International Journal of GEOMATE, Nov., 2018 Vol.15, Issue 51, pp.240-251
Geotec., Const. Mat. & Env., DOI: https://doi.org/10.21660/2018.51.35376
ISSN: 2186-2982 (Print), 2186-2990 (Online), Japan

# CRACK DETECTION IN HISTORICAL STRUCTURES BASED ON CONVOLUTIONAL NEURAL NETWORK

*Krisada Chaiyasarn[1], Mayank Sharma[2], Luqman Ali[2], Wasif Khan[2] and Nakhon Poovarodom[1]

[1]Department of Civil Engineering, Thammasat University, Thailand
[2]Department of Electrical and Computer Engineering, Thammasat University, Thailand

*Corresponding Author, Received: 17 April 2018, Revised: 8 June 2018, Accepted: 20 June 2018

**ABSTRACT:** Regular inspection and maintenance work is required to ensure the structural integrity of historic structures, especially the masonry structures which are deteriorating due to ageing and man-made activities. The structures are typically examined by visual inspection, which is a costly and laborious procedure, and often, the inspection results are subjective. In this study, an automatic image-based crack detection system using Convolutional Neural Network (CNN) for masonry structures is proposed to aid the inspection procedure. Previous crack detection systems generally involve handcrafted features, which are then classified by classification algorithms. This approach relies heavily on feature extraction stage, which may not offer accurate results as some hidden features may not be extracted. In this paper, the feature extraction process is done by CNN from RGB images, and then the softmax layer is replaced by other classifiers to improve classification accuracy. Three classifiers are studied, namely the CNN itself, Support Vector Machines (SVM) and Random Forest (RF). A dataset containing images of cracks from masonry structures was created using a digital camera and an unmanned aerial vehicle. The images were used in training and validating the proposed system. The collected images were also used to build a 3D model using the technique based on Structure from Motion (SFM), which allowed images containing cracks to be located in 3D world coordinates. It was found that the combined CNN and SVM model performs the best among other methods with the detection accuracy of approximately 86% in the validation stages and 74% in the testing stage. As shown in this paper, the integration of CNN and other classifiers can improve detection accuracy. In addition, it was shown that the system can be used to detect cracks automatically for the images of masonry structures, which is useful for the inspection of heritage structures.

*Keywords: Convolutional Neural Networks, Support Vector Machine, Random Forest, Crack Detection, Masonry Structures.*

## 1. INTRODUCTION

Historic structures are vital to Thailand's tourist industry as they are the country's cultural heritage. With the threat of changing the environment (e.g. flooding) and unforeseen natural disasters (e.g. earthquake), the country requires expertise and research to establish ways to properly maintain historical structures. A number of heritage structures in Thailand urgently need constant inspection and monitoring as they are currently being exposed to damages from environmental changes, such as flooding. Some damages can be found in Wat Chai Wattanaram (shown in Fig. 1(a)), a temple dated back to 16th century in a historic province of Ayutthaya, a former capital of Thailand. Fig. 1(b) and (c) exemplified damages commonly found in the masonry structures in Ayutthaya. Some temples have been deteriorating due to natural causes, such as ground subsidence, or manmade activities such as vibration from the nearby roads. Monitoring and inspection of heritage structures in Thailand are in urgent need as it requires a systematic approach to maintain the structures. Failure in doing so can lead to the loss of these important heritage properties forever, which will result in the loss of Thailand's culture and its national identity.

Visual inspection is a common procedure to examine and assess the current state of historical buildings. However, this procedure is laborious and time-consuming as it normally involved inspectors traveling to interesting sites to assess the structures conditions based on the visual appearance of structures. Hence, the process cannot be conducted frequently due to high labor cost, and it is also prone to human-error. Often, sites cannot be inspected due to inaccessibility, especially the temples in Ayutthaya, where the tops of many stupas are extremely high and cannot be easily accessed for inspection. Failure to detect problems can lead to disastrous effects such as temple collapse. Many temples in Ayuthaya are visibly tilting, possibly due to the settlement from the flood. The tilt angle of Wat Yai Chai Mongkol was studied in [5] and these structures require frequent monitoring and inspection. In this paper, an automatic image-based crack detection system based on the convolutional neural network is proposed for the inspection of masonry structures.

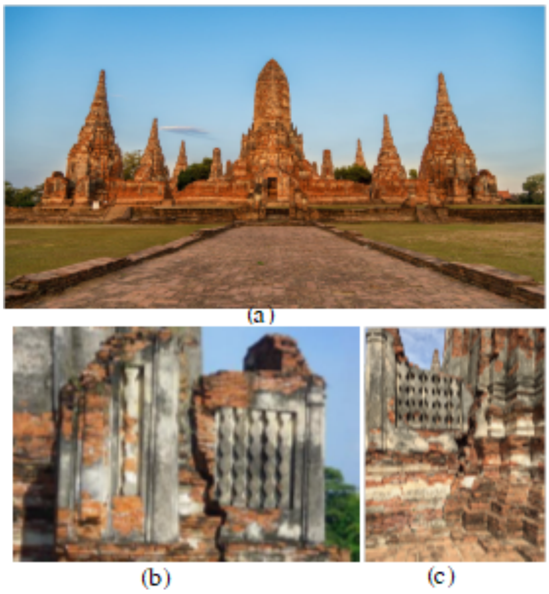


Fig. 1 (a) Wat Chai Wattanaram, a temple located in the historical park in Ayutthaya, Thailand, (b) and (c) Example images of cracks found around Wat Chai Wattanaram.

