AUTOMATIC ASSESSMENT OF STRUCTURAL DAMAGE OF MASONRY STRUCTURES BY VISUAL ANALYSIS OF SURFACE CRACKS

by

Nhan Trong Nguyen

HONORS THESIS

Submitted to Texas State University
in partial fulfillment
of the requirements for
graduation in the Honors College
December 2021

Thesis Supervisor:

Vangelis Metsis

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

I would like to thank all who have helped me through this process, I could never have completed this thesis without the unrelenting support of those around me.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

ABSTRACT

Crack detection on the road or building surface is normally done using manual inspection by specialists. The process consumes a lot of time, and the inspection result might differ depending on the specialist's experience and knowledge. This work will propose an automated detection and rating of cracks on concrete surfaces based on convolutional neural networks (CNNs). Our method also provides a visualization of how the model learns the crack by directing the attention of the model to the different parts of the image by utilizing Gradient-weighted Class Activation Mapping (grad-cam) library. Finally, we show how combining two different data types, such as raw images and manually extracted features, into a hybrid convolutional neural network can increase the accuracy of the model.

I. INTRODUCTION

Cracking is a common phenomenon that is no stranger to structures such as bridges, ports, roads, houses, and public works. Over time, if these cracks are left untreated, they are likely to develop with unpredictable consequences. This is because the cracks are likely to become more widespread over time and can make the house more and more subsidence, so it directly affects the structure of the house leading to loss of aesthetics, but it can also be life-threatening if the building collapses. Moreover, it also affects the durability and longevity of the structure of infrastructure works.

In recent years, there has been significant research promoting image processing-based algorithms for semi-automatic or automatic crack detection modes to solve this problem. These techniques are also used to detect cracks in concrete, potholes, and crevices in the pavement. The main advantage of the image processing-based crack detection technique is that not only can it be used to detect cracks from captured images, but it can also measure the width and direction of the cracks. Crack detection methods often use histogram-based matching techniques, edge detection techniques, etc. The input images are pre-processed and enhanced by image processing and then segmentation, followed by a local binary coding method used to measure the histogram of each foreground and background image to classify cracks and non-cracks in the input image. However, the robustness of these image processing techniques is poor as it depends on the quality of the captured image, which is often affected by several factors such as low light, shadows, rusty surfaces, etc., in real situations. To improve the accuracy of the image-based crack detection system and make it fully automated, machine learning (ML)-based techniques have been proposed by different authors. ML-based techniques first extract crack features from input images captured by image processing and then classify them into images with or without cracks based on the parts. Points are extracted. Artificial Neural Networks (ANNs) detect concrete cracks, pavements cracks, wall cracks, and other structural cracks/damages. However, the success of machine learning-based crack detection techniques is highly dependent on the accurate representation of cracks by hand-selected features for mining. Therefore, the performance of these methods depends on the extracted crack features, and often, the results of these methods are inevitably affected by the extraction of false features.

In this work, we experiment with a novel approach for crack detection using a Convolutional Neural Network with Visual Attention to automatically localize the crack and produce an estimate of the structural damage as measured by the "Max drift" metric, commonly used in the civil engineering domain. The accuracy of the model is evaluated by performing a cross-validation experiment on a dataset previously labeled by human experts. The mean absolute error (MAE) between the real and predicted drift values is estimated and plotted. Finally, to further improve the accuracy of the model, we develop a hybrid neural network that can accept both raw images and features extracted manually from the images, such as the length, width, angle of the crack, etc. Our experimental results show that such a hybrid model can achieve a lower error rate. The extraction of features used in the model can be automated in future work.

II. RELATED WORK

This section discusses the most relevant literature related to our problem. In particular, we will present two previous works that focus on the problem of crack analysis on masonry structures.

