Contents lists available at SCCE

Journal of Soft Computing in Civil Engineering

Journal homepage: www.jsoftcivil.com

Improved Image Based Super Resolution and Concrete Crack Prediction Using Pre-Trained Deep Learning Models

K. Sathya[1*], D. Sangavi[2], P. Sridharshini[2], M. Manobharathi[2], G. Jayapriya[2]

1. Assistant Professor, Department of CSE, Coimbatore Institute of Technology, Coimbatore, India
2. UG Student, Coimbatore Institute of Technology, Coimbatore, India

Corresponding author: *sathya.k@cit.edu.in*

https://doi.org/10.22115/SCCE.2020.229355.1219

ARTICLE INFO

Article history:
Received: 30 April 2020
Revised: 15 July 2020
Accepted: 24 July 2020

Keywords:
Generative adversarial network (GAN);
Crack prediction;
Super-resolution generative adversarial network (SRGAN);
Highly resolution image;
VGG19;
ResNet50;
Crack and non-crack images.

ABSTRACT

Detection and prediction of cracks play a vital role in the maintenance of concrete structures. The manual instructions result in having images captured from different sources wherein the acquisition of such images into the network may cause an error. The errors are rectified by a method to increase the resolution of those images and are imposed through Super-Resolution Generative Adversarial Network (SRGAN) with a pre-trained model of VGG19. After increasing the resolution then comes the prediction of crack from high resolution images through Convolutional Neural Network (CNN) with a pre-trained model of ResNet50 that trains a dataset of 40,000 images which consists of both crack and non-crack images. This work makes a comparative analysis of predicting the crack after and before the super-resolution method and their performance measure is compared. Compared with other methods on super-resolution and prediction, the proposed method appears to be more stable, faster and highly effective. For the dataset used in this work, the model yields an accuracy of 98.2%, proving the potential of using deep learning for concrete crack detection.

How to cite this article: Sathya K, Sangavi D, Sridharshini P, Manobharathi M, Jayapriya G. Improved Image Based Super Resolution and Concrete Crack Prediction Using Pre-Trained Deep Learning Models. J Soft Comput Civ Eng 2020;4(3):40–51. https://doi.org/10.22115/scce.2020.229355.1219.

1. Introduction

The existence of Concrete Structures highly relies on the method of crack detection. Manual inspection of these cracks involves a small handheld magnifier with a built-in measuring scale that clearly identifies the width of the crack. However, the structures like skyscrapers, bridges, dams make it difficult to undergo manual inspection and do not ensure safety. Thus, to overcome these drawbacks Convolutional Neural Network for predicting the cracks through images was introduced [1]. However, the process of image capturing is made through various sources like cameras of different pixels. In that case, the image may be affected by various factors like motion blur, camera mode noise, and optical distortion. Thus, the acquisition of such images may cause an error in prediction. Hence a Super-resolution method is introduced to increase the resolution of the images. This method includes Generative Adversarial Network with a pre-trained model of VGG19.

High resolution images then enter the Convolutional Neural Network with a pre-trained model of ResNet50. CNN is deep learning algorithms that are highly specified with image classification and feature recognition. Because of partial connection, sharing weights, and pooling process between the neurons, Convolutional Neural Network is able to recognize the features within the images [2]. The dataset has 40,000 images of equally separated crack and non-crack images. For these datasets, we use pre-trained networks as a starting point and image augmentation to perform operations like vertical or horizontal flip, brightness and rotation to improve accuracy.

The main contributions of this study are an efficient classification framework is proposed for the effective categorization of cracks and non-cracks based on a crack candidate region (CCR), comparative analysis between SRGAN-based and CNN-based methods is conducted for evaluating the classification performances and a comprehensive crack identification is conducted in the presence of crack-like no cracks for practical applications.

The outline of the paper is as follows. Section 2 depicts the outline of the project along with the short description of the techniques used. Followed by a brief explanation of the existing works that face various minor and major limitations. Section 3 depicts our analysis of the existing methods that support our project. A brief explanation of our techniques is presented in Section 4. Details on datasets and our implementations are detailed in Section 5 and 6 respectively. Finally, conclusions are presented in Section 7.

