

A Deep learning Framework for Cultural Heritage Damage Detection for Preservation; Based on the case of Heunginjimun and Yeongnamnu in South Korea

Sang-Yun Lee*, Daekyeom Lee**

**Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea*

** *SEASON Co., Ltd., Sejong, Republic of Korea*

syllee@etri.re.kr, daek29@season.co.kr

Abstract—In general, there are many restrictions on investigations for safety inspection due to the uniqueness of cultural heritages. Methods such as visual inspection and non-destructive inspection, which are mainly used as inspection methods, are regularly carried out, but there are limitations on time and cost. This is insufficient to identify and respond quickly when an abnormal symptom appears in cultural heritage. As a basic study of system development for rapid abnormal detection of architectural, cultural properties through Deep Learning, this paper organized a Deep Learning framework for detecting tilt in buildings for the roof of Heunginjimun Gate (Korea Treasure No. 1) and Yeongnamnu Pavilion (Korea Treasure No. 147). A framework was developed using a Convolutional Neural Network (CNN). As a result of an application, EfficientnetB0 and EfficientnetB2 models showed excellent accuracy in detecting the tilt of the roof of Heunginjimun with an average accuracy of 99.66% and 99.69%, respectively. In addition, EfficientnetB0, EfficientnetB2, and Shufflenet_v2 models showed excellent accuracy in detecting tilt of the roof of Yeongnamnu with 98.81%, 99.80%, and 98.48% accuracy. Additionally, the Grad-CAM experiment was conducted as a basis for whether the model made the proper judgment to confirm the criteria for determining abnormal detection according to the results of each model. These findings quickly detect abnormalities occurring in cultural heritages from the perspective of cultural heritage management and preservation, enabling rapid response, and are valuable for research on artificial intelligence technology related to cultural heritages.

Keyword—Conventional Neural Network, Cultural Heritage, Grad-CAM, Preservation

I. INTRODUCTION

IN recent years, disastrous incidents have occurred during the attempts to preserve cultural assets, such as the tilting of Cheomseongdae in Gyeongju due to the earthquake in Pohang in 2016, the burning of Sungnyemun Gate, the attempted arson of Heunginjimun Gate in 2018, and the

collapse of the Gongsanseong Fortress Wall due to torrential rain in July 2020.

According to the data on the status of emergency repair of cultural properties by cause of damage that occurred for 6 years in the data of 'Cultural heritage in statistics (2021, 2022)' published by the Cultural Heritage Administration of South Korea, total 319 cases of damage occurred for 6 years, storm and flood damage (typhoon, strong wind, heavy rain), biological damage (termites, pests), others (collapse, fall, unknown cause, etc.), cold waves, and earthquakes accounted for 168 cases, 48 cases, 36 cases, 19 cases, and 16 cases, respectively [1, 2].

In line with the situation above, the cost of maintenance and repair of cultural properties is also increasing. According to the current status of cultural asset repair and maintenance by the Cultural Heritage Administration of the Republic of Korea, the amount spent on cultural property repair was about 432 billion won in 2018, about 526 billion won in 2019, about 570 billion won in 2020, about 577 billion won in 2021, and about 574 billion won in 2022. It shows a gradually increasing trend [3].

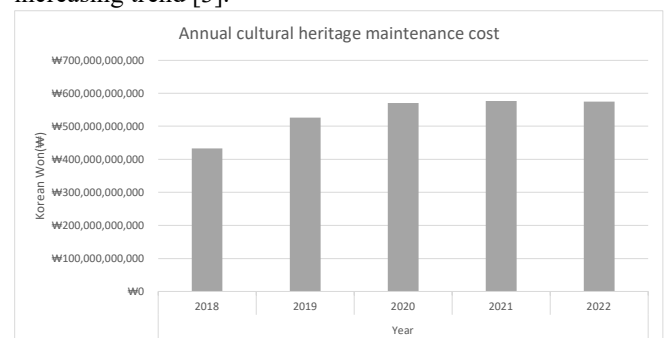


Fig. 1. Annual cultural heritage maintenance cost in South Korea (Reorganized based on KOSIS, 2022)

South Korea's cultural heritages are valuable assets that embodies human culture, so it is imperative that they be

Manuscript received Dec. 13, 2022. This paper was funded by the government (Cultural Heritage Administration) in 2023 with the support of the National Research Institute of Cultural Heritage (Grant ID: No.2021A01D06-001), and a follow-up of the invited journal to the accepted & presented paper entitled "Damage Detection and Safety Diagnosis for Immovable Cultural Assets Using Deep Learning Framework" of the 25th International Conference on Advanced Communication Technology (ICTACT2023).

