Deep Architecture Based Spalling Severity Detection System Using Encoder-Decoder Networks

Tamanna Yasmin, Chuong Le, and Hung Manh La(✉)

University of Nevada, Reno, NV 89557, USA
hla@unr.edu
https://ara.cse.unr.edu/

Abstract. Proper maintenance of concrete structures is a significant issue to avoid any hazardous situation in civil infrastructure. Spalling is a significant surface concrete distress in bridges and buildings. Correctly detecting the severity level of spalling can make it happen to detect and maintain the harmful spalling promptly to avoid any accidents [10]. While previous works have been on surface defects, like cracks and spallings, few have addressed spalling severity detection. In this paper, we have proposed a deep learning-based approach to detect the exact location of spalling according to severity level by using pixel-by-pixel classification. Our network labels each pixel as no-spalling, small, or large spalling. To get the optimal proposed deep architecture, we tested several encoder-decoder networks to compare and analyze the performance of the detection processes.

Keywords: Spalling detection · Spalling severity · Deep architecture · Encoder-decoder

1 Introduction

Structural health inspection is vital to civil infrastructure, and concrete is essential to that [6]. 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 [24]. Therefore, proper inspection and timely maintenance should be done to avoid unwanted events. Another crucial part is detecting the severity of spalling and ensuring proper maintenance based on the detection result. Different types of concrete spalling are shown in Fig. 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 concretes, especially if spallings happen at a crucial point like under the breeze, or underwater beams [5]. We need an autonomous system with little or no human intervention that can solve the issues with traditional methods.

G. Bebis et al. (Eds.): ISVC 2022, LNCS 13599, pp. 332–343, 2022.
https://doi.org/10.1007/978-3-031-20716-7_26

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 [13], they need a filtration process to remove unwanted features. Sometimes, the filtration process can remove useful features and keep the one that is not necessary. Approaches combining machine learning and image processing are computationally expensive and require images' prepossessing. 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 [5]. The challenges that lie in detecting and classifying the severity level of spalling are extracting features and implementing appropriate methods into real-life applications. Moreover, one of the main challenges is managing large amounts of data to extract the feature from the environment of concretes. Our proposed work leads to overcoming these challenges.

(a) (b)

Fig. 1. Concrete spalling: (a) Large or very severe. (b) Small or less severe.

1.1 Related Works

Several spalling detection approaches have been developed in recent years. The proposed methods classify and detect spalling in metro tunnel, subway networks, railway surfaces [7,8,12,20,23,27]. However, there are very few approaches for detecting and classifying the severity of spalling in concrete. For metro tunnels, a concrete automatic spalling detection method has been proposed [8]. The proposed method can detect spalling damage on the tunnel surface using a 3D cloud point that contains information about the inner wall and outlier points. To detect and measure the quantity of the spalling, a machine learning and vision-based method for subways has been developed [23]. This work is a hybrid process of extracting important features about spalling by removing noise from the images and applying the process of detecting distress in subway surfaces. An optical detection algorithm based on visual salience has been proposed to detect spalling on the rail surface [12]. The detection algorithm removes the impact of unnecessary surroundings and increases the difference between the spalling and the neighborhood. Then, it detects the spalling by using a threshold value.

Methods for spalling localization, evaluation, and detection based on machine vision, laser scanning, Deep learning, infrared thermography have enticed the attraction of researchers [1,4,9,11,15,22,25]. Three variations of Mask R-CNNs have been proposed for detecting two major types of structural damage, namely, spalling and crack [4]. This work is designed to segment the damaged area in the brides and buildings and continuously check for the damages. A framework has been proposed to detect spalling and the damaged properties of the spalling area using point cloud data [25]. This point cloud data has been used in the framework to extract spalling features from Reinforced concrete and to detect spalling semi-automatically. Image texture and a piece-wise linear stochastic gradient descent logistic regression are used to detect automatically spalling in concrete [11]. The regression model is used for pattern detection, and image textures are used to extract features from images. The pattern detection approach can classify spalling and non-spalling areas depending on the image features. A novel idea using a hybrid machine learning approach and image texture analysis has been proposed to categorize the severity of concrete spalling [10]. The machine learning approach has been optimized using jellyfish search. Depending on this method, a shallow spall or deep spall can be identified from the images. The effect of specimen size, aggregate size, and aggregate type on spalling in concrete under hydrocarbon fire exposure are crucial [18]. The investigation on different specimen sizes, aggregate sizes, and types indicates that these properties impact concrete spalling under hydrocarbon fire exposure. To classify concrete spall severity, a solution based on computer vision has been proposed [19]. The authors categorize two spall classes: shallow and deep spall, using Extreme Gradient Boosting Machine and Deep Convolutional Neural Network. Detection and severity of the spalling are crucial points for reinforced concrete bridges. An entropy-based automated method has been introduced, developing three significant parts: detecting spall, assessing spall, and rating severity of spalling in concrete [17]. Therefore, this proposed work includes three models; a segmentation model for images to detect spall, a feature extracting model for retrieving important properties of the image, and a rating model to compute the severity of the spalling depending on its area and depth.

