



## Vision Based Crack Detection in Concrete Structures Using Cutting-Edge Deep Learning Techniques

Nirmala Paramanandham<sup>1\*</sup>, Deepali Koppad<sup>2</sup>, Sasithradevi Anbalagan<sup>1</sup>

<sup>1</sup> School of Electronics Engineering, Vellore Institute of Technology, Chennai Campus, Chennai 600127, India

<sup>2</sup> Electronics and Communication Engineering, Ramaiah Institute of Technology, Bangalore 560054, India

Corresponding Author Email: [nirmala.p@vit.ac.in](mailto:nirmala.p@vit.ac.in)

<https://doi.org/10.18280/ts.390210>

### ABSTRACT

**Received:** 14 November 2021

**Accepted:** 20 March 2022

#### Keywords:

*accuracy, concrete images, crack, deep learning, detection*

Concrete crack detection is the process of inspecting the concrete structures. If the defects present in any structures could not be detected in time, it may have a severe impact. The cracks can be detected using destructive as well as non-destructive testing (NDT) techniques. This article presents image based NDT techniques for detecting the concrete cracks using the cutting edge deep learning techniques. NDT is the process of analysing the materials, components, structures etc. without causing any damage to it. In this paper, a transfer learning technique is proposed for detecting the cracks in the concrete structures. A dataset of 40000 images of concrete which is collected from METU Campus is analysed in NVIDIA Tesla V100 12 GB GPU servers using various recent deep learning techniques and the results are tabulated. Performance of four pre trained network architectures such as Alexnet, VGG16, VGG19 and ResNet-50 is assessed for categorizing the images. From the results, it is revealed that the residual neural network technique is successful in detecting the cracks with high accuracy and less complexity.

## 1. INTRODUCTION

Over a period of time, infrastructures such as bridges, roads, and dams have experienced cracks while being used either due to environmental, loading effects, mis-handling, etc. Extending crack in the structures may be the primary reason for the terrible failure and it leads to the severe damage to the environment. Detection of cracks plays a significant role in the structural wellbeing and maintaining the reliability of architecture. Early stage detection of cracks guarantees the protection and helps in arranging the corrective maintenance activity.

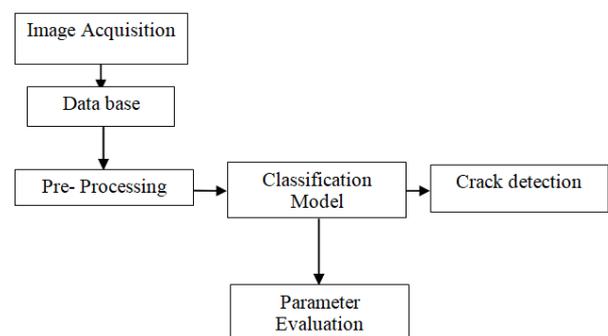
The manual crack detection techniques are time-consuming, low-accuracy, error-prone, risky and subjective [1, 2]. This also requires more labour power and cost. Most of the time, it has less efficiency and causes security dangers. With the growth of numerous techniques, analysts have done various studies on the crack detection. Dimarogonas [3] and Wauer [4] reviewed the vibration techniques for detecting the cracks.

As the cracks are typically thin and lengthy, the unique feature is its continuous property. Inspections that are accomplished using robots and unmanned aerial vehicle (UAV) is mostly image processing based, which means that the inspection platform will obtain the images, and then these images are tested by various techniques. Image processing techniques are extensively used for detecting the cracks [5]. These techniques extract the required features from the images and the deep learning algorithms used for implementing the detection and classification.

Zalama et al. [6] used the filtering based method for determining the spatial and frequency features and classifier for detecting the cracks. Hu et al. [7] developed a map for detecting cracks automatically. This algorithm minimizes

noises and it helps the map for deciding the results. However, automatic crack detection is challenging due to the variation in the intensity, structure, complexity and contrast [7, 8].

Koch et al. [9] reviewed various machine learning techniques for detecting the defects. In the recent years, many deep learning algorithms were implemented for detecting the cracks in civil infrastructures [10-14]. Chen and Jahanshahi [15] implemented a detection procedure using convolutional neural network (CNN) that analyses each frame. Fan et al. [16] trained a CNN model as a multilevel classifier for detecting the cracks in the pavements.



**Figure 1.** General block diagram

Recent progression in automatic tools and advanced Graphical Processing Units (GPUs), it is imperative to use automated techniques for detection of cracks with altering the usage of the architecture. Four such methods have been tried in this article for detecting defects in the concrete material. The crack images are obtained using a camera and a dataset is

developed. After performing pre-processing, these images are automatically processed to state if there is a crack present in the image or not. Figure 1 shows the schematic diagram of crack detection process in the concretes.

This paper concentrates on crack detection using Alexnet, VGG16, VGG19 and ResNet-50. Various related techniques are discussed in section 2. Section 3 describes the 4 models in brief. The results are analysed and discussed in Section 4 and section 5 concludes the paper.

