

# Damage Detection and Safety Diagnosis for Immovable Cultural Assets Using Deep Learning Framework

Sang-Yun LEE\*, Ho-Hyun CHO\*\*

\*ETRI (Electronics and Telecommunications Research Institute), South Korea

\*\*SEASON Co., Ltd., Sejong, Korea

sylllee@etri.re.kr, cghgus93@season.co.kr

**Abstract**— Cultural assets are recognized as having cultural values among the products of the cultural activities our ancestors carried out and of great historical or artistic value as well. Since cultural assets, which are also those of the nation, cannot be restored once damaged, both preservation and transmission of the cultural assets are critical. When some changes occur in cultural properties by any chance, if they can be identified as quickly as possible by using Deep Learning, rapid initial response and management will be possible, which will significantly help the management and preservation of cultural properties. This paper aims to create a Deep Learning Framework that can effectively detect minute inclinations that may be difficult to identify with human eyes. For Heunginjimun, the No.1 national treasure of South Korea, four types of pretraining-based Deep Learning models were used, which were EfficientNetB0, EfficientNetB2, ShuffleNet\_v2, and AlexNet. The experiment was conducted using a typical dataset in the environment constructed by using Heunginjimun CCTV images, and then an abnormal dataset was regenerated based on the typical dataset. As a result of applying the Deep Learning model to each environment we have built, the average prediction accuracies of EfficientNetB0, EfficientNetB2, and ShuffleNet\_v2 among the four models were 99.69%, 99.66%, and 93.46%, respectively, showing high prediction accuracies. Therefore, it is judged that the results of the experiment will be of great help in the field of management and preservation of cultural properties in the future.

**Keywords**— Deep Learning, Cultural Heritage, Machine Learning, Anomaly Detection, Displacement detection

## I. INTRODUCTION

### A. Research Background and Purpose

Regular inspection and maintenance work is required to secure the structural stability of wooden cultural properties that are deteriorating due to deterioration and artificial activities. However, there are many restrictions on investigation for safety inspection due to the uniqueness of cultural heritage itself. Therefore, the inspection process is usually conducted by methods such as visual inspection and non-destructive ones. However, this type of regular inspection has limitations in terms of time and cost. In addition, it may

not be enough to identify abnormal symptoms of cultural properties or to take prompt and effective initial responses when problems occur. Recently, various studies have been conducted to predict damage or deformation of cultural properties in advance by using Deep Learning technology and coping with the possible problems in the early stages. In recent studies [1] and [2], an Ensemble-based Deep Learning model was presented to detect the displacement of fullness for wooden cultural properties. In recent studies [3], [4], [5], a Deep Learning model that measures the amount of displacement using a generative model was presented. In [6], a model using an object-oriented augmentation technique was proposed to detect displacement in images of wooden cultural properties efficiently. Suppose a building anomaly detection system using Deep Learning is applied along with the existing inspection system. In that case, it will be of great help to the preservation and transmission of cultural assets. In this paper, we propose a method for normal-abnormal binary classification using CCTV image data of the roof of Heunginjimun, Korea's treasure No. 1. To this end, six image datasets were created, and by using them, we propose a Deep Learning framework that can detect subtle tilts that are difficult to be identified even with the human eye.

## II. MODEL DESIGN AND IMPLEMENTATION

### A. The Proposed Scheme

The image data is used as the input value of the Neural Network model by loading normal data and abnormal data from the dataset constructed through the extraction process from Heunginjimun CCTV images. As a Neural Network model for the classification task, four types of pre-trained-based Deep Learning models, EfficientNetB0, EfficientNetB2, ShuffleNet\_v2, and AlexNet, were used. EfficientNet [6] was introduced in a paper at the ICML conference in May 2019. It is a width scaling technique that increases the number of channels to improve model performance, a depth scaling technique that increases the number of layers, and a resolution scaling technique that increases the resolution of an input

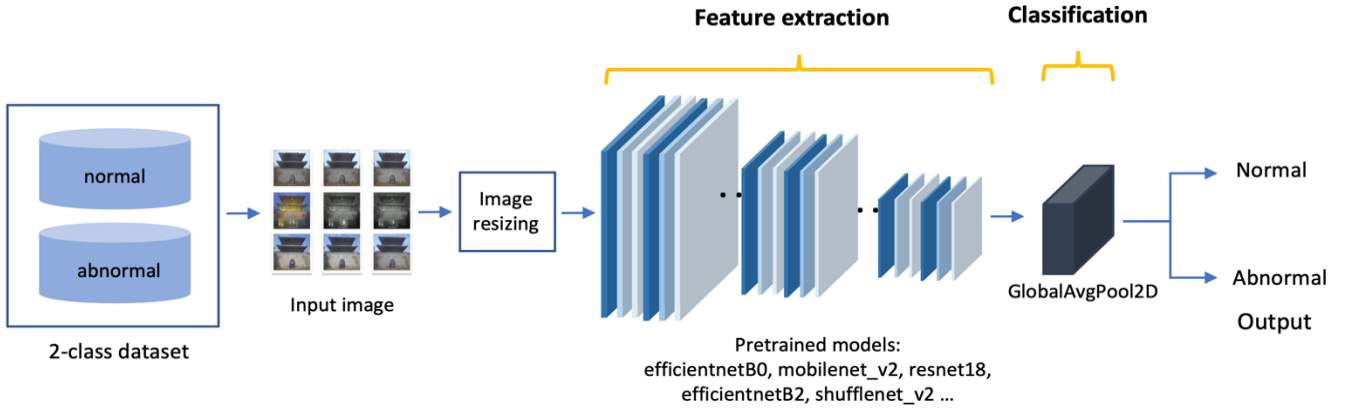


Figure 1. Deep Learning model architecture

image. This is an architecture that proposes a method that simultaneously considers two approaches.

Despite the small number of model parameters and inference speed, it performed better than existing models. ShuffleNet\_v2 [7] is an efficient convolutional Neural Network architecture designed for mobile devices presented at the 2018 ECCV conference. Pointwise group convolution and channel shuffling were proposed architectures, which enabled information communication between different channel groups. AlexNet is a Neural Network that generalizes the CNN structure's GPU implementation and the Dropout application.

Figure 1 is a structured picture of the overall architecture of Heunginjimun roof tilt detection. The dataset used for pre-learning is ImageNet [8]. The ImageNet dataset is a large-scale dataset with more than 14 million pieces and 1000 classes. When the image of the dataset generated from Heunginjimun CCTV is input into the model, the image is readjusted according to the input size of each model. Then the model's feature extraction and classification parts are classified into normal and abnormal classes.

### B. Data Pre-processing

The first step is to convert the CCTV recording file into the AVI video format. The conversion work is performed through IRAS, a CCTV monitoring software installed on the front of Heunginjimun Gate. The compression method was set to the MPEG-4-part Microsoft variant version3, the resolution of the video was set to 1920\*1080, the bitrate was set to 8000 kbps, and the key frame interval was set to 60 Frames. The converted AVI video was divided into JPG images. A dataset was constructed with 5,000 normal data, composed of JPG images generated through segmentation, and 5,000 abnormal data generated by using the normal data. Abnormal data was created in the form of drooping eaves in the normal state by using the action function of the Adobe Photoshop program. A total of 10 displacement scenarios were conceived by configuring one displacement scenario for each one of the 10 target parts to generate abnormal data. For data learning and verification, the ratios of the Train set, Validation set, and Test set were set to 50%, 25%, and 25%, respectively. Figure 2 is a schematic diagram of the process of building an image dataset.

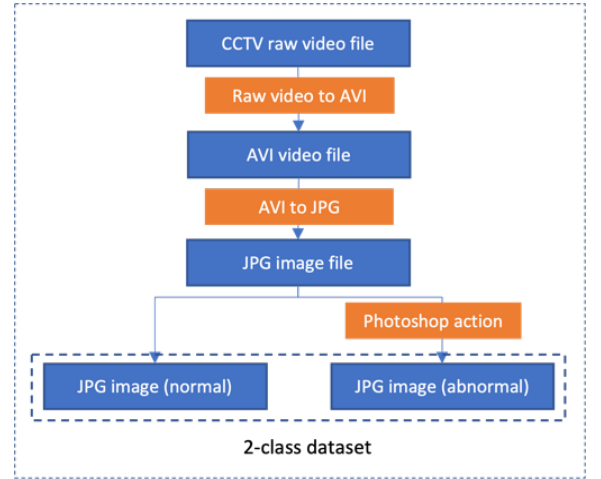


Figure 2. Preprocessing process

For domain generalization, six different environment datasets were constructed. This is to ensure that the model can perform the image classification tasks well even if the surrounding environment of Heunginjimun Gate changes, such as the time zone changes or it is snowing or raining. Six datasets were created according to the environments of a sunny day, clear night, cloudy day, rainy day, rainy night, and snowy night (Figure 3).

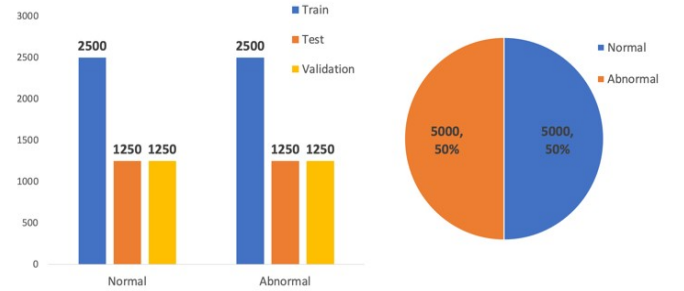
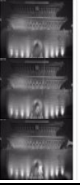


Figure 3. The composition of the dataset

Table 1 is a sample of the six datasets produced.

TABLE 1. EXAMPLES OF DATASETS BY ENVIRONMENT

| Environment  | Clear-day                                                                         |                                                                                   |                                                                                   | Clear-night                                                                       |                                                                                   |                                                                                    | Cloudy-day                                                                          |                                                                                     |                                                                                     |
|--------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Data Example |  |  |  |  |  |  |  |  |  |
| Environment  | Rainy-day                                                                         |                                                                                   |                                                                                   | Rainy-night                                                                       |                                                                                   |                                                                                    | Snowy-night                                                                         |                                                                                     |                                                                                     |
| Data Example |  |  |  |  |  |  |  |  |  |

### III. EXPERIMENTS AND RESULTS

#### A. Parameter Setting

Table 2 shows the parameters of the model set during the experiment.

TABLE 2. MODEL PARAMETERS

| Model          | Image Input size | Optimizer                                  | Batch Size | Epochs |
|----------------|------------------|--------------------------------------------|------------|--------|
| EfficientNetB0 | 224              | SGD<br>(learning rate=0.001, Momentum=0.9) | 8          | 15     |
| EfficientNetB2 | 260              | SGD<br>(learning rate=0.05, Momentum=0.9)  |            |        |
| ShuffleNet_v2  | 224              | SGD<br>(learning rate=0.05, Momentum=0.9)  |            |        |
| AlexNet        | 224              | SGD<br>(learning rate=0.002, Momentum=0.9) |            |        |

We adopted SGD momentum as the optimizer. This is an algorithm that solves the problems of the existing SGD, such as slow learning speed and ending learning at the saddle point, by bringing the concept of inertia. The learning rate parameters are set to 0.002 for AlexNet, 0.001 for EfficientNetB0, and 0.05 for EfficientNetB2 and ShuffleNet\_v2.

Momentum was set to 0.9, batch size was set to 8, and epoch was set to 15 each.

#### B. Experiment Result

For the entire dataset for each environment, the average prediction performance accuracy of each model is shown in Table 3 below.

TABLE 3. ACCURACY RESULTS BY MODEL

| Model          | Recall | Specificity | Accuracy |
|----------------|--------|-------------|----------|
| EfficientNetB0 | 99.97% | 99.25%      | 99.61%   |
| EfficientNetB2 | 99.96% | 99.13%      | 99.55%   |
| ShuffleNet_v2  | 93.09% | 87.93%      | 90.51%   |
| AlexNet        | 87.12% | 45.75%      | 67.10%   |

The overall prediction accuracy was 99.61%, 99.55%, 90.51%, and 67.10%, respectively, in the order of EfficientNetB0, EfficientNetB2, ShuffleNet\_v2, and AlexNet. As for the recall value, it is the prediction accuracy when a normal image is an input. 99.97%, 99.96%, 93.09%, and 87.12%, respectively, in the order of EfficientNetB0, EfficientNetB2, ShuffleNet\_v2, and AlexNet. As for the specificity value, the prediction accuracy when an abnormal image was input, was 99.25%, 99.13%, 87.93%, and 45.75%, respectively, in the order of EfficientNetB0, EfficientNetB2, ShuffleNet\_v2, and AlexNet.

### IV. CONCLUSIONS

In this paper, CCTV images of Heunginjimun were collected, and a dataset was constructed to determine the displacement of the roof. As for cultural properties, it is important to identify damage before it occurs preemptively, and even if damage occurs, efforts to minimize damage are given priority. For the reasons above, it is challenging to secure damaged images of cultural properties. Even if secured, it is challenging to construct a sufficient dataset to be used in

analysis due to the minute differences. Therefore, this study applied artificial displacements to create a virtual database. In addition, to secure the damaged images' validity, some experts reviewed them in the process. We tested the model's performance by training the model using the built dataset. Among the models tested, the EfficientNetB0 and EfficientNetB2 models have shown high prediction accuracy of 99.61% and 99.55%, respectively, for detecting the roof tilts of Heunginjimun Gate. This experiment presented a method for securing datasets to overcome difficulties in obtaining damage data due to the nature of cultural properties and a method for quickly detecting abnormal signs for wooden cultural properties.

On the other hand, since the difference in accuracy between the two models is insignificant, it has a limitation in that it cannot explain what kind of difference they show. Nevertheless, if an additional experiment is performed to compare the two models in future studies, it is judged that an optimal roof displacement discrimination algorithm can be derived. In addition, the framework's robustness can be further improved by adding more diverse displacement scenarios in future experiments and applying a dataset that mixes environments for learning.

Lastly, suppose models are designed and tested for roof inclination and other types of displacement such as fullness and separation. In that case, it is expected to help significantly manage and preserve cultural properties.

#### ACKNOWLEDGMENT

This paper was funded by the government (Cultural Heritage Administration) in 2022 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 Immoveable cultural heritage)

#### REFERENCES

- [1] Woo-Yeon Choi, Jung-Woo Park, Sang-Yun Lee, "Anomaly Detection of Wooden Cultural Assets Using F-AnoGAN Model," pp. 168-169, 2021
- [2] Jungwoo Park, and Sang-Yun Lee, "Detection of Displacement in Wooden Cultural Properties Using f-AnoGAN Model," pp. in Summer Annual Conference of IEIE, pp 1379-1380, 2022
- [3] Woo-Yeon Choi, Jungwoo Park, and Sang-Yun Lee. "GAN based Deep Learning Model for Detecting Damage and Displacement of Cultural Asset", ICCE-Asia, pp.82-85, Nov 2021
- [4] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs and Ursula Schmidt-Erfurth "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks", PMID: 30831356, Jan 2019
- [5] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, Georg Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery", arXiv:1703.05921, Mar 2017
- [6] Kang, Jaeyong, Kim, Inki, Lim, Hyunseok, Gwak, Jeonghwan, "An Displacement Detection Model in Cultural Asset Images using Object-centric Augmentation", Conference of KSCI, 2021
- [7] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional Neural Networks" In International Conference on Machine Learning (pp. 6105-6114). PMLR, 2019.
- [8] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 116-131, 2018.
- [9] J.Deng, W.Dong, R. Sochere, L.J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition (pp.248-255).IEEE, 2009.



**Sang-Yun Lee**(B'94-M'96-D'07) is a principal researcher at Police Science & Public Safety ICT Research Center of Intelligent Convergence Research