Sang-Yun Lee, he is a Ph. D. and working for Electronics and Telecommunications Research Institute (ETRI) as a principal researcher in Daejeon, 34129, Republic of Korea (corresponding author to provide phone: +82-10-2995-0928; E-mail: syllee@etri.re.kr)

Daekyeom Lee, he is working for Season Co., Ltd in Sejong-si, 30127, Republic of Korea. He is now leading the AI Convergence Technology Research Team. (phone: +82-10-3287-9100; E-mail: daek29@season.co.kr)

handed down to future generations. Therefore, regular inspection and maintenance work is required to sufficiently maintain the structural stability of aging cultural properties due to issues such as climate change, Artificial activity, and deterioration. In addition, since damage to cultural heritage is irreversible once incited, efforts should be made to prevent them from being destroyed or damaged. However, due to the unique qualities of cultural heritage, there are many restrictions on safety inspections, so inspection is usually conducted by methods such as visual inspection and other non-destructive methods. Methods such as visual inspection and non-destructive inspection are usually performed on a regular basis, but there are limitations on time and cost. In addition, when abnormal symptoms appear in cultural properties, identifying them and taking immediate initial responses is not enough. In accordance to this issue, attempts are being made to combine cultural heritage conservation and artificial intelligence technology to respond to cultural heritage damage [4-14]. As a representative case, Mishra et al implemented a case study on Dadi - Poti tom in New Deli Hauz Khas Village. They developed a You Only Look Once (YOLOv5) real-time object detection algorithm and ResNet 101-based R-CNN through customized defect detection and localization. A study was conducted using the model to detect four types of defects: discoloration, brick exposure, cracks, and delamination [4]. Mansuri and Patel used the R-CNN model to build an automatic visual inspection system in Surat, India, and British and Dutch cemeteries as well have found 'breakthroughs', 'exposed brick', 'cracks' spalling exposed bricks, and cracks existing. A study was conducted to detect these three types of defects in heritage structures [5]. Yu et al generated data and used deep networks to carry out the Dunhuang cultural heritage protection project known as Thousand Buddha Caves [7]. The papers above performed machine learning mainly on stone structures, unlike this paper, which is for wooden buildings.

In this paper, Deep Learning video data taken by CCTV of the roof of Heunginjimun (Treasure No. 1), and Yeongnamnu (Treasure No. 147), were used for Deep Learning. In the case of Heunginjimun, it is the only one among the gates of the capital city surrounding Hanyang, the capital of the Joseon Dynasty, that maintains the form of Onseong (a double wall surrounding the gate to protect it). Like Sungnyemun, which is National Treasure No. 1, the gatehouse has multiple levels (double-layer), and like Gyeongbokgung Palace, the existence of Jabsang (Small figures on the roof. People believed that they protected the building from evil spirits, especially fire spirits.) on the roof indicates that the building holds significant historical value [15,16]. Also, in the case of Yeongnamnu, it is a pavilion-style building that is a representative form of South Korean traditional architecture. It is also a representative wooden structure of the late Joseon Dynasty. It is a cultural property with high preservation value, as it is being called one of the three significant pavilions in Korea [17,18,19]. In addition, as a preceding study, [20] proposed a Deep Learning Framework that can detect fine gradients for Heunginjimun. However, unlike this paper, since the experiment was performed with only a single subject of Heunginjimun, this paper additionally included Yeongnamru as a subject, and confirmed the results with Grad-CAM to verify that Deep Learning is performed well.

By using the data from the aforementioned cultural heritages, we will identify an optimal Deep Learning model and propose a building abnormality detection system by constructing a framework. The Deep Learning framework developed in this study is believed to be of great help for the preservation and transmission of cultural assets. The structure of this thesis is as follows. In Chapter 2, the data set composition and preprocessing process for generating abnormal data were explained, and in Chapter 3, the methodology used in this experiment and the parameter setting of the used model were described. In Chapter 4, the experimental results of each model and Grad-CAM results for verify were presented, and finally, the conclusion is made in Chapter 5.

II. DATA COLLECTION AND PREPROCESSING

A. The composition of the dataset

In order to construct a dataset for use in the damage detection Deep Learning framework of cultural heritage, images of cultural heritage filmed through CCTV were used. CCTV images installed at the front of Heunginjimun and Yeongnamnu were used, and the data recorded by CCTV are converted into AVI video format in IRAS (IDIS CCTV monitoring SW). The converted AVI video was divided into 600 frame units, saved as JPG pictures, and input to the pre-trained model. The dataset constructed using CCTV images as above was divided into five categories: Clear-Day, Clear-Night, Rainy-Day, Rainy-Night, and Cloudy-Day by environment. The composition of each environment dataset consists of 10000 pieces of data and consists of 5000 pieces of the training set, 2500 pieces of the validation set, and 2500 pieces of the test set. The ratio of the experiment's train set, validation set, and test set was 50% 25% 25%. Table. 1 shows the composition and ratio of each data set of Heunginjimun and Yeongnamnu.