The first is “Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning” by Dimitris Dais. This work assessed and identified cracks based on photographs, using simple CNC and manual polishing of the cracks. This project has potential but marking each image will take much time in the information processing stage. Moreover, the result is a general assessment of the crack’s appearance on the available pictures [6].

The second related project is “Detection of Surface Cracks in Concrete Structures using Deep Learning” by Priya Dwivedi. The project also has the topic of predicting whether there are cracks in the wall and dividing the wall into parts and evaluating crack marks on that wall. This project has potential and is easy to implement but showing whether a wall has cracks is a simple outcome that does not have too much practical application [10]. This project will take a new approach. That is to build a CNN regression model with the output that will be a number to evaluate the damage level of the wall. This number is used to evaluate a lot in the working environment of road and bridge engineers, so the practical application of the model will be enhanced.

III. METHODOLOGY

Convolutional Neural Networks (CNNs) have shown great success in various image analysis tasks. To solve the problem of estimating structural damage based on surface cracks, a good solution is choosing to focus on the crack learning process in the wall using a CNN implemented by TensorFlow library. TensorFlow is one of the most popular deep learning packages today.   With excellent classes and functions in image processing. Besides the flexibility in choosing different variables to get the best results, the layers can be combined with other functions to form a model with higher accuracy. Therefore, Tensor Flow will be the first choice in the construction process for this training process. The process of doing this project is divided into several phases:

3.1 Learn about CNN and image processing

In the process, I began to learn and build my knowledge to better understand the tools I will use to solve the problems in this essay. Since then, when I have problems during the test, it will help me solve the problem faster. Convolutional Neural Network (CNN) is a Deep Learning algorithm that can take in an input image, assign importance to various aspects/objects in the image, and differentiate one from the other.

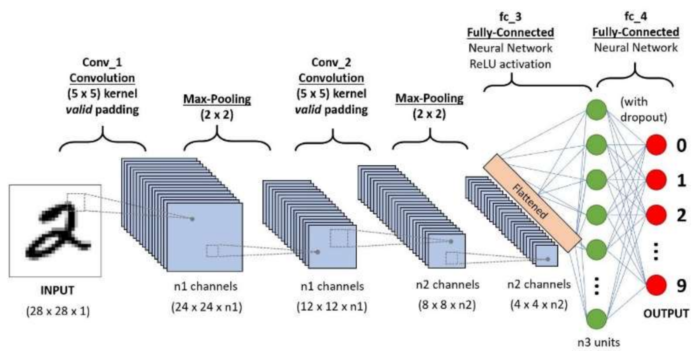
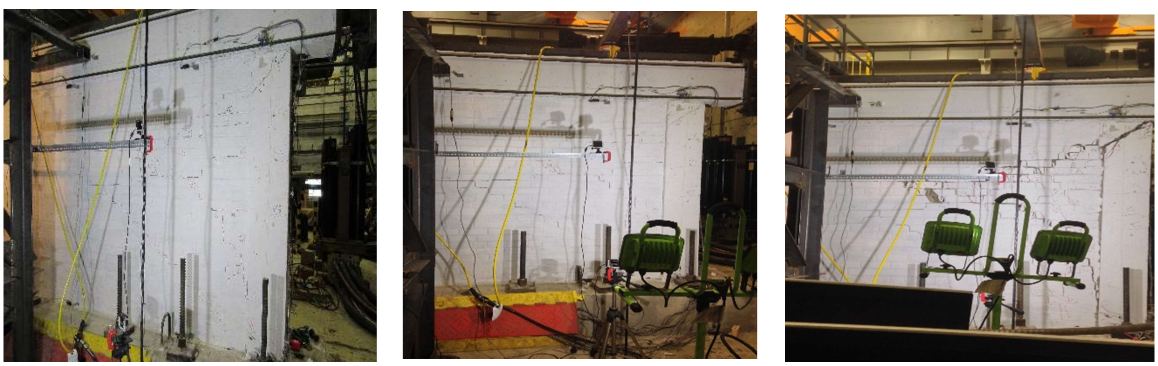