## 2. RELATED WORKS

Generally, vision based crack detection techniques are built based on setting the multiple threshold levels, edge detection methods, and machine learning techniques. Intensity based methods were broadly used due to their simplicity. Even though, these techniques are simple, it is prone to noise and it is leading to false results. Furthermore, setting the proper threshold is not easy [17].

Considering the Edge detection techniques, that use i.e. edge detectors in the vision based algorithm for finding the cracks in the image so that they can be precisely located. In edge detection based techniques, cracks in the image were treated as edges and found by the detectors. Though these techniques are detecting the cracks mostly, fail if the image is not having good resolution or contrast. Morphological methods utilize the connectivity across the pixels and have been effectively exploited in detecting the pavement crack [17]. Though their performances were usually based upon the factor selection, it required manual process for every data set.

Merazi Meksen et al. [18] proposed a split-spectrum based automated method for detecting the cracks using the ultrasonic images. In this technique, instead of storing the images, the sparse matrix of the images are stored (i.e.) only few non-zero coefficients alone were stored. From the sparse matrix, the curves were calculated. For automation, randomized Hough transform is utilized. Split spectrum processing technique was used if the signals were close to each other.

Rodríguez-Martin et al. [19] proposed a technique for finding the cracks in the welding. The cracks in the welding were determined using the infrared image. While capturing the image, the camera is controlled using the personal computer which will take care of setting the parameters such as interval, duration, humidity and temperature. The parameter setting plays a major role in correcting the environmental effects. In this technique, the evaluation has been done by comparing the manual measurement of crack.

Adhikari et al. [20] implemented a technique for the bridge inspection. This technique has a combined model which is based on vision processing. This model consists of quantification, networks as well as visualization models for finding the cracks in the bridges. The branch points in the image and crack parts can be detected by deducting branch pixel values from the crack skeleton. The length of the crack was also computed using the crack skeleton perimeter.

Arena et al. [21] developed a technique for finding the defects using scanning electron microscope images. Initially the images were segmented and converted in binary image. Then using the separation algorithm, filtering and quantification, the cracks were detected. For implementing the algorithm, the authors used the Matlab platform.

Rabha et al. [22] implemented a hybrid algorithm for detecting the cracks in buildings. Initially the correction was

done in the testing image using shading by using the median filter. It will highlight the crack pixels and suppress the non-crack pixels, so that the crack has been easily detected. Once the crack is detected, crack mapping has been done by converting the pixel co-ordinate to terrestrial laser scanner co-ordinate system. Shan et al. [23] developed a detection technique by combining the correlative property and emissivity. The previous technique generated an accurate dimension of surface movements, so that the openings and spacing of cracks were computed. The authors also utilized various dissimilar beam proportionalities for improving the accuracy.

Iyer and Sinha [24] proposed a technique for finding the cracks in pipes which is having noise. This technique is based on the evaluation of curvature and the morphological techniques. The authors also have done segmentation based on geometric model for defining the crack patterns. Initially, the cracks were highlighted using morphology based on their properties. For differentiating the cracks from its background patterns, curvature based assessment have been followed by linear filtering.

To overcome the limitations of traditional techniques, the cracks are detected using artificial intelligent techniques. Recently, the artificial intelligence techniques became more famous because the features were extracted automatically and most of the time, it gives the accurate results.

Dung [25] implemented a methodology centred on fully convolution network for detecting the cracks. In this technique, the authors evaluated the performance of three different pretrained network architecture and VGG16 based encoder was also trained on the subset of 500 images. The authors stated that they achieved around 90% in the precision. For validating the algorithm, the implemented system was tested with a sample which was obtained from a video of cyclic loading.

Yang et al. [26] have proposed a technique using transfer learning based on deep convolution neural network VGG16. The authors stated that, two connected layers itself were sufficient for detecting the cracks and the performance is not improved even if the number of connected layers is increased. Parameter transfer learning has benefits when compared to full learning model, and the achieved improvement was 2.333% and 5.06%. The training time is also reduced by 10 times when compared to full learning for achieving the equivalent results.

The following section explains the methodology of the proposed technique and the selection of dataset in detail with the required block diagrams.

## 3. METHODOLOGY

For performing the automatic crack detection, a model has to be selected. In this article, four different deep Learning models are considered for performing the classification. The first step is either creating a dataset or acquiring an existing dataset. Then the model should be trained and validated using the data. Finally, some of the data is reserved for testing the model efficiency.

### 3.1 Dataset

An existing dataset [27] is utilized for training and testing the four models. The dataset contains images of concrete material which have cracks in them. They are collected from

the METU Campus Buildings. The images are captured and a dataset is created from 458 high-resolution camera (4032X4032 pixels). No data augmentation is done on the images. There are a total of 40000 images categorised as positive, i.e. with cracks and negative, i.e. without cracks. The 40000 images are split into 20000 for each positive and negative category respectively. These images will be the input to the four models and classification is performed. Few of the images with cracks are displayed in Figure 2.

