International Journal of High-Rise Buildings
December 2020, Vol 9, No 4, 351-360
https://doi.org/10.21022/IJHRB.2020.9.4.351

International Journal of
High-Rise Buildings
www.ctbuh-korea.org/ijhrb/index.php

Transfer Learning Based Real-Time Crack Detection Using Unmanned Aerial System

N. Yuvaraj[1], Bubryur Kim[2†], and K. R. Sri Preethaa[3]

[1,2]*Department of Architectural Engineering, Kyungil University, 50 Gamasilgil, Hayangeup, Gyeongsan, Gyeongbuk, 38428, South Korea*
[3]*Department of Artificial Intelligence and Data Science, KPR Institute of Engineering and Technology, Coimbatore, 641407, India*

Abstract

Monitoring civil structures periodically is necessary for ensuring the fitness of the structures. Cracks on inner and outer surfaces of the building plays a vital role in indicating the health of the building. Conventionally, human visual inspection techniques were carried up to human reachable altitudes. Monitoring of high rise infrastructures cannot be done using this primitive method. Also, there is a necessity for more accurate prediction of cracks on building surfaces for ensuring the health and safety of the building. The proposed research focused on developing an efficient crack classification model using Transfer Learning enabled EfficientNet (TL-EN) architecture. Though many other pre-trained models were available for crack classification, they rely on more number of training parameters for better accuracy. The TL-EN model attained an accuracy of 0.99 with less number of parameters on large dataset. A bench marked METU dataset with 40000 images were used to test and validate the proposed model. The surfaces of high rise buildings were investigated using vision enabled Unmanned Arial Vehicles (UAV). These UAV is fabricated with TL-EN model schema for capturing and analyzing the real time streaming video of building surfaces.

Keywords: High rise buildings, transfer learning, EfficientNet, Unmanned Arial Vehicle (UAV), crack classification

1. Introduction

Cracks are developed on building surfaces when stress on the building exceeds its strength. Crack is the important indicator for well-being of a building. All infrastructures such as houses, skyscrapers, roads, bridges and dams, gets affected by internal and environmental parameters, challenging its structural safety (Kim et al, 2018). Inspection of buildings is essential for safety of its residents (National Transportation Safety Board, Minneapolis, 2007). Periodical monitoring of cracks on building and taking counteractive measures is required for improving the longevity of the building (Asakura T and Kojima Y, 2003). The longevity of a building is exponential to the maintenance of the infrastructures for ensuring structural safety (Lecun etal., 1998; Chuan-Wei Zhang et al., 2019). Earlier, the human visual inspection of building surfaces was carried out for evaluating the surface quality of the buildings. This manual procedure cannot be used for high rise buildings where human cannot reach with ease. The method is time consuming and expensive. Also, the results predicted by human visualization may not conform to the quality standards framed by the regulating bodies and also the nature of decision varies depending on human skills (Kim et al, 2018). Though there are many parameters for ensuring the structural safety, cracks on surface of the building directly intimates the condition of the structure (Billie F Spencer et al., 2019). Statistical analysis had reported that about 46% of the structural weakness are due failure in monitoring and treating of surface cracks (Cao Vu Dung and Le Duc Anh, 2019). Many accidents have been reported as caused by structural damages (Asakura T and Kojima Y, 2003). To prevent such accidents in future, it is essential to build an efficient model to inspect surface cracks on buildings (Christian Koch et al., 2015). Vibration based SHM systems were applied to enhance the accuracy of conventional human visual inspection method (Fathi H et al., 2015; Yan J et al., 2019). The vibration based methods has a many challenges in monitoring large scale civil infrastructures, structural uncertainties and environmental factors (Cha Y.J. and Wooram Choi, 2017). The challenges in vibration based methods can be overcome using computer vision technologies. Recent developments in computer vision techniques plays key role in civil applications (Billie F Spencer et al., 2019). Computer vision techniques are furnished in revealing numerous hidden insights from image or video data (Thikra Dawood et al., 2017; Ali R et al., 2018; Prasanna P et al., 2016). Most commonly used methods for analyzing images are segmentation, fuzzy clustering (Noh Y et al., 2017), image filtering, pattern recognition, histogram

†Corresponding author: Bubryur Kim
Tel: +82-53-6000-5456, Fax: +82-53-600-5459
E-mail: brkim1@kiu.kr

analysis (Prasanna P et al., 2016; Dinh et al., 2016) and Fourier transformation (Abdel-Qader et al., 2003). The drawback of these traditional methods are they failed to handle noisy image data. Initially, various edge detection algorithms were used by the researches to improve the accuracy of handling noisy data (Gibb, S et al., 2018; Islam, M. M. M and Kim J. M, 2019). Methods like wavelet, Laplace Gaussian and Canny edge detector were used for concrete crack detection, among which wavelet is more widely used in the literature (Stefania C. Radopoulou and Ioannis Brilakis, 2015). (Abdel-Qader et al., 2003) used edge-detection algorithms namely Haar, Canny, Fourier and Sobel transform for crack detection, but these methods failed handle huge volume of images or video data effectively. This paved a way for implementation machine learning models for identification of cracks on building surfaces (Li G et al., 2017). Machine learning algorithms are recommended for handling images with simple features. Huge volume of data with complex images can be efficiently handled using artificial neural networks (ANNs) (Bishop C.M, 2006) and Convolutional Neural Networks (CNNs) (Chen F and Jahanshahi M. R., 2018). Computer vision based end to end learning models using ANN and CNN, can be built which uses parameterized nonlinear activation functions for relating complex input and output.