Generally, crack detection is done by either technique related to feature extraction or by automatic feature learning, i.e. a deep learning, which is a technique proposed in this paper. The technique based on feature extraction generally consists of two main steps, feature extraction, and classification. In the feature extraction step, various types of crack features are extracted, such as edges [1], percolation-based features [31,32] or multi-features [24], and then classifiers, including Support Vector Machines (SVM) [7] and Neural Networks [29] are used in classifying the extracted features. This approach, however, can fail as some hidden crack features may not be extracted due to their complexity. Therefore, deep learning technique is a better approach in the crack detection problem as crack features are learned automatically from raw images, and complex features can be built from low-level features. Deep learning has been applied to many problems as it proves to be a better technique in the classification task [6,27].

In this paper, an image-based system to inspect masonry structures is proposed. The system integrates Convolutional Neural Network (CNN) with a classifier to improve classification accuracy. Three types of classifiers are explored, CNN, Support Vector Machine (SVM) and Random Forest (RF). The image data was collected using an Unmanned Aerial Vehicle (UAV) and a handheld DSLR camera from various locations around historical temples in Ayutthaya. The images were also used in creating a 3D model so that the locations of detected cracks can be located in 3D space.

The contribution of this paper is two-fold. Firstly, an automatic inspection system for historical structures is demonstrated. Secondly, a CNN-based system is shown to be a good technique for detecting cracks in masonry structures. The rest of the paper is organized as follows, Section 2 presents literature review about automatic crack detection. Section 3 and 4 describe the methodology, and Section 5 shows the experiments of the proposed system and image-based 3D modeling. Discussion and conclusion are drawn in Section 6 and Section 7, respectively.

## 2. LITERATURE REVIEW

### 2.1 Inspection of heritage structures

Non-invasive inspection is required when assessing damage in vulnerable historic buildings to prevent further damage [27]. Fregonese et. al. [18] applied terrestrial laser scanner (TLS) to monitor out-of-plane displacement of an ancient building by registering two sets of laser scan data to geo-referenced control points. Mogahed et. al. [20] transformed control points from a laser scan data to a total station data using a series of transformation to find a discrepancy between two sets of data and displacements for heritage structures. Tapete et. al. [8] integrated ground-based synthetic aperture radar interferometry (GBInSAR) to detect deformation of objects from SAR images. Armesto et. al. [15] applied TLS to a masonry bridge and estimated the bridge deformation an algorithm based on the arch symmetry. Bhakapong et. al. [5] applied photogrammetry technique to compare the cross-sectional profile of temples in order to assess the building inclination. Costanzo et. al. [3] presented a methodology that combined TLS and infrared thermal images for inspecting St. Augustine Monumental in Calabria (South Italy). Achille et. al. [2] demonstrated the application of photogrammetry using unmanned aerial vehicle (UAV) to survey the historical structure “Santa Barbara” bell tower in mantua (Italy). Pesci et. al. [4] applied TLS with digital images to detect the trace of restoration in the ancient part of palazzo d’Accursio in Bologna, Italy.

Image-based photogrammetry techniques were used to create 3D models of historical buildings. The techniques rely on automatic control points and Structure from Motion (SfM) allowing 3D models to be created from images taken with arbitrary motions. Bhakapong et. al. [5] applied an image-based technique to estimate the tilt angle of Wat Yai Chai Monkol from 3D point cloud.

### 2.2 Autonomous crack detection system

Visual inspection is a common procedure for examining the current conditions of structural components. The guideline from Federal highway administration and federal transit administration [13] suggests that inspection should be carried out within an inspector arm's length, which can be impossible in inaccessible areas. Crack detection is the first stage in inspection, and once cracks are found, they will require systematic monitoring. Current visual inspection procedures have a number of issues, including high labor costs, time-consuming and subjectivity. Visual inspection can be highly subjective as it relies on inspectors' experience. Crack detection remains one of the most difficult tasks, which is a major concern in inspection as discussed in [33].

The robotic system can be used to automatically acquire images as shown in [26], who used a robotic system for bridge inspection. The system is pre-programmed to navigate around a pavement to detect cracks using automatic crack detection algorithm. Loupos et. al. [16] built an intelligent robotic system for tunnel inspection. The system incorporated many sensors, including laser scanner, infrared and vision camera, to improve the robot's navigation and detection system although the system was still not completely autonomous.

Unmanned Aerial Vehicles (UAVs) are ideal tools in applications that require rapid and effective views of the site, for example, in archaeological sites or in a post-earthquake zone [19]. The use of image-based techniques with UAVs for the condition assessment of infrastructure, particularly crack detection is exemplified in [27]. In his work, a UAV was used for image acquisition, then image processing algorithm was applied to inspect cracks across the building surface. Similarly, Pereira et. al. [11] applied embedded an image-based system for the automatic recognition of crack in building facades using UAVs.

Many automatic crack detection systems are based on extracting handcrafted features. Ellenberg et. al. [10] discussed several algorithms, including percolation approach, fractal method and tensor voting for crack detection. Abdel-Qader et. al. [1] applied four different edge detection techniques, i.e. Fast Haar Transform (FHT), Fast Fourier Transform, Sobel and Canny detectors for concrete bridges. The FHT was the best one among other detectors in the study. The limitation of edge detection algorithms is generally due to noise. Liu et. al. [17] applied image intensity features and Support Vector Machine (SVM) for tunnel crack detection. This method is prone to error due to noise. For different types of cracks and images containing noise, the techniques based on handcrafted features fail to perform. Hence, automatic feature extraction based on learning techniques such as deep learning performs well when compared to the techniques based on handcrafted features. Zhang et. al. [34] applied deep convolutional neural network for road crack detection from images collected using a low-cost smart phone. Cha et. al. [6] used the deep convolutional neural network (DCNN) for automatic concrete crack detection and presented 98% accuracy, which is much better and more accurate than techniques relying on handcrafted feature extraction techniques.

