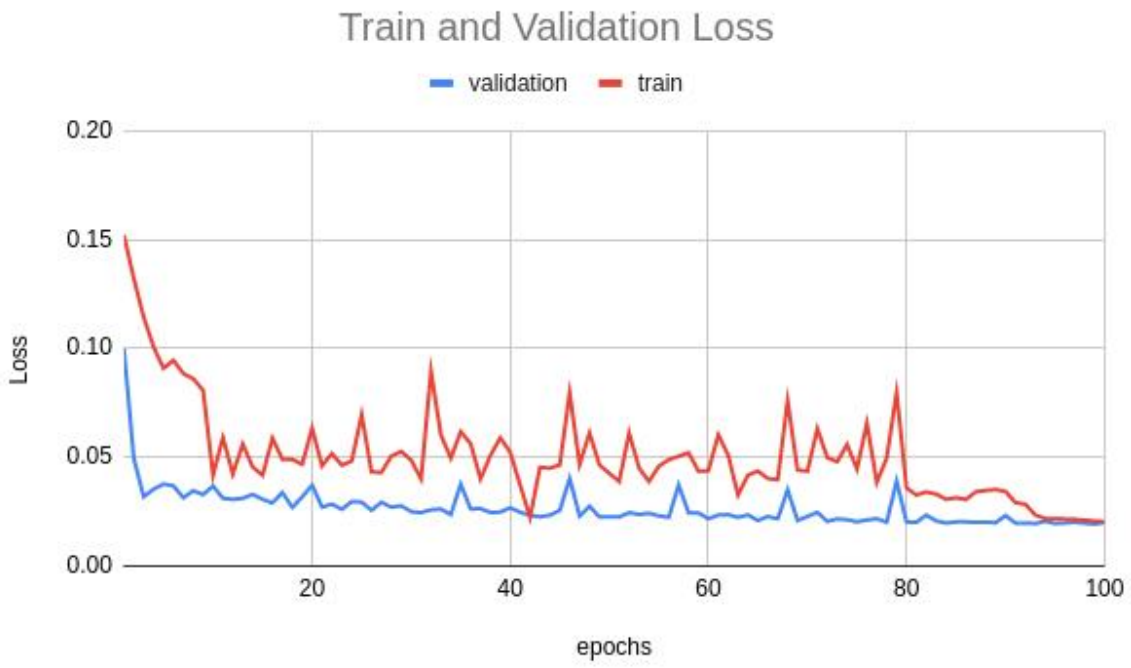

Figure 2.5: Training and validation loss for UNet with backbone network ResNet-50.

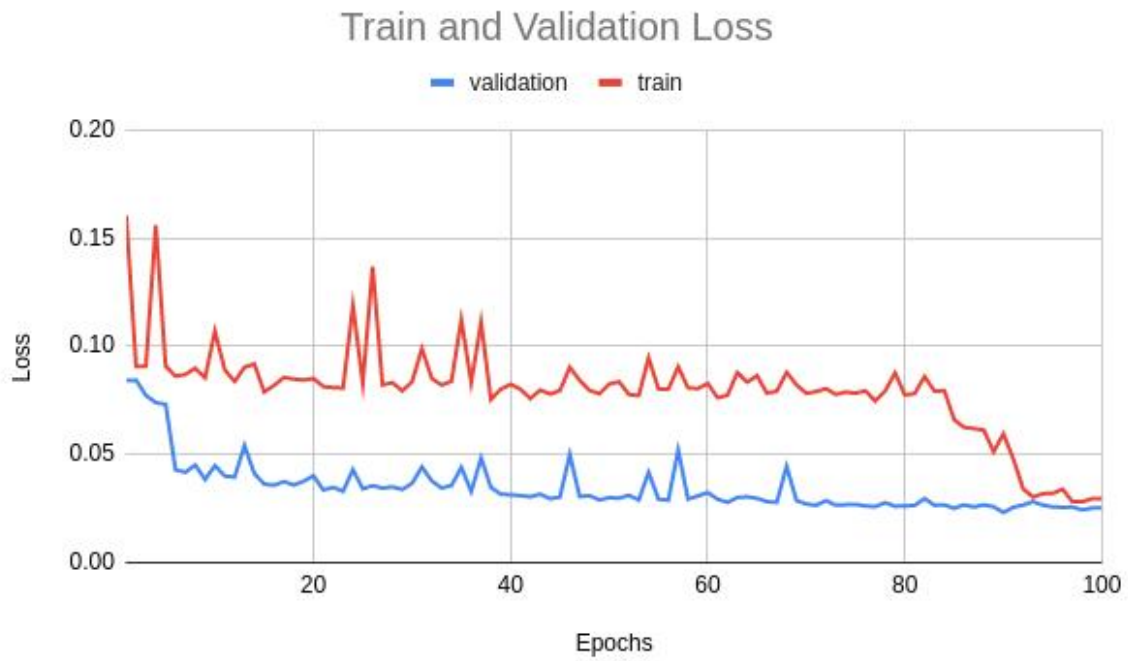

Figure 2.6: Training and validation loss for UNet with backbone network VGG19.

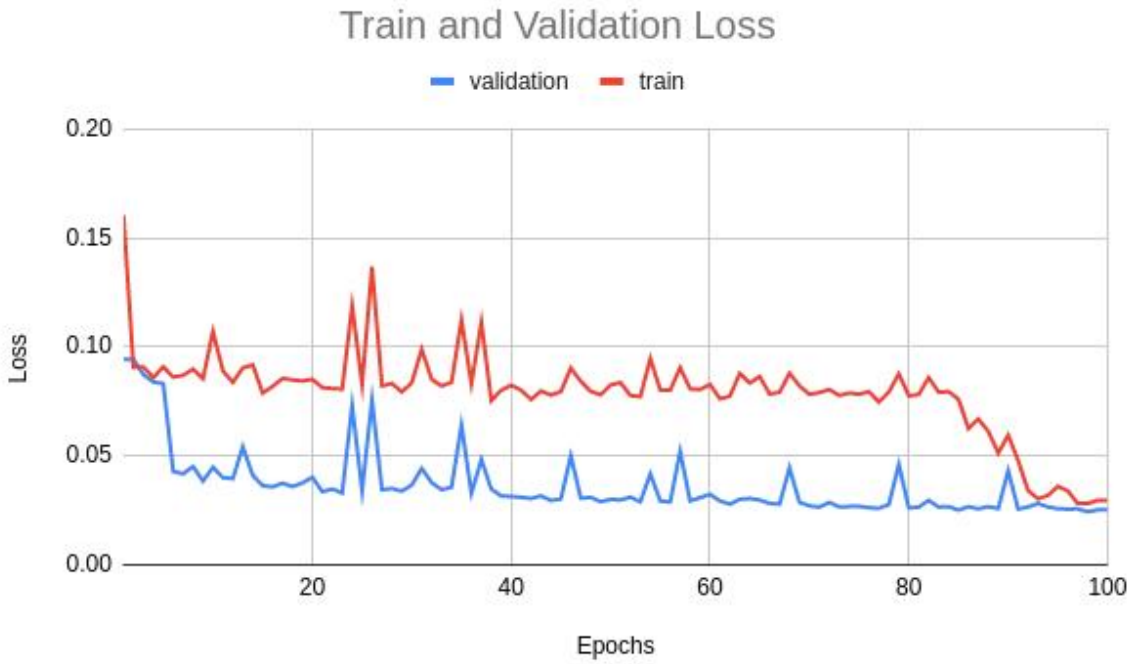

Figure 2.7: Training and validation loss for SegNet with backbone network VGG19.

2.3.3 Qualitative Analysis

This section shows the non-statistical evaluation of our proposed architecture for spalling detection and the severity level. Figure 2.8 presented the qualitative performance for the deep UNet architecture for different encoder-decoder networks. Figure 2.9 shows the performance analysis for SegNet architecture with the encoder-decoder networks.

Table 2.1 already shows that the UNet architecture with ResNet-50 shows better results than all the architectures. In Figure 2.8, from the left, the images are from our dataset, then the pixel map for the images (shown as ground truth), then the multi-class classification for each image. For a better view, we have provided separate images for the spalling severity classification as Large spalling and Small spalling. The area labeled with the color black has been considered as an area with no spalling. The qualitative and quantitative analysis shows that the accuracy for detecting spalling and severity classification is better than others. For that reason, the UNet architecture with ResNet-50 shows results similar to ground truth.

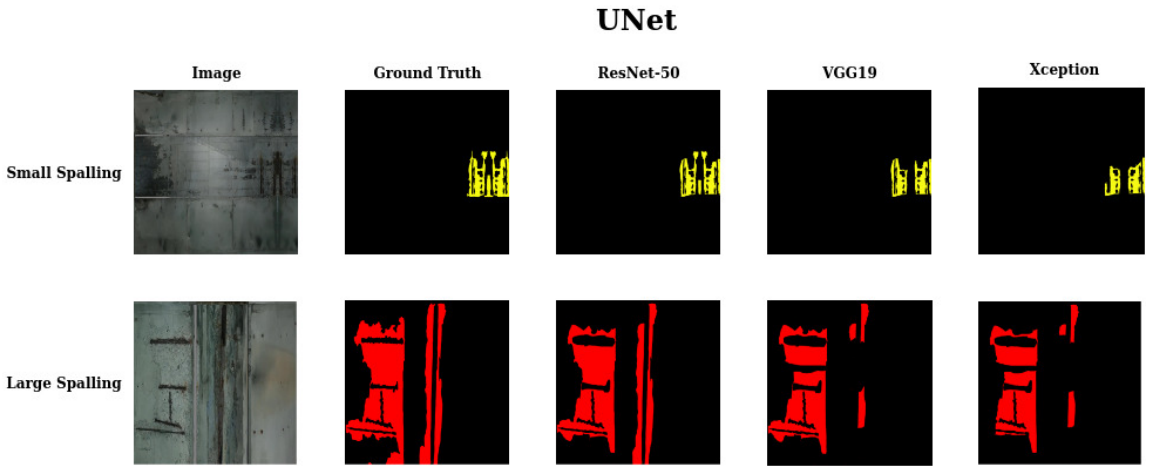

Figure 2.8: Results shown for UNet framework with different encoder-decoder networks