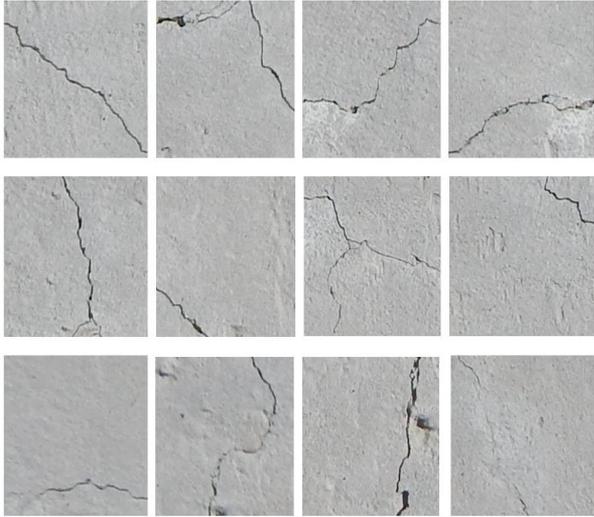


Figure 2. Few examples of some crack images

### 3.2 Deep learning models

#### Convolution layer:

Basic building block of a CNN model is a convolution layer which aids in reducing the dimensions of the original picture using kernels and filters. Even though the size is reduced, all the important features of the image are still maintained for identifying the image. The size of the image is reduced to lessen the data processing and for reducing the overall processing time.

#### MaxPooling layer:

The convolved image is further reduced by using the Pooling mechanism. The frequently used pooling methods are maximum and average pooling. The pooling layer helps in minimizing the needed computational power for processing

the data as the image size is reduced. In the Max Pooling method the maximum value from the image covered by the filter kernel is returned. Though information is lost due to this, the main features can still be extracted and the quality of the image is not lost.

#### ReLU activation function:

Many activation functions namely tanh, Rectified Linear Activation function (ReLU), sigmoid, etc. are commonly used in a neural network. All the models in this paper, use of the ReLU activation function. In this function, the return parameter value will be either the input itself or a 0 depending on if the input is greater than 0.0 or any negative value respectively [28].

#### Fully connected Layer:

The convolution or pooling layer output is fed into a fully connected layer. This layer aids in classifying the image into a specific class.

The final output layer uses the Softmax function or Sigmoid function. Softmax is used for multiclass classification whereas sigmoid is used for binary. Either of the two can be used in the final output layer. In this paper, all the four models are using the Softmax function in the output layer by setting the number of classes to 2.

### 3.3 CNN model

AlexNet is the most commonly used CNN architecture. It is the most commonly used CNN models for classification of images. The significance of this model is the ability to extract practical information regarding the images and also learn about the added features in every layer. It performs the classification of the input images which belongs to 1 among the 1000 classes. The output vector has been modified in the current paper to have only 2 classes since we are dealing with only 2 categories. The generic schematic of AlexNet architecture is as shown in Figure 3 [27].

Alexnet architecture totally consists of eight learned layers in which 5 layers are convolutional with a combination of max pooling and the other three are fully connected. The activation function ReLu is used in all layers. It consists of two drop out layers. Softmax activation function is utilized in the output layer with 2 classes.

The authors assessed the Alexnet model on the same dataset using a Nvidia Titan X GPU with 12 GB memory. The results of this assessment are given by Koppad and Paramanandham [29] for further reading.