2. Related works

2.1. Data augmentation

Image Data augmentation is used to artificially enlarge the size of the training dataset by creating varied versions of the image in the dataset. This technique is performed by applying domain-specific to some samples from the training data that leads to creation of new and different training samples [3]. The augmentation technique creates variations in the images that leads to improvement in the ability of the fit models in order to generalize the learned new images [4]. For training the datasets the specific data augmentation technique is chosen carefully and within the context of the dataset and knowledge of the problem domain. It can be used in isolation and check if they result in a measurable improvement to model performance [4,5]. It is only applicable for the training dataset and not to the validation or test dataset.

2.2. Generative adversarial network (GAN)

GAN uses the technique of a max-min player game to generate the real-like images which the generator produces. The GAN technique was improvised by providing information to the generator to produce a particular output for the given input [6]. Though GAN produces good results, Deep Convolutional GAN was implemented to make it more stable. It is more effective than GAN in producing real-like images by reducing noises [7]. DC-GAN includes transposed convolution techniques to perform the upsampling of 2D image size [8]. DC-GAN also utilizes the encoding-decoding network, where encoding takes input, and output as a feature [9]. While decoder is a network that takes the feature from the encoder and gives the best closest match to the input as an output.

2.3. Super resolution generative adversarial networks (SRGAN)

Super Resolution GAN produces higher resolution image by applying a deep network along with an adversary network. From low resolution image SRGAN can generate super resolution image with finer details and high quality [10]. Earlier CNNs were used to produce high resolution images, but SRGANs are more preferable as CNNs often produce blurry images. During the training, a low-resolution image is obtained by down sampling the high-resolution image. Later the low-resolution images are unsampled to super resolution images by the GAN generator [11].

2.4. Pre-trained deep learning models

VGG19. In Deep Neural Network, the multi-layered operation is performed by the Visual Geometry Group network (VGG) [12]. It uses a 3x3 convolutional layer to increase the depth level, max pool layers to reduce the volume size, and two fully connected layers with 4096 neurons [13]. The pre-processing involves subtracting the RGB mean from each pixel computed over a large training set. The spatial resolution of the images is preserved through Spatial Padding [13,14]. Max Pool layer is performed over 2 * 2 pixels. Later, nonlinearity is introduced by Residual Learning (ReLu) to boost the feature classification and computational time.

Resnet50. In a deep convolutional neural network, layers such as a convolutional layer, ReLu, fully connected layer, Maxpool layer, flatten layer are stacked and they are trained [15]. The network learns several features such as low, medium and high at the end of its layers [16]. In residual learning, instead of trying to learn some features, we learn some residuals. Residual is the reduction of features learned from the input of that layer [17]. ResNet performs this by using input of the nth layer and some $(n+x)^{th}$ layer [17]. This shows that the training of this network is easier than simple deep convolutional neural networks and also resolves the accuracy degradation issue [9]. This is the fundamental concept of ResNet. ResNet50 is a 50-layer Residual Network and there are other variants like ResNet101 and ResNet152 also.

In ResNet50, Skip Connection is its main innovation. Due to backpropagation in the model, the gradient becomes smaller and smaller gradients may cause learning difficulty. Thus, Skip

Connection allows the network to pass the input through the block without interpreting through other weight layers [9]. Also, the layers would be skipped over if it is not useful for the model. Therefore, the addition of layers would not affect the performance of the model.

3. Proposed method

The proposed technique was able to build a network for concrete crack prediction. The proposed system was capable of classifying the images based on the classes like crack(positive) and non-crack(negative). To achieve the classification and prediction, the system was implemented in a proper way for better understanding and implementation.

3.1. Problem statement

The problem statement includes upscaling of the captured real images in order to predict and trace the cracks from the concrete crack images.

3.2. Methodology

The process of crack prediction involves both the architecture of pre-trained CNN models and SRGAN. Each of which plays its own major role in the prediction of crack. The process of training the datasets, data splitting, data augmentation, etc. are performed in the CNN network, while the role of increasing the resolution of an image is performed by the VGG19 model in the SRGAN network, as shown in Fig .1. Both networks get collaborated at the level of prediction.