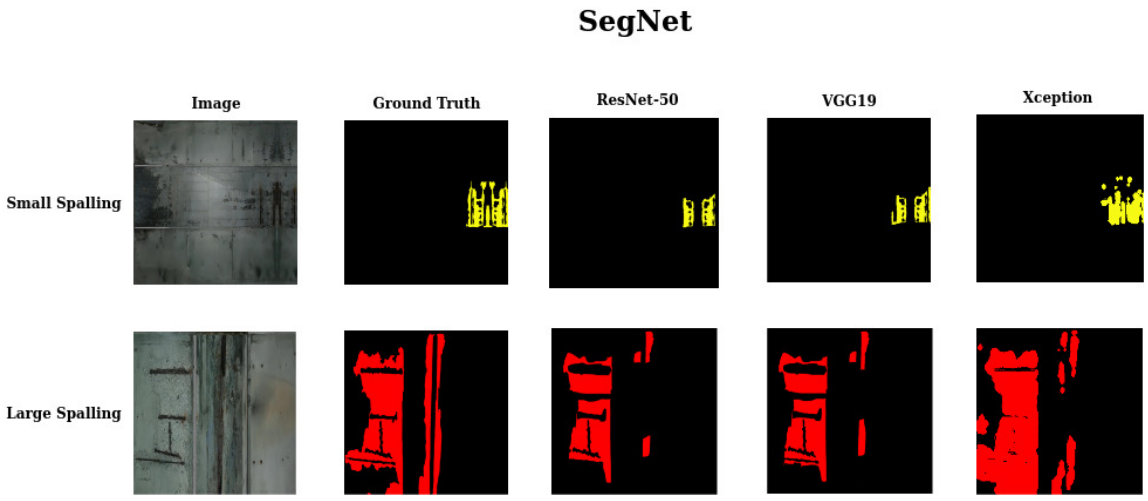

Figure 2.9: Results shown for SegNet framework with different encoder-decoder networks

In Figure 2.9, we have compared the results of SegNet architecture with encoder-decoder networks and the image's pixel map (ground truth). For SegNet architecture, the VGG19 encoder shows better results than others. We have separated the pixel map for spalling severity classification as Large spalling and Small Spalling in Figure 2.9. The no-spalling area has been labeled as color black.

We can infer from our quantitative and qualitative analysis that the UNet architecture shows comparatively better results with ResNet-50, VGG19, and Xception encoders. The SegNet architecture with Xception encoder-decoder network shows that the result was not very accurate compared to the ground truth.

Chapter 3

Spalling Severity Detection System-2

3.1 Introduction

The maintenance of the structural health of concrete is crucial. It is required to conduct regular inspections to ensure structural integrity [81]. Therefore, the detection of spalling and severity level classification have a significant impact on ensuring the wellness of civil infrastructure. From the above discussion in the literature review, we can infer that there are several methods for spalling detection. For the detection of spalling severity levels, there are very few approaches. The severity level of spalling is important as well, since it helps to prioritize the process of spalling maintenance. There are certain shortcomings in the previous proposed approaches for detecting spalling severity levels. "No-spalling" areas were not categorized as a level of severity. To properly identify distressed surfaces and non-affected areas, these "no-spalling" areas should be included in the severity level. Moreover, another important part is the discrete segmentation of spalling classes with proper visual pixel mapping based on severity level. As a result, we present two major criteria in this thesis to de-

tect the severity level of spalling. Chapter 2 categorizes the severity level as large spalling, small spalling, or shallow spalling. In Chapter 3, we have classified the level of severity as 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. As a consequence, we have proposed a method to detect spalling and classify the severity level. We have designed our proposed method around the use of deep architecture using different encoder-decoder networks. Our proposed approach used pixel-by-pixel multiclass semantic segmentation to classify the severity levels of spalling as no-spalling, shallow spalling, or deep spalling. Figure 3.1 shows the level of severity according to depth of the spalling.

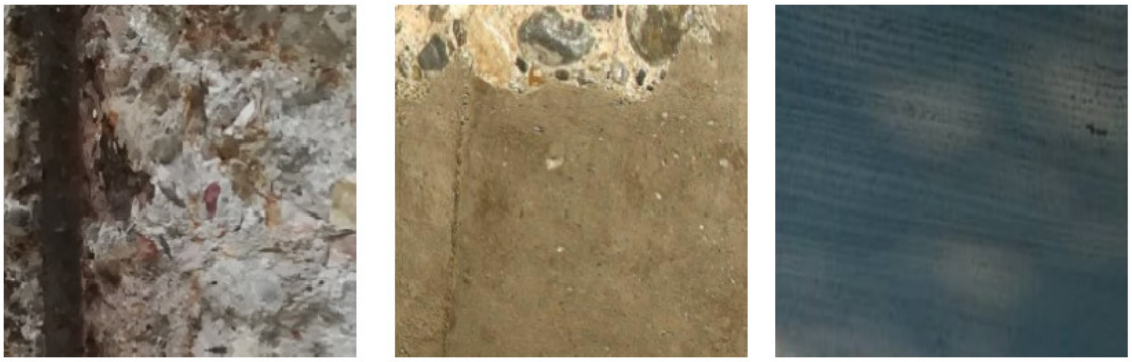

Figure 3.1: Different level of spalling: (a) Deep Spalling (b) Shallow Spalling (c) No-Spalling.

To determine the best result and combination of the deep architecture with encoder-decoder networks, we have conducted a comparative analysis. According to the severity level, deep spallings are more crucial since they affect the concrete surfaces more alarmingly. Therefore, it is essential to analyze the affected deep spalling area more. The affected area of a deep spalling may vary according to different sizes. Hence, we have proposed an approach to calculate the pixel-wise ratio of the deeply affected

spalling area with respect to the overall image area.

Our main contributions in the proposed approach are:

- A deep architecture based method with different backbone networks to detect the spalling and the severity level.
- Multi-class segmentation using pixel-by-pixel categorization to determine the severity level of spalling as no-spalling, shallow spalling, or deep spalling.
- A pixel-wise severity ranking method to calculate the ranking of severity for deep spalling area.
- Qualitative and quantitative analysis to get the best result and identify the best deep architecture combined with a backbone network.

3.2 Research Methodology

In this section, we have presented a comprehensive analysis of the different aspects of our proposed method for detecting spalling and severity level. The approach for detecting spalling and spalling severity level is based on Deep encoder-decoder networks. In recent years, several encoder-decoder based deep convolutional networks have been proposed; SegNet [3], UNet [4], PSPNet [5], FCN [37], DeepLab [38], DeepCrack [39]. 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 include ResNet-50 [40] [41], VGG-19 [42], Xception [43], and MobileNet [44].

We have discussed several concepts related to our proposed approach. First, we will go over the deep encoder-decoder architectures that we used for our proposed architecture. 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.

3.2.1 Deep Encoder-Decoder Architecture

SegNet: The SegNet is an encoder-decoder network based architecture [3]. SegNet architecture based image segmentation process has been used to extract abnormal skin lesion from dermoscopy image [82], for gland segmentation from colon cancer histology images [83], to detect dark spots in oil spill areas [84], for automated brain tumor segmentation on multi-modal MR image [85], to detect pixel level crack detection [86], and for the inspection and evaluation of bridge decks [87].

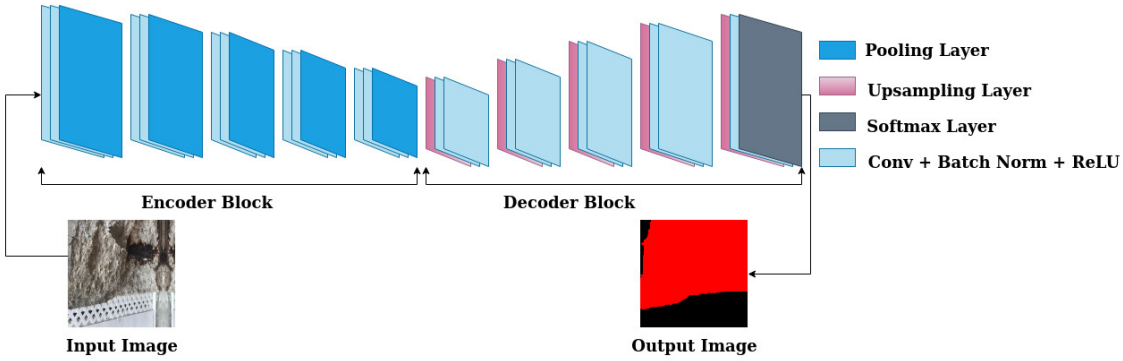

Figure 3.2: Overview of SegNet Architecture [3].

This architecture was proposed for pixel-wise semantic segmentation. The architecture for SegNet with encoder-decoder block is shown in Figure 3.2. 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 [76]. The max pooling is performed with a (2×2) window. The SegNet architecture avoid 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 [3]. The decoder conducts the up-sampling and call 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.

UNet: Several works used the UNet architecture for image segmentation; brain tumor image segmentation using UNet [88] and UNet-VGG16 [89], COVID-19 lung CT image segmentation [90], crack detection model [91], and dental panoramic image segmentation [92].