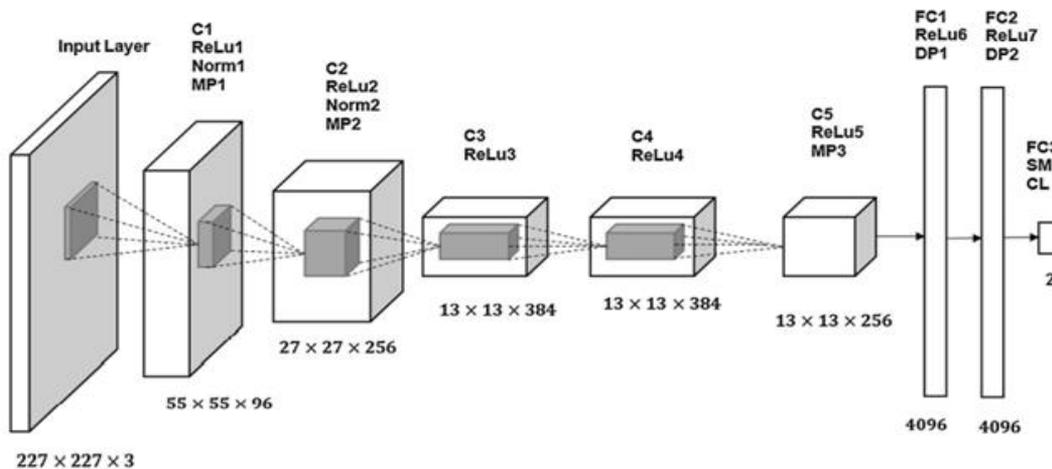


Figure 3. AlexNet architecture

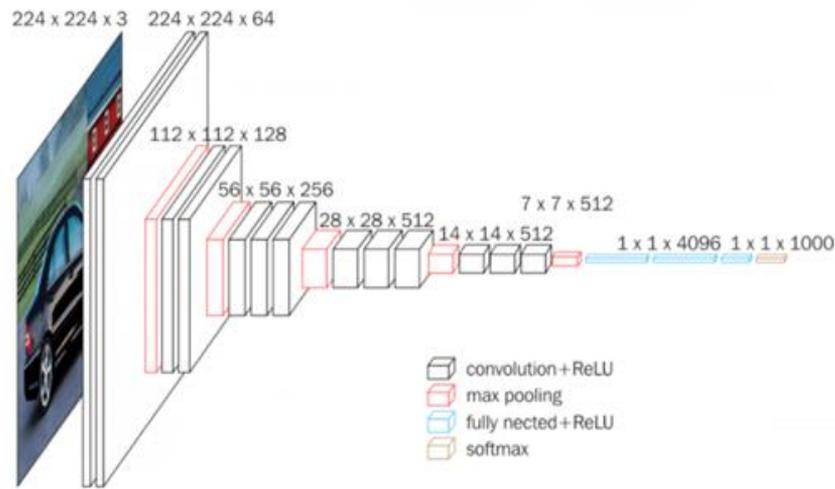


Figure 4. VGG16 model

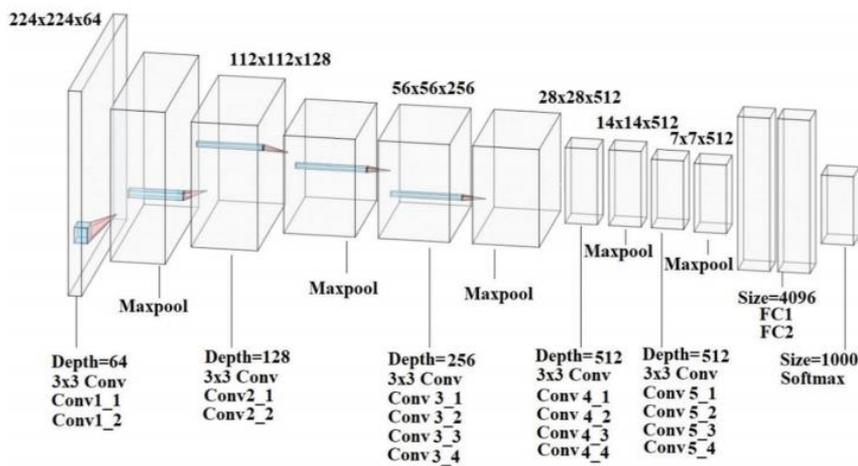


Figure 5. VGG19 model

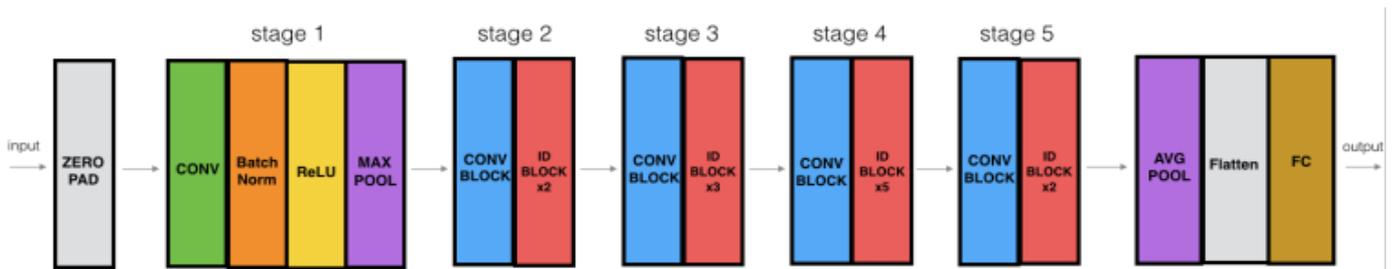


Figure 6. ResNet-50 architecture

### 3.4 VGG16 model

A VGG network was first proposed by Visual Geometry Group (by Oxford University). The standard VGG16 model is as shown in Figure 4 [30]. Simonyan and Zisserman [31] proposed an architecture named VGG16: This architecture is a convolutional neural network model. The model achieves 92.7% in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. This model makes the enhancement over AlexNet by substituting bigger kernel-sized filters using many 3x3 sized kernel filters one by one. VGG16 has been trained for multiple days together and utilized by NVIDIA Titan Black GPU's.

In this model, instead of having a bigger number of hyper-parameters, it is targeted on having convolution layers of 3 x

3 filter with stride 1 and max pool layer of 2 x 2 filter of stride 2. In VGG16, 16 refer the number of layers. It consists of 138 million parameters approximately.