1.2 Contributions

From the above discussion, we can refer that spalling detection and severity level classification are crucial for maintaining the structural health of concretes. There are several approaches for spalling detection. However, very few methods discussed the severity level of spalling. The severity level of spalling can be measured using depth (deep or shallow) and size (Large or small). The most crucial part of the severity level falls into 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, or large spalling. Moreover, we have provided a comparative analysis of the deep architectures with different encoder-decoder networks to predict the best result among the proposed approaches. The main contributions of our proposed work include:

1. An encoder-decoder-based deep architecture to detect the spalling in concrete.
2. Classification of the level of spalling severity using multi-class semantic segmentation.
3. Comparative analysis between deep architectures with different encoder-decoder networks for spalling detection and severity level classification.

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 [2,21,26]. In this paper we have selected SegNet [2] and UNet [21] for our proposed Architecture. Moreover, later in this paper, we discussed the comparison between these two architectures based on spalling detection and severity level classification.

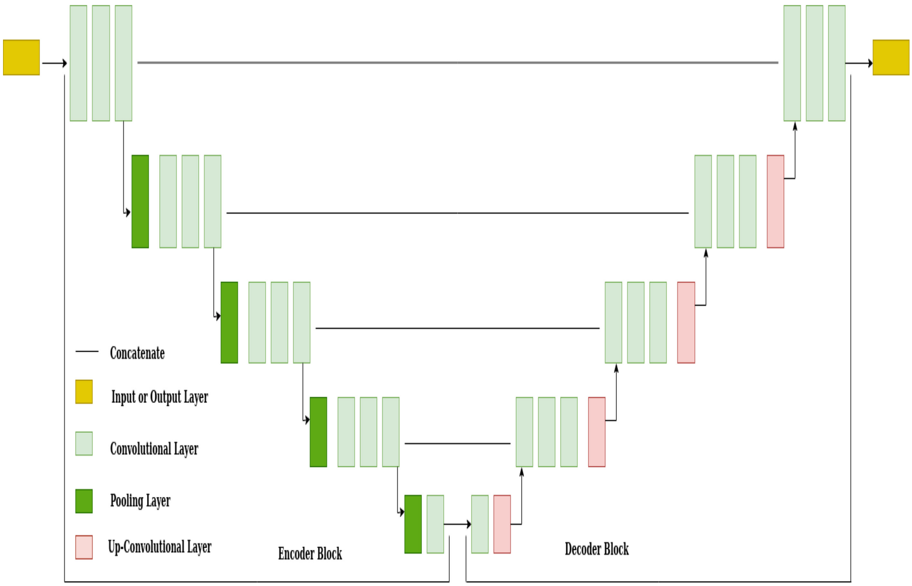

Fig. 2. Architecture for UNet with encoder and decoder [16]

2.1 UNet

The UNet architecture consists of four encoder and four decoder blocks. Each encoder block contains two 3 × 3 convolutions [21]. 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 ““ 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. The architecture for UNet is shown in Fig. 2.

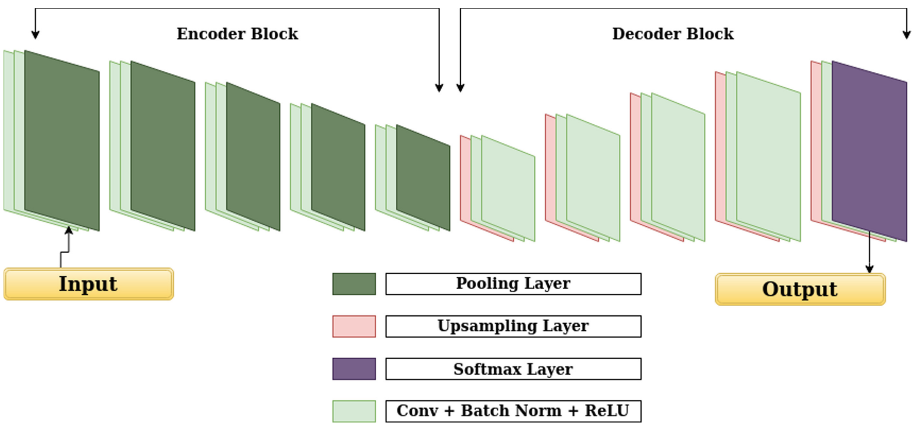

Fig. 3. Architecture for SegNet with encoder and decoder [14]

2.2 SegNet

The SegNet architecture consists of encoder and decoder networks and is proposed for pixel-wise semantic segmentation. 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 [2]. The max pooling is done with a 2 $\times$ 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 3 shows the architecture for SegNet with encoder-decoder block.