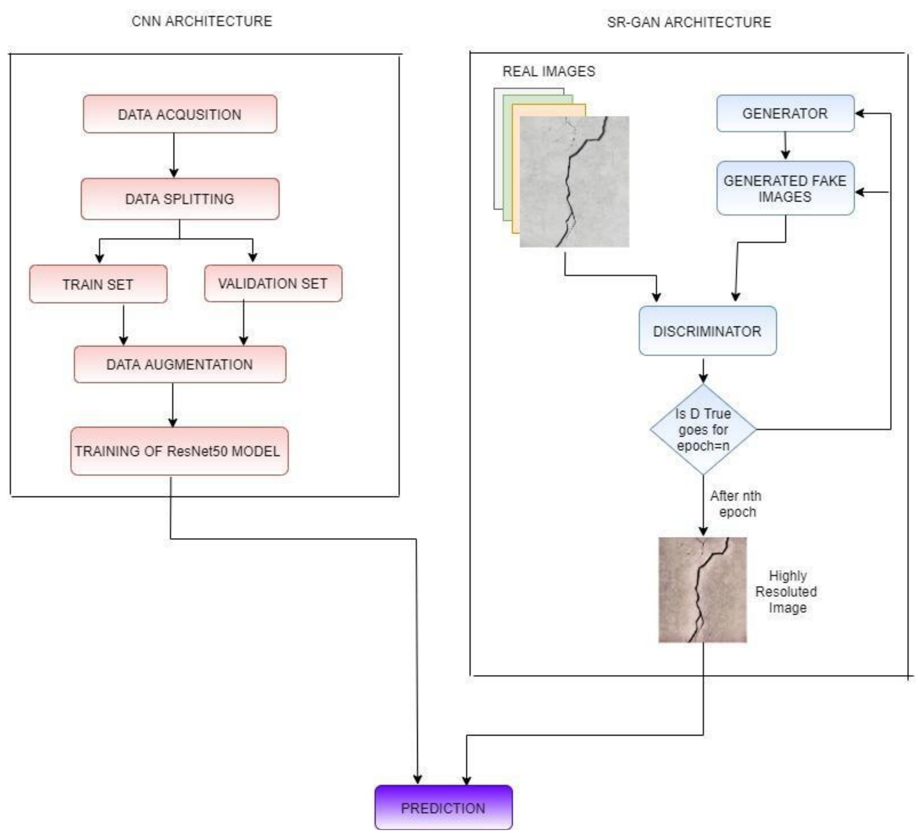


Fig. 1. Concrete Crack Detection architecture.

3.3. Experiments

The experimental program comprises four phases that includes: i) acquisition of a classified dataset; ii) apply transformation to augment datasets; iii) implement transfer-learning approach; and iv) run the training experiments. These phases are briefed as follows.

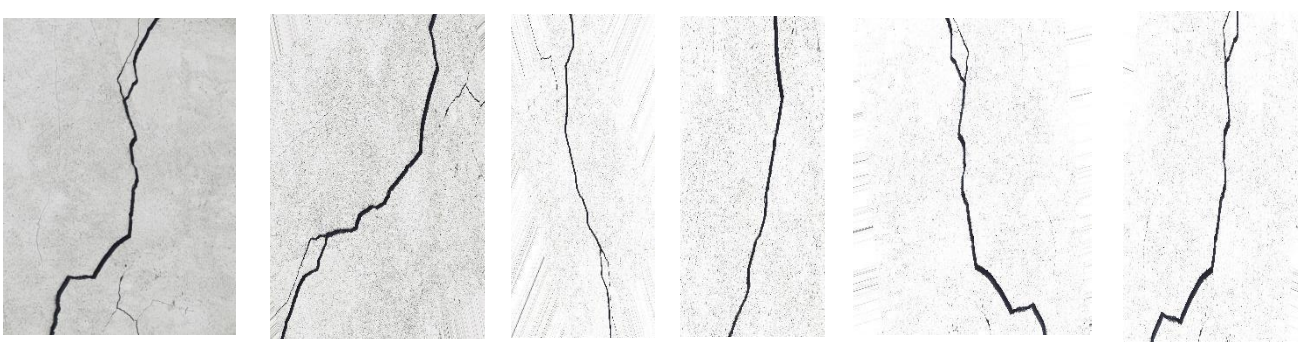
3.4. Datasets

Data acquired from Mendeley's Concrete Crack Images for Classification has 40,000 images including both non-crack and crack images. The dataset is generated from 458 high-resolution images (4032x3024 pixel) [18]. Each image in the data set is a 227 x 227 pixels RGB image. The datasets that are used are as follows: i) Surface Crack Detection: Positive and Negative ii) Real images iii) Testing set of Surface Crack Detection and iv) Generated set of high-resolution images. The dataset can be found in https://data.mendeley.com/datasets/5y9wdsg2zt/2.