### 3.5 VGG19 model

VGG19 [32] is a variant of the basic VGG network which has 19 layers. The network depth of this architecture is improved by using 3 x 3 convolution layers. For reducing the size, max-pooling layers were used. VGG is trained using separate lesions and for testing all types of lesions taken into account for reducing the false positives. Maxpooling and down sampling is used for extracting the features. Every stacked convolution layer is included with Rectified Linear Unit (ReLU) layer as well as max-pooling layer. ReLU is one of the

finest non-linear activation functions that let only the positive part. The down sampling layer is utilized for improving the anti-distortion ability of the network to the image, in the same time for retaining the key features of the image. The general structure of the VGG19 model is as shown in Figure 5 [32]. The network has 47 layers. In VGG 19, the number 19 represents the number of deep layers with learnable parameters. The output layer has softmax activation function.

### 3.6 ResNet-50 model

ResNet-50 (i.e.) Residual network -50 model [33] is made up of 5 stages each with an identity block and convolution block, which takes an image of size 224x224. Each convolution block is again made up of 3 convolution layers and each identity block has 3 convolution layers which are shown in Figure 6. In ResNet-50, the number 50 represents the number of deep layer. It has more variants that run on the similar concept but have different numbers of layer. It is used to describe the variant that can work with 50 neural network layers.

## 4. RESULTS AND DISCUSSIONS

This section describes the experimental set up of the dataset and the parameters for the 4 different models used for classification of the surface cracks in concrete material. The initial dataset was split into three: Training dataset, validation dataset and testing dataset. The set of 40000 images (including both positive and negative samples) was split into 32000 for training, 4000 each for validation and testing respectively.

### Pre-trained models

A pretrained model is a model in which parameters are trained on a large training set from one domain. The most commonly used is the ImageNet for image classification. This pretrained model can be applied to other problems by applying transfer learning. Transfer learning is a method used to adapt the pretrained model to another dataset. This model is then fine-tuned for the new dataset and the parameters are tweaked as per requirement.

For the CNN model, the dataset images were used in the original size format of 227 X 227. But for the VGG16, VGG19 and ResNet-50 models, the images were resized to 224 X 224 as this is the required input image size of these models. All the four models (CNN, VGG16, VGG19 and ResNet-50) were trained on NVIDIA Tesla V100 12 GB GPU servers.

For all the 4 models, the following parameters were kept consistent: Adam optimizer with a learning rate of 0.001, loss function used was sparse categorical cross entropy. The batch size for CNN was set at 1000 whereas for VGG16, VGG19 and ResNet-50 it was set to 700. The batch size of both the VGG and resNet-50 models was reduced since the GPU was running out of memory if the batch size was maintained at 1000. All the four models were run for 20 epochs and the number of epoch for which maximum accuracy was obtained for validation dataset was noted. Table 1 represents confusion matrix used in this article.

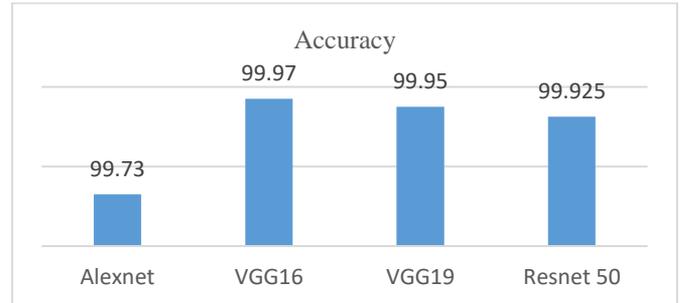
**Table 1.** Confusion matrix

Parameters	Predicted No	Predicted Yes
Actual No	True Negative	False Positive
Actual Yes	False Negative	True Positive

*Accuracy:*

$$accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

From the above Eq. (1), the accuracy of the model can be detected. Accuracy is the number of correctly classified images among all the images. The validation accuracy for the 4 models is as shown in Figure 7.

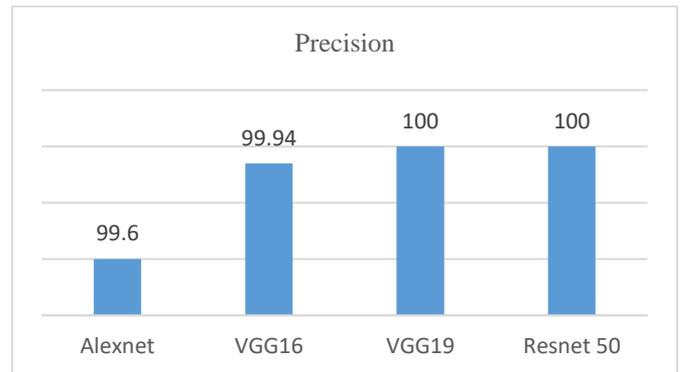


**Figure 7.** Accuracy of the four models

*Precision:*

$$Precision = \frac{(TP)}{(TP+FP)} \quad (2)$$

Precision is a parameter which identifies how many images are actually positive among those that are predicted positive. The precision for the four models is as shown in Figure 8.