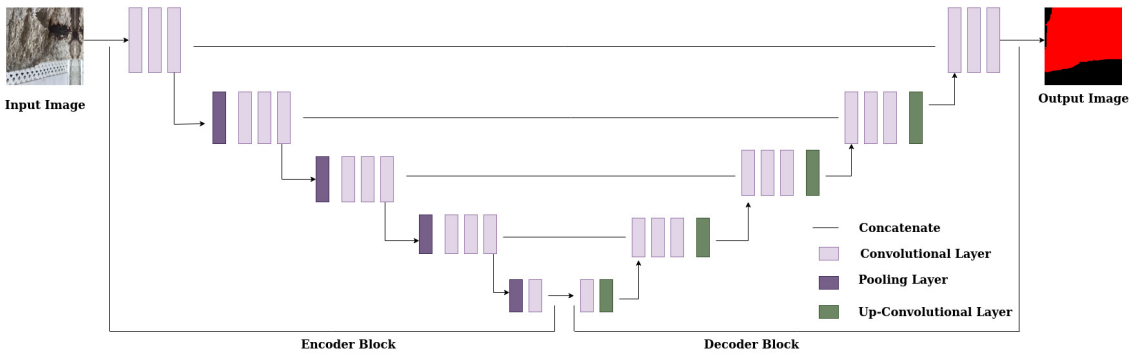

Figure 3.3: Overview of UNet Architecture [4] [1].

UNet is an encoder-decoder based architecture consists of encoder and decoder blocks. Figure 3.3 shows the overview of UNet architecture. The encoder block contains two 3 × 3 convolutions [4]. 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 the 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 Figure 3.4. PSPNet is one of the most well-recognized image segmentation models. PSPNet based semantic segmentation process used in image semantic segmentation [93], pavement distress detection [94] and crack detection [95] [96], arms and hands segmentation for egocentric perspective using image segmentation [97], and image segmentation for coronary angiography [98].

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 [99] 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.

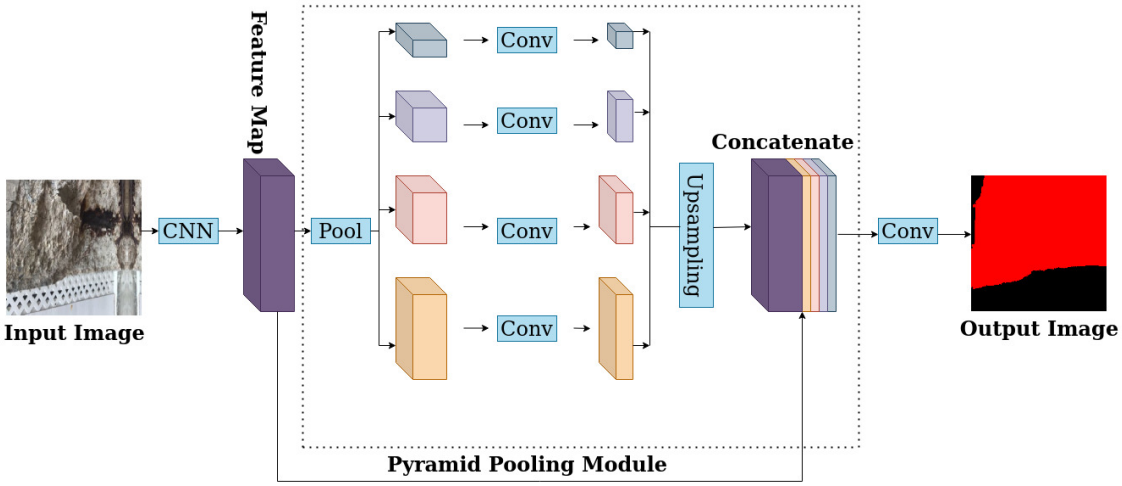

Figure 3.4: Overview of PSPNet Architecture [5].

In comparison to standard convolution, dilated convolution has a broader receptive field. The size of the used context information 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 it's layers once the encoder has extracted the image's features. The decoder is another network that processes inputted characteristics to provide predictions.

3.2.2 Backbone Network

Several CNN's backbone networks have made significant advancements with 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 [45]. We have used ResNet-50, VGG-19, MobileNet, and Xception as feature extractor in our deep architectures.

ResNet-50, VGG-19, and Xception are CNN with 50 layers, 19 layers and 71 layers deep, respectively [41] [42] [43]. MobileNet is one kind of CNN designed for mobile and embedded vision applications [44].

3.2.3 Preparation of Dataset

Large volumes of data are required to train, validate, and test the models in deep network architectures [33]. 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. We also collected the image 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 no-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 no-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 [34]. We have annotated each images according to the spalling area; Deep spalling area, shallow spalling area, and no-spalling area. The example of the annotation process of an image is shown in Figure 3.5.

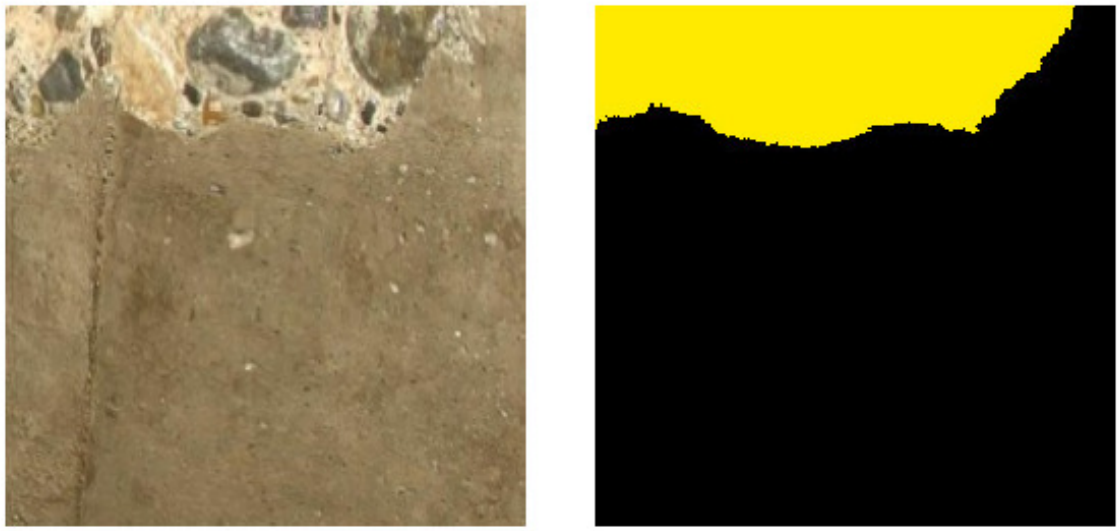

Figure 3.5: Data Annotation (a) Image of shallow spalling (b) Annotation of shallow spalling

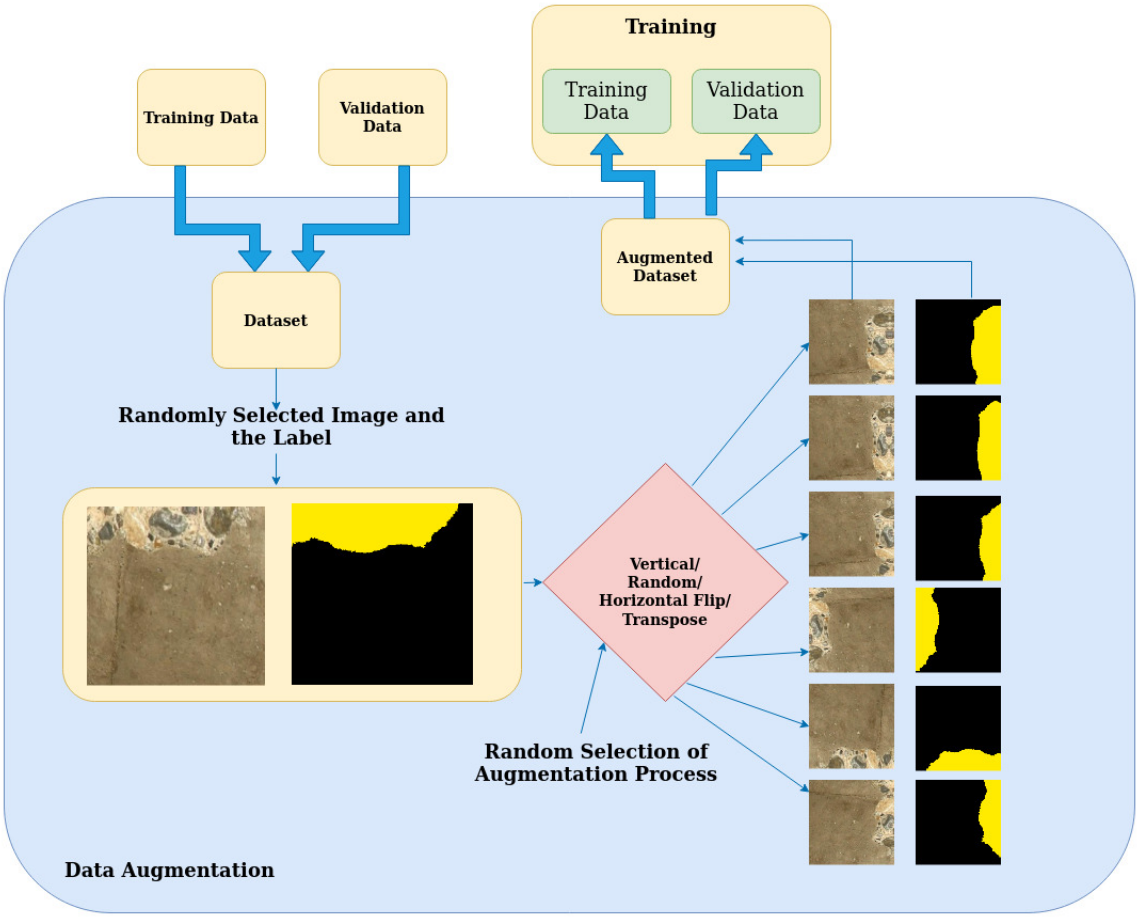

Figure 3.6: Data Augmentation Process.

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 are 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 Figure 3.6.

3.2.4 Proposed Architecture

We have proposed the method using three different types of deep architectures with different backbone networks for detecting the spalling and level of severity. The Encoder-Decoders model used in deep learning-based image segmentation technology is trained from start to end [80]. 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 SegNet (Figure 3.2), UNet (Figure 3.3), and PSPNet (Figure 3.4).