3.5. Experimental setup

The training details includes 40000 random samples of images from Surface Crack database for training and trained on NVidia K80 GPUs.

Data Augmentation. In deep learning data augmentation plays its role as an internal process since it has to deal with large amounts of data. It is the concept of creating new data at different orientations [19]. It helps to introduce variability within the datasets. As shown in Fig. 2, Operations performed by data augmentation includes rotation, flipping, zooming, cropping, and changing the brightness level.



a) Original Image b) Rotation c) Shear d) Zoom e) Horizontal Flip f) Brightness

Fig. 2. Traditional data augmentation techniques.

Super-Resolution Generative Adversarial Network (SRGAN).SRGAN focuses on enhancing images with low resolution by applying Generative Adversarial Network to generate images of high resolution [20]. The structure of SRGAN includes 16 Residual blocks, Pixel Shuffler x2, 2 subpixel CNN in Generator. Parameterized Relu (PReLu) is used to learn the negative part coefficient. The process of producing high resolution images is depicted in Fig. 3. The main work in here is multi-task loss function [21] which comprises of three modules as, (i). Encoding pixel- wise similarities using Mean Square Error (MSE) loss.

(ii). Perceptual similarity metric is defined over high level image representation in terms of distance metric.

(iii). Adversarial Loss

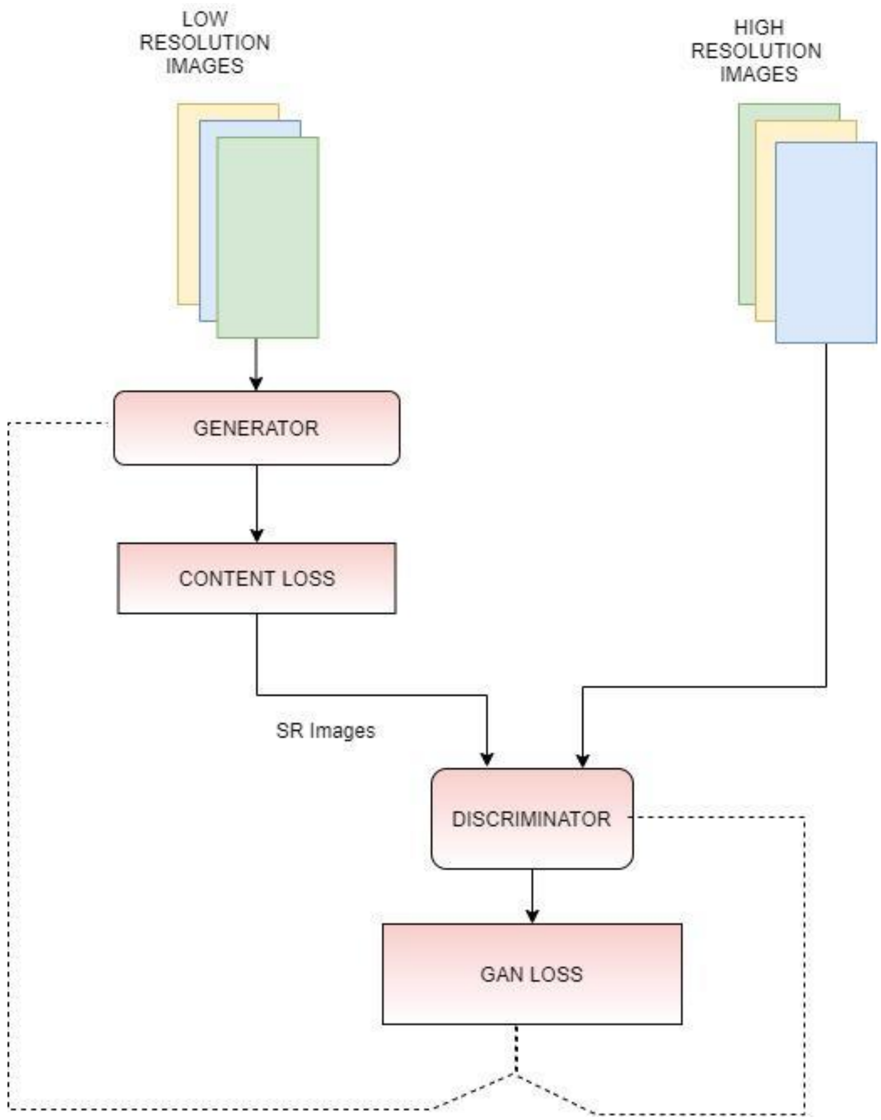


Fig. 3. Block Diagram of Super-Resolution Generative Adversarial Network

As given in Fig. 3, this network has two main blocks as, i) Generator and ii) Discriminator.

In the process of training, the low-resolution images are produced through gaussian filtering of high-resolution images and later undergoing the process of down sampling with the down sampling factor **"r = 4".**

1. Low Resolution Images (ILR) - W **x H x C**
2. High Resolution Images (IHR) **-rW x rH x C**

It also has a pre-trained model of VGG19 which is a Convolutional Neural Network that makes use of 19 layers in it that are trained on several image samples. It also utilizes the Architectural style of Zero-Centre normalization, convolution, ReLU, Max Pooling, Fully Connected layer, etc.

Computation of the content loss pixel-wise using the mean square error (MSE) between the HR and SR images. Nevertheless, when it determines the distance mathematically, that is not easily identified by humans. SRGAN measures the MSE of features extracted by a VGG-19 network using perceptual loss. Wherein some features are to be expected to be matched for certain special layers in VGG-19. It is also calculated based on probabilities provided by Discriminator.