2.3 Data Prepossessing and Augmentation

In Deep network architectures, we need large amounts of data [6] 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.

For each image in our dataset, we labeled 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 labeled 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.

2.4 Proposed Deep Architecture

We have proposed using two different types of deep architectures with different encoder-decoder networks for detecting the severity of spalling. First, we detect spalling and the severity level using SegNet and UNet architectures. Finally, we discuss the comparative analysis of their performances. Our proposed architecture is very similar to SegNet and UNet architectures. We have considered spalling detection and severity level classification as multi-class semantic segmentation. Therefore, our proposed method follows the architectures of SegNet and Unet. In the encoder phase for both architectures, we have used three different encoders and provided a comparative analysis for the proposed architectures.

The proposed architecture has an encoder block and a decoder block, making it an encoder-decoder network. This architecture is similar to SegNet architecture, as shown in Fig. 2. Therefore, two main blocks are being used here. Besides the main blocks, the encoder block has network and pooling layers. The decoder block differs from the encoder block with the upsampling layer instead of the pooling layers. For better results in segmentation, the encoder block does subsampling which gives better classification results but reduces the map sizes of the features. For this reason, the decoder block uses upsampling to recreate the

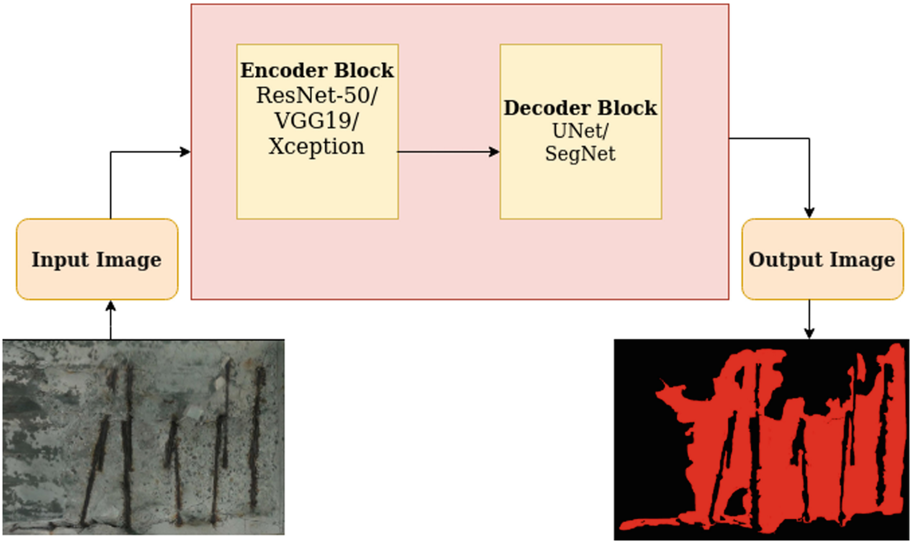

Fig. 4. Proposed encoder-decoder based deep architecture

output resolution as the input images. In this paper, since we are classifying the severity of spalling as no-spalling, small, and large spalling, we have used multi-class segmentation to classify the image pixel by pixel. The proposed Deep architecture is shown in Fig. 4

3 Result

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.

3.1 Data Processing and Experimental Setup

In this study, we have collected images of different types of spalling from buildings, bridges, and roads to train the model for classifying the severity level. These images contain various noises like faded colors, stones, and oil spills. 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 $\times$ 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 $\times$ 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 with different encoder-decoder networks.

3.2 Quantitative Analysis

This section has prepared a performance-based quantitative analysis for two deep architectures with different encoder-decoder networks. Table 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 Eq. 1, 2, and 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.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

With two deep architectures UNet [21] and SegNet [2], we used VGG19, Xception and ResNet50 as encoder-decoder networks. The results from Table 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 performances (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 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 [10,19] and predicted severity rating according to area and depth [17]. 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.