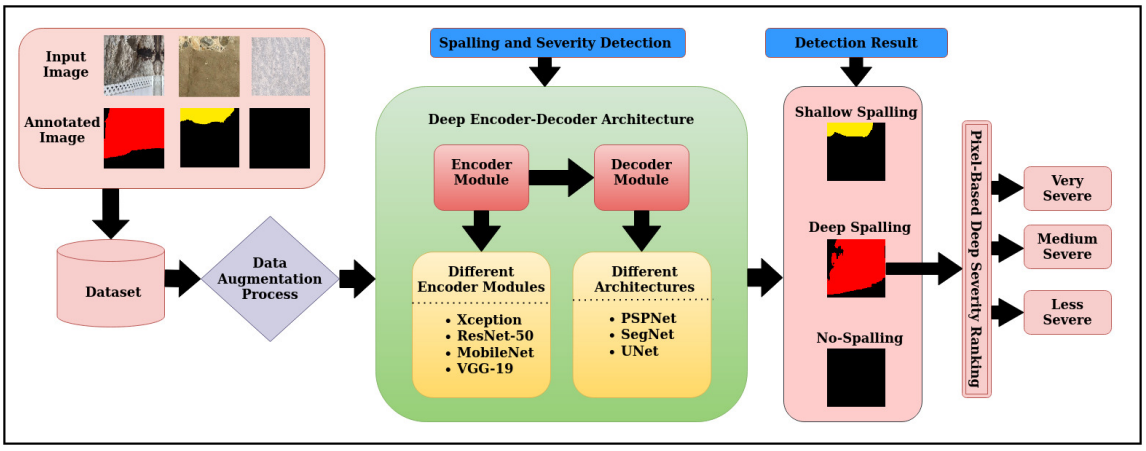

Figure 3.7: The Proposed Network Architecture Overview.

For the encoder part, we have employed ResNet-50, VGG-19, Xception, and Mo-

bileNet. 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 encoder to form integrated feature points after the feature 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 different category at various stages into the pixel space learned by the encoder to fully segment the target region [80]. Moreover, using the concatenation process the decoder connects to the corresponding encoder and help to reduce the loss happened during down-sampling process. Therefore, due to the advantage and performance of deep learning-based image segmentation architecture in several fields [77] [78] [79], 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; no-spalling, deep spalling, or shallow spalling. Figure 3.7 shows the overview of proposed methodology to detect spalling severity.

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 about 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. Afterwords, 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 3.1, where D_pixel provides the number of total pixel of deep spalling area, and T_pixel counts the total number of pixel 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.

$$Ratio = \frac{D\_pixel}{T\_pixel} \tag{3.1}$$

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 comparison analysis of severity ranking for three selected image categories in the Result and Discussion section.

3.3 Result and Discussion

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

3.3.1 Experimental Setup

Our 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 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.

Our proposed work trained and tested on a system with a GTX 1080 GPU. The size of the image was (256×256). We have used GIMP(GNU Image Manipulation Program) to annotate our images in pixel-by-pixel map. GIMP is one of the most popular illustration and image editing programs available [100]. 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 category of images: only deep spalling, deep spalling with no-spalling area, only shallow spalling, shallow spalling with no-spalling area, and no-spalling. The spalling area are 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 no-spalling. Moreover, the spalling without exposed steel bars was considered shallow spalling for our proposed approach.

We have used an augmentation process (described in Figure 3.6) during the training and validation phases. The use of the augmentation process during the training and validation phases has the advantage of avoiding overfitting problem [34]. The dataset contains 10000 images for training and another 2000 images for validation. For the multi-class classification problem, we used categorical cross-entropy. The Adam optimizer was used to optimize the architecture with a learning rate of 0.001. We have calculated training and validation loss like we mentioned in our first proposed methodology. However, we have shown the graphs for training and validation

loss only for our first methodology in Chapter 2. We conducted the training for 100 epochs. 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. We present the performance-based statistical analysis for both of the datasets, with 100 (Table 3.3) and 300 (Table 3.2) images in the quantitative analysis section. However, we have described the performance analysis in the quantitative analysis section, only for the dataset with 300 images, since the difference between the two datasets is negligible. The non-statistical comparison in the qualitative analysis section is shown only for the dataset with 300 images (Figure 3.8, 3.9, 3.10), since the difference in performance for the two datasets is approximately 0.01% for all metrics. The proposed approach trained for 100 epochs. Therefore, on each epoch, it was trained on a different dataset because of the augmentation process.

Because our proposed work detects spalling and the severity level of spalling, we have presented non-statistical qualitative analysis as well as quantitative analysis with statistical measurements. In the quantitative analysis subsection, we have 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.

3.3.2 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 3.2. We have used Dice loss, mIoU, Precision, Recall, and Accuracy metrics for the perfor-

mance 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 3.2, 3.3, and 3.4 to find out the Accuracy, Precision, and Recall respectively. From Equation 3.5, we get the calculation for IoU for each class which helps to calculate the mean value of IoU for all classes. Table 3.1 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.

Table 3.1: Quantitative measures used for evaluating the performance of deep network architectures. Spalling pixels belong to the positive class and no-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 no-spalling pixels
FN	False Negative	the number of pixels detected as non-spalling erroneously

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.3}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.4}$$

$$IoU = \frac{TP}{TP + FN + FP} \tag{3.5}$$

The statistical performance for PSPNet is shown in Table 3.2; the result for PSP-Net 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.

In Table 3.2 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 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 3.2: Performance comparison among Deep architectures with different backbone networks (Dataset with 300 images).

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

The comparative analysis of statistical performance of SegNet architecture with default encoder-decoder and Xception, ResNet-50, MobileNet, and VGG-19 are shown in Table 3.2. 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 architec-

Table 3.3: Performance comparison among Deep architectures with different encoder-decoder networks (Dataset with 100 images).

Base Model	**Encoder**	**Dice Loss(%)**	**mIoU(%)**	**Precision(%)**	**Recall(%)**	**Accuracy(%)**
SegNet	-	9.59	82.49	93.99	95.36	92.39
"	**Xception**	**6.80**	**87.97**	**90.86**	**98.43**	**95.19**
"	ResNet-50	6.97	86.97	93.78	96.39	94.56
"	MobileNet	8.49	84.35	95.36	95.79	94.23
"	VGG-19	16.30	71.96	93.83	92.34	92.0
UNet	-	11.79	78.90	92.73	91.41	89.65
"	**Xception**	**7.28**	**86.43**	**92.75**	**97.23**	**93.79**
"	ResNet-50	11.58	79.23	90.68	92.89	92.03
"	MobileNet	7.36	86.29	93.91	96.13	93.52
"	VGG-19	18.16	69.26	89.47	86.53	88.57
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	7.45	86.13	91.19	93.45	93.58
"	VGG-19	7.82	85.49	94.44	92.95	92.97

ture, 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 3.2 infer that PSPNet gives comparatively good performance for detecting spalling and severity level 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 level 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 3.4: 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

Table 3.4 shows the results for the severity ranking of deep spalling for three image categories. We have defined the ranking as very severe, medium severe, and

less severe.

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 3.4 displays the ratio, which is expressed in terms of 100%. 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. We have selected three different categories of images from the 100 deep spalling images. In Table 3.4, 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 pixel of the image, which is same for all the images, **Spalling Size** refers the number of pixels affected by deep spalling, **Ratio** is calculated using Equation 3.1 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.

3.3.3 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 Figure 3.8, Figure 3.9, and Figure 3.10. The results are highlighting overall performance spalling and severity detection based on deep spalling, shallow spalling, and no-spalling images.