Figure 1: How Convolutional Neural Network works

3.2 Prepare data

The raw data are photos taken during the simulated earthquake in a controlled environment (lab). The data includes 117 photos divided into nineteen groups based on the experiment's drift level. Drift is a term used in civil engineering. It measures the relative translational displacement difference between two consecutive floors. A datasheet contains the image's name, path, the drift score with each image, and value to determine the location of the crack by binary value. Given this amount of data, I would judge it to be a relatively small amount of data. The shooting angle and size between the pictures are different. In addition, the level of unnecessary objects in the photo is also quite a lot. Therefore, using this image source directly for the CNN model will cause a lot of noise to the model. Therefore, the first solution I choose is to process the image before putting it into the model.

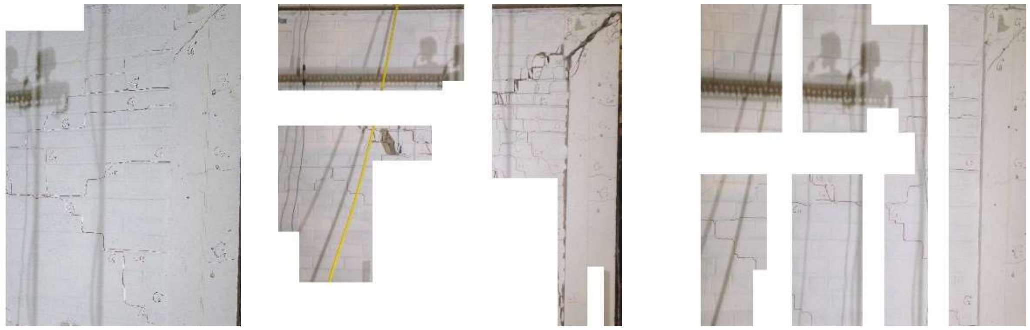


(a) (b) (c)

Figure 2: Raw image before pre-processing

3.3 Data pre-processing

The processing step includes removing blurry photos, cropping, and focusing on the walls and cracks, adjusting the photos so that they can be roughly the same size, masking unnecessary and distracting objects in the image to ensure that the training process will focus on the wall and cracks in the image. The image processing does not stop here. After each training with the model, depending on the results, I will continue to process the image to improve the model's prediction results.



(a) (b) (c)

Figure 3: Raw image after pre-processing

3.4 Build a simple test model

After consolidating my knowledge and preparing data, I started to build a simple CNN model. To evaluate data usage and the next steps to take. In the first model, I use the split train-test to split the data into two training and testing groups. The model consists of Conv2D, MaxPool2D, and three Dense layers ending with linear Dense to form a CNN regression model. The input of the model includes the image, and the label is the Drift score. The output of the model will be the prediction of the Drift of the image.

3.5 Building a cross-validation model to assess the model's reliability

Cross-validation is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand and implement. It results in skill estimates that generally have a lower bias than other methods [11].

After a preliminary evaluation of the model, I started making tweaks to the data to make sure the data was usable with the least amount of noise. Then apply a cross-validation model as a tool to evaluate the model's accuracy from which to make necessary modifications to the model.

3.6 Build more attention model and grad-cam

Attention models, or attention mechanisms, are input processing techniques for neural networks that allow the network to focus on specific aspects of a complex input, one at a time. The purpose of using attention mechanisms is to break down complicated tasks into smaller areas, helping the model can focus on the crack more.