**Figure 8.** Precision of the four models

*Recall:*

$$Recall = \frac{(TP)}{(TP+FN)} \quad (3)$$

Recall is related to the sensitivity of the models. It represents the total number of positive cases among all the true positive cases. This parameter for the 4 models is as shown in Figure 9.

*F1 Score:* A good F1 score, i.e. near to 1, directs that there are low FP and FN. This shows that there are fewer false alarms. It can be computed through Eq. (4)

$$F1 = 2X \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

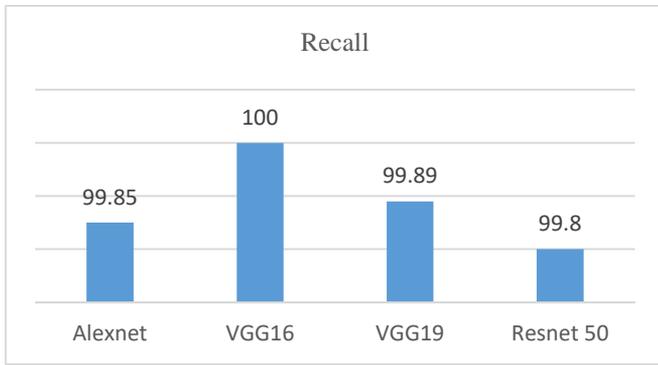


Figure 9. Recall for the four models

The F1 score for the 4 models is as shown in Figure 10.

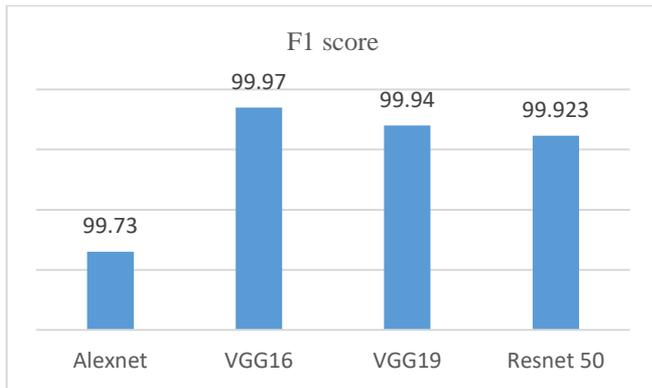


Figure 10. F1 score for the four models

### Confusion Matrix

Table 2, 3, 4 and 5 represents the confusion matrix values of the various networks such as Alexnet, VGG16, VGG 19 and ResNet-50.

Table 2. Alexnet

n=4000	Predicted No	Predicted Yes
Actual No	1979	8
Actual Yes	3	2010

Table 3. VGG16

n=4000	Predicted No	Predicted Yes
Actual No	2019	1
Actual Yes	0	1980

Table 4. VGG19

n=4000	Predicted No	Predicted Yes
Actual No	2027	0
Actual Yes	2	1971

Table 5. ResNet -50

n=4000	Predicted No	Predicted Yes
Actual No	2030	0
Actual Yes	3	1967

The summary of all the results obtained using the 4000 test samples, for all the parameters mentioned above are given in Table 6.

Table 6. Parameters evaluating the effectiveness of the CNN model

Parameters	Alexnet	VGG16	VGG19	ResNet-50
Accuracy (%)	99.73	99.97	99.95	99.925
Precision (%)	99.60	99.94	100	100
Recall (%)	99.85	100	99.89	99.8
F1 score (%)	99.73	99.97	99.94	99.923

The Table 7 shows the results for the test data obtained from the four models. The test data are those images which the model has not been trained or validated in the first phase.

Table 7. Comparison of the models

Model	Parameters				Time taken for training the model (mins)
	Batch Size	Training Accuracy (%)	Epoch	Testing Accuracy (%)	
Alexnet	1000	99.70	3	98.03	187.19
VGG16	700	98.93	20	99.98	849.47
VGG19	700	99.95	6	99.55	1246.55
ResNet-50	700	100	11	99.92	14

There is no single learning model that unanimously performs best across all domains. Hence a number of models need to be explored [34-36].

This article is an attempt in order to evaluate the performance of four different models for identifying cracks in concrete materials. From the Table 7, it can be observed that, Alexnet model has the least test accuracy of 98.03% but it executes faster (approximately 3 hours) when compared to VGG16 & VGG19 models. ResNet-50 has the testing accuracy (i.e.) 99.92% and time taken to train the samples is also very less (i.e. only 14 minutes) when compared to all the other models which are taken for comparison.

VGG16 has the highest test accuracy of 99.98 and takes approximately 14 hours to train the model. The results obtained in this article are in conjunction with the claims presented by Khan et al. [37]. It is mentioned that the ResNet model is 20 and 8 times deeper when compared to AlexNet and VGG models [37]. The computational complexity of the network is also less when compared to Alexnet and VGG Models. All the models were trained using NVIDIA Tesla V100 12 GB GPU Server.