As observed in the previous literature, the research trend in the area of infrastructure inspection is to achieve a completely automated system based on automatic defects detection incorporated with an autonomous robotic system. However, the performance of these systems, such as accuracy, reliability, robustness and efficiency, still requires a great deal of improvement if they are to be adopted by the industry. A combination of the detection system, sensors and robotic systems are the key success and to also ensure that most defects are detected and monitored.

## 3. PROPOSED METHODOLOGY

The outline of the proposed system is shown in Fig. 2. The system consists of 3 modules as explained in detail below, (1) Image acquisition via a drone and DSLR camera (2) Crack detection using CNN, and (3) Image-based 3D modeling. The output from the system is a 3D model and its associated images containing cracks.

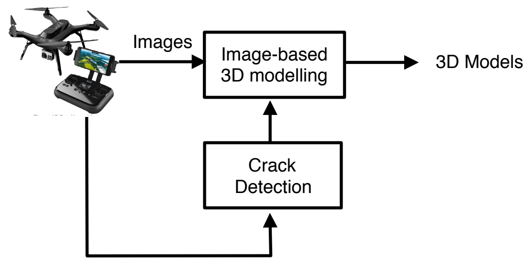


Fig. 2 Outline of the proposed system

### 3.1 Image acquisition

In recent years, UAVs have been utilized in surveying as an alternative to conventional surveying methods since they are faster, simpler and cheaper. In this research, a drone DJI Phantom 4 and DSLR was used to acquire all images, the specification of the drone's camera. In order to obtain the full coverage of a structure as well as achieving a detailed 3D model, two pre-planned flight path strategies have been adopted. Fig. 3(a) demonstrates the first strategy, the sweeping strategy, in which the drone flew in a zig-zag motion to sweep an entire area from a specified height from the ground level.

The drone was pre-programmed to take pictures every 2-3 seconds to ensure that an overlap between consecutive images is at least 50%. Fig. 3(b) shows the second strategy, the Point of Interest (POI) strategy, in which the drone flew around a fixated object in a circular motion. This strategy is suitable for capturing images of an individual building, and in this method, the drone's camera was programmed to fixate its viewing angle to the main pagoda located in the center of the temple.

The drone took pictures every 2-3 seconds when it moved around the temple at the pre-determined radius. Pictures were collected at two levels of pre-determined height, and the images at the ground level were collected manually by researchers who took pictures around the temple in approximate a circular path, similar to the drone motion. In this work, the drone flight paths were pre-programmed in an IOS application call Auto Flight Logic. The application requires input parameters, including altitude, radius, velocity, and camera viewing angles. It is important to note that the drone flight path should be planned carefully so that it can complete the tasks within the flight time of 28 minutes for a single battery pack. It is also worth mentioning that a 3D model created from images can be registered with GPS, unlike the models created from videos. In addition, some close-up images are manually collected by using DSLR camera so that crack features are more visible in images. These images can also be combined with the 3D model. The sample images collected using UAVs are shown in Fig. 4.

### 3.2 IMAGE-BASED 3D MODELING

Agisoft is an image-based 3D reconstruction system, which takes input as a set of images and the output will be a 3D point cloud of an interested scene or objects. The software is based on Structure from Motion (SFM) and interested readers can refer to [21] for more detail of the theory and technology.

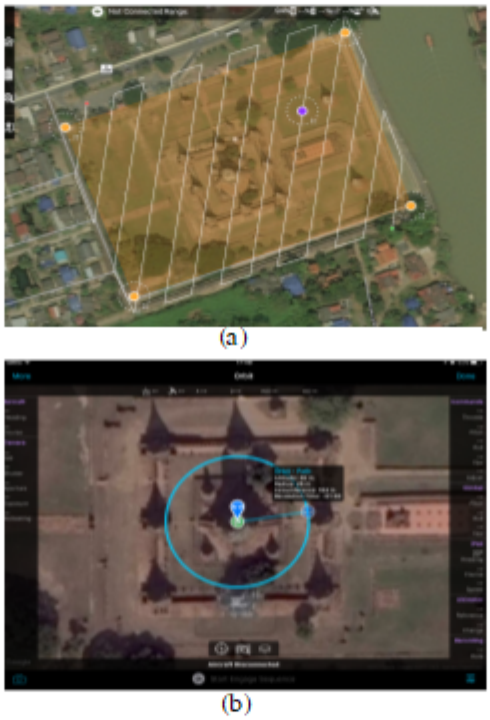


Fig. 3 Shows proposed pre-planned flight paths, (a) the sweeping strategy showing the drone flies in a zig-zag motion, and (b) circular motion or the POI strategy, where the drone flew around the Point of

Fig. 4 Sample images (a) acquired using a UAV (Ayutthaya Temple Thailand)

## 4. CRACK DETECTION

Fig. 5 shows an overview of a crack detection system. The detection system consists of two main modules, feature extraction, and classification. In this paper, feature extraction was conducted using the convolutional neural network (CNN). CNN is currently a state of the art for most vision problem and it is expected to outperform the traditional feature extraction technique. In the classification step, once features are extracted from the previous step, a classifier is used to determine whether the extracted features belong to a crack class. The classification process is also done by using three different classifiers in this study, a softmax layer in the CNN itself, Randomized Forest (RF) and Support Vector Machine (SVM) for comparison purpose. In this paper, there are two labels, crack and non-crack classes. As shown in the figure, the input to the system is an image patch, and the output is a label to determine if a patch belongs to as a crack or non-crack class.