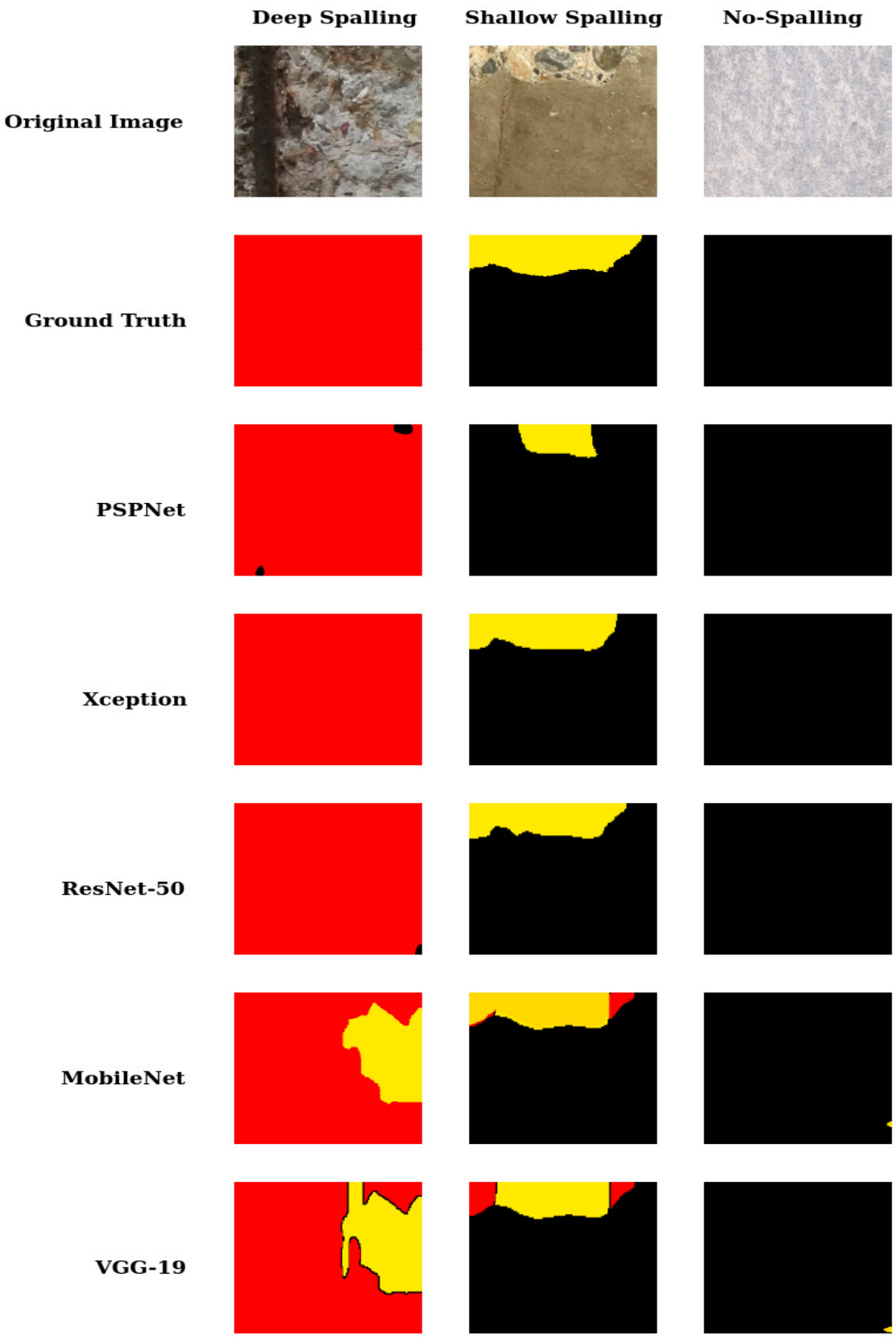

Figure 3.8: Results shown for PSPNet framework with different encoder-decoder networks.

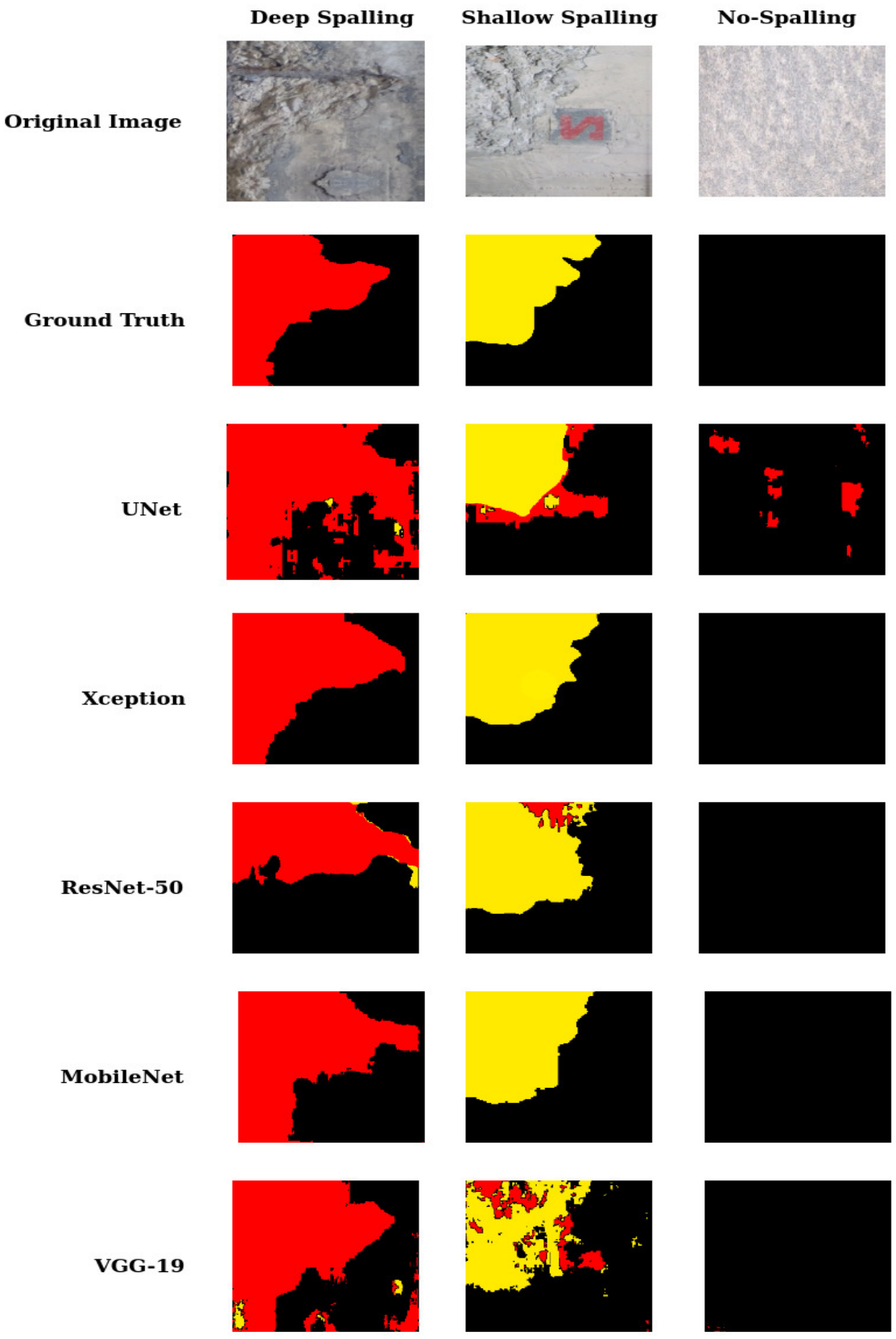

Figure 3.9: Results shown for UNet framework with different encoder-decoder networks.

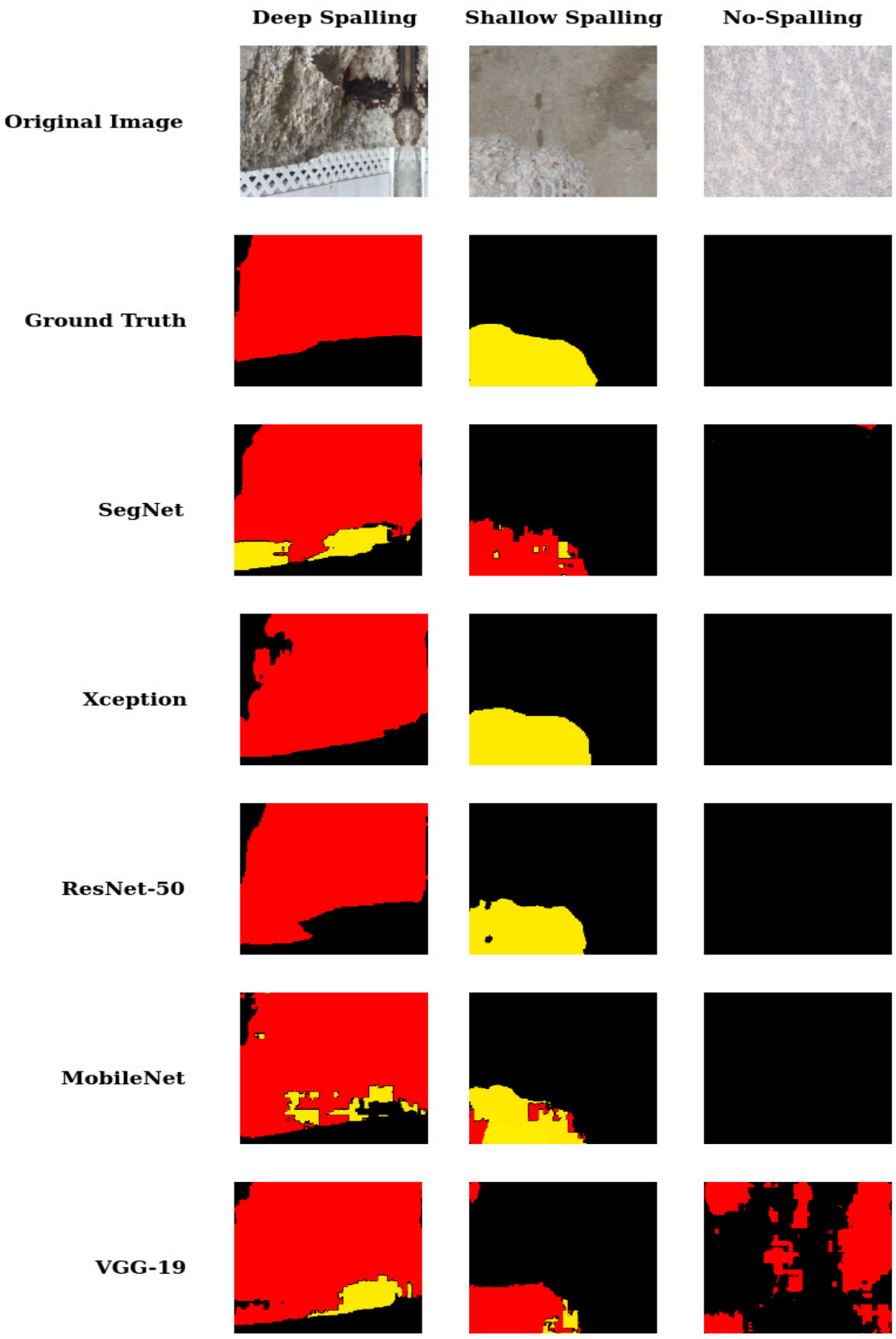

Figure 3.10: Results shown for SegNet framework with different encoder-decoder networks.

In Figure 3.8, 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 no-spalling area, only shallow spalling, shallow spalling with no-spalling area, and no-spalling. For the PSPNet framework, only deep spalling, shallow spalling with no-spalling, and no-spalling areas were chosen as input to present the performance evaluation.

In Figure 3.8, we have original image, ground truth which is pixel-by-pixel 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 Figure 3.8 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 no-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, no-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 Figure 3.9. For the UNet framework, we have chosen here deep spalling with no-spalling area, shallow spalling with no-spalling area, and no-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 Figure 3.9, respectively. Figure 3.9 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 no-spalling images.