From the deep learning models, it can be observed that ResNet-50 is suitable for the classifying the crack images as the time taken is very less when compared to the other models. The minimal number of epochs aids in dropping the computational time [38] of the model. Furthermore, a greater number of epochs might be the cause of overfitting of the models.

## 5. CONCLUSIONS AND FUTURE WORK

Crack detection in concrete material is performed using Deep Learning techniques. In this paper, the performance of four different models such as Alexnet, VGG16, VGG19 and ResNet-50 was compared in terms of validation and testing accuracy and also the time taken to train the model. All the

models were trained on an NVIDIA Tesla V100 12GB GPU server. All the models were trained for 20 epochs with Adam optimizer with a learning rate of 0.001.

Alexnet model has the least test accuracy of 98.03% at the third epoch but it executes faster (approximately 3 hours) when compared to VGG16 & VGG19 models. ResNet-50 has the validation accuracy (i.e.) 99.92% at 11<sup>th</sup> epoch and time taken to train the samples are also very less (i.e. only 14 minutes) when compared to all the other models which are taken for comparison. VGG16 has the highest test accuracy of 99.98 at 20<sup>th</sup> epoch and takes approximately 14 hours to train the model. VGG19 attained 99.95% accuracy at 6<sup>th</sup> epoch.

From the analysis, it can be concluded that ResNet-50 is suitable for the classifying the crack images as the time taken is very less when compared to the other models. The models in this paper were trained, validated and tested using the concrete crack images which are taken from METU Campus.

## REFERENCES

- [1] Kim, H., Sim, S.H., Cho, S. (2015). Unmanned aerial vehicle (UAV)-powered concrete crack detection based on digital image processing. In International Conference on Advances in Experimental Structural Engineering.
- [2] Wang, P., Hu, Y., Dai, Y., Tian, M. (2017). Asphalt pavement pothole detection and segmentation based on wavelet energy field. *Mathematical Problems in Engineering*, 2017: 1-13. <https://doi.org/10.1155/2017/1604130>
- [3] Dimarogonas, A.D. (1996). Vibration of cracked structures: A state of the art review. *Engineering Fracture Mechanics*, 55(5): 831-857. [https://doi.org/10.1016/0013-7944\(94\)00175-8](https://doi.org/10.1016/0013-7944(94)00175-8)
- [4] Wauer, J.R. (1990). On the dynamics of cracked rotors: a literature survey. *Appl. Mech. Rev.*, 43(1): 13-17. <https://doi.org/10.1115/1.3119157>
- [5] Yang, Q., Shi, W., Chen, J., Lin, W. (2020). Deep convolution neural network-based transfer learning method for civil infrastructure crack detection. *Automation in Construction*, 116: 103199. <https://doi.org/10.1016/j.autcon.2020.103199>
- [6] Zalama, E., Gómez-García-Bermejo, J., Medina, R., Llamas, J. (2014). Road crack detection using visual features extracted by Gabor filters. *Computer-Aided Civil and Infrastructure Engineering*, 29(5): 342-358. <https://doi.org/10.1111/mice.12042>
- [7] Hu, Y., Zhao, C.X., Wang, H.N. (2010). Automatic pavement crack detection using texture and shape descriptors. *IETE Technical Review*, 27(5): 398-405.
- [8] Tang, J., Gu, Y. (2013). Automatic crack detection and segmentation using a hybrid algorithm for road distress analysis. In 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 3026-3030. <https://doi.org/10.1109/SMC.2013.516>
- [9] Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, 29(2): 196-210. <https://doi.org/10.1016/j.aei.2015.01.008>
- [10] Noh, Y., Koo, D., Kang, Y.M., Park, D., Lee, D. (2017). Automatic crack detection on concrete images using segmentation via fuzzy C-means clustering. In 2017 International Conference on Applied System Innovation (ICASI), pp. 877-880. <https://doi.org/10.1109/ICASI.2017.7988574>
- [11] Dinh, T.H., Ha, Q.P., La, H.M. (2016). Computer vision-based method for concrete crack detection. In 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), pp. 1-6. <https://doi.org/10.1109/ICARCV.2016.7838682>
- [12] Sato, Y., Bao, Y., Koya, Y. (2018). Crack detection on concrete surfaces using V-shaped features. *World of Computer Science & Information Technology Journal*, 8(1): 1-6.
- [13] Ali, R., Gopal, D.L., Cha, Y.J. (2018). Vision-based concrete crack detection technique using cascade features. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, 10598: 105980L. <https://doi.org/10.1117/12.2295962>
- [14] Prasanna, P., Dana, K.J., Gucunski, N., Basily, B.B., La, H.M., Lim, R.S., Parvardeh, H. (2014). Automated crack detection on concrete bridges. *IEEE Transactions on Automation Science and Engineering*, 13(2): 591-599. <https://doi.org/10.1109/TASE.2014.2354314>
- [15] Chen, F.C., Jahanshahi, M.R. (2017). NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion. *IEEE Transactions on Industrial Electronics*, 65(5): 4392-4400. <https://doi.org/10.1109/TIE.2017.2764844>
- [16] Fan, Z., Wu, Y., Lu, J., Li, W. (2018). Automatic pavement crack detection based on structured prediction with the convolutional neural network. *arXiv preprint arXiv:1802.02208*. <https://doi.org/10.48550/arXiv.1802.02208>
- [17] Li, H., Song, D., Liu, Y., Li, B. (2018). Automatic pavement crack detection by multi-scale image fusion. *IEEE Transactions on Intelligent Transportation Systems*, 20(6): 2025-2036. <https://doi.org/10.1109/TITS.2018.2856928>
- [18] Merazi Meksen, T., Boudraa, B., Draï, R., Boudraa, M. (2010). Automatic crack detection and characterization during ultrasonic inspection. *Journal of Nondestructive Evaluation*, 29(3): 169-174. <https://doi.org/10.1007/s10921-010-0074-4>
- [19] Rodríguez-Martín, M., Lagüela, S., González-Aguilera, D., Martínez, J. (2016). Thermographic test for the geometric characterization of cracks in welding using IR image rectification. *Automation in Construction*, 61: 58-65. <https://doi.org/10.1016/j.autcon.2015.10.012>
- [20] Adhikari, R.S., Moselhi, O., Bagchi, A. (2014). Image-based retrieval of concrete crack properties for bridge inspection. *Automation in Construction*, 39: 180-194. <https://doi.org/10.1016/j.autcon.2013.06.011>
- [21] Arena, A., Delle Piane, C., Sarout, J. (2014). A new computational approach to cracks quantification from 2D image analysis: Application to micro-cracks description in rocks. *Computers & Geosciences*, 66: 106-120. <https://doi.org/10.1016/j.cageo.2014.01.007>
- [22] Rabah, M., Elhattab, A., Fayad, A. (2013). Automatic concrete cracks detection and mapping of terrestrial laser scan data. *NRIAG Journal of Astronomy and Geophysics*, 2(2): 250-255. <https://doi.org/10.1016/j.nrjag.2013.12.002>
- [23] Shan, B., Zheng, S., Ou, J. (2016). A stereovision-based crack width detection approach for concrete surface assessment. *KSCE Journal of Civil Engineering*, 20(2):