Table 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

3.3 Qualitative Analysis

This section shows the non-statistical evaluation of our proposed architecture for spalling detection and the severity level. Figure 5 presented the qualitative performance for the deep UNet architecture for different encoder-decoder networks. Figure 6 shows the performance analysis for SegNet architecture with the encoder-decoder networks.

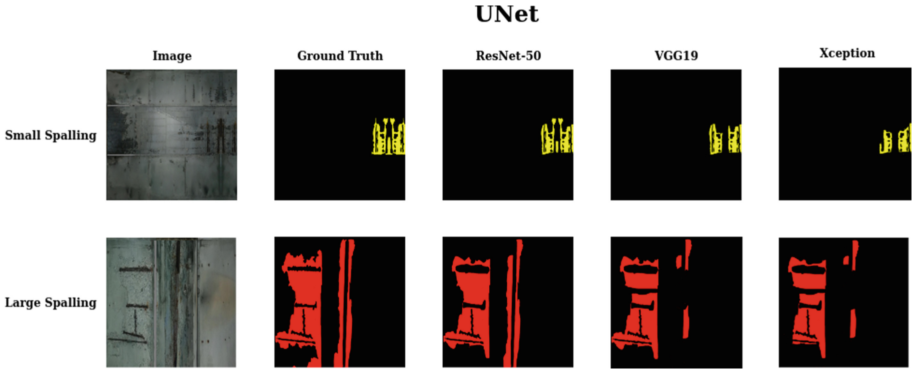

Fig. 5. Results shown for UNet framework with different encoder-decoder networks

Table 1 already shows that the UNet architecture with ResNet-50 shows better results than all the architectures. In Fig. 5, 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.

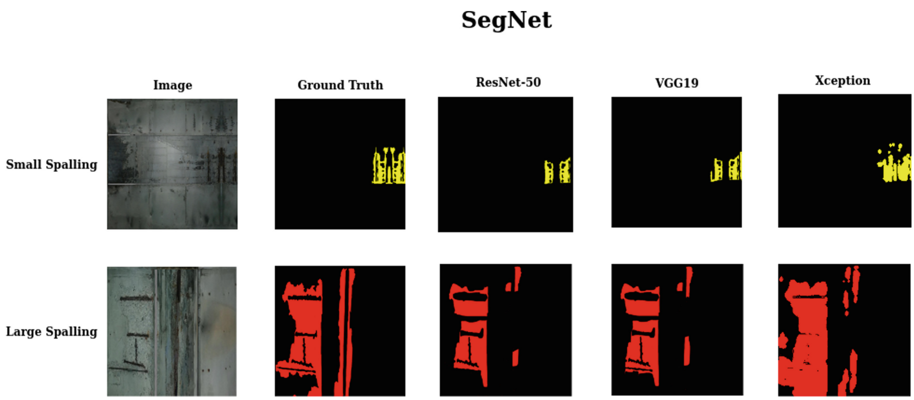

Fig. 6. Results shown for SegNet framework with different encoder-decoder networks

In Fig. 6, 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 Fig. 6. 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.

4 Conclusion

We presented an encoder-decoder-based deep architecture for detecting and classifying the severity level of spalling. Our architecture shows we can classify the spalling levels using semantic segmentation for multi-class. In past studies, there are very few methods to detect the severity level of spalling. In civil infrastructure, it is crucial to know the exact position of the spalling and how severe it is. Therefore, in this study, we have proposed a semantic segmentation-based process with an encoder-decoder. To get a better result and compare them, we have used two different encoder-decoder-based architectures with different encoder-decoder combinations. The most promising result we get for UNet architecture with ResNet-50 encoder. Our future direction will be designing a novel lightweight deep architecture to reduce power consumption and memory while achieving better performance.

Acknowledgment. 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, the Vingroup Innovation Foundation (VINIF) in project code VINIF.2020.NCUD.DA094.