TABLE I
THE COMPOSITION AND RATIO OF THE DATASET

Heunginjimun								
Dataset	Normal/Abnormal (Ratio)	Train/Validation/ Test (Quantity)	Train set (Quantity)		Validation set (Quantity)		Test set (Quantity)	
			Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Clear-Day	5:5	5000/2500/2500	2500	2500	1250	1250	1250	1250
Clear-Night								
Rainy-Day								
Rainy-Night								
Cloudy-Day								

Yeongnamnu								
Dataset	Normal/Abnormal (Ratio)	Train/Validation/ Test (Quantity)	Train set (Quantity)		Validation set (Quantity)		Test set (Quantity)	
			Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Clear-Day	5:5	5000/2500/2500	2500	2500	1250	1250	1250	1250
Clear-Night								
Rainy-Day								
Rainy-Night								
Cloudy-Day								

B. Configuring the data set's Environments

In order to predict the tilt of the roof part, prediction experiments were conducted in various environments. We tried to reflect on the most common environments according to our data. The roof of Heunginjimun was divided into five environments: Clear-Day, Clear-Night, Rainy-Day, Rainy-Night, and Cloudy-Day. Also, it was divided into Day and

Night based on the time when CCTV converted to night camera. Yeongnamnu also divided the environment into same. Table. 2 shows examples of data for each environment.

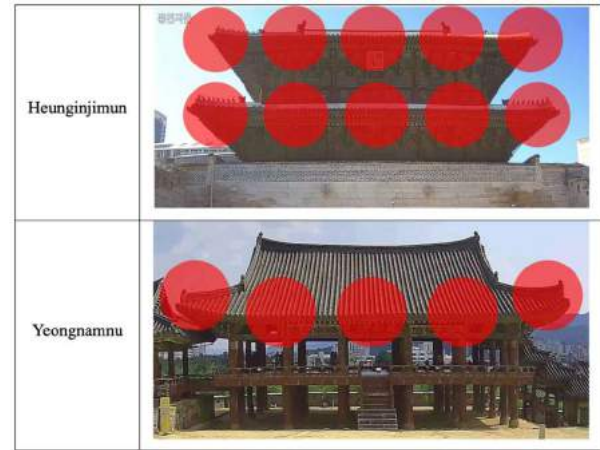
TABLE II
EXAMPLES OF DATASETS BY ENVIRONMENT

Data Example		
Environment	Heunginjimun	Yeongnamnu
Clear-Day		
Clear-Night		
Rainy-Day		
Rainy-Night		
Cloudy-Day		

C. Generation of Abnormal data

As with the normal data we needed for Deep Learning, an example of abnormal data was needed to explore the abnormal state of architecture. However, once a cultural heritage is damaged, it is often difficult to restore it to its original state. Even if it can be restored, it takes a tremendous amount of money and time, so collecting enough abnormal data for machine learning is very difficult. Therefore, as an alternative method, in the case of abnormal data, the Adobe Photoshop action function was used to generate abnormal data for each scenario in each environment by stretching the eaves. In the case of the roof of Heunginjimun, it is composed of two layers, and displacements were applied to a total of 10 areas. The area where the displacement was applied is marked by a red circle. Unlike Heunginjimun, the roof of Yeongnamnu is composed of a single layer, and displacements were applied to five areas accordingly. As with Heunginjimun, the area marked with a red circle is the area where the displacement was used. Table. 3 shows the displacement points of Heunginjimun and Yeongnamnu.

TABLE III
DISPLACEMENT POINTS FOR GENERATING ABNORMAL DATA



The samples of the abnormal data set generated by applying displacement to the points of 10 areas of the roof of Heunginjimun and 5 areas of the roof of Yeongnamnu are shown in Fig. 2 and Fig. 3 below. When checking the abnormal data compared to the original, it can be seen that the tilt of the displaced area (the red rectangle in Fig. 2 and Fig. 3) is slightly different.

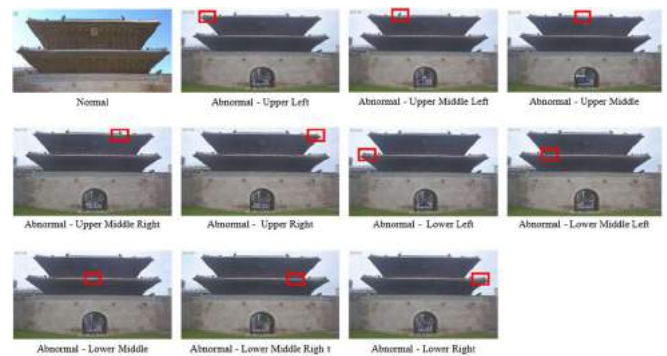


Fig. 2. Samples of abnormal images of Heunginjimun



Fig. 3. Samples of abnormal images of Yeongnamnu

III. METHODOLOGY

A. Convolutional Neural Network (CNN)

Convolution Neural Network (CNN) is a technique announced by LeCun and Bengio in 1995 [21]. It is an algorithm that shows excellent performance in image recognition or voice recognition among Deep Learning methods, and its structure is designed according to neurobiological principles [22-25]. In the case of CNN, it has the ability to analyze the specific characteristics of the object in question regardless of the location and the direction in which it is placed. CNN consists of a convolution layer that

extracts features from an input image and a pooling layer that compresses the extracted information. Convolution and pooling are repeated to extract the features [25]. Furthermore, CNN uses a filter in the convolution step to extract the features of an image. As shown in Fig. 4, the Filter composed of matrices in the Image Matrix scans the image from top left to bottom right.