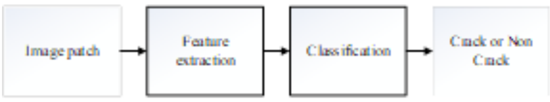


Fig. 5 shows an overview of a crack detection system

### 4.1 Feature Extraction

In this paper, CNN is used for feature extraction. Convolutional Neural Network (CNN) is a type of multilayer feed forward biologically inspired or an influenced variant of the artificial neural network, which has shown their significance in solving real-world problems. CNN architecture can be considered as a combination of a multi-level deep feature extractor and a classifier. As shown in Fig. 7, the feature extractor contains multiple layers, the input for the first layer comes in the form of three 2D arrays containing image pixel intensity values in the RGB color channels, and these layers retrieve the variance information at each level in the form of discriminative features. The extracted features in feature extractor are used to train the classifier. In this work, a softmax layer, which is the classification step in CNN, is modified by different classifiers. The major advantage of the Deep Convolutional Neural Network is that it does not require any intervention in the design of multiple stage layers. These multiple layers are learned from raw data using a learning procedure. The architecture of CNN consists of a series of layers. In our model, we used the Keras sequential model [25], for the CNN architecture, which is three-stage layers as shown in Fig. 6. The first few stages of CNN architecture consists of three types of layers, convolutional layers, activation layer (ReLU) and max-pooling layers.

As shown in Fig. 7, in the CNN architecture in the proposed work, each stage of convolutional layers consists of a convolution as a filter layer, non-linearity as an activation layer (ReLU) and max-pooling as a down-sampling layer. They are stacked together in each convolutional layer and fully connected layers are used in computing the class scores. In the proposed system, the output from a fully connected layer is the input as feature vectors for classifiers.

### 4.2 Classification

#### 4.2.1 Softmax classifier

The Softmax classifier provides normalized class probabilities, where the hinge loss is replaced with cross-entropy loss with

$$\boldsymbol{L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right)}$$

where $\boldsymbol{f_j}$ is the j-th element of the vector of class scores $\boldsymbol{f}$. The softmax function takes a vector of arbitrary real-valued scores and squashes it to a vector to values between 0 and 1. In our proposed system, the feature vector obtained from the fully connected layer is classified by the softmax classifier.

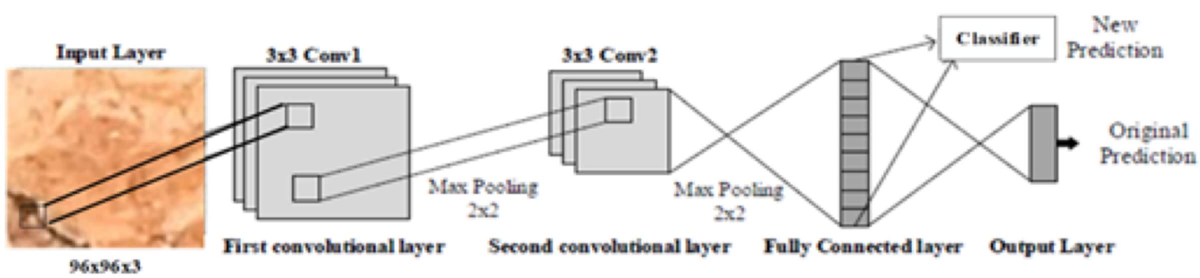


Fig. 6 The diagram shows the architecture of convolutional neural network in a feature extraction step. In the fully connected layers, extracted features are input to classifiers. In this project, the extracted features are classified by a classifier

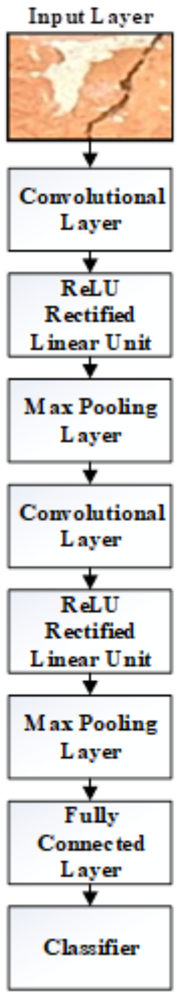


Fig. 7 The architecture of the convolutional neural network in the proposed work

### 4.2.2 Randomized Forest

A forest $T$ can be looked as a composition of decision trees $f$. Each decision tree $f_t(x)$ produces a prediction of a sample $x$, which is achieved by classifying a sample $x$ by recursively branching left or right down the tree until the last node of decision tthe ree is reached. Randomized Forests (RF) are a combination of bagging and random feature space algorithms. The random forests are supervised learning method for classification, which is applied to classify features into different class clusters so that the same features within the same class clusters are meant to have similar characteristic features in order to qualify to be in the same clusters as shown in Figure 8. And interestedthe reader can read [23] for more detail.

### 4.2.3 Support Vector Machine

The main objective in SVM is to find a hyperplane that separates the largest fraction of a labeled dataset for binary classification. The training data is a set of training samples pairs$\{(x_1, y_1), \dots, (x_i, y_i)\}$, where $x_i$the observation is or input feature for the $i^{th}$ sample and $y_i \in \{1, 0\}$ is the associated class label .The SVM classifier is the discriminant function that maps an input feature space $x_i$ into a class label $y_i$. An interested reader is referred to read [7] for the detail of Support Vector Machines.

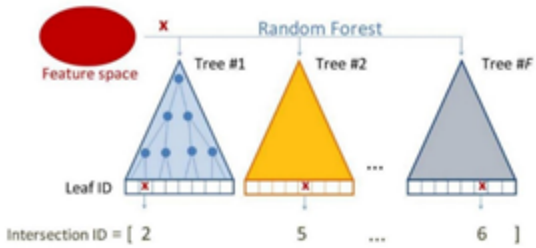


Fig. 8 shows an example of Random Forest classification (taken from [3])

## 5. EXPERIMENTS AND RESULTS

### 5.1 Image-based 3D Modelling

Fig. 9(a) shows a 3D model of the stupa with camera locations. This 3D model can be used as part of an inspection report. The crack detection system can be applied to images, then the locations of images containing cracks can be identified in this model. This enables inspectors to know the locations of cracks in a 3D sense, which is useful for inspection. Three types of collection strategies are applied to obtain images to demonstrate the use of the image-based 3D modeling technology, including sweeping strategy (see Figure 10(a)) and (b), POI strategy (see Figure 9(a) and (b)), and close-up strategy. The close-up strategy was done by taken pictures close to the surface of structures to allow more surface details. The model created from the sweeping strategy provides an overview of an entire site and the model from the POI strategy is used to obtain more information from each stupa in the site. The models from these two strategies can be combined but it is beyond the scope of the study.

### 5.2 Crack detection

#### 5.2.1. Training and Validation Data

To provide the robustness for the crack detection system, multiple sources of data were used. This includes images of masonry surface from various sites. Sample images of cracks on different types of surface enable the crack detection system

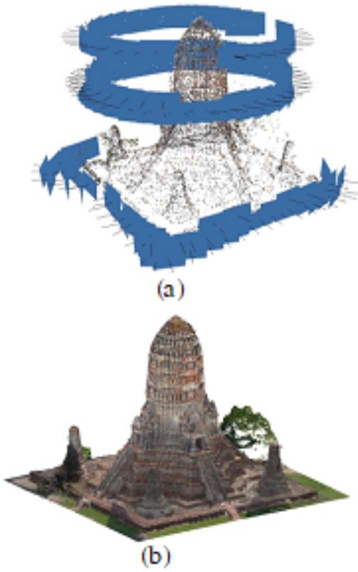

Fig. 9 shows the 3D point cloud created from the image-based 3D modelling techniques using the POI flight path strategy, (a) sparse point cloud, and (b) dense point cloud.

to be applied to many types of surface. A DSLR digital camera and images from the drone were employed to capture images near the surface of structures from various locations, mainly from around Ayutthaya historical park. The crack images of structure surface are collected based on their visual appearance. The system devised the classification framework by having two classes, crack and non-crack classes. Moreover, the database includes different types of commonly found cracks, i.e. longitudinal, transversal and others. Fig. 11 (a), (b) and (c) shows an example of crack patches on masonry surface and Fig. 11 (d), (e) and (f) are non-crack patches. The crack images of structure surface are collected based on their visual appearance. These images have been used as a database for a system. In the proposed system, image patches are used instead of an original size. For the evaluation of the proposed model, 6002 crack and non-crack image patches have been constructed to classify and detect the cracks in the concrete structure. The image patches have been split into three sets, i.e. training, validation and testing data, with a split ratio 6:2:2. 60% image samples were randomly selected for training database, 20% for validation database and 20% for the testing database. The number of cracks and non-crack image patches are set equally in all three datasets.

To train CNN, manual labeling operation was done on input images. For testing data, images that were not in the training and validation data were selected. For labeling, one is assigned to patches containing crack and zero for non-crack patches. The RGB values are used as features in CNN input.

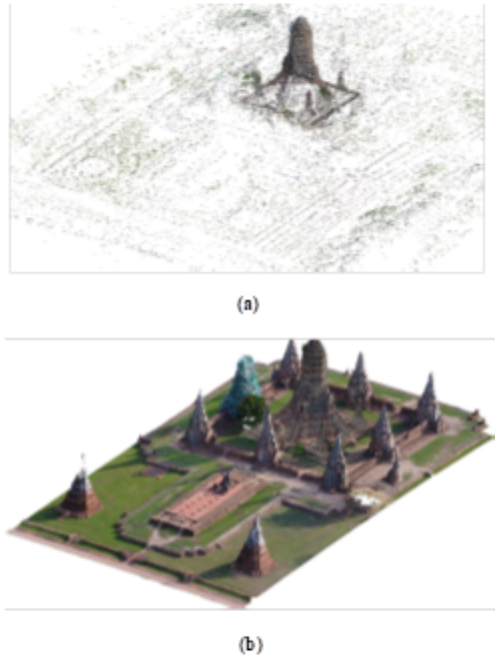

Fig. 10 shows the 3D point cloud created from the image-based 3D modeling techniques using the POI flight path strategy, (a) sparse point cloud, and (b) dense point cloud.

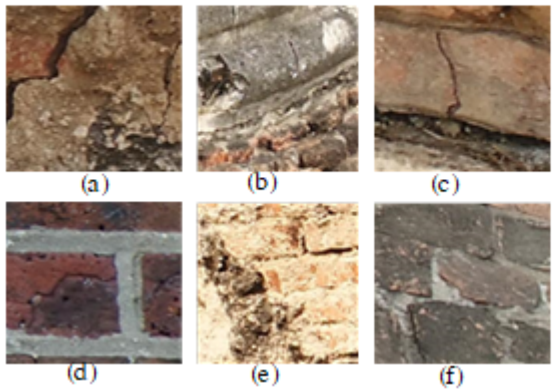

Fig. 11 (a), (b), (c) shows sample images of crack patches. (d), (e), (f) shows Sample images of non-crack patches.

### 5.2.2. Parameters Estimation

**Convolutional neural network**

CNN was implemented using the Keras library in python. Charles [31] developed the Keras library that provides a framework for deep convolutional neural network and a collection of backend libraries. Training CNN increases features variation and can avoid the over-fitting problem. The dropout method is used for CNN training to reduce an over fitting problem as suggested in [22]. The input to CNN is an r × r × d image patch, where r is the height and width of the patch and d = 3, which is the RGB channel. The training dataset is $\{I_n, y_n\}$, for n = 1, 2, ..., m and m is the total number of image patches, $I_n$ is a 28×28 patch and $i_n$ is {1, 0}, a class label. The first convolutional layer consists of 32 3x3 convolutional filters and the max pooling filters have a ratio of 2. CNN training was stopped after 17 epochs when the loss converged to a fixed value as shown in Fig. 12.