Figure 3.9 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 Figure 3.10. The categorization of images for performance evaluation of the SegNet framework was chosen here as deep spalling with no-spalling area, shallow spalling with no-spalling, and no-spalling. In Figure 3.10, we have original image of deep spalling, shallow spalling and no-spalling area, ground truth which is pixel-by-pixel mapping of 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 no-spalling and shallow spalling classes. The no-spalling is predicted pretty accurately by most of the architecture combinations, except for the SegNet architecture with VGG-19. In Figure 3.10, 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 Figure 3.8, 3.9, and 3.10, it can be concluded that most deep architectures with encoder-decoder networks provide comparatively good results for no-spalling areas.

The performance evaluation for predicting deep spalling and shallow spalling closely follows the performance evaluation for predicting no-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 no-spalling with SegNet, UNet, and PSPNet deep architectures. Comparatively, VGG-19 shows poor performance for detecting spalling and severity level with UNet and SegNet architectures. For PSPNet, the VGG-19 encoder closely follows the performance of PSPNet with the default encoder-decoder network.

Chapter 4

Conclusion and Future Works

In this thesis, we have discussed a concrete distress type, spalling. This distress in concrete occurs due to exposure to extreme environmental hazards and a lack of preventive measures, which creates a surface abnormality and destroys the surface look. Aside from its appearance, this surface abnormality can lead to dangerous situations and make the civil infrastructure unusable. Therefore, proper inspection and maintenance in these areas are required to avoid any unwanted situations. Keeping this in mind, this thesis presents deep learning and encoder-decoder-based architectures for detecting spalling and its severity levels.

The proposed methodologies show that deep learning-based semantic segmentation technology can be employed to detect spalling and severity levels. According to the studies presented in the literature, there are very few methods for detecting the severity of spalling. Due to the good performance of deep learning based architectures with encoder-decoder networks in biology image segmentation, agricultural image segmentation, and medical image segmentation, we have employed this technology in civil infrastructure to detect spalling and severity levels. The exact position of spalling and the level of severity are crucial information for civil infrastructure. The proposed

works introduce multi-class semantic segmentation-based approaches incorporating different encoder-decoder networks to detect the severity level of spalling. The after-inspection process can be different depending on the condition of the spalling. We need precise information to detect the area of the spalling. The spalling can be large or deep enough that it needs to be fixed immediately, or shallow or small spalling can get priority after the large or deep spalling. Therefore, a correct prediction of the priority of fixing the spalling is required for a cost-effective solution.

We have proposed two deep architecture-based methods to detect spalling and its severity level. In the first method, we have detected spalling and severity levels according to spalling size. The categorization for this spalling severity level is large spalling, shallow spalling, and no-spalling. Two deep convolutional architectures, SegNet and UNet, have been used for the proposed approach with three different encoder-decoder network combinations. From the quantitative and qualitative analyses, it is shown that the UNet architecture shows comparatively better results. We have incorporated three different deep architectures in the second proposed methodology: SegNet, UNet, and PSPNet. This proposed approach detects spalling and determines the severity level based on the depth of the spalling. The spalling severity is categorized as deep spalling, shallow spalling, and no-spalling. We have used four different backbone networks with three different deep architectures to get the best performance and the best combination of deep architecture with the backbone network. For the second approach, we get the best results from PSPNet. The proposed methodologies' performance has been demonstrated using statistical and non-statistical analysis. Both of the proposed methodologies achieved high precision and recall rates. In the second proposed method, we have added further analysis for the deep spalling areas. Along with the depth, the ratio of the affected deep spall areas helps provide more insight about 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 with respect to the overall area. From the comparative analysis, we infer that the results of the severity ranking follow the ground truth ranking very closely.

4.1 Published Research Article

Our proposed approach titled "Deep Architecture Based Spalling Severity Detection System Using Encoder-Decoder Networks" described in Chapter 2 has been accepted in the 17th International Symposium on Visual Computing (ISVC'22).

4.2 Future Work

The knowledge gained from earlier approaches served as the foundation for these proposed works, so new ideas should also be drawn from it. We came up with the proposal included in this thesis out of the need for an autonomous system to detect the exact location and condition of the spalling. Our proposed approaches provide good results for critical positions in civil infrastructure. As a result, the proposed work can be used in autonomous systems such as climbing robots and drones for the automatic detection of spalling in critical places of civil infrastructure.

In the future, we want to further improve our proposed deep architectures to reduce power consumption and memory requirements while achieving better performance in detecting spalling and severity levels. Another potential future direction for this paper is to explore various concrete distresses using a deep architecture-based combined detection process. Cracking and spalling are major concrete defects in civil

infrastructure. These two types of distress are very different in nature. As the future direction, we want to propose a deep architecture-based multi-class segmentation approach to detect the two major defects of concrete.

Bibliography

[1] J. Li, W. Li, C. Jin, L. Yang, and H. He, "One view per city for buildings segmentation in remote-sensing images via fully convolutional networks: A proof-of-concept study," *Sensors*, vol. 20, no. 1, p. 141, 2019. [Online]. Available: https://doi.org/10.3390/s20010141

[2] B. Khagi and G.-R. Kwon, "Pixel-label-based segmentation of cross-sectional brain mri using simplified segnet architecture-based cnn," *Journal of healthcare engineering*, vol. 2018, 2018. [Online]. Available: https://doi.org/10.1155/2018/3640705

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017. [Online]. Available: https://doi.org/10.1109/tpami.2016.2644615

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241. [Online]. Available: https://doi.org/10.1007/978-3-319-24574-4_28

[5] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer*

vision and pattern recognition, 2017, pp. 2881–2890. [Online]. Available: https://doi.org/10.48550/arXiv.1612.01105

[6] H. Ahmed, H. M. La, and N. Gucunski, "Review of non-destructive civil infrastructure evaluation for bridges: State-of-the-art robotic platforms, sensors and algorithms," *Sensors*, vol. 20, no. 14, p. 3954, 2020. [Online]. Available: https://doi.org/10.3390/s20143954

[7] M. Zhou, W. Cheng, H. Huang, and J. Chen, "A novel approach to automated 3d spalling defects inspection in railway tunnel linings using laser intensity and depth information," *Sensors*, vol. 21, no. 17, p. 5725, 2021. [Online]. Available: https://doi.org/10.3390/s21175725

[8] H. M. La, R. S. Lim, B. B. Basily, N. Gucunski, J. Yi, A. Maher, F. A. Romero, and H. Parvardeh, "Mechatronic systems design for an autonomous robotic system for high-efficiency bridge deck inspection and evaluation," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 6, pp. 1655–1664, 2013. [Online]. Available: https://doi.org/10.1109/tmech.2013.2279751

[9] S. Gibb, H. M. La, T. Le, L. Nguyen, R. Schmid, and H. Pham, "Nondestructive evaluation sensor fusion with autonomous robotic system for civil infrastructure inspection," *Journal of Field Robotics*, vol. 35, no. 6, pp. 988–1004, 2018. [Online]. Available: https://doi.org/10.1002/rob.21791

[10] S. T. Nguyen and H. M. La, "A climbing robot for steel bridge inspection," *Journal of Intelligent & Robotic Systems*, vol. 102, no. 4, pp. 1–21, 2021. [Online]. Available: https://doi.org/10.1007/s10846-020-01266-1

[11] H. M. La, T. H. Dinh, N. H. Pham, Q. P. Ha, and A. Q. Pham, "Automated robotic monitoring and inspection of steel structures and

bridges," *Robotica*, vol. 37, no. 5, pp. 947–967, 2019. [Online]. Available: https://doi.org/10.1017/S0263574717000601

[12] H. M. La, N. Gucunski, S.-H. Kee, and L. V. Nguyen, "Data analysis and visualization for the bridge deck inspection and evaluation robotic system," *Visualization in Engineering*, vol. 3, no. 1, pp. 1–16, 2015. [Online]. Available: https://doi.org/10.1186/s40327-015-0017-3

[13] H. M. La, N. Gucunski, K. Dana, and S.-H. Kee, "Development of an autonomous bridge deck inspection robotic system," *Journal of Field Robotics*, vol. 34, no. 8, pp. 1489–1504, 2017. [Online]. Available: https://doi.org/10.1002/rob.21725

[14] T. Le, S. Gibb, N. Pham, H. M. La, L. Falk, and T. Berendsen, "Autonomous robotic system using non-destructive evaluation methods for bridge deck inspection," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 3672–3677. [Online]. Available: https://doi.org/10.1109/icra.2017.7989421

[15] S. T. Nguyen, A. Q. Pham, C. Motley, and H. M. La, "A practical climbing robot for steel bridge inspection," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9322–9328. [Online]. Available: https://doi.org/10.1109/icra40945.2020.9196892

[16] N. Gucunski, S. Kee, H. La, B. Basily, and A. Maher, "Delamination and concrete quality assessment of concrete bridge decks using a fully autonomous rabit platform," *Structural Monitoring and Maintenance*, vol. 2, no. 1, pp. 19–34, 2015. [Online]. Available: http://dx.doi.org/10.12989/smm.2015.2.1.019

[17] Y. Otsuki, S. T. Nguyen, H. M. La, and Y. Wang, "Autonomous ultrasonic thickness measurement using a steel climbing mobile robot integrated with martlet wireless sensing," 2022. [Online]. Available: https://doi.org/10.32548/rs.2022.029

[18] S. T. Nguyen, H. Nguyen, S. T. Bui, V. A. Ho, and H. M. La, "Multi-directional bicycle robot for steel structure inspection," *arXiv preprint arXiv:2103.11522*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2103.11522

[19] C. Motley, S. T. Nguyen, and H. M. La, "Design of a high strength multi-steering climbing robot for steel bridge inspection," in *2022 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2022, pp. 323–328. [Online]. Available: https://doi.org/10.1109/sii52469.2022.9708750