$$l_{Gen}{}^{SR} = \sum_{a=1}^{N} - \log \mathrm{D}_{\theta_{\mathrm{D}}}\left(G_{\theta_{\mathrm{G}}}\left(\mathrm{I}^{\mathrm{LR}} \right) \right) \tag{1}$$

where D –Discriminator, $l_{Gen}{}^{SR}$ – adversarial loss, $G_{\theta_G}\left(I^{LR}\right)$ –pixel in generated image**.**

CNN. Convolution Neural Network is built in Pytorch. Since we are having a limited number of images, we can use a pretrained network as a starting point and for improving accuracy we use image augmentations [22]. Image augmentations are involved in allowing us to do transformations like vertical and horizontal flip, rotation and brightness changes significantly increasing the sample and helping the model generalize [23].

Resnet 50 model pretrained on ImageNet to jump start the model. The ResNet50 model consists of 5 stages. Each of these stages are provided with a convolution and Identity block. Each convolution block is provided with 3 convolution layers and similarly each identity block has 3 convolution layers. The ResNet-50 has over 23 million trainable parameters. All these weights along with 2 more fully connected layers are freeze. The first layer has 128 neurons in the output and the second layer has 2 neurons in the output which are the final predictions.

The model works on the validation data but it should be made sure that it also works on unseen data from the internet. For testing this, we take random images of cracked concrete structures and cracks from road surfaces. These images are much bigger than our training images. Our model was trained on crops of 227, 227 pixels. Now, the input image is broken into small patches and runs the prediction on it. If the model is predicted as crack, then colour the patch with red (cracked) else colour the patched green. The model does very well on images that it has not seen before. As shown in the image below, the model is able to detect a very long crack in concrete by processing 100s of patches on the image.

4. Results

The ultimate outcome of this report is image wherein the cracks are traced after the process of increasing its resolution using SRGAN. The highly resoluted image of a real image is shown in the Fig. 4.

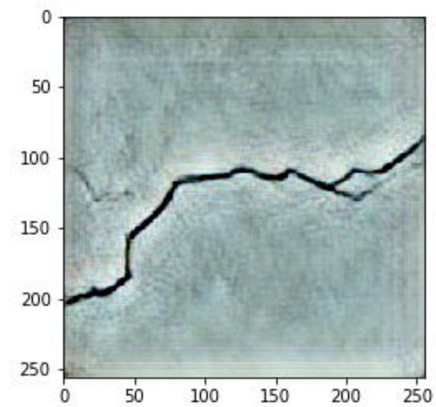


Fig. 4. Generated Image with High Resolution.