CNN integrated with machine learning algorithms to analyze the video frames for crack and non-crack classifications (Zhang L et al., 2016). CNN models were also used in various applications like automatic road crack detection (Wang, K.C et al., 2017), cracks on asphalty surfaces (Ren S et al., 2017), steel delamination and corrosion (Fan Z et al., 2018), pavement crack detection (Pauly L et al., 2017), cracks on mechanical steel structures (Cha Y.J et al., 2017), and concrete damage detection (Rawat W and Wang, Z, 2017). The vision based techniques on video data increase the volume of data and handling of voluminous data remains challenging (Sri Preethaa K. R. and Sabari A, 2020). There exists different pre-trained CNN models like Resnet50 (Young?Jin Cha et al., 2017), VGG16 (Sri Preethaa K. R. and Sabari A, 2020), AlexNet (LeCun Y et al., 1998), and InceptionV3 (Zhang C et al., 2019), for crack identification. The main challenges with these pre-trained models were its dependency on high quality data for feature extraction, high computational cost and difficulty in optimizing hyper parameters for complex tasks (Rey N.J, 2016). The Unmanned Arial vehicles (UAV) have extended to operate and monitor the remote locations. The UAV has capability to record multi-dimensional photography and it can be controlled manually for flying at different altitudes. The interpretation of multi-dimensional images needs machine learning algorithms (Asharul Islam Khan and Yaseen Al-Mulla, 2019).

The blending of machine learning algorithms and UAVs has resulted in fast and reliable outputs (Khairunniza-Bejo et al., 2014). UAV and ANN models are used in crop yield prediction (Wu et al., 2007), plant classification based on leaves (Zaman B et al., 2017), wetland mitigation for highway management (Sudheer K.P et al., 2002), rainfall-runoff process, and prediction of evaporation (Douglas Reward K et al., 2018). UAV image data is used for structural modelling (Nahar P et al., 2017). The graffiti clean-up systems were built using the UAV platform and machine learning algorithms (Siewert et al., 2018). The machine learning algorithms were used in detecting, tracking, and classifying flying objects whether un-manned or piloted (Alipour-Fanid et al., 2019). Recent years, the civil societies were focusing on UAV based surveillance. Inspection on buildings using UAV can be implemented widely to capture video of building surfaces and detecting cracks automatically. This method would be reliable and effective than manual inspection of HRBs. Though used in many applications, the machine learning techniques faces complications in handling huge data sets, which are expensive and time consuming.

Our research work focuses on building UAV based reliable model for automatic detection of cracks on building surfaces using deep learning and evaluating quality of the building. A novel method with transfer learning using pre-trained EfficientNet (TN-EL) architecture is proposed for detecting cracks from the live stream video captured by UAV.

2. Overview of the proposed method

This section summarizes the entire process involved in the process of surface crack detection using TL-EN model. Cracks on the concrete surfaces are formed due to various reasons like premature drying, concrete mixture and settlement shrinkage, setting shrinkage chemical reactions, thermal changes, stress concentrations and structural designs. Not all cracks are vulnerable to cause damage to the structures. Some type of crazy cracks on concrete surfaces are very fine and not of structural concern. It is necessary to classify the images into crack and non- crack based on the vulnerability of cracks. The METU dataset used in the proposed work is a collection of both vulnerable and non-vulnerable types of concrete crack images, among which the crazy cracks are labeled under non- cracks. The proposed TN-EL model trained with METU dataset was enabled to classify crazy cracks on dry exterior concrete surfaces as non-cracks. The process flow of crack identification on concrete surfaces of high rise buildings using UAV is depicted in Figure 1. Transfer learning enabled Efficient Net (TL-EN) was built for detecting surface concrete cracks.

A bench marking METU dataset with total 40000 images was used for building the classification model. The dataset is divided for model training and for validating the model. Transfer learning is employed on Efficient Net for building classification model. The

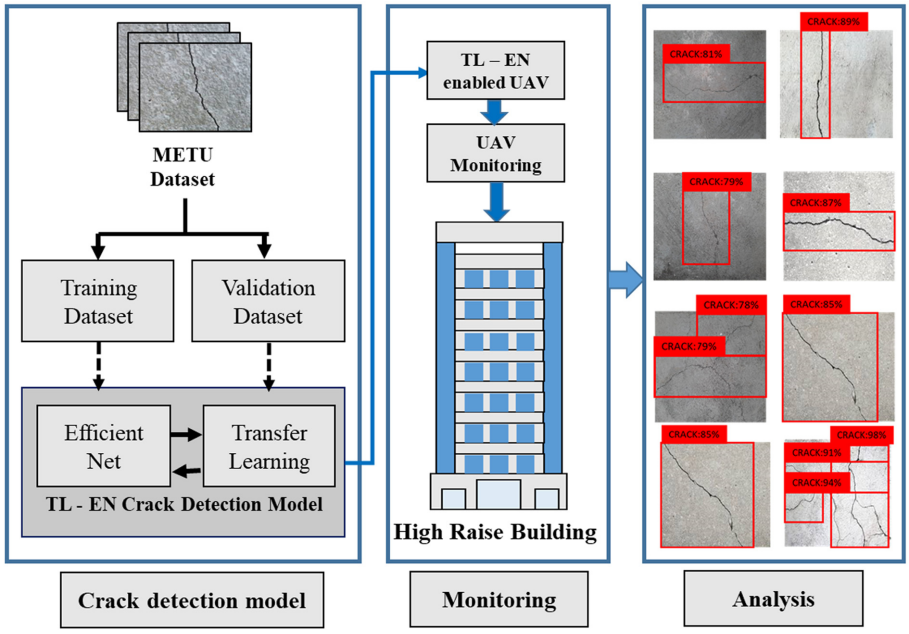


Figure 1. Process flow of crack identification.

developed TL-EN model was integrated with UAV for examining the HRBs. The TL-EN enabled UAV is capable of monitoring the internal and external concrete surfaces of the buildings, where the live streaming video is captured and given as input to the TL-EN model. Using an UI, the results for classification of cracks from the streaming video was analysed. The cracks caused on bridges, tunnels and any other building structures can also be effectively identified using the proposed TN-EL model.