[20] A. Q. Pham, C. Motley, S. T. Nguyen, and H. M. La, "A robust and reliable climbing robot for steel structure inspection," in *2022 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2022, pp. 336–343. [Online]. Available: https://doi.org/10.1109/sii52469.2022.9708747

[21] A. Q. Pham, A. T. La, E. Chang, and H. M. La, "Flying-climbing mobile robot for steel bridge inspection," in *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 2021, pp. 230–235. [Online]. Available: https://doi.org/10.1109/ssrr53300.2021.9597676

[22] N. Harris, S. Liu, S. J. Louis, and J. H. La, "A genetic algorithm for multi-robot routing in automated bridge inspection," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2019, pp. 369–370. [Online]. Available: https://doi.org/10.1145/3319619.3321917

[23] A. Q. Pham, H. M. La, K. T. La, and M. T. Nguyen, "A magnetic wheeled robot for steel bridge inspection," in *International Conference on Engineering Research and Applications*. Springer, 2019, pp. 11–17. [Online]. Available: https://doi.org/10.1007/978-3-030-37497-6_2

[24] N. H. Pham and H. M. La, "Design and implementation of an autonomous robot for steel bridge inspection," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2016, pp. 556–562. [Online]. Available: https://doi.org/10.1109/allerton.2016.7852280

[25] N. H. Pham, H. M. La, Q. P. Ha, S. N. Dang, A. H. Vo, and Q. H. Dinh, "Visual and 3d mapping for steel bridge inspection using a climbing robot," in *ISARC 2016-33rd International Symposium on Automation and Robotics in Construction*, 2016. [Online]. Available: https://doi.org/10.22260/isarc2016/0018

[26] H.-D. Bui, S. Nguyen, U. H. Billah, C. Le, A. Tavakkoli, and H. M. La, "Control framework for a hybrid-steel bridge inspection robot," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2585–2591. [Online]. Available: https://doi.org/10.1109/iros45743.2020.9340637

[27] S. T. Nguyen and H. M. La, "Development of a steel bridge climbing robot," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1912–1917. [Online]. Available: https://doi.org/10.1108/ir-09-2015-0186

[28] L. Van Nguyen, S. Gibb, H. X. Pham, and H. M. La, "A mobile robot for automated civil infrastructure inspection and evaluation," in *2018 IEEE International Symposium on Safety, Security, and Rescue*

Robotics (SSRR). IEEE, 2018, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ssrr.2018.8468642

[29] N. Gucunski, S.-H. Kee, H. La, B. Basily, A. Maher, and H. Ghasemi, "Implementation of a fully autonomous platform for assessment of concrete bridge decks rabit," in *Structures Congress 2015*, 2015, pp. 367–378. [Online]. Available: https://doi.org/10.1061/9780784479117.032

[30] R. S. Lim, H. M. La, and W. Sheng, "A robotic crack inspection and mapping system for bridge deck maintenance," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 2, pp. 367–378, 2014. [Online]. Available: https://doi.org/10.1109/tase.2013.2294687

[31] P. Prasanna, K. J. Dana, N. Gucunski, B. B. Basily, H. M. La, R. S. Lim, and H. Parvardeh, "Automated crack detection on concrete bridges," *IEEE Transactions on automation science and engineering*, vol. 13, no. 2, pp. 591–599, 2014. [Online]. Available: https://doi.org/10.1109/tase.2014.2354314

[32] T. H. Dinh, Q. P. Ha, and H. M. La, "Computer vision-based method for concrete crack detection," in *2016 14th international conference on control, automation, robotics and vision (ICARCV)*. IEEE, 2016, pp. 1–6. [Online]. Available: https://doi.org/10.1109/icarcv.2016.7838682

[33] U. H. Billah, A. Tavakkoli, and H. M. La, "Concrete crack pixel classification using an encoder decoder based deep learning architecture," in *International Symposium on Visual Computing*. Springer, 2019, pp. 593–604. [Online]. Available: https://doi.org/10.1007/978-3-030-33720-9_46

[34] U. H. Billah, H. M. La, and A. Tavakkoli, "Deep learning-based feature silencing for accurate concrete crack detection," *Sensors*, vol. 20, no. 16, p. 4403, 2020. [Online]. Available: https://doi.org/10.3390/s20164403

[35] N.-D. Hoang, Q.-L. Nguyen, and X.-L. Tran, "Automatic detection of concrete spalling using piecewise linear stochastic gradient descent logistic regression and image texture analysis," *Complexity*, vol. 2019, 2019. [Online]. Available: https://doi.org/10.1155/2019/5910625

[36] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2021. [Online]. Available: https://doi.org/10.1109/tpami.2021.3059968

[37] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440. [Online]. Available: https://doi.org/10.1109/cvpr.2015.7298965

[38] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017. [Online]. Available: https://doi.org/10.1109/tpami.2017.2699184

[39] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "Deepcrack: Learning hierarchical convolutional features for crack detection," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2018. [Online]. Available: https://doi.org/10.1109/tip.2018.2878966

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. [Online]. Available: https://doi.org/10.1109/cvpr.2016.90

[41] B. Koonce, "Resnet 50," in *Convolutional neural networks with swift for tensorflow*. Springer, 2021, pp. 63–72. [Online]. Available: https://doi.org/10.1007/978-1-4842-6168-2_6

[42] W. Gómez-Flores and W. C. de Albuquerque Pereira, "A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound," *Computers in Biology and Medicine*, vol. 126, p. 104036, 2020. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2020.104036

[43] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258. [Online]. Available: https://doi.org/10.1109/cvpr.2017.195

[44] D. Sinha and M. El-Sharkawy, "Thin mobilenet: An enhanced mobilenet architecture," in *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*. IEEE, 2019, pp. 0280–0285. [Online]. Available: https://doi.org/10.1109/uemcon47517.2019.8993089

[45] Y. Xing, L. Zhong, and X. Zhong, "An encoder-decoder network based fcn architecture for semantic segmentation," *Wireless Communications and Mobile Computing*, vol. 2020, 2020. [Online]. Available: https://doi.org/10.1155/2020/8861886

[46] J. König, M. D. Jenkins, M. Mannion, P. Barrie, and G. Morison, "Optimized deep encoder-decoder methods for crack segmentation," *Digital Signal Processing*, vol. 108, p. 102907, 2021. [Online]. Available: https://doi.org/10.1016/j.dsp.2020.102907

[47] D. Theckedath and R. Sedamkar, "Detecting affect states using vgg16, resnet50 and se-resnet50 networks," *SN Computer Science*, vol. 1, no. 2, pp. 1–7, 2020. [Online]. Available: https://doi.org/10.1007/s42979-020-0114-9

[48] T. Carvalho, E. R. De Rezende, M. T. Alves, F. K. Balieiro, and R. B. Sovat, "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, pp. 866–870. [Online]. Available: https://doi.org/10.1109/icmla.2017.00-47

[49] C. Bi, J. Wang, Y. Duan, B. Fu, J.-R. Kang, and Y. Shi, "Mobilenet based apple leaf diseases identification," *Mobile Networks and Applications*, pp. 1–9, 2020. [Online]. Available: https://doi.org/10.1007/s11036-020-01640-1

[50] M. Chhabra and R. Kumar, "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, pp. 161–167. [Online]. Available: https://doi.org/10.1109/com-it-con54601.2022.9850612

[51] E. Ayan and H. M. Ünver, "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, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ebbt.2019.8741582

[52] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, "Transfer learning for image classification using vgg19: Caltech-101 image data set," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2021. [Online]. Available: https://doi.org/10.1007/s12652-021-03488-z

[53] E. Rezende, G. Ruppert, T. Carvalho, F. Ramos, and P. De Geus, "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, pp. 1011–1014. [Online]. Available: https://doi.org/10.1109/icmla.2017.00-19

[54] W. W. Lo, X. Yang, and Y. Wang, "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, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ntms.2019.8763852

[55] B. Li and D. Lima, "Facial expression recognition via resnet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 57–64, 2021. [Online]. Available: https://doi.org/10.1016/j.ijcce.2021.02.002

[56] H. Ahmed, H. M. La, and K. Tran, "Rebar detection and localization for bridge deck inspection and evaluation using deep residual networks," *Automation in Construction*, vol. 120, p. 103393, 2020. [Online]. Available: https://doi.org/10.1016/j.autcon.2020.103393

[57] T. Dawood, Z. Zhu, and T. Zayed, "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, pp. 607–615. [Online]. Available: https://doi.org/10.1007/978-981-10-6190-5_55

[58] ——, "Machine vision-based model for spalling detection and quantification in subway networks," *Automation in Construction*, vol. 81, pp. 149–160, 2017. [Online]. Available: https://doi.org/10.1016/j.autcon.2017.06.008

[59] D. Pham, M. Ha, and C. Xiao, "A novel visual inspection system for rail surface spalling detection," in *IOP Conference Series: Materials Science and Engineering*, vol. 1048, no. 1. IOP Publishing, 2021, p. 012015. [Online]. Available: https://dx.doi.org/10.1088/1757-899X/1048/1/012015

[60] H. Wu, X. Ao, Z. Chen, C. Liu, Z. Xu, and P. Yu, "Concrete spalling detection for metro tunnel from point cloud based on roughness descriptor," *Journal of Sensors*, vol. 2019, 2019. [Online]. Available: https://doi.org/10.1155/2019/8574750

[61] Z. Hu, H. Zhu, M. Hu, and Y. Ma, "Rail surface spalling detection based on visual saliency," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 13, no. 3, pp. 505–509, 2018. [Online]. Available: https://doi.org/10.1002/tee.22594

[62] M. Bai and H. Sezen, "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. [Online]. Available: https://doi.org/10.5194/isprs-annals-V-2-2021-161-2021

[63] T. Ghosh Mondal, M. R. Jahanshahi, R.-T. Wu, and Z. Y. Wu, "Deep learning-based multi-class damage detection for autonomous post-disaster

reconnaissance," *Structural Control and Health Monitoring*, vol. 27, no. 4, p. e2507, 2020. [Online]. Available: https://doi.org/10.1002/stc.2507