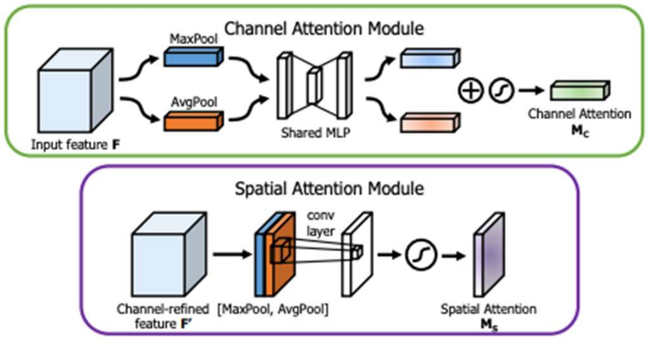


Figure 4: Model shows how channel attention and spatial attention work

Gradient-weighted Class Activation Mapping (Grad-CAM) is a technique for visualizing and making more transparent decisions from a large class of CNN-based models. The purpose of using Grad-cam is to show which part of the images is noticed by the model and weigh it, so that I can tell if the model is learning from the correct parts of the image.

In this step, the model consists of Conv2D, MaxPool2D, Channel attention, Spatial attention, and three Dense layers ending with linear Dense to form a CNN regression model. The input will include the image and the Drift score. The output will also be the predicted Drift score of the image. At the same time, I will combine with cross-validation to evaluate the completeness of the model, and I will also adjust the parameters to get the best model applicable to the next step.

3.7 Simultaneously combine verified data and images to build a model and evaluate the reliability of the model again.

To increase the accuracy of the model, I chose to combine the images and collected data together. I will build two different training models. Then combine the output of both and a new model and train the model one more time. Thereby the results of the two models will complement each other to increase the accuracy of the total model. This model will be a combination of two different models. The first is the multilayer perceptron (mlp) model, I started to process the data and build the best model to combine with the CNN model. The second model is the CNN model combined with the attention layer. The two models will receive input to the last layer of the two models that will be combined in a common model. The given output will also predict the Drift score of the image. Besides, I will combine with cross validation to evaluate the reliability of the model to ensure the best results.

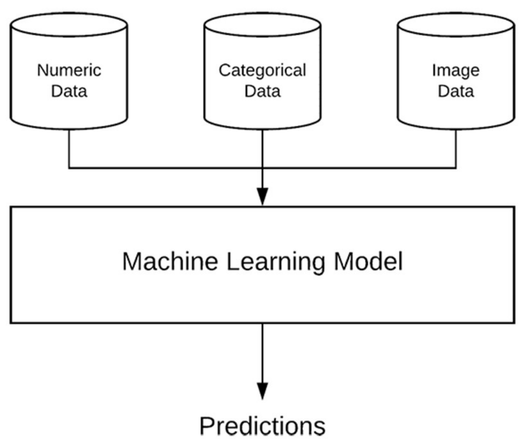


Figure 5: Model multiply input and mixed data

IV. EXPERIMENTAL RESULTS

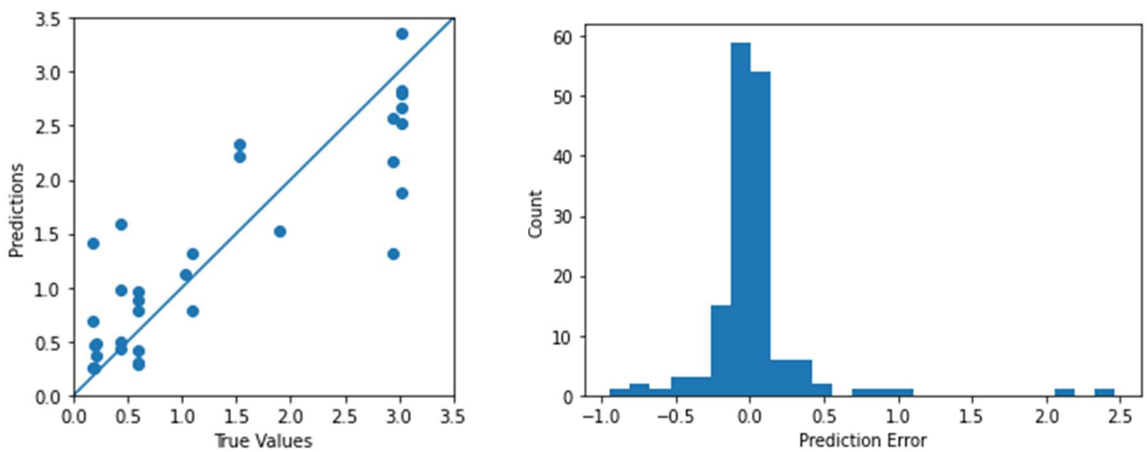
The results of the experiment will be evaluated on three factors