3. Methodlogy

This section explains the overall layer stack of EfficientNet (EN) architecture enabled with transfer learning. The different layers used in this proposed EN architecture and the background of each layer is also discussed. Basically EN is a Convolution Neural Network (CNN) based architecture. A fully connected CNN pre trained neural network is used for classifying crack and non-crack images. CNN is very efficient in processing image to perform computer vision tasks namely segmentation, classification and object detection. Typically, the developments on CNN architectures for accuracy were made by scaling different dimension of the models such as number of channels, kernel size, number of filters and stride size. Inception-v2, ResNet-50, VGG-16, Xception, and DenseNet-121 are few of the enhanced CNN architectures structured for improving the accuracy and computational efficiency. The convolutional filter is applied on convolutional layer to receive the input. The sub-sampling layer obtains feature maps with a lower dimension using average pooling or max pooling in the receptive field. The existing CNNs have the vanishing gradient problem that prevents the front layers of the model from learning in cases where many layers exist. The previous research has shown that scaling any one of the dimensions produced better accuracy up to 80% and on further increase in scaling, the gain gets saturated rapidly (Tan, M. and Le, Q., 2019). In EfficientNet architecture, scaling is done balancing the combination of dimensions for enhanced performance. The EfficientNet proved to produce about 6% higher accuracy and better computational efficiency over other deep CNN architectures namely ResNet-50, Xception, DenseNet-121 and Inception-v2 (Tan, M. and Le, Q., 2019). In this work, EfficientNet architecture with optimized dimensions is proposed to classify the surface crack and non-crack images.

3.1. Overall EfficientNet Architecture

The EfficientNet method uniformly scales each dimension with a fixed set of scaling coefficients. The scaling of individual element dimension improves the model performance by balancing all dimensions of the network like width, depth and image resolution. The effectiveness of model scaling relies on the baseline network. The main building block for EN is MBConv, an inverted bottleneck of convolution layer, originally known as MobileNetV2. By using shortcuts between bottlenecks by connecting a much smaller number of channels, it was combined with an in-depth separable convolution, which reduced the calculation by almost k^2 compared to traditional layers. Where k denotes the kernel size, it specifies the height and width of the 2-dimensional convolution window. The Figure 2 depicts the block representation of multi-layered EfficientNet.

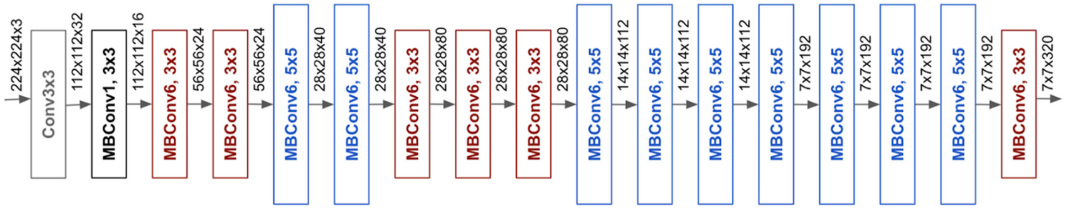


Figure 2. Block representation of EfficientNet

Table 1. Layered architecture of EfficientNet

Stage	Operator	Resolution	Channels
1	Conv3 × 3	224 × 224	3
2	MBConv1,3 × 3	112 × 112	32
3	MBConv6,3 × 3	112 × 112	16
4	MBConv6,3 × 3	56 × 56	24
5	MBConv6,5 × 5	28 × 28	40
6	MBConv6,5 × 5	28 × 28	40
7	MBConv6,3 × 3	28 × 28	80
8	MBConv6,3 × 3	28 × 28	80
9	MBConv6,3 × 3	28 × 28	80
10	MBConv6,5 × 5	14 × 14	112
11	MBConv6,5 × 5	14 × 14	112
12	MBConv6,5 × 5	14 × 14	112
13	MBConv6,5 × 5	7 × 7	192
14	MBConv6,5 × 5	7 × 7	192
15	MBConv6,5 × 5	7 × 7	192
16	MBConv6,5 × 5	7 × 7	192
17	MBConv6,3 × 3	7 × 7	320
18	Conv1x1, Pooling	7 × 7	1280

The baseline of the model by leveraging a multi objective neural architecture search which optimizes the accuracy and flops. The EfficientNet is bigger due the larger layers. Table 1 represents the architecture of EfficientNet with operator, resolution with number of channels. The mobile inverted bottleneck MBConv is the main building block this model. With the building block squeeze and excitation optimization is added to accommodate large number of parameters.

The original convolution layer is divided to reduce the coast of calculation with a minimum loss of accuracy. The MBConv block consist of a layer that squeezes the channels, then a layer that extends the channels. In MBConv, however, blocks consist of a layer that first extends channels and then compresses them to form a multiple layer. The model uses a linear activation function in the last layer to prevent the information loss.

3.2. Transfer learning enabled Efficient net (TL-EN) for crack image classification

Transfer learning (TL) model is a knowledge transmitting technique, where a knowledge gained by model in previous training process is applied for new related tasks. TL does not require repeated training process, thus quickening the learning process of the model. TL is used when a new task does not have sufficient number of parameters for building a new model. Our proposed model uses TL with pre-trained EfficientNet architecture to handle large dataset with minimal parameters better than other CNN models.

The developed TL-EN model was built on multi layered Efficient Net architecture, where training for crack classification is done on one layer and it is transferred to its subsequent layers. The TL-EN architecture consists of 244 × 244 × 3 convolutional layer and the second layer is mobile convolutional layer of 112 × 112 × 32 is used with mobile net's as inverted as res bottleneck. The important hyper-parameters of the deep learning architectures are total number of layers, number of hidden units, Activation function, Epochs and learning rate. In EfficientNet B7 architecture the more number of input parameters are enhanced to provide the efficient results. This shallow architecture relays on minimum computational resource and produces high performances with maximum number of parameters. Overall computational architecture for crack and non-crack detection model performs model training and validation.