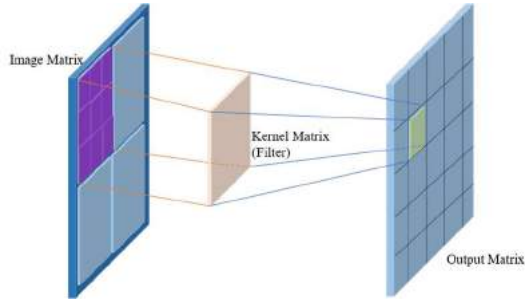


Fig. 4. Feature extraction in convolution

Subsequently, after going through the feature extraction step in which convolution and pooling are repeated, the classification step is passed as shown in Fig.5. The feature matrix extracted in the flattening step is flattened into vectors. Then the sorted vectors go through a fully connected layer and an activation function such as softmax or sigmoid is ordered to output a class and classify the image and create a result

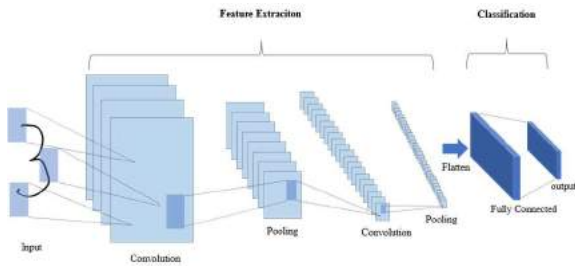


Fig. 5. Convolutional Neural Network Architecture

B. Experimental Environment

This experiment was performed on Ubuntu 20.040.2. and under the LTS environment. The specifications and other packages of the devices used in this study are as follows.

- Memory 64GB
- CPU 24 core
- GPU 1080(GeForce GTX 1080ti)
- CUDA 11.3
- Python 3.9
- efficientnet-pytorch==0.7.1
- matplotlib==3.6.0
- nnpack==0.1.0
- opencv-python==4.6.0.66
- Pillow==9.2.0
- scikit-learn==1.1.2
- tqdm==4.64.1
- numpy==1.23.3
- decord==0.6.0
- imgaug==0.4.0
- ttach
- tqdm

C. Framework design and implementation

A Deep Learning framework was designed to find the tilt that occurs on the roof of Heunginjimun and the roof of Yeongnamnu. The Deep Learning framework is shown in Fig. 6

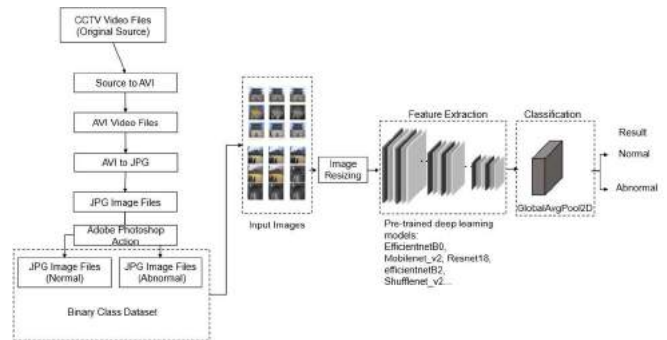


Fig. 6. The framework for cultural heritage damage detection

The Deep Learning framework first converts the video generated by CCTV filming cultural heritage into an AVI video format, cuts them in frame units, and extracts them into JPG images. The extracted image is used as normal data, and in the case of abnormal data, it is created by applying distortion to the normal data. This method of applying distortion to normal data uses the action cam function of Adobe Photoshop. Thereafter, an image is adjusted through a resizing procedure to use the data in a neural network model. After image resizing, the dataset is divided into training, testing, and verification sets and used as input values for neural network models. Finally, through the feature extraction part and classification part of the model, it is classified into two classes: normal and abnormal.

D. Model selection and parameter settings

The neural network model used in machine learning loads image data from the dataset created through the preprocessing process and uses normal and abnormal images as input values for the neural network model. For the neural network model used for classification, the most commonly used models were selected: efficientnetB0 [26], mobilenet_v2 [27], resnet18 [28], efficientnetB2, shufflenet_v2 [29], alexnet [30], mnasnet [31], inception_v3[32], densenet161[33], efficientnetb4, and 10 types of pre-learning-based Deep Learning models were used. The dataset used for pre-learning is ImageNet. Learning is conducted to classify it into normal and abnormal classes through the feature extraction and classification parts of the model. Table. 4 shows the parameters used for machine learning. Optimal parameters were used as long as they were versatile for each environment. In the case of the optimizer, it was adopted through Stochastic Gradient Descent (SGD). The initial learning rate was set to 0.002 for Alexanet and 0.05 for EfficientnetB0, EfficientnetB2, Shufflenet_v2, and Mnasnet. The momentum value was set to 0.9, and 10 to 20 epochs were performed.