- The factor that ensures that the model correctly represents the regression training is shown in the first diagram. The closer the predictions are to the straight line, the higher the accuracy of the model is to be training.

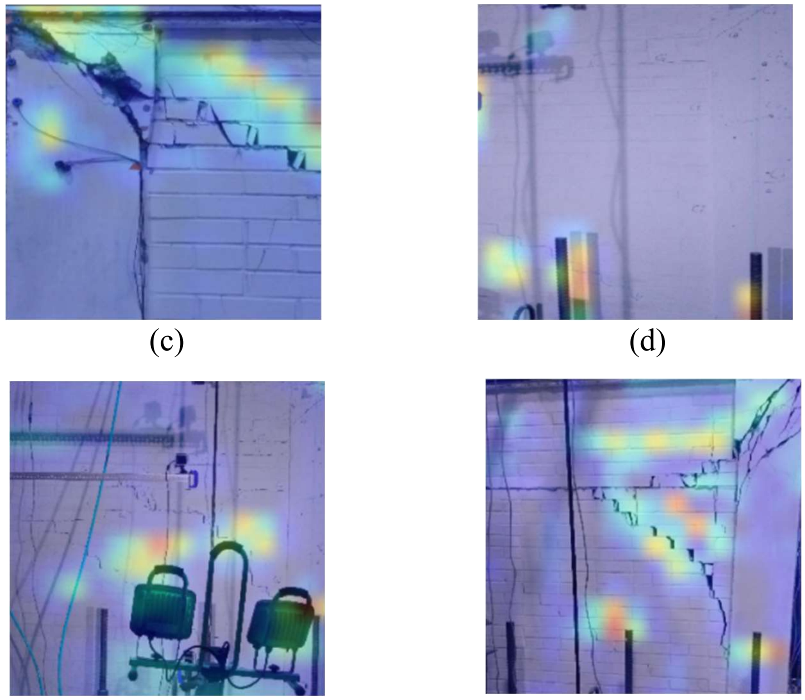
- The factor that represents the accuracy of the model compared to the actual results. The more accurate the result, the higher the count of values close to zero.

- The element of expression is based on the emphasis that the model gives to the photo. If the crack has a red color, it means that the model has learned and correctly evaluated the crack

4.1 Result of simple model



(a) (b)



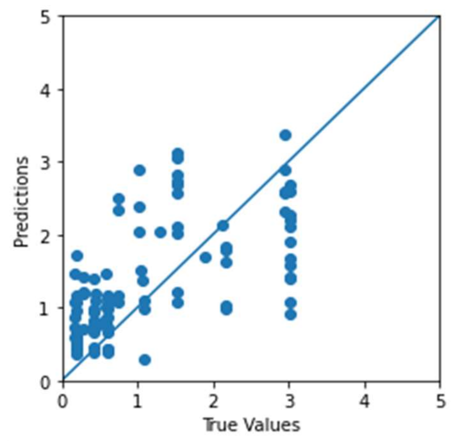
(c) (d)

(e) (f)