The dataset pre-processing is made before building and training the convolutional neural network model. High quality images have variance in terms if surface finish, brightness, contrast and illumination conditions. No data augmentation such as random rotation and flipping is applied. Input images are read and stores in NumPy array along with its label. Data transformation is done by processing the input images which are in NumPy array. The save array can be loaded directly without any pre-processing. The training data is used to construct the EfficientNet model. Epochs shows the sum of one forward pass for all training samples. The model is trained with 20 epochs for better time optimization. The proposed model is shallow network model which uses maximum parameter and minimum resource for computation. The test data is used to ensure the optimization and accuracy of the crack detection model. The performance of the EfficientNet model is validated with other pre-trained CNN models.

4. Experimental results on METU data

4.1. Dataset details

The open sourced Middle East Technical University (METU) dataset is used in this proposed work for crack image classification on the high raise building. The dataset consists a total of 40,000 images of 227 × 227 pixels and were equally grouped into "crack" and "non-crack" for the task of binary classification. The details about the dataset is represented in Table 2.

METU dataset consists of different types of concrete crack images like shrinkage crack, expansion crack, hairline crack, crazy crack, heavy crack and multiple crack. Not all cracks are susceptible to structural damage. Considering the insignificance, such cracks are classified as non - cracks. This diversified collection of crack images as shown in Figure 3 will be helpful in building the enhanced intelligent crack detection model. Among the total of 40000 images, a total of 32000 images were used for model building and 8000 images were used for model validation. Among the 32000 images used in model building consists of 16000 crack and non-crack images respectively. Out of the 8000 images used for model validation, 4000 images were of crack images and 4000 non-crack images.

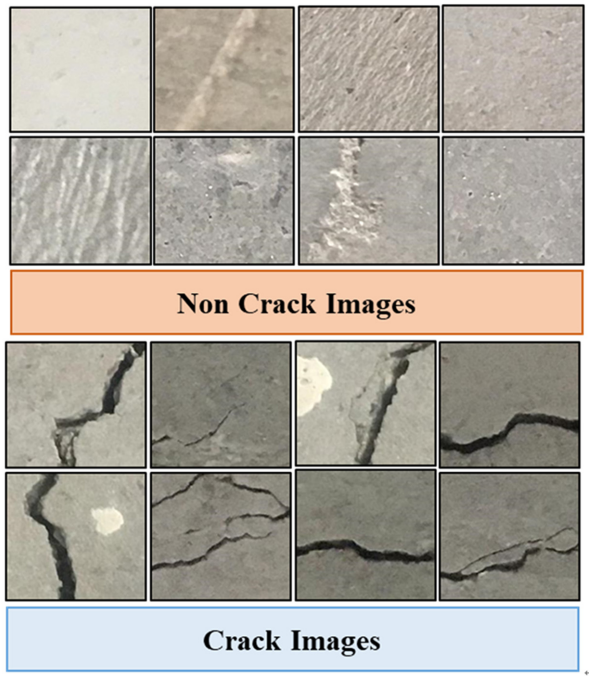


Figure 3. Sample collection of images

4.2. Performance of Transfer Learning enabled EfficientNet algorithm

In this proposed work the crack detection model is developed using the transfer learning enabled efficient net algorithm (TL-EN). Since efficient net is a CNN based model, the performance of TL-EN is validated by comparing its performance with other existing pre-trained CNN classifiers like Inception V3 and ResNet. The classifiers were trained for 50 epochs with a uniform batch size of 16. In addition, to improve performance of the model, parameter sharing concept has been enabled with the classifiers. During training phase During model training TL-EN, produces the maximum accuracy of 99% followed up by inception V3 and ResNet with 98% and 94% of accuracy. The same trend exists during the model validation also where TN-EL, manages to produce the maximum accuracy of 99% followed up by Inception V3 and ResNet.

4.2.1 Crack identification using TL-EN model

The classifier of TL-EN was trained for 50 epochs with batch size of 16. TL-EN model is validated for crack classification using Inception V3 and ResNet pre-trained models. The classifier of TL-En model produces almost perfect classification of crack with only 8 false positives (FP) and 10 false negatives (FN) over total of 8000 test images (Figure 4a). The Inception V3 classifier produces 104 FP and 96 FN (Figure 4b), whereas ResNet classifier produced 266 FP and 199 FN (Figure 4c).

It is observed that TL-EN classifier produces very fewer FP and FN than other two classifiers. Therefore, TL-EN classifier has achieved is high accuracy with 0.997 followed by Inception v3 with 0.975 and ResNet with 0.943.

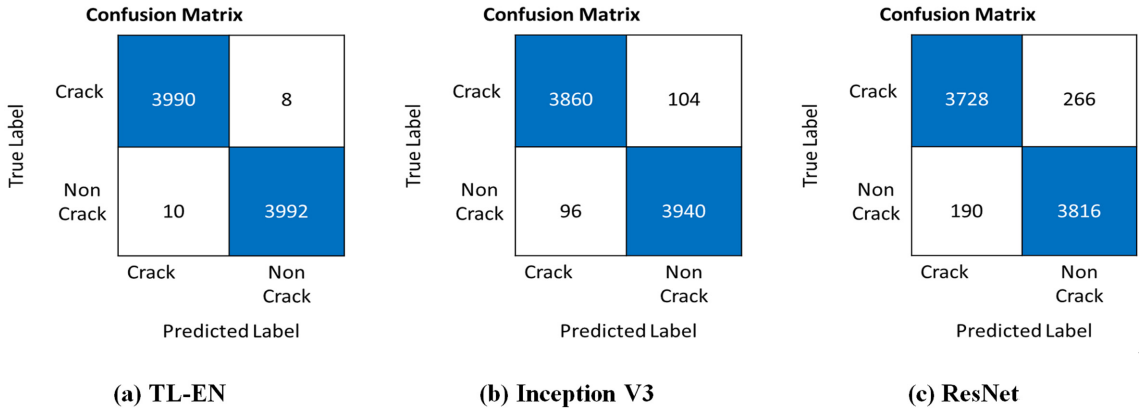
The performance of TL-EN is measured using training accuracy, training loss, validation accuracy and validation loss with respective 50 epochs. TL-EN manages to produce and maintain the maximum training accuracy of 0.998 and validation accuracy 0.998 at 50th epoch (Figure 5).