References

1. Abdelkader, E.M., Moselhi, O., Marzouk, M., Zayed, T.: Evaluation of spalling in bridges using machine vision method. In: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 37, pp. 1136–1143. IAARC Publications (2020)
2. Badrinarayanan, V., Handa, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. arXiv preprint arXiv:1505.07293 (2015)
3. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
4. Bai, M., Sezen, H.: Detecting cracks and spalling automatically in extreme events by end-to-end deep learning frameworks. In: ISPRS Annals of Photogrammetry and Remote Sensing Spatial Information Science, XXIV ISPRS Congress, International Society for Photogrammetry and Remote Sensing (2021)
5. Billah, U.H., La, H.M., Tavakkoli, A.: Deep learning-based feature silencing for accurate concrete crack detection. Sensors **20**(16), 4403 (2020)
6. Billah, U.H., Tavakkoli, A., La, H.M.: Concrete crack pixel classification using an encoder decoder based deep learning architecture. In: Bebis, G., et al. (eds.) ISVC 2019. LNCS, vol. 11844, pp. 593–604. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33720-9_46
7. Dawood, T., Zhu, Z., Zayed, T.: Detection and quantification of spalling distress in subway networks. In: Chau, K.W., Chan, I.Y.S., Lu, W., Webster, C. (eds.) Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate, pp. 607–615. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6190-5_55
8. Dawood, T., Zhu, Z., Zayed, T.: Machine vision-based model for spalling detection and quantification in subway networks. Autom. Constr. **81**, 149–160 (2017)
9. Ghosh Mondal, T., Jahanshahi, M.R., Wu, R.T., Wu, Z.Y.: Deep learning-based multi-class damage detection for autonomous post-disaster reconnaissance. Struct. Control. Health Monit. **27**(4), e2507 (2020)
10. Hoang, N.D., Huynh, T.C., Tran, V.D.: Concrete spalling severity classification using image texture analysis and a novel jellyfish search optimized machine learning approach. Adv. Civil Eng. **2021** (2021)
11. Hoang, N.D., Nguyen, Q.L., Tran, X.L.: Automatic detection of concrete spalling using piecewise linear stochastic gradient descent logistic regression and image texture analysis. Complexity **2019** (2019)
12. Hu, Z., Zhu, H., Hu, M., Ma, Y.: Rail surface spalling detection based on visual saliency. IEEJ Trans. Electr. Electron. Eng. **13**(3), 505–509 (2018)
13. Joshi, D., Singh, T.P., Sharma, G.: Automatic surface crack detection using segmentation-based deep-learning approach. Eng. Fract. Mech. **268**, 108467 (2022)

14. Khagi, B., Kwon, G.R.: Pixel-label-based segmentation of cross-sectional brain MRI using simplified SegNet architecture-based CNN. J. Healthc. Eng. **2018**, 1–8 (2018)
15. Kim, M.K., Sohn, H., Chang, C.C.: Localization and quantification of concrete spalling defects using terrestrial laser scanning. J. Comput. Civ. Eng. **29**(6), 04014086 (2015)
16. Li, J., Li, W., Jin, C., Yang, L., He, H.: One view per city for buildings segmentation in remote-sensing images via fully convolutional networks: a proof-of-concept study. Sensors **20**(1), 141 (2019)
17. Mohammed Abdelkader, E., Moselhi, O., Marzouk, M., Zayed, T.: Entropy-based automated method for detection and assessment of spalling severities in reinforced concrete bridges. J. Perform. Constr. Facil. **35**(1), 04020132 (2021)
18. Mohd Ali, A., Sanjayan, J., Guerrieri, M.: Specimens size, aggregate size, and aggregate type effect on spalling of concrete in fire. Fire Mater. **42**(1), 59–68 (2018)
19. Nguyen, H., Hoang, N.D.: Computer vision-based classification of concrete spall severity using metaheuristic-optimized extreme gradient boosting machine and deep convolutional neural network. Autom. Constr. **140**, 104371 (2022)
20. Pham, D., Ha, M., Xiao, C.: A novel visual inspection system for rail surface spalling detection. In: IOP Conference Series: Materials Science and Engineering, vol. 1048, p. 012015. IOP Publishing (2021)
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
22. Tanaka, H., Tottori, S., Nihei, T.: Detection of concrete spalling using active infrared thermography. Q. Rep. RTRI **47**(3), 138–144 (2006)
23. Wu, H., Ao, X., Chen, Z., Liu, C., Xu, Z., Yu, P.: Concrete spalling detection for metro tunnel from point cloud based on roughness descriptor. J. Sensors **2019** (2019)
24. Yang, L., Li, B., Li, W., Liu, Z., Yang, G., Xiao, J.: Deep concrete inspection using unmanned aerial vehicle towards CSSC database. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 24–28 (2017)
25. Zhang, H., Zou, Y., del Rey Castillo, E., Yang, X.: Detection of RC spalling damage and quantification of its key properties from 3D point cloud. KSCE J. Civ. Eng. **26**(5), 2023–2035 (2022)
26. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)
27. Zhou, M., Cheng, W., Huang, H., Chen, J.: A novel approach to automated 3d spalling defects inspection in railway tunnel linings using laser intensity and depth information. Sensors **21**(17), 5725 (2021)