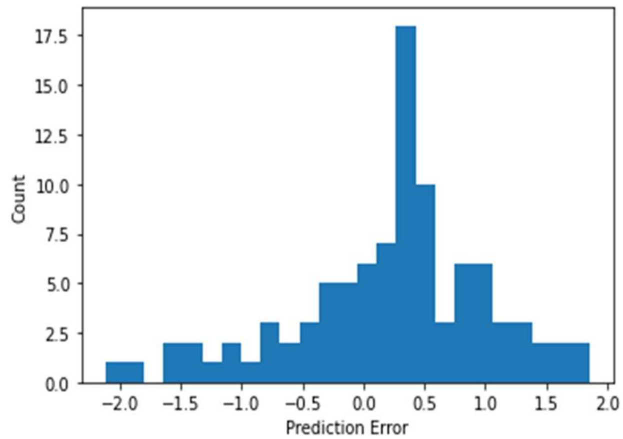
Figure 6: Graph and Grad cam of simple CNN model

Although the results obtained are very good when predicting the results of the model has quite good results. But when looking at the image from grad-cam, it can be concluded that the model does not learn much from the cracks and only makes a general assessment from the image and the noise information on the image. Therefore, the accuracy of the results is not reliable in practice. Based on that, I started to adjust in the image to make sure to remove all the unnecessary objects as well as adjust the angle of the pictures to be as close as possible.

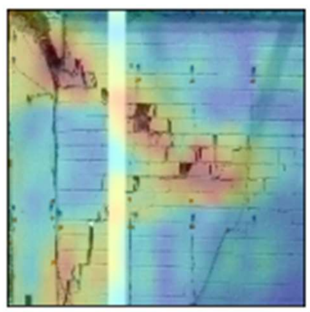
4.2 Result after image processing and attention model



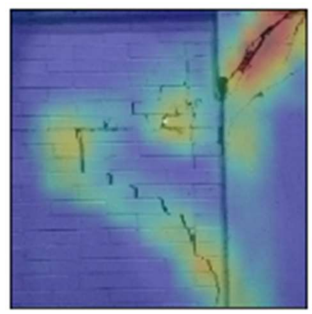
(a)



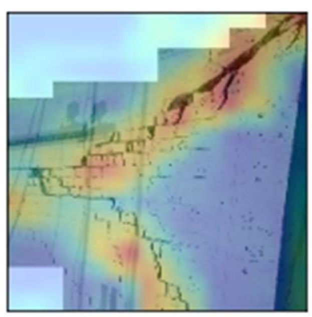
(b)



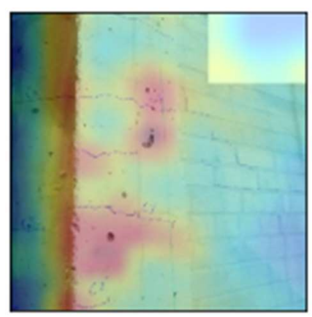
(c)



(d)



(e)



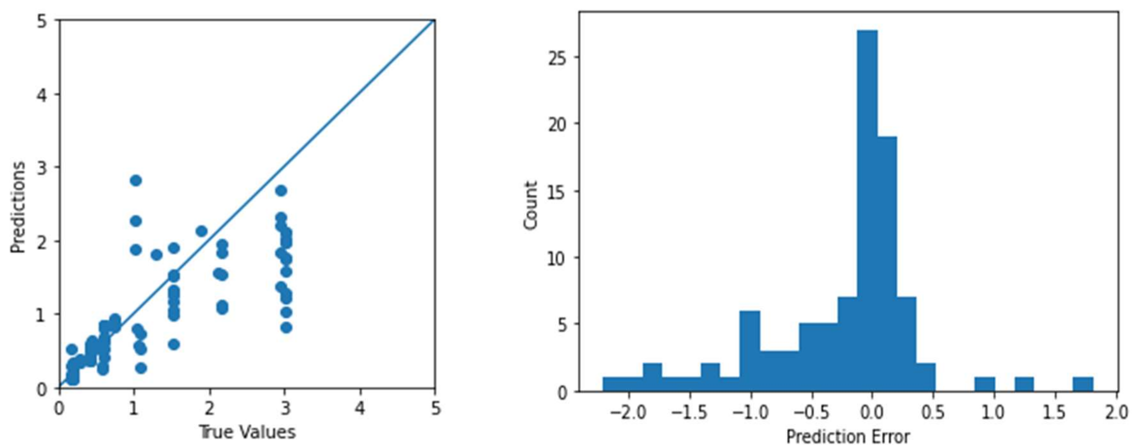
(f)