TABLE IV
PARAMETERS OF HEUNGINJIMUN AND YEONGNAMNU

	Heunginjimun				Yeongnamnu			
	Models	Input size	Optimizer	Epochs	Input size	Optimizer	Epochs	
Clear-Day	efficientnetB0	224	SGD (lr=0.05, momentum = 0.9)	20	224	SGD (lr=0.05, momentum = 0.9)	10	
	mobilenet_v2	224	SGD (lr=0.05, momentum = 0.9)	20	224	SGD (lr=0.05, momentum = 0.9)	10	
	resnet18	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnetB2	260	SGD (lr=0.05, momentum = 0.9)	15	260	SGD (lr=0.05, momentum = 0.9)	10	
	shufflenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	alexnet	224	SGD (lr=0.002, momentum = 0.9)	15	224	SGD (lr=0.002, momentum = 0.9)	10	
	mnasnet	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	inception_v3	299	SGD (lr=0.05, momentum = 0.9)	15	299	SGD (lr=0.05, momentum = 0.9)	10	
	densenet161	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnet_v2_s	384	SGD (lr=0.05, momentum = 0.9)	15	384	SGD (lr=0.05, momentum = 0.9)	20	
Clear-Night	efficientnetB0	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	mobilenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	resnet18	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnetB2	260	SGD (lr=0.05, momentum = 0.9)	15	260	SGD (lr=0.05, momentum = 0.9)	10	
	shufflenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	alexnet	224	SGD (lr=0.002, momentum = 0.9)	15	224	SGD (lr=0.002, momentum = 0.9)	10	
	mnasnet	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	inception_v3	299	SGD (lr=0.05, momentum = 0.9)	15	299	SGD (lr=0.05, momentum = 0.9)	10	
	densenet161	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnet_v2_s	384	SGD (lr=0.05, momentum = 0.9)	15	384	SGD (lr=0.05, momentum = 0.9)	20	
Rainy-Day	efficientnetB0	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	mobilenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	resnet18	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnetB2	260	SGD (lr=0.05, momentum = 0.9)	15	260	SGD (lr=0.05, momentum = 0.9)	10	
	shufflenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	alexnet	224	SGD (lr=0.002, momentum = 0.9)	15	224	SGD (lr=0.002, momentum = 0.9)	10	
	mnasnet	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	inception_v3	299	SGD (lr=0.05, momentum = 0.9)	15	299	SGD (lr=0.05, momentum = 0.9)	10	
	densenet161	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnet_v2_s	384	SGD (lr=0.05, momentum = 0.9)	15	384	SGD (lr=0.05, momentum = 0.9)	20	
Rainy-Night	efficientnetB0	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	mobilenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	resnet18	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnetB2	260	SGD (lr=0.05, momentum = 0.9)	15	260	SGD (lr=0.05, momentum = 0.9)	10	
	shufflenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	alexnet	224	SGD (lr=0.002, momentum = 0.9)	15	224	SGD (lr=0.002, momentum = 0.9)	10	
	mnasnet	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	inception_v3	299	SGD (lr=0.05, momentum = 0.9)	15	299	SGD (lr=0.05, momentum = 0.9)	10	
	densenet161	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnet_v2_s	384	SGD (lr=0.05, momentum = 0.9)	15	384	SGD (lr=0.05, momentum = 0.9)	10	
Cloudy-Day	efficientnetB0	224	SGD (lr=0.05, momentum = 0.9)	10	224	SGD (lr=0.05, momentum = 0.9)	10	
	mobilenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	resnet18	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnetB2	260	SGD (lr=0.05, momentum = 0.9)	15	260	SGD (lr=0.05, momentum = 0.9)	10	
	shufflenet_v2	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	alexnet	224	SGD (lr=0.002, momentum = 0.9)	15	224	SGD (lr=0.002, momentum = 0.9)	10	
	mnasnet	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	inception_v3	299	SGD (lr=0.05, momentum = 0.9)	15	299	SGD (lr=0.05, momentum = 0.9)	10	
	densenet161	224	SGD (lr=0.05, momentum = 0.9)	15	224	SGD (lr=0.05, momentum = 0.9)	10	
	efficientnet_v2_s	384	SGD (lr=0.05, momentum = 0.9)	15	384	SGD (lr=0.05, momentum = 0.9)	10	

IV. CONCLUSION OF EXPERIMENTAL RESULTS

A. Performance results of each model

Using datasets for each environment, a normal and abnormal classification experiment was conducted on the roof of Heunginjimun and the roof of Yeongnamnu. The average values are shown in Table. 5 and Table. 6 by synthesizing the evaluation index results of each model that performed environmental classification. In the case of Heunginjimun, among ten machine learning models, EfficientnetB0, EfficientnetB2, and Shufflenet_v2 models showed excellent values of 99.61%, 99.57%, and 90.31%, respectively, based on the accuracy value. In the case of Yeongnamnu, among ten machine learning models, EfficientnetB2, EfficientnetB0, and Shufflenet_v2 showed values of 99.90%, 98.81%, and 98.48% in that order.