Also, TL-EN has shown minimum training loss of 0.02 value and minimum validation loss of 0.012 at 50th epoch (Figure 6). It is observed that TL-EN is able to maintain the minimum loss and maximum accuracy in both during training and validation.

The performance comparison of TL-EN with performance of Inception V3 and ResNet is shown in (Figure 7). It is witnessed the training accuracy and validation accuracy of Inception V3 at epoch 40 is 0.975 and 0.97 respectively. The training accuracy and validation

Table 2. Dataset details

Dataset details				Model building		Model validation	
Total images	Image size	Crack	Non crack	Crack	Non crack	Crack	Non crack
40000	227 × 227	20000	20000	16000	16000	4000	4000



(a) TL-EN **(b) Inception V3** **(c) ResNet**

Figure 4. Confusion matrices of pre-trained models

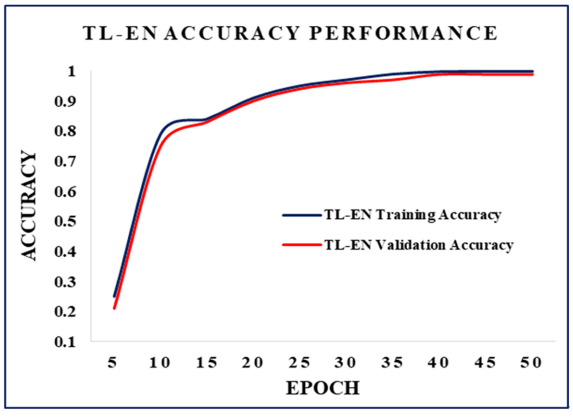


Figure 5. Accuracy of TL-EN during training and validation.

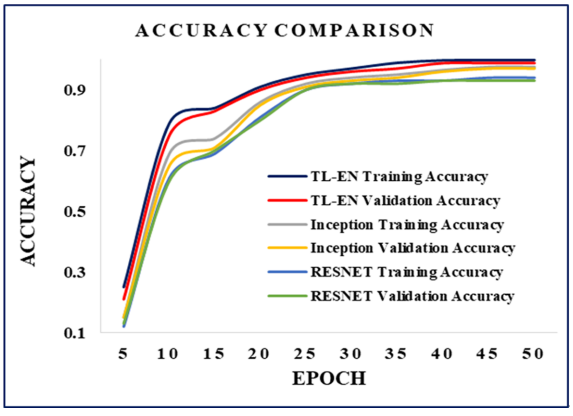


Figure 7. Accuracy comparison during training and validation.

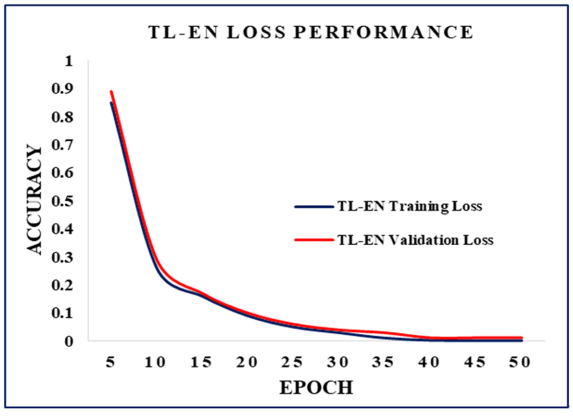


Figure 6. Loss of TL-EN during training and validation.

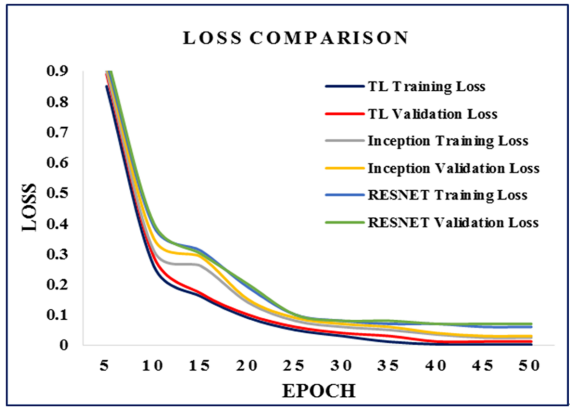


Figure 8. Loss comparison during training and validation

accuracy of ResNet is 0.94 and 0.93 respectively which much lesser when compared to TL-EN and Inception V3 models. From the above results proposed TL-EN is found to perform better than other two modes for classification of cracks.

Similarly, the loss comparison during training and validation phases of the three models is analysed (Figure 8). The TL-EN model produces minimum training and validation loss with 0.002 and 0.012 respectively 50th epoch. The Inception V3 classifier produces a little greater loss than TL-EN with training loss of 0.024 and validation loss of 0.03 at 50th epoch. The ResNet is found to have greater loss in comparison with both TL-EN and Inception V3 with values 0.06 and 0.07 for training and validation loss respectively. From the results it is observed that the pre-trained model with transfer

learning could produce better performance than pre-trained models for classification of cracks.

4.3. Real time crack detection

The longevity of the building depends on the regular maintenance of its structure. Cracks play a vital role in deciding the healthiness of the building. Periodical inspection of building for surface cracks is a key role for maintenance. Conventional method of crack detection is difficult in high rise buildings. With recent developments in computer vision technologies and UAVs, crack detection is made at ease. The real time application of TL-EN embedded vision system enabled UAV integrated with crack detection model for surface crack analysis and maintenance of High Raise Building (HRB) is shown in Figure 9.