Figure 7: Graph and Grad cam of CNN model combine attention

The model detected the crack much better than the first simple model. This proves that the attention classes have performed well in the role of recognizing the position that needs to be evaluated in the images. But even so, the results of the data evaluation of the photos have quite large errors. Therefore, it shows that the model still needs to be improved on the data side to ensure more accuracy of the model. (Fig.7)

4.3 Result of multilayer perceptron model

The results to evaluate the reliability of the model only use pure data before combining with the image model. The results show that the model has high accuracy and can be used. (Fig.8)

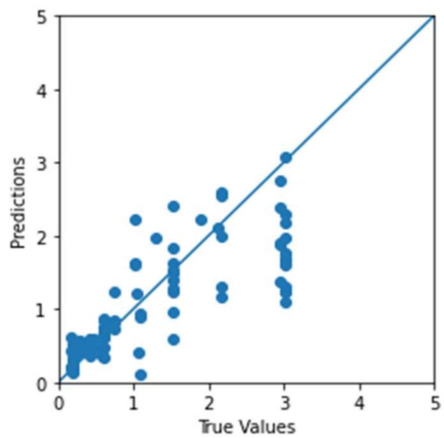


(a) (b)

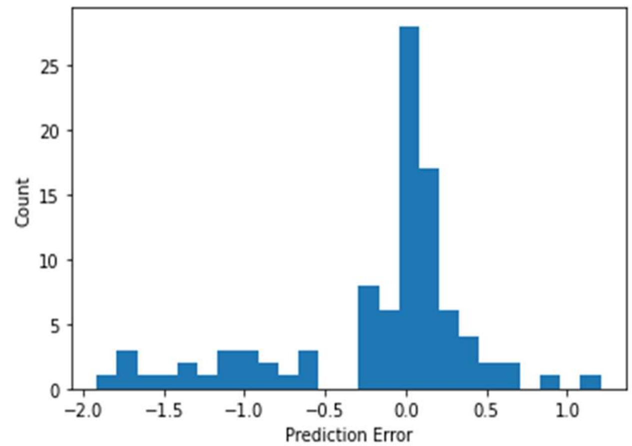
Figure 8: Graph of data structure mode

4.4 Result Multiple Inputs and Mixed Data

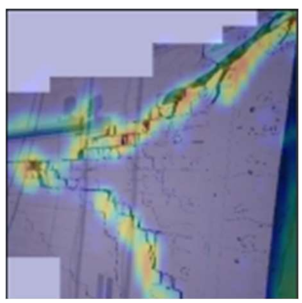
The numerical results of this model obtained are the best of all three models, which proves that the combination of images and data provides more accurate assessment prediction. Although in terms of images, the Multiple Inputs and Mixed Data model with focus on cracks is not as good as the pure CNN model. That can be explained because it is a hybrid model. If the multilayer perceptron model predicts more accurate results, the aggregate model will tend to use that model’s weight, leaving less emphasis on the CNN model. (Fig.9)



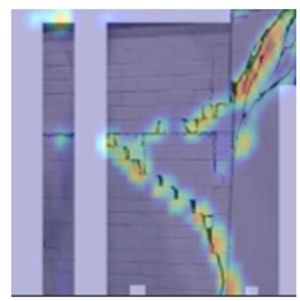
(a)