Perceptual loss function is used when comparing two different images that look similar, like the same photo but shifted by one pixel. The function is used to compare high level differences, like content and style discrepancies, between images.

$$l^{sr} = l_x^{\ sr} + 10^{-3} l_{Gen}^{\ sr} \tag{2}$$

where l_x^{sr} – context loss, l_{Gen}^{sr} – adversarial loss

Content loss is the mean squared error calculated between each pixel value from the real image and each pixel value from the generated image. Content loss can be of two types. They are pixel-wise MSE and VGG loss. Pixel-wise MSE loss is mean squared error between each pixel in real image and a pixel in generated image.

$$l_{MSE}^{\ SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \left(\mathrm{I}_{x,y}^{\ \mathrm{HR}} - \mathrm{G}_{\theta\mathrm{G}}\left(\mathrm{I}^{\mathrm{LR}} \right)_{x,y} \right)^2 \tag{3}$$

where MSE – mean square error, $I_{x,y}^{HR}$ – pixel in real image, $G_{\theta_G}\left(I^{LR}\right)$ –pixel in generated image.

VGG loss applied over generated images and real images. It is calculated as the Euclidean distance between the feature maps of the generated image and the real image.

$$l^{SR}_{\ VGG/i.j} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}\left(\mathrm{I}^{\mathrm{HR}} \right)_{x,y} - \phi_{i,j}\left(G_{\theta_G}\left(\mathrm{I}^{\mathrm{LR}} \right) \right)_{x,y} \right)^2 \tag{4}$$

Further the image gets into the Convolutional Neural Network to be processed to determine the trace of the crack [24]. Also, the cracks are traced without undergoing the process of resolution in order to determine the accuracy. The image after crack prediction process with and without undergoing resolution process is shown in Fig. 5 and Fig. 6 respectively.

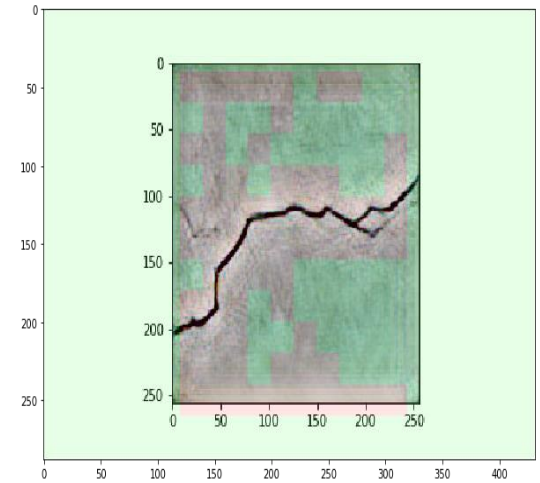


Fig. 5. Prediction with increased Resolution.

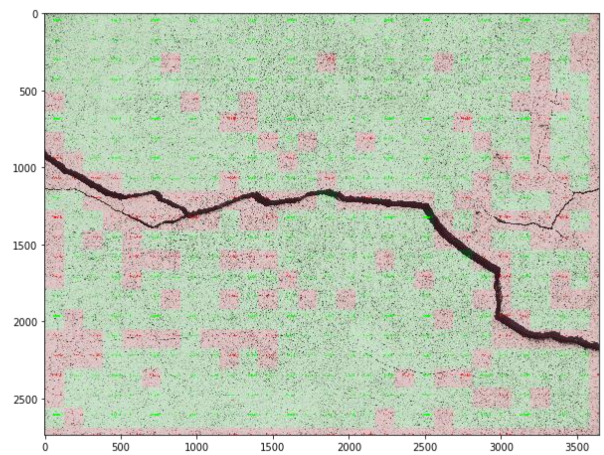


Fig. 6. Prediction with low Resolution.

4.1. Performance measures

Transfer learning is used for training the model on the training dataset and the measuring loss and accuracy is done on the validation set. As shown in Fig. 7, after the 1st epoch, train accuracy is 87% and validation accuracy is 97%. This is the power of transfer learning. Our final model has a validation accuracy of 98.2%. Therefore, as the loss decreases, the accuracy of the model increases gradually.