The competent model for surface crack detection was developed using transfer learning based on Efficient Net architecture. This model is enabled to process multi-dimensional video data with varied resolutions efficiently. HRB structures are captured in vision enabled UAVs. Earlier versions of UAVs are capable of autonomous navigation and video capturing only, with no processing memory. Vision enabled UAVs are capable of integrating the model schema for analysing the captured video. The structures of HRB were live streamed using vision enabled UAVs, the TL-EN schema in UAV analyse the live video for cracks. The results are analysed through

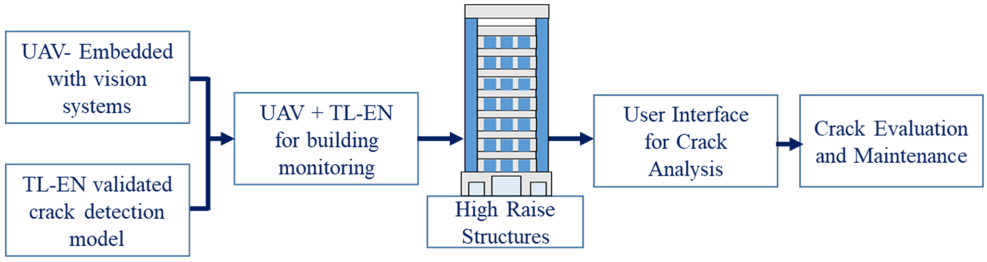
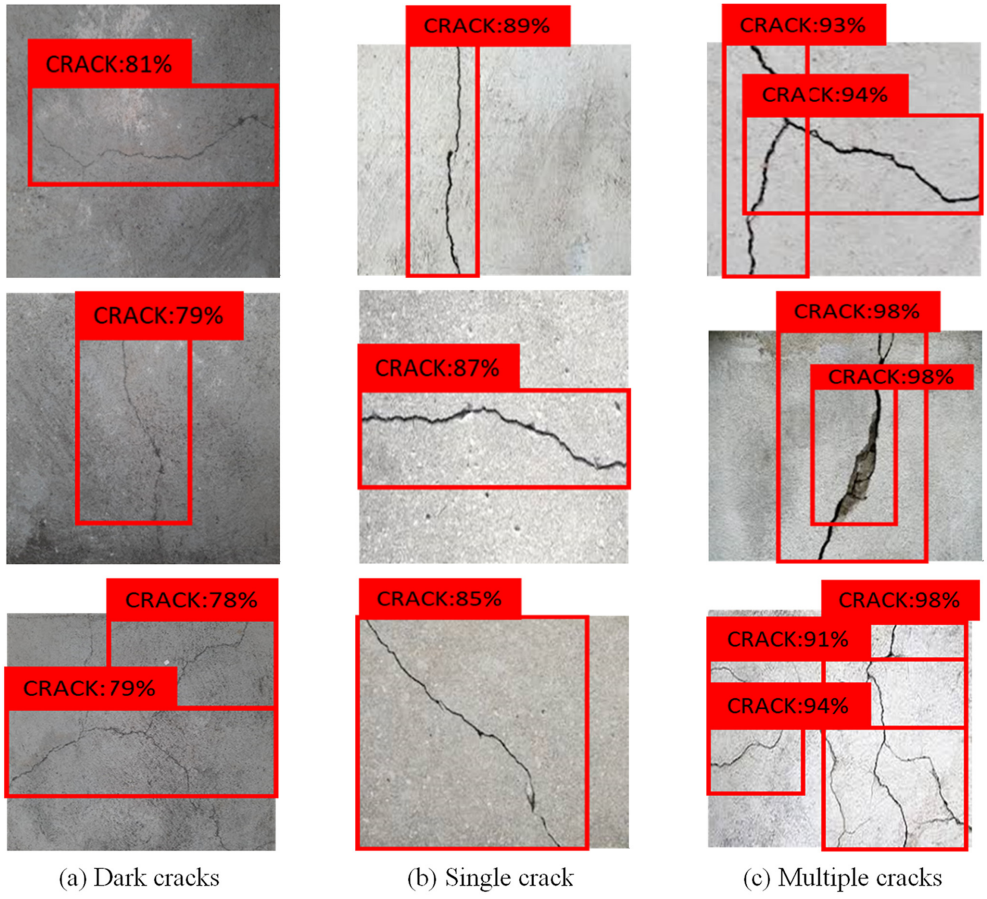


Figure 9. Real time crack detection by TL-EN.



(a) Dark cracks (b) Single crack (c) Multiple cracks

Figure 10. Results of real time crack detection by TL-EN.

user interfaces for crack vulnerabilities and recommended for maintenance.

The UAV integrated with TL-EN model is used to capture video on inner and outer concrete surfaces of the high rise building. The TL-EN model operates on the video as captured by UAV and classifies the cracks within the bounding box. The depth and width of the cracks are the key indicators to evaluate the vulnerability of the crack. The depth and width of the cracks on the surface varies based on the distance and angles at which the images are captured by the UAV. This factor greatly affects the accuracy in crack detection and classification. The proposed work uses UAV equipped with four proximity sensors, to regulate the distance between the concrete surface and camera. The proximity sensors are capable of calculating the distance from the nearest obstacle in a linear direction. In addition, the UAV is configured with high end camera mounted on a 360-degree rotatable head that enabled UAV to track the surfaces at fixed distance and multiple angles. The proposed model was capable of tracking the concrete surface at fixed distance and resolution uniformity among the captured images was ensured.

It is found that performance of the model varies with respective to illumination and complexity of the cracks. Eliminating illumination effect from the real time image data is a complex task. The over exposure to light degrades the accuracy of algorithms. The average classification accuracy of single concrete crack with low illumination is observed to be around 79% (Figure 10a), whereas with sufficient light is 87% (Figure 10b). Similarly, multiple concrete cracks with sufficient illumination manages to produce around 95 % (Figure 10b) accuracy whereas with low illumination is only around 78% (Figure 10c). From the results it is observed that TL-EN model performs well on classifying single or multiple cracks with sufficient illumination on concrete surfaces.