- 803-812. <https://doi.org/10.1007/s12205-015-0461-6>
- [24] Iyer, S., Sinha, S.K. (2005). A robust approach for automatic detection and segmentation of cracks in underground pipeline images. *Image and Vision Computing*, 23(10): 921-933. <https://doi.org/10.1016/j.imavis.2005.05.017>
- [25] Dung, C.V. (2019). Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, 99: 52-58. <https://doi.org/10.1016/j.autcon.2018.11.028>
- [26] Yang, Q., Shi, W., Chen, J., Lin, W. (2020). Deep convolution neural network-based transfer learning method for civil infrastructure crack detection. *Automation in Construction*, 116: 103199. <https://doi.org/10.1016/j.autcon.2020.103199>
- [27] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.
- [28] Nair, V., Hinton, G. E. (2010, January). Rectified linear units improve restricted Boltzmann machines. In *Icml*.
- [29] Koppad, D., Paramanandham, N. (2020). Non-destructive testing for cracks in concrete. In: Jayakumari, J., Karagiannidis, G., Ma, M., Hossain, S. (eds) *Advances in Communication Systems and Networks. Lecture Notes in Electrical Engineering*, vol 656. Springer, Singapore. [https://doi.org/10.1007/978-981-15-3992-3\\_56](https://doi.org/10.1007/978-981-15-3992-3_56)
- [30] <https://neurohive.io/en/popular-networks/vgg16/>, accessed on 3 March 2021.
- [31] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- [32] Sudha, V., Ganeshbabu, T.R. (2021). A convolutional neural network classifier VGG-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning. *CMC-Computers Materials & Continua*, 66(1): 827-842.
- [33] <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>, accessed on 3 March 2021.
- [34] Deng, J., Lu, Y., Lee, V.C.S. (2021). Imaging-based crack detection on concrete surfaces using You Only Look Once network. *Structural Health Monitoring*, 20(2): 484-499. <https://doi.org/10.1177/1475921720938486>
- [35] Ali, L., Harous, S., Zaki, N., Khan, W., Alnajjar, F., Al Jassmi, H. (2021). Performance evaluation of different algorithms for crack detection in concrete structures. In *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pp. 53-58. <https://doi.org/10.1109/ICCAKM50778.2021.9357717>
- [36] Frias, D., Hidalgo, J. (2021). A high accuracy image hashing and random forest classifier for crack detection in concrete surface images. *arXiv preprint arXiv:2106.05755*. <https://doi.org/10.48550/arXiv.2106.05755>
- [37] Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8): 5455-5516. <https://doi.org/10.1007/s10462-020-09825-6>
- [38] Ali, L., Alnajjar, F., Jassmi, H.A., Gochoo, M., Khan, W., Serhani, M.A. (2021). Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors*, 21(5): 1688. <https://doi.org/10.3390/s21051688>