The loss is calculated on training and validation and its interpretation is based on how well the model is doing in these two sets. Loss value implies how poorly or well a model behaves after each iteration of optimization. An accuracy metric is used to measure the algorithm's performance in an interpretable way.

The discriminator seeks to maximize the probability assigned to real and fake images. While a discriminator neural network tries to differentiates between real samples and the ones generated by the Generator Network, and the Generator Network trying to fool the discriminator.

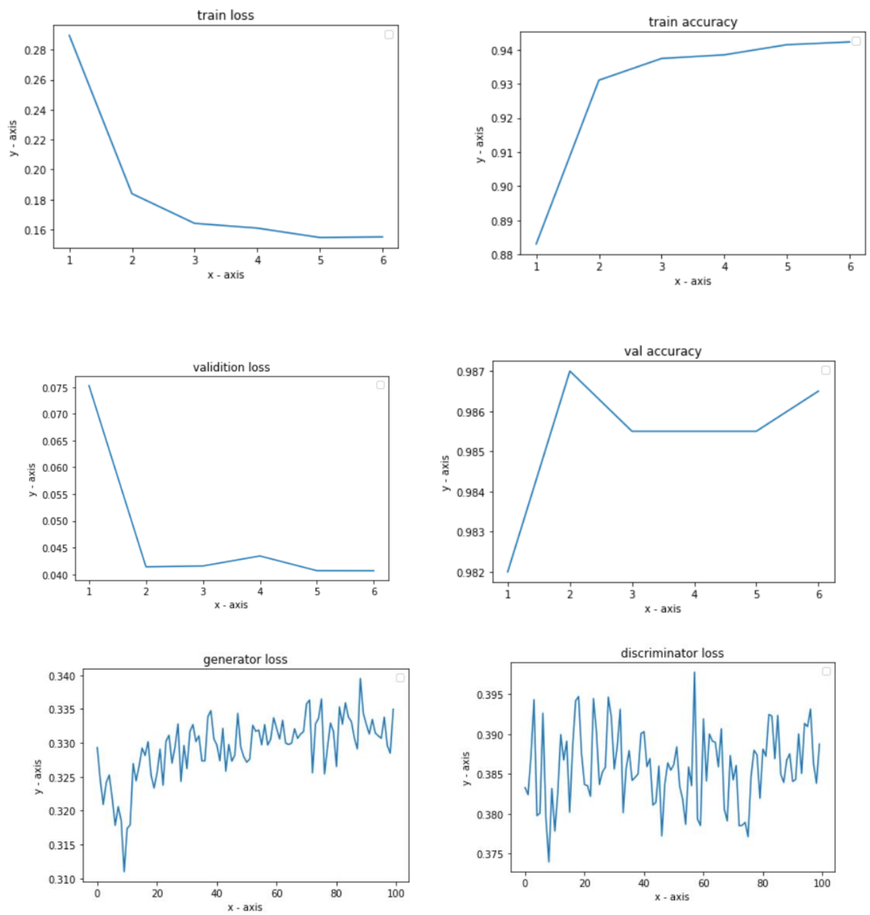


Fig. 7. Accuracy of Training and Validation.

5. Conclusion and future directions

The main intention of this work is to detect the cracks using the image processing techniques. Firstly, the dataset which is utilized were analyzed and concluded that most of the system uses real data sets for convenience as well as efficiency. Then the analysis is done based on the accuracy level. Then the resolution of the crack and non-crack images is increased by Super Resolution Generative Adversarial Networks. Then the transformation of the images to improve accuracy is done by the image augmentation. Then the cracks are differentiated in the images by the resnet50 model.

References

[1] Li S, Zhao X. Image-Based Concrete Crack Detection Using Convolutional Neural Network and Exhaustive Search Technique. Adv Civ Eng 2019;2019:1–12. doi:10.1155/2019/6520620.

[2] Yang X, Li H, Yu Y, Luo X, Huang T, Yang X. Automatic Pixel-Level Crack Detection and Measurement Using Fully Convolutional Network. Comput Civ Infrastruct Eng 2018;33:1090–109. doi:10.1111/mice.12412.