[64] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015. [Online]. Available: https://doi.org/10.1109/tpami.2016.2577031

[65] M.-K. Kim, H. Sohn, and C.-C. Chang, "Localization and quantification of concrete spalling defects using terrestrial laser scanning," *Journal of Computing in Civil Engineering*, vol. 29, no. 6, p. 04014086, 2015. [Online]. Available: https://doi.org/10.1061/(ASCE)CP.1943-5487.0000415

[66] A. Mohd Ali, J. Sanjayan, and M. Guerrieri, "Specimens size, aggregate size, and aggregate type effect on spalling of concrete in fire," *Fire and Materials*, vol. 42, no. 1, pp. 59–68, 2018. [Online]. Available: https://doi.org/10.1002/fam.2457

[67] H. Tanaka, S. Tottori, and T. Nihei, "Detection of concrete spalling using active infrared thermography," *Quarterly Report of RTRI*, vol. 47, no. 3, pp. 138–144, 2006. [Online]. Available: https://doi.org/10.2219/rtriqr.47.138

[68] H. Zhang, Y. Zou, E. del Rey Castillo, and X. Yang, "Detection of rc spalling damage and quantification of its key properties from 3d point cloud," *KSCE Journal of Civil Engineering*, vol. 26, no. 5, pp. 2023–2035, 2022. [Online]. Available: https://doi.org/10.1007/s12205-022-0890-y

[69] E. M. Abdelkader, O. Moselhi, M. Marzouk, and T. Zayed, "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, pp. 1136–1143. [Online]. Available: https://doi.org/10.22260/ISARC2020/0156

[70] V. K. Reja, K. Varghese, and Q. P. Ha, "Computer vision-based construction progress monitoring," *Automation in Construction*, vol. 138, p. 104245, 2022. [Online]. Available: https://doi.org/10.1016/j.autcon.2022.104245

[71] N.-D. Hoang, T.-C. Huynh, and V.-D. Tran, "Concrete spalling severity classification using image texture analysis and a novel jellyfish search optimized machine learning approach," *Advances in Civil Engineering*, vol. 2021, 2021. [Online]. Available: https://doi.org/10.1155/2021/5551555

[72] E. Mohammed Abdelkader, O. Moselhi, M. Marzouk, and T. Zayed, "Entropy-based automated method for detection and assessment of spalling severities in reinforced concrete bridges," *Journal of Performance of Constructed Facilities*, vol. 35, no. 1, p. 04020132, 2021. [Online]. Available: https://doi.org/10.1061/(ASCE)CF.1943-5509.0001544

[73] H. Nguyen and N.-D. Hoang, "Computer vision-based classification of concrete spall severity using metaheuristic-optimized extreme gradient boosting machine and deep convolutional neural network," *Automation in Construction*, vol. 140, p. 104371, 2022. [Online]. Available: https://doi.org/10.1016/j.autcon.2022.104371

[74] H. AHMED and H. M. LA, "Steel defect detection in bridges using deep encoder-decoder networks," *STRUCTURAL HEALTH MONITORING 2021*, 2021. [Online]. Available: https://doi.org/10.12783/shm2021/36335

[75] D. Joshi, T. P. Singh, and G. Sharma, "Automatic surface crack detection using segmentation-based deep-learning approach," *Engineering*

Fracture Mechanics, vol. 268, p. 108467, 2022. [Online]. Available: https://doi.org/10.1016/j.engfracmech.2022.108467

[76] V. Badrinarayanan, A. Handa, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *arXiv preprint arXiv:1505.07293*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1505.07293

[77] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "Dense-inception u-net for medical image segmentation," *Computer methods and programs in biomedicine*, vol. 192, p. 105395, 2020. [Online]. Available: https://doi.org/10.1016/j.cmpb.2020.105395

[78] H. Peng, C. Xue, Y. Shao, K. Chen, J. Xiong, Z. Xie, and L. Zhang, "Semantic segmentation of litchi branches using deeplabv3+ model," *IEEE Access*, vol. 8, pp. 164 546–164 555, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.3021739

[79] S. Liu, M. Li, M. Li, and Q. Xu, "Research of animals image semantic segmentation based on deep learning," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, p. e4892, 2020. [Online]. Available: https://doi.org/10.1002/cpe.4892

[80] Y. Liu, Z. Zhang, X. Liu, L. Wang, and X. Xia, "Efficient image segmentation based on deep learning for mineral image classification," *Advanced Powder Technology*, vol. 32, no. 10, pp. 3885–3903, 2021. [Online]. Available: https://doi.org/10.1016/j.apt.2021.08.038

[81] H. Ahmed, H. M. La, and G. Pekcan, "Rebar detection and localization for non-destructive infrastructure evaluation of bridges using deep residual networks,"

in *International Symposium on Visual Computing*. Springer, 2019, pp. 631–643. [Online]. Available: https://doi.org/10.1007/978-3-030-33720-9_49

[82] S. Kadry, D. Taniar, R. Damaševičius, V. Rajinikanth, and I. A. Lawal, "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, pp. 1–5. [Online]. Available: https://doi.org/10.1109/icbsii51839.2021.9445180

[83] J. Tang, J. Li, and X. Xu, "Segnet-based gland segmentation from colon cancer histology images," in *2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2018, pp. 1078–1082. [Online]. Available: https://doi.org/10.1109/yac.2018.8406531

[84] H. Guo, G. Wei, and J. An, "Dark spot detection in sar images of oil spill using segnet," *Applied Sciences*, vol. 8, no. 12, p. 2670, 2018. [Online]. Available: https://doi.org/10.3390/app8122670

[85] S. Alqazzaz, X. Sun, X. Yang, and L. Nokes, "Automated brain tumor segmentation on multi-modal mr image using segnet," *Computational Visual Media*, vol. 5, no. 2, pp. 209–219, 2019. [Online]. Available: https://doi.org/10.1007/s41095-019-0139-y

[86] C. Song, L. Wu, Z. Chen, H. Zhou, P. Lin, S. Cheng, and Z. Wu, "Pixel-level crack detection in images using segnet," in *International Conference on Multi-disciplinary Trends in Artificial Intelligence*. Springer, 2019, pp. 247–254. [Online]. Available: https://doi.org/10.1007/978-3-030-33709-4_22

[87] H. AHMED, H. M. LA, and A. TAVAKOLLI, "Use of deep encoder-decoder network for sub-surface inspection and evaluation of bridge decks,"

STRUCTURAL HEALTH MONITORING 2021, 2021. [Online]. Available: https://doi.org/10.12783/shm2021/36334

[88] M. Aghalari, A. Aghagolzadeh, and M. Ezoji, "Brain tumor image segmentation via asymmetric/symmetric unet based on two-pathway-residual blocks," *Biomedical Signal Processing and Control*, vol. 69, p. 102841, 2021. [Online]. Available: https://doi.org/10.1016/j.bspc.2021.102841

[89] A. A. Pravitasari, N. Iriawan, M. Almuhayar, T. Azmi, I. Irhamah, K. Fithriasari, S. W. Purnami, and W. Ferriastuti, "Unet-vgg16 with transfer learning for mri-based brain tumor segmentation," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1310–1318, 2020. [Online]. Available: http://doi.org/10.12928/telkomnika.v18i3.14753

[90] A. Saood and I. Hatem, "Covid-19 lung ct image segmentation using deep learning methods: Unet vs. segnet," 2020. [Online]. Available: https://doi.org/10.21203/rs.3.rs-56882/v3

[91] F. Liu and L. Wang, "Unet-based model for crack detection integrating visual explanations," *Construction and Building Materials*, vol. 322, p. 126265, 2022. [Online]. Available: https://doi.org/10.1016/j.conbuildmat.2021.126265

[92] S. Sivagami, P. Chitra, G. S. R. Kailash, and S. Muralidharan, "Unet architecture based dental panoramic image segmentation," in *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*. IEEE, 2020, pp. 187–191. [Online]. Available: https://doi.org/10.1109/wispnet48689.2020.9198370

[93] C. Yang and H. Guo, "A method of image semantic segmentation based on pspnet," *Mathematical Problems in Engineering*, vol. 2022, 2022. [Online]. Available: https://doi.org/10.1155/2022/8958154

[94] J. Zhong, J. Zhu, J. Huyan, T. Ma, and W. Zhang, "Multi-scale feature fusion network for pixel-level pavement distress detection," *Automation in Construction*, vol. 141, p. 104436, 2022. [Online]. Available: https://doi.org/10.1016/j.autcon.2022.104436

[95] J. Shu, J. Li, J. Zhang, W. Zhao, Y. Duan, and Z. Zhang, "An active learning method with difficulty learning mechanism for crack detection," *Smart Structures and Systems*, vol. 29, no. 1, pp. 195–206, 2022. [Online]. Available: https://doi.org/10.12989/sss.2022.29.1.195

[96] J.-J. Wang, Y.-F. Liu, X. Nie, and Y. Mo, "Deep convolutional neural networks for semantic segmentation of cracks," *Structural Control and Health Monitoring*, vol. 29, no. 1, p. e2850, 2022. [Online]. Available: https://doi.org/10.1002/stc.2850

[97] H. Sarah, E. Clua, and C. N. Vasconcelos, "Arms and hands segmentation for egocentric perspective based on pspnet and deeplab," in *International Conference on Human-Computer Interaction*. Springer, 2020, pp. 152–170. [Online]. Available: https://doi.org/10.1007/978-3-030-49695-1_11

[98] X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, "Coronary angiography image segmentation based on pspnet," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105897, 2021. [Online]. Available: https://doi.org/10.1016/j.cmpb.2020.105897

[99] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 472–480. [Online]. Available: https://doi.org/10.1109/cvpr.2017.75

[100] A. C. Sparavigna, "A method for the segmentation of images based on thresholding and applied to vesicular textures," *arXiv preprint arXiv:1612.01131*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1612.01131