5. Conclusions

Inspecting surfaces of HRB for cracks manually and accurately is essential for improving the longevity of the buildings. Reluctant act in monitoring HRB may result unsafe for the residents of the buildings as well the pedestrians. The conventional methods for crack detection using human visual inspection and evaluations were expensive, time consuming and unreliable. In this work the issues of inspecting HRBs and inaccuracy in detecting surface cracks were addressed. The developed TL-EN model for classifying the cracks on the building surfaces had produced the classification accuracy of 0.998. The model was built with the concept of transfer learning on pre trained EfficientNet architecture. The transfer learning enabled the model to be trained efficiently and quickly on dataset with smaller number of parameters. The inspection of building structures at higher altitudes were carried out using vision enabled UAV. The UAV is integrated with the schema of TL-EN classification model to identify the cracks in the live stream video captured by the UAV. Further, evaluation and analysis of the crack identification results can be done through user interfaces. The METU dataset with 40000 images was used for building TL-EN model, of which 32,000 images were used for training and 8000 images were used for validating the model. The classifier of TL-EN was trained in batches of size 16 for 50 epochs. The performance of the TL-EN model was validated in comparison with performance of pre-trained Inception V3 and ResNet architectures. On measuring the performance of the model using confusion matrix, TL-EN produced more accurate classification of cracks with only 8 FP and 10 FN over total of 8000 test images, whereas Inception V3 classifier produced 104 FP, 96 FN and ResNet with 266 FP and 199 FN. It is observed that TL-EN had performed accurate classification than Inception V3 and ResNet on METU dataset. The performance of the TL-EN model was also measured using other parameters namely training accuracy, training loss, validation accuracy and validation loss. The TL-EN had produced improved validation accuracy with 0.988 whereas Inception V3 and ResNet models obtained 0.97 and 0.93 as accuracy values respectively. Also TL-EN model have proved to produce minimal validation loss with 0.012, whereas loss validation loss of Inception V3 and ResNet is 0.03 and 0.07 respectively. It is witnessed that TL-EN is able to maintain the minimum loss and maximum accuracy both during training and validation. From the results, the TL-EN model is validated to produce accurate, reliable classification of surface cracks. It is also noted that illumination and complexity of cracks on building surfaces during inspection by UAV had varied effects on the performance of the model. The surfaces with sufficient light and simple cracks are found to get classified more accurately by TL-EN model. The elimination of lighting effects from the real time stream video remains a challenging task. The developed TL-EN schema embedded UAV is recommended for real time monitoring and evaluation of cracks in maintaining the structural health of HRBs.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2019R1G1A1095215). This work was supported by Korea Research Fellowship Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science and ICT (No. 2019H1D3A1A01101442).

References

Abdel-Qader, Abudayyeh, O. and Kelly, (2003). Analysis of Edge-Detection Techniques for Crack Identification in

Bridges. *Journal of Computing in Civil Engineering*, 17, 255-63.

Ali, R., Gopal L.G. and Cha Y.J. (2018). Vision-based concrete crack detection technique using cascade features, Proc. SPIE, Sensors and Smart Structures Technologies for Civil Mechanical, and Aerospace Systems, 10598.

Alipour-Fanid, Amir, Dabaghchian, Monireh, Wang, Ning, Wang, Pu & Zhao, Liang & Zeng, Kai. (2019). Machine Learning-Based Delay-Aware UAV Detection and Operation Mode Identification over Encrypted Wi-Fi Traffic.

Asakura, T and Kojima, Y. (2003). Tunnel maintenance in Japan, Tunnelling and Underground Space Technology. 18, 161-169.

ASCE, 2017, American Society of Civil Engineers (ASCE), Infrastructure Report Card.

Asharul Islam Khan and Yaseen Al-Mulla. (2019). Unmanned Aerial Vehicle in the Machine Learning Environment, *Procedia Computer Science*, 160, pp. 46-53.

Billie F Spencer, Vedhus Hoskere and Yasutaka Narazaki (2019). Advances in Computer Vision-Based Civil *Infrastructure Inspection and Monitoring. Engineering*, 5(2), 199-222.

Bishop, C.M. (2006) Pattern recognition and machine learning. New York City, Springer-Verlag.

Cao Vu Dung and Le Duc Anh (2019). Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, 52-58.

Cha, Y.J. and Wooram Choi. (2017). Deep learning-based crack damage detection using convolutional neural networks, *Computer-Aided Civil and Infrastructure Engineering*, 32, 361–378.

Cha, Y.J., Choi, W., Suh, G., Mahmoudkhani, S and Buyukozturk, O. (2017) Autonomous structural visual inspection using region based deep learning for detecting multiple damage types, *Computer Aided Civil Infrastructure Engineering*, 33(9), 731-747.

Chen, F. and Jahanshahi, M. R. (2018). 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.

Christian Koch, Kristina Georgieva, Varun Kasireddy, Burcu Akinci and Paul Fieguth. (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.

Chuan-Wei Zhang , Meng-Yue Yang, Hong-Jun Zeng, Jian-Ping Wen. (2019). Pedestrian detection based on improved LeNet-5 convolutional neural network. *Journal of Algorithms & Computational Technology*, 13, 1-10.

Dinh, T.H. ,Ha, Q. and La, H. M. (2016) "Computer vision-based method for concrete crack detection." Proc. 14th International Conference on Control, Automation, Robotics and Vision, August, pp.1-6.

Dongho Kang and Young-Jin Cha (2018). Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging. *Computer-Aided Civil and Infrastructure Engineering*, 33 885-902.

Douglas, Reward K., Said Nawar, M Carmen Alamar, AM Mouazen, and Frederic Coulon. (2018). Rapid Assessment of Petroleum Hydrocarbon Contamination in Soils: Spectroscopy Approach. *Journal of Engineerng and Applied Sciences*, 4(3), 66-73.

Fan, Z., Wu, Y., Lu, J. and Li, W. (2018) Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network, Computer Visions and Pattern Recognition, arXiv:1802.02208.

Fathi, H., Fei Dai and Manolis Lourakis. (2015). Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges. *Advanced Engineering Informatics*, 29, 49-161.

Gibb, S., La H. M. and Louis, S. (2018). A genetic algorithm for convolutional network structure optimization for concrete crack detection. IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, 1-8.

Islam, M. M. M., Kim J. M. (2019). Vision-Based Autonomous Crack Detection of Concrete Structures Using a Fully Convolutional Encoder-Decoder Network, Sensors, 19.