[3] Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. 2018 Int Interdiscip PhD Work, IEEE; 2018, p. 117–22.

[4] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. ArXiv Prepr ArXiv171204621 2017.

[5] Inoue H. Data augmentation by pairing samples for images classification. ArXiv Prepr ArXiv180102929 2018.

[6] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Adv Neural Inf Process Syst, 2014, p. 2672–80.

[7] Wang C, Xu C, Yao X, Tao D. Evolutionary Generative Adversarial Networks. IEEE Trans Evol Comput 2019;23:921–34. doi:10.1109/TEVC.2019.2895748.

[8] Yang W, Zhang X, Tian Y, Wang W, Xue J-H, Liao Q. Deep Learning for Single Image Super-Resolution: A Brief Review. IEEE Trans Multimed 2019;21:3106–21. doi:10.1109/TMM.2019.2919431.

[9] Hoang N-D. Detection of Surface Crack in Building Structures Using Image Processing Technique with an Improved Otsu Method for Image Thresholding. Adv Civ Eng 2018;2018:1–10. doi:10.1155/2018/3924120.

[10] Wang M, Chen Z, Wu QMJ, Jian M. Improved face super-resolution generative adversarial networks. Mach Vis Appl 2020;31:22. doi:10.1007/s00138-020-01073-6.

[11] Zhang T. Research and Improvement of Single Image Super-Resolution Based on Generative Adversarial Network. J Phys Conf Ser, vol. 1237, IOP Publishing; 2019, p. 32046.

[12] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv Prepr ArXiv14091556 2014.

[13] Mikami H, Suganuma H, Tanaka Y, Kageyama Y. Imagenet/resnet-50 training in 224 seconds. ArXiv Prepr ArXiv181105233 2018:1–8.

[14] Dhankhar P. ResNet-50 and VGG-16 for recognizing Facial Emotions. Int J Innov Eng Technol 2019;13:126–30.

[15] Panella F, Boehm J, Loo Y, Kaushik A, Gonzalez D. Deep learning and image processing for automated crack detection and defect measurement in underground structures. ISPRS - Int Arch Photogramm Remote Sens Spat Inf Sci 2018;XLII–2:829–35. doi:10.5194/isprs-archives-XLII-2-829-2018.

[16] Nandepu R. Understanding and implementation of Residual Networks (ResNets) 2019. https://medium.com/analytics-vidhya/understanding-and-implementation-of-residual-networks-resnets-b80f9a507b9c (accessed October 31, 2019).

[17] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conf Comput Vis Pattern Recognit, IEEE; 2016, p. 770–8. doi:10.1109/CVPR.2016.90.

[18] Zhang K, Cheng H-D, Gai S. Efficient Dense-Dilation Network for Pavement Cracks Detection with Large Input Image Size. 2018 21st Int Conf Intell Transp Syst, IEEE; 2018, p. 884–9. doi:10.1109/ITSC.2018.8569958.

[19] Zhang K, Cheng HD, Zhang B. Unified Approach to Pavement Crack and Sealed Crack Detection Using Preclassification Based on Transfer Learning. J Comput Civ Eng 2018;32:04018001. doi:10.1061/(ASCE)CP.1943-5487.0000736.

[20] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. Proc IEEE Conf Comput Vis pattern Recognit, 2017, p. 4681–90.

[21] Zhu X, Zhang L, Zhang L, Liu X, Shen Y, Zhao S. GAN-Based Image Super-Resolution with a Novel Quality Loss. Math Probl Eng 2020;2020:1–12. doi:10.1155/2020/5217429.

[22] Cha Y-J, Choi W, Büyüköztürk O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Comput Civ Infrastruct Eng 2017;32:361–78. doi:10.1111/mice.12263.

[23] Sultana F, Sufian A, Dutta P. Advancements in Image Classification using Convolutional Neural Network. 2018 Fourth Int Conf Res Comput Intell Commun Networks, IEEE; 2018, p. 122–9. doi:10.1109/ICRCICN.2018.8718718.

[24] Zhang L, Yang F, Daniel Zhang Y, Zhu YJ. Road crack detection using deep convolutional neural network. 2016 IEEE Int Conf Image Process, IEEE; 2016, p. 3708–12. doi:10.1109/ICIP.2016.7533052.