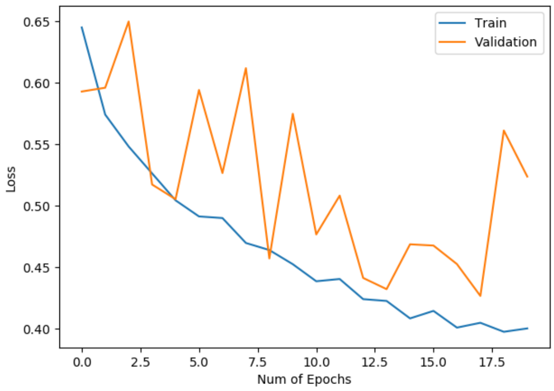


Fig. 12 shows the plot of a loss against the number of Epochs for CNN training

**Randomized Forest**

In the proposed system, the fully connected layers of CNN are input as feature vectors for Random Forest. To choose suitable parameters of the RF classifier, including the number of estimators E, the maximum depth of tree D and the minimum sample split M, a different combination of the values E, D, and M was conducted using the training data. The results are shown in Table 1, the best accuracy was obtained when E = 70, D = 20 and M = 2.

**Support Vector Machine**

For SVM, the Radial Basis Function (RBF) was used as a kernel, hence a cross-validation technique was employed to obtain the optimal values for the kernel. Table 2 shows a parametric study for SVM, where a different combination of C and gamma values were tried in the validation dataset to obtain the maximum accuracy. As shown in Table 2, the best accuracy occurred when C = 4 and gamma = 1.

Table 1**:** Parametric study for Random Forest

| Estimator | Maximum Depth | Minimum Split | Accuracy |
|---|---|---|---|
| 30 | 10 | 2 | 69.15 |
| 50 | 10 | 2 | 69.15 |
| 70 | 10 | 2 | 69.39 |
| 75 | 10 | 2 | 69.32 |
| 80 | 10 | 2 | 69.18 |
| 70 | 15 | 2 | 69.83 |
| 70 | 20 | 2 | 70.51 |
| 70 | 25 | 2 | 70.10 |
| 70 | 20 | 4 | 70.00 |
| 70 | 20 | 6 | 69.66 |

Table 2: Parametric study for SVM

| C | gamma | Accuracy |
|---|---|---|
| **1** | 0.5 | 0.770 |
| **1** | 1 | 0.72 |
| **2** | 1 | 0.72 |
| **3** | 1 | 0.73 |
| **4** | 1 | 0.73 |
| **5** | 1 | 0.71 |

### 5.2.3. Validation Results

**Performance metrics**

All methods were compared using the ROC curves, and classification results using the definition shown below. Table 3 shows the definition of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The classification report and ROC curves were obtained based on the confusion matrix, which can be explained as shown in Table 3.

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

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

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

$$F1score = \frac{2 \times Prec \times Recall}{Prec + Recall} \tag{4}$$

Table 3. Confusion matrix for class classification

| Ground Truth Label | Predicted Label | |
|---|---|---|
| | *Positive (Crack)* | *Negative (Non-crack)* |
| **Positive (Crack)** | True Positive (TP) | False Negative (FN) |
| **Negative (Non-crack)** | False Positive (FP) | True Negative (TN) |

**Results**

Table 4 shows the results from the validation dataset. The experiments were conducted on three different methods,

- Convolutional neural network for feature extraction and classification (CNN).
- Convolutional neural network for feature extraction and Random forest classification (CNN-RF).
- Convolutional neural network for feature extraction and support vector machine classification (CNN-SVM).

It is clear that the combined model, CNN-SVM, outperforms both methods, CNN and CNN-RF with the accuracy up to 85.94% in the validation step. The CNN-SVM method outperforms the other two methods in all evaluation metrics. The CNN-SVM method performs better than the CNN method alone and CNN-RF. This implies that the CNN method is good to extract useful information from images and a classifier can be used to help to boost the classification performance. It is clear from this results that the machine learning technique is a much better way for the classification problem. Fig. 13 shows the ROC plot for all the methods. It can be seen that the CNN-SVM method has the best performance as the ROC curve goes towards the top left corner of the graph.

Table 4 Performance of crack detection system on validation data

| Method | Validation Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| **CNN** | 82.94 | 0.83 | 0.71 | 0.74 |
| **CNN-RF** | 83.11 | 0.84 | 0.85 | 0.84 |
| **CNN-SVM** | 85.94 | 0.84 | 0.79 | 0.79 |

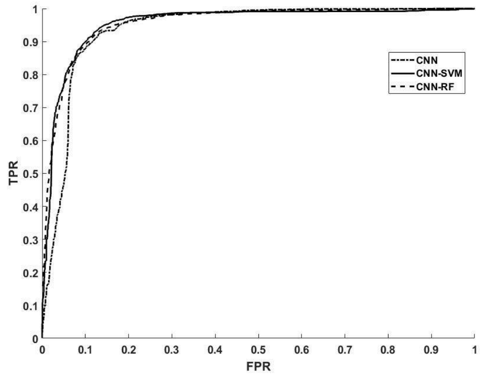


Fig. 13 shows The ROC curves between the CNN technique and the combined CNN-RF, CNN-SVM model for validation data.

**5.2.4. Application to testing dataset**

In this section, the crack detection system was conducted on the testing dataset, which was images of one of a stupa in Wat Chai Wattanaram that contains many visible cracks as shown in Fig. 14. 2934 image patches were used in this experiment, which was not included in the training of CNN and classifiers. Table 5 shows the performance of each method on the testing dataset. It can be seen that the CNN-SVM method performs the best with the accuracy of up to 74.90%. All other metrics by the CNN-SVM method are also better than any other methods.

Fig. 15 shows the results of images with cracks detected. The areas inside red boxes indicate that cracks are detected inside the area. It can be seen that, on the top pair of images, most crack regions are correctly identified, and similarly for the middle pair of images. However, the middle pair of images contains many false negative areas, which may be due to the system was confused the grout lines with cracks. The bottom pair of images also has many false negatives, especially around the grout lines. This suggests that the inaccuracy of the system may be due to these regions as their appearance are very similar to the appearance of cracks. Nevertheless, with more training dataset, the result should improve.

Table 5 Performance of crack detection system on testing data

| Method | Validation Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| **CNN** | 67.5 | 0.80 | 0.68 | 0.73 |
| **CNN-RF** | 72.05 | 0.79 | 0.72 | 0.70 |
| **CNN-SVM** | 74.90 | 0.82 | 0.78 | 0.78 |

Fig. 14 shows a sample of cracks image on the testing dataset.

Fig. 15 shows crack localization of selected sample images

## 6. DISCUSSION

The combined model, i.e. CNN-SVM, show an increase in accuracy as shown in Table 3 and 5, which is in line with the previous research as shown in Xue et al. [9]. Xue proved that better classification performance can be achieved when combining CNN with a classifier, in which in their paper, they combined CNN with Support Vector Machine.

It can be seen from the results of this research work that crack can be detected automatically using images and are useful for inspecting heritage structures. The system is based on images, in which cracks and damages can be detected based on images. This cannot be achieved by the laser scan system as the point cloud created by this technology is never dense enough for a crack to be detected from the 3D point cloud.

This research shows the usefulness of an automated system in inspecting heritage structures. The system keeps the database of heritage structures as well as providing the current state of the structures.

The system allows structures to be inspected more frequently. This system demonstrates the use of technology in the inspection of heritage structures and shows the advancement in automation in this area. The system proposed here can be a prototype in a robotic system using a video for inspection.

CNN is capable of extracting discriminative features from a large number of images without any pre-processing. However, a large amount of training images is required, which can be considered as a disadvantage. For example, a large amount of digital data requires high system specification to process a large amount of data.

Although CNN is extremely good at extracting features and classification, it relies on good datasets. Good datasets can be difficult to create as they rely on human as a gold standard. Images of cracks cannot be easily identified and need to be verified by experienced inspectors. Therefore, the datasets for crack and non-crack classes may need to verify by multiple sources, and this process can be laborious and time-consuming. In this project, due to limited sources, datasets were not verified by multiple people, which may result in the accuracy of CNN. Therefore, the proposed system can be improved by using better training datasets.

Crack detection on masonry structures is difficult as cracks cannot be easily identified in images. Cracks in masonry structures have a similar appearance to grout lines, which can be mistaken. Therefore, it can be difficult to create datasets as the scene is quite complex and confusing. Nevertheless, good datasets are required for any CNN system.

## 7. CONCLUSION

As demonstrated in this work, it can be concluded that the proposed automated crack detection system together with the drone technology can be applied in inspecting damages, especially cracks, from images in heritage structures in Thailand. The system employed the drone technology and a standard digital camera to acquire images, which are then processed by the crack detection system. The 3D model created from acquired images can be used for archiving heritage structures as well as for creating an inspection report in a form of 3D model with associated image locations to provide a better view and sense of space for inspectors. This can increase the effectiveness in the long-term monitoring of heritage structures. Such procedure provides an opportunity for practitioners to systematically archive and maintain heritage structures in Thailand.

It can be concluded that, for the crack detection problem, the method based on CNN-SVM is better than CNN alone and CNN-RF. This is clear in the present results that the CNN-SVM method has better accuracy in both validation and testing dataset. The system also performs better than various hand crafted feature based system in which feature extraction techniques are usually based on some assumptions which sometimes are not a generalization and latent variables may not be extracted or will be missed. Therefore, it can be said that, for the classification problem, CNN should be employed instead of the feature extraction techniques.

As shown in this work, CNN is best to be used as a feature extractor, and these features can be classified by any classifiers. In the future work, different classifiers can be explored to see if the accuracy of the system can be improved. The accuracy of the system can also be improved by providing better and bigger training datasets.

## 8. Acknowledgments

The authors gratefully acknowledge the financial support provided by Thammasat University under the TU New Research Scholar Contract No. 18/2559

## 8. REFERENCES

[1] Abdel-Qader I., Osama A. and Kelly M. Analysis of edge-detection techniques for crack identification in bridges, Journal of Computing in Civil Engineering, 17 (October 2003), pp. 255-263.

[2] Achille, C., Adami, A., Chiarini, S., Cremonesi, S., Fassi, F., Fregonese, L. and Taffurelli, L. (2015). UAV-Based Photogrammetry and Integrated Technologies for Architectural Applications—Methodological Strategies for the After-Quake Survey of Vertical Structures in Mantua (Italy). Sensors, 15(7), pp.15520-15539.

[3] Antonio Costanzo, Mario Minasi, Giuseppe Casula, Massimo Musacchio, and Maria Buongiorno. Combined use of terrestrial laser scanning and IR thermography applied to a historical building. Sensors, 15(1):194–213, Dec 2014.

[4] Arianna Pesci, Elena Bonali, Claudio Galli, and Enzo Boschi. Laser scanning and digital imaging for the investigation of an ancient building: Palazzo d'accursio study case (bologna, italy). Journal of Cultural Heritage, 13(2):215–220, apr 2012.

[5] Bhakapong Bhadrakom and Krisada Chaiyasarn. As-built 3d modeling based on structure from motion for deformation assessment of historical buildings. Int. J. GEOMATE, 11(24):2378–2384, 2016.

[6] Cha Y., Wooram C., and Oral B. Deep Learning Based Crack Damage Detection Using Convolutional Neural Networks. Computer-Aided Civil and Infrastructure Engineering 32, no. 5 (2017): 361-378.

[7] Cortes C. and Vapnik V. (1995). Support vector machine. Machine learning, 20(3), 273-297.

[8] Deodato Tapete, Nicola Casagli, Guido Luzi, Riccardo Fanti, Giovanni Gigli, and Davide Leva. Integrating radar and laser-based remote sensing techniques for monitoring structural deformation of archaeological monuments. Journal of Archaeological Science, 40(1):176–189, jan 2013.

[9] Di-Xiu Xue, Rong Zhang, Hui Feng, and Ya-LeiWang. Cnn-svm for microvascular morphological type recognition with data augmentation. Journal of medical and biological engineering, 36(6):755–764, 2016.

[10] Ellenberg A., Kontsos A., Bartoli I. and Pradhan A. Masonry crack detection application of an unmanned aerial vehicle. In Computing in Civil and Building Engineering (2014), pp. 1788-1795. 2014.

[11] Fábio Celestino Pereira and Carlos Eduardo Pereira. Embedded image processing systems for automatic recognition of cracks using uavs. IFAC-PapersOnLine, 48(10):16–21, 2015.

[12] Fabio Remondino, Silvio Del Pizzo, Thomas P Kersten, and Salvatore Troisi. Low-cost and opensource solutions for automated image orientation–a critical overview. In Euro-Mediterranean Conference, pages 40–54. Springer, 2012.

[13] Federal Highway and Transit Administration. Highway and rail transit tunnel inspection manual. US Department of Transportation, 2005.

[14] Fujita Y. and Yoshihiko H. A robust automatic crack detection method from noisy concrete surfaces. Machine Vision and Applications 22, no. 2 (2011): 245-254.

[15] Julia Armesto-Gonzlez, Beln Riveiro-Rodrguez, Diego Gonzlez-Aguilera, and M. Teresa Rivas-Brea. Terrestrial laser scanning intensity data applied to damage detection for historical buildings. Journal of Archaeological Science, 37(12):3037–3047, Dec 2010.

[16] Konstantinos Loupos, Angelos Amditis, Christos Stentoumis, Philippe Chrobocinski, Juan Victores, MaxWietek, Panagiotis

Panetsos, Alberto Roncaglia, Stephanos Camarinopoulos, Vassilis Kalidromitis, et al. Robotic intelligent vision and control for tunnel inspection and evaluation-the robinspect ec project. In Robotic and Sensors Environments (ROSE), 2014 IEEE International Symposium

[17] Liu Z., Shahrel A., Takeshi O., and Toshiaki E. Tunnel crack detection and classification system based on image processing. In Machine Vision Applications in Industrial Inspection X, vol. 4664, pp. 145-153. International Society for Optics and Photonics, 2002.

[18] Luigi Fregonese, Gaia Barbieri, Luigi Biolzi, Massimiliano Bocciarelli, Aronne Frigeri, and Laura Taffurelli. Surveying and monitoring for vulnerability assessment of an ancient building. Sensors, 13(8):9747–9773, Jul 2013.

[19] Matthew M Torok, Mani Golparvar-Fard, and Kevin B Kochersberger. Image-based automated 3d crack detection for post-disaster building assessment. Journal of Computing in Civil Engineering, 28(5): 4014004, 2013.

[20] Mogahed Y, El-Tokhey ME, Abdel-Gawad AK and El-Maghraby AM. Accuracy assessment of laser scanner in measuring and monitoring deformations of structures. World Applied Sciences, (2):144–151, 2013.

[21] Snavely N, Seitz S, and Szelisk R, Photo tourism :exploring photo collections in 3d ." ACM Transactions on Graphics, 25)3(, 835–846, 2006.

[22] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. Journal of machine learning research, 15(1):1929–1958, 2014.

[23] Perbet F, Stenger B, and Maki A. Random forest clustering and application to video segmentation. In Proceedings of the British Machine Vision Conference (BMVC), 2009

[24] Prateek Prasanna, Kristin J Dana, Nenad Gucunski, Basily B Basily, Hung M La, Ronny Salim Lim, and Hooman Parvardeh. Automated crack detection on concrete bridges. IEEE Transactions on Automation Science and Engineering, 13(2):591–599, 2016.

[25] Charles PWD. Keras. https://github.com/charlespwd/project-title, accessed 2017.

[26] Prasanna P., Kristin J., Nenad G., Basily B., Hung M., Ronny L, and Hooman P. "Automated crack detection on concrete bridges." IEEE Transactions on Automation Science and Engineering13, no. 2 (2016): 591-599.

[27] Sankarasrinivasan S, Balasubramanian E, Karthik K, Chandrasekar U, and Rishi Gupta. Health monitoring of civil structures with integrated uav and image processing system. Procedia Computer Science, 54:508–515, 2015.

[28] Suykens AK. and Joos V. Least squares support vector machine classifiers. Neural processing letters9, no. 3 (1999): 293-300.

[29] Rowley H., Shumeet B. and Takeo K. Neural network-based face detection. IEEE Transactions on pattern analysis and machine intelligence 20, no. 1 (1998): 23-38.

[30] Weesakul U, Fujii Y, Watanabe K, Bhadrakom B, Poovarodom N and Kawano S. Application of photogrammetry for the conservation of historical remains. NCCE 18, 2012.

[31] Yamaguchi T., Suzuki K., Hartono P. and Hashimoto S. Percolation approach to image-based crack detection. In Proceedings of the 7th international conference on Quality Control by Artificial Vision, Vol. 11, pp. 291-296. 2005.

[32] Yamaguchi T., Shingo N. and Shuji H. An efficient crack detection method using percolation-based image processing. In Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on, pp. 1875-1880. IEEE, 2008.

[33] Yu S, Jang J, and Han C. Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel. Automation in Construction, 16(3):255–261, 2007.

[34] Zhang L., Fan Y., Yimin D. and Ying J. Road crack detection using the deep convolutional neural network. In Image Processing (ICIP), 2016 IEEE International Conference on, pp. 3708-3712. IEEE, 2016.