TABLE V
CLASSIFICATION EVALUATION METRICS VALUE OF THE MODELS FOR ROOF OF HEUNGINJIMUN

Models	Specificity	Recall	Accuracy
EfficientnetB0	99.27%	99.97%	99.61%
Mobilenet_v2	50.35%	65.60%	62.97%
Resnet18	32.43%	68.28%	55.36%
EfficientnetB2	99.16%	99.98%	99.57%
Shufflenet_v2	87.25%	93.00%	90.31%
AlexNet	29.13%	87.15%	63.81%
Mnasnet	58.71%	50.23%	59.47%
Inception_v3	75.47%	58.15%	71.81%
Densenet161	47.49%	71.23%	64.35%
Efficientnet_v2_s	56.53%	75.35%	70.94%

TABLE VI
CLASSIFICATION EVALUATION METRICS VALUE OF THE MODELS FOR ROOF OF YEONGNAMNU

Models	Specificity	Recall	Accuracy
EfficientnetB0	99.46%	98.16%	98.81%
Mobilenet_v2	70.70%	30.10%	50.40%
Resnet18	55.38%	44.46%	49.92%
EfficientnetB2	99.63%	99.97%	99.80%
Shufflenet_v2	99.22%	97.74%	98.48%
AlexNet	79.06%	20.94%	50.00%
Mnasnet	20.00%	80.00%	50.00%
Inception_v3	58.54%	42.08%	50.31%
Densenet161	62.02%	38.05%	50.03%
Efficientnet_v2_s	75.17%	75.60%	75.38%

B. Grad-CAM result of optimal model

To verify the machine learning experiment, Gradient-weighted Class Activation Mapping (Grad-CAM) visualization was performed to find the part that is the grounds of the Machine Learning results. Visualization results were shown in Table. 7 and Table .8. In the case of Heunginjimun Gate, it can be seen that the characteristics of the 10 areas with displacement are also reflected in the Grad-CAM results, which serves as the basis for abnormal judgment. Likewise, Yeongnamnu also, it was confirmed that the area of the five areas to which the displacement was applied was referred to as the basis for abnormal judgment.

TABLE VII
 HEUNGINJIMUN GRAD-CAM RESULTS

Inclined area	Lower Left	Lower Middle Left	Lower Middle	Lower Middle Right	Lower Right
Input Image					
Grad-CAM Result					
Inclined Area	Upper Left	Upper Middle Left	Upper Middle	Upper Middle Right	Upper Right
Input Image					
Grad-CAM Result					

 TABLE VIII
 YEONGNAMNU GRAD-CAM RESULTS

Inclined Area	Left	Middle Left	Middle	Middle right	Right
input image					
Grad-CAM Result					

V. CONCLUSION

In this paper, basic research was conducted on the development of a system for detecting anomalies in architectural, cultural properties so that abnormal symptoms can be identified or prompt initial responses can be made for the preservation of cultural properties. Heunginjimun and Yeongnamnu, treasures of the Republic of Korea, were used as subjects of basic research. CCTV images of Heunginjimun and Yeongnamnu were collected to build a dataset to determine the roof's displacement. A Deep Learning model was trained with the built data to test the abnormal detection performance so that the model can detect the degree of tilt corresponding to damage to cultural heritage.

The experimental results for 10 models on the roof of Heunginjimun were as follows. Summarizing only the top 3 models with optimal results, the EfficientnetB0, EfficientnetB2, and Shufflenet_V2 models showed excellent values of 99.61%, 99.57%, and 90.31% based on the accuracy. Specificity also showed 99.27%, 99.16%, and 87.25% values for EfficientnetB0, EfficientnetB2, and Shufflenet_V2. Finally, based on the recall, EfficientnetB0, EfficientnetB2, and Shufflenet_V2 showed values of 99.97%, 99.98%, and 87.15% in order. In the case of the roof of Yeongnamnu, as with the roof of Heunginjimun, a comparative experiment was conducted on 10 models. Based on the accuracy value, the EfficientnetB0, EfficientnetB2, and Shufflenet_V2 models showed values of 98.81%, 99.80%, and 98.48%, and based on the specificity, they showed values of 99.46%, 99.63%, and 99.22%. In addition, based on the recall, values of 98.16%, 99.97%, and 97.74% were shown. As a result, the

Efficientnet series models of EfficientnetB0 and EfficientnetB2 showed optimal performance for tilt detection.

Then, using Grad-CAM, we tried to check whether the EfficientnetB0 model, which represents the Efficientnet algorithm, produced results based on appropriate judgment grounds. As a result of Grad-CAM, it was confirmed that the model made a judgment based on the change in tilt displacement of the roof. The framework proposed in this experiment was proposed to perform rapid damage detection of cultural heritage and to help apply artificial intelligence, which remains at the introductory stage in managing and preserving cultural heritage. First, a normal dataset was constructed for Deep Learning and abnormal data based on normal data was generated. In addition, in order to find the most effective and universally usable Deep Learning model, we tried to derive a model that is optimal for all environments filmed by CCTV using 10 commonly used models. In conclusion, the Efficientnet series algorithm performed the best in detecting the tilt of cultural properties, which means that the Efficientnet model is the most universally effective model in situations where parameter tuning is limited. However, this study did not reflect seasonal characteristics due to an insufficient data collection period. The five environments covered the most basic environments, so learning by applying more displacement scenarios and mixed environments in future experiments is necessary. Complementing these limitations will further improve the robustness of the framework.

ACKNOWLEDGMENT

This paper was funded by the government (Cultural Heritage Administration) in 2023 with the support of the National Research Institute of Cultural Heritage (No.2021A01D06-001, Development of damage detection and alarm technology based on intelligent image analysis for safety diagnosis of Immovable cultural heritage)

REFERENCES

- [1] E. C. Choi, "Cultural heritage in statistics," *Cultural Heritage Administration of the Republic of Korea*, 2022, Available: https://www.cha.go.kr/cop/bbs/selectBoardArticle.do?nttId=85272&bbId=BBSMSTR_1020&pageIndex=1&pageUnit=10&searchCnd=&searchWrd=&ctgryLrcls=&ctgryMdcls=&ctgrySmcls=&ntcStartDt=&ntcEndDt=&searchUseYn=&mn=NS_03_07_04. Accessed 31 March 2023
- [2] H.M. Kim, "Cultural heritage in statistics," *Cultural Heritage Administration of the Republic of Korea*, 2023, Available: https://www.cha.go.kr/cop/bbs/selectBoardArticle.do?nttId=82245&bbId=BBSMSTR_1020&pageIndex=1&pageUnit=10&searchCnd=&searchWrd=&ctgryLrcls=&ctgryMdcls=&ctgrySmcls=&ntcStartDt=&ntcEndDt=&searchUseYn=&mn=NS_03_07_04. Accessed 01 July 2023
- [3] "Current Status of Maintenance and Maintenance of Cultural Heritage," *Korean Statistical Information Service(KOSIS)*, 2023, Available: https://kosis.kr/common/meta_oneDepth.jsp?vwcd=MT_OTITLE&listid=150. Accessed 01 July 2023
- [4] M. Mishra, T. Barman, and G. V. Ramana, "Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage," *Journal of Civil Structural Health Monitoring*, pp. 1-18, 2022, <https://doi.org/10.1007/s13349-022-00643-8>
- [5] L.E. Mansuri, D.A. Patel, "Artificial intelligence-based automatic visual inspection system for built heritage," *Smart and Sustainable Built Environment*, vol. 11(3), pp. 622-646, 2022, <https://doi.org/10.1108/SASBE-09-2020-0139>
- [6] M. Mishra, "Machine learning techniques for structural health monitoring of heritage buildings: A state-of-the-art review and case

- studies,” *Journal of Cultural Heritage*, vol. 47, pp. 227-245, 2021, <https://doi.org/10.1016/j.culher.2020.09.005>
- [7] T. Yu, C. Lin, S. Zhang, C. Wang, X. Ding, H. An, and J. Zhang, “Artificial Intelligence for Dunhuang Cultural Heritage Protection: The Project and the Dataset,” *International Journal of Computer Vision* vol.130(11), pp.2646-2673, 2022, <https://doi.org/10.1007/s11263-022-01665-x>
- [8] A. Belhi, H. Gasmi, A. Bouras, T. Alfaqheri, A. S. Aondoakaa, A. H. Sadka, and S. Fofou, “Machine learning and digital heritage: the CEPROQHA project perspective,” *Springer Singapor*, 2020, [In 4th International Congress on Information and Communication Technology: ICICT 2019 London, 2019, vol.2, pp.363-374], https://doi.org/10.1007/978-981-32-9343-4_29
- [9] N. Wang, X. Zhao, P. Zhao, Y. Zhang, Z. Zou, and J. Ou, “Automatic damage detection of historic masonry buildings based on mobile deep learning,” *Automation in Construction*, vol.103, pp.53-66, 2019, <https://doi.org/10.1016/j.autcon.2019.03.003>
- [10] D. Bienvenido-Huertas, JE. Nieto-Julián, JJ. Moyano, JM. Macías-Bernal, J. Castro, “Implementing artificial intelligence in h-bim using the J48 algorithm to manage historic buildings,” *Int J Archit Herit*, vol.14(8), pp.1148–1160, 2019, <https://doi.org/10.1080/15583058.2019.1589602>
- [11] T. Bakirman, B. Kulavuz, and B. Bayram, “Use of Artificial Intelligence Toward Climate-neutral Cultural Heritage,” *Photogrammetric Engineering & Remote Sensing*, vol. 89(3), pp.163-171, 2023, <https://doi.org/10.14358/PERS.22-00118R2>
- [12] I. Garrido, J. Erazo-Aux, S. Lagüela, S. Sfarra, C. Ibarra-Castanedo, E. Pivarčiová, and P. Arias, “Introduction of deep learning in thermographic monitoring of cultural heritage and improvement by automatic thermogram pre-processing algorithms,” *Sensors*, vol. 21(3), pp. 750, 2021, <https://doi.org/10.3390/s21030750>
- [13] Z. Zou, X. Zhao, P. Zhao, F. Qi, and N. Wang, “CNN-based statistics and location estimation of missing components in routine inspection of historic buildings,” *Journal of Cultural Heritage*, vol. 38, pp. 221-230, 2019, <https://doi.org/10.1016/j.culher.2019.02.002>
- [14] T. Sharma, P. Agrawal, and N. K. Verma, “Detection of dust deposition using convolutional neural network for heritage images,” *Springer Singapore*, 2019 [In Computational Intelligence: Theories, Applications and Future Directions-Volume II: ICCI-2017, pp.347-359], https://doi.org/10.1007/978-981-13-1135-2_27
- [15] E. J. Jeong, “Heunginjinmun and Heritages around the Gate,” *Art History & Cultural Heritages*, vol. 5, pp. 79-109, 2016.
- [16] S. A. Park, K. W. Min, and J. S. Choi, “Ambient vibration analysis of Heunginjinmun,” *Journal of The Architectural Institute of Korea: Structure & Construction*, vol. 27(5), pp. 19-26, 2011.
- [17] T. G. Eom, S. J. Kim, J. L. Park, H. M. Kang, and W. K. Sim, “Interpretation of Cultural Landscape at the Geumsidang (今是堂) sibigyung (12 Landscapes) in Miryang, Gyeongnam,” *Journal of the Korean Institute of Traditional Landscape Architecture*, vol. 29(2), pp. 1-18, 2011.
- [18] S. L. Ryoo, “A Study on the Changes of the Government Pavilion, Miryang Yeongnamnu in terms of Function and Spatiality,” *Journal of the Architectural Institute of Korea Planning & Design*, vol. 34(8), pp. 69-76, 2019, https://doi.org/10.5659/JAIK_PD.2018.34.8.69
- [19] H. Y. Lee, “A Study on the Historic Changes of Yungnam-Ru in Historic Periods and Architectural Building Forms,” *Journal of architectural history*, vol. 9(1), pp. 7-25, 2000.
- [20] S. Y. Lee, H. H. Cho, “Damage Detection and Safety Diagnosis for immovable Cultural Assets Using Deep Learning Framework,” *ICTACT2023*, pp. 310-313, 2023.
- [21] Y. LeCun, Y. Bengio, “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks* 3361(10), 1995.
- [22] J. Wu, “Introduction to convolutional neural networks,” *National Key Lab for Novel Software Technology*, vol. 5(23), pp. 495, 2017, Nanjing University. China.
- [23] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, “Convolutional neural network (CNN) for image detection and recognition,” *2018 first international conference on secure cyber computing and communication (ICSCCC 2018)*, pp. 278-282, 10.1109/ICSCCC.2018.8703316.
- [24] M. Jogin, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, “Feature extraction using convolution neural networks (CNN) and deep learning,” *2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*, pp. 2319-2323, 10.1109/RTEICT42901.2018.9012507.
- [25] C. D. James, J. B. Aimone, N. E. Miner, C. M. Vineyard, F. H. Rothganger, K. D. Carlson, and S. J. Plimpton, “A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications,” *Biologically Inspired Cognitive Architectures* vol. 19, pp.49-64, 2017, <https://doi.org/10.1016/j.bica.2016.11.002>.
- [26] M. Tan, Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” *International conference on machine learning 2019*, pp. 6105-6114, 2019, May.
- [27] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv*, 2017, preprint arXiv:1704.04861, <https://doi.org/10.48550/arXiv.1704.04861>.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE conference on computer vision and pattern recognition 2016*, pp. 770-778, 2016.
- [29] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, “ShuffleNet v2: Practical guidelines for efficient cnn architecture design,” *2018 Proceedings of the European conference on computer vision (ECCV)*, pp. 116-131, 2018.
- [30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60(6), pp. 84-90, 2017, <https://doi.org/10.1145/3065386>
- [31] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, “Mnasnet: Platform-aware neural architecture search for mobile,” *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019*, pp. 2820-2828.
- [32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *Proceedings of the IEEE conference on computer vision and pattern recognition 2019*, pp. 2818-2826.
- [33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, (2017) “Densely connected convolutional networks,” *Proceedings of the IEEE conference on computer vision and pattern recognition 2017*, pp. 4700-4708.



etc.

His main research interests have been in Computational Sciences. Currently, he is involved in developing a technique to detect displacement using Artificial Intelligence technology for CCTV images of cultural assets. He also leads several projects including disaster management for cultural heritage. Since 2016, He acts as an editor of the Study Group 16 (SG16) in the ITU-T and has been developing international standards. He is an international standards expert at TTA and a member of the Korean ITU-T Research Committee. He has been carrying out more than 10 government-funded projects. He is author of more than 70 scientific papers and has registered more than 10 patents.



Daekyeom Lee (B'16 – M'18) is a Lead of AI Convergence Technology Research Team at Season. From 2021 to present, he has been working at Season Co., Ltd. He graduated from Dongguk University (Rep. of Korea) with a double major in energy systems and economics in 2016 and received a master's degree in engineering from Sejong University (Rep. of Korea) in 2018. Since 2018, he has conducted research in statistics, data analysis, and artificial intelligence.

His main research interest was big data analysis and data science. Additionally, after joining Season Co., Ltd., he has been working as a team leader developing artificial intelligence technology in terms of data utilization. He mainly works on the application of artificial intelligence technology using manufacturing data, object detection using deep learning, and improvement of learning speed through transfer learning. In addition, he is conducting joint research on technology utilization in cooperation with related research institutes and companies and has authored 8 domestic and foreign journals.