Khairunniza-Bejo, Siti, Samihah Mustaffha, and Wan Ishak Wan Ismail. (2014). Application of artificial neural network in predicting crop yield: A review. *Journal of Food Science and Engineering*, 4(1), p. 1.

Kim, I.H, Jeon, H., Baek, S.C.,Hong W.H and Jung H.J. (2018). Application of Crack Identification Techniques for an Aging Concrete Bridge Inspection Using an Unmanned Aerial Vehicle. Sensors, *18.*

L. Zhang, F. Yang, Y. Daniel Zhang and Y. J. Zhu, "Road crack detection using deep convolutional neural network," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 3708-3712, doi:10.1109/ICIP.2016.7533052.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) "Gradient-based learning applied to document recognition." *Proc. the IEEE*, 86, pp. 2278-2324.

Lecun,Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) "Gradient-based learning applied to document recognition." *Proc of the IEEE*, 86(11), 2278-2324.

Li, G., Zhao, X., Du, K., Ru, F. and Zhang, Y. (2017). Recognition and evaluation of bridge cracks with modified active contour model and greedy search-based support vector machine. *Automation in Construction*, 78, 51-61.

Nahar, P., Wu, K., Mei, S., Ghoghari, H., Srinivasan, P., Lee, Y., Gao, J.Z., & Guan, X. (2017). Autonomous UAV forced graffiti detection and removal system based on machine learning. 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, *Internet of People and Smart City Innovation*, 1-8.

National Transportation Safety Board, Collapse of I-35W Highway Bridge, Minneapolis, Minnesota (2007).

Noh, Y., Koo, D., Kang, Y., Park, D. and Lee. D. (2017) "Automatic crack detection on concrete images using segmentation via fuzzy C526 means clustering." Proc. of the 2017 International Conference on Applied System

Innovation (ICASI), Sapporo, Japan, April, pp. 877-880.

Pauly, L., Hogg, D., Fuentes, R., Peel, H. (2017) "Deeper networks for pavement crack detection." Proc. the 34th International Symposium in Automation and Robotics in Construction, IAARC, Taipei, Taiwan, pp. 479-485.

Prasanna, P., Dana, K. J., Gucunski, N., Basily, B. B., La, H.M, Lim, R.S. and Parvardeh. (2016). Automated crack detection on concrete bridges. *IEEE Transactions on Automation Science and Engineering*, 13, 591-599.

Prasanna, P., Kristin J. Dana, Nenad Gucunski, Basily B. Basily, Hung M. La, Ronny Salim Lim, and Hooman Parvardeh. (2016). Automated Crack Detection on Concrete Bridges, *IEEE Transactions on Automation Science and Engineering*, 13(2), 591-599.

Rawat, W. and Wang, Z. (2017) Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computing*, 29(9), 2352-2449.

Ren, S., He, K., Girshick, R. and Sun, J. (2017), Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 39, 1137-1149.

Rey, N.J. (2016). Combining UAV-imagery and machine learning for wildlife conservation.

Road Bureau, Ministry of Land, Infrastructure, Transportation, and Tourism, Roads in Japan, Retrieved from: http:// www.mlit.go.jp/road/road_e/index_e.html.

Siewert, Sam, Mehran Andalibi, Stephen Bruder, Iacopo Gentilini, Aasheesh Dandupally, Soumyatha Gavvala, Omkar Prabhu, Jonathan Buchholz, and Dakota Burklund. (2018). Drone Net, a passive instrument network driven by machine vision and machine learning to automate UAS traffic management. AUVSI Xponential, Denver, Colorado.

Sri Preethaa, K. R. and Sabari, A. (2020), Intelligent video analysis for enhanced pedestrian detection by hybrid metaheuristic approach, *Soft Computing*, 24(16), 12303-12311.

Stefania C. Radopoulou and Ioannis Brilakis. (2015). Patch detection for pavement assessment, *Automation in Construction*, 53, 95-104.

Sudheer, K.P., Gosain, A.K., Mohana Rangan, D and Saheb, S.M. (2002) Modelling evaporation using an artificial neural network algorithm. *Hydrological Processes*, 16(16), 3189-3202.

Tan, M. and Le, Q. (2019). "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." Proceedings of the 36th International Conference on Machine Learning, in PMLR, Long Beach, California, 97, pp.6105-6114.

Thikra Dawood, Zhenhua Zhu and Tarek Zayed. (2017). Machine vision-based model for spalling detection and quantification in subway networks. *Automation in Construction*, 81, 149-160.

Wang, K.C., Zhang, A., Li, J.Q., Fei, Y., Chen, C. and Li, B. (2017). Deep learning for asphalt pavement cracking recognition using convolutional neural network, *Airfeld and Highway Pavements*, 166–177.

Wu, Stephen Gang, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, and Qiao-Liang Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. in 2007 IEEE international symposium on signal processing and information technology. 2007: IEEE.

Yan, J., Downey, A., Cancelli, A., Laflamme, S., Chen, A., Li, J. and Ubertini, F. (2019). Concrete Crack Detection and Monitoring Using a Capacitive Dense Sensor Array. Sensors, 19, 1843.

Young?Jin Cha, Wooram Choi and Oral Buyukozturk (2017). Deep learning-based crack damage detection using convolutional neural networks, *Computer Aided Civil and Infrastructure Engineering*, 32, 361-378.

Zaman, B., Jensen, A.M., and McKee, M. (2017). UAV, Machine Learning, and GIS for Wetland Mitigation in Southwestern Utah, USA.

Zhang, C., Patras, P. and Haddadi, H. (2019). Deep Learning in Mobile and Wireless Networking: A Survey. I*EEE Communications Surveys & Tutorials*, 21(3), pp. 2224-2287.

Zhang, L., Yang, F., Zhang, Y.D. and Zhu, Y.J. (2016). "Road crack detection using deep convolutional neural network." Proc. in IEEE International conference on Image Processing, Phoenix, AZ, pp. 3708-3712.