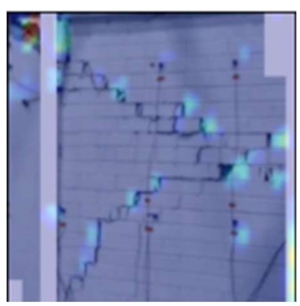
(b)



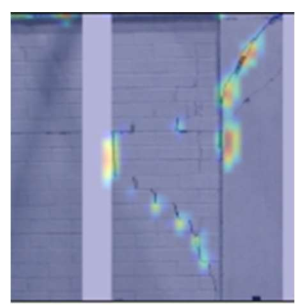
(c)



(d)



(e)



(f)

Figure 9: Graph and Grad cam of Multiple Inputs and Mixed Data

4.5 Table mean absolute error of all models:

Because this is a regression model. The accuracy of each model also will base on MAE (mean absolute error).

Model	Mean absolute error
Simple CNN	0.4729

CNN with Attention	0.6764
Structure data	0.4235
Multiple Inputs and Mixed Data	0.3919

Table 1. Mean absolute error of all models

V. CONCLUSIONS AND FUTURE WORK

Due to the relatively small amount of data and the different shooting angles between the images, the accuracy is not high. Irrelevant objects in the pictures also add significant noise in the training process of the model. Nevertheless, the obtained results show that the model has potential for future development. CNNs with visual attention can be a good choice for crack detection and assessment. The combination of images and other features extracted from the data will be a more suitable choice for more accurate rating results.

Future work can experiment with a fixed photo angle to reduce variability between data samples. It is expected that with such an approach, the model's learning will be improved, and the accuracy will be further improved. Another problem that can be improved is increasing the amount of data. With more images, the accuracy of the model will be improved.

REFERENCES

[1] Build, train and evaluate models with tensorflow decision forests.

[2] Grad-cam: Visual explanations from deep networks via gradient-based localization.

[3] Scipy lecture notes.11

[4] Keras imagedatagenerator for image augmentation: Python use case, Aug2020.

[5] K-fold cross-validation with tensorflow keras, Aug 2021.

[6] Ryan Allred. Image augmentation for deep learning using keras and his-togram equalization, Jul 2018.

[7] Jason Brownlee. A gentle introduction to k-fold cross-validation, Aug 2020.

[8] Francois Chollet. The keras blog.

[9] Chris, Devidas, Keerti, Josseline, Rebeen Ali, Rebeen, John, Ming, Zay,Laura, and et al. K-fold cross validation with tensorflow and keras, Mar2021.

[10] Dimitris Dais, İhsan Engin Bal, Eleni Smyrou, and Vasilis Sarhosis. Au-tomatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning, Feb 2021.

[11] Priya Dwivedi. Detection of surface cracks in concrete structures using deep learning, Dec 2019.

[12] Vladimir Lyashenko Young AI enthusiast who is passionate about EdTech, Computer Vision in medicine. I want to make the world a better place by helping other people to study, Vladimir Lyashenko, Young AI enthusiast who is passionate about EdTech, Computer Vision in medicine. I want to make the world a better place by helping other people to study and follow me on. Cross-validation in machine learning:

How to do it right, Jul 2021.

[13] Laura Lewis. Building a mixed-data neural network in keras, Sep 2021.

[14] Mid, Adrian Rosebrock, Anas Bin Iftikhar, Chakib BELAFDIL, Luke, Abu-Abdurrahman, Gary Cao, Sri, Steve Frank, El Jefe, et al. Grad-cam: Visualize class activation maps with keras, TensorFlow, and deep learning, Apr2021.

[15] Victor Perez. Transformers in computer vision: Farewell convolutions! Dec2020.

[16] Rich, Adrian Rosebrock, Hendrick, Khine, Riyaz, Arslan, Laxmi, Yash Rathod, Henry, Henrik T., et al. Keras: Multiple inputs and mixed data, Jul 2021.

[17] Sumit Saha. A comprehensive guide to convolutional neural networks-the eli5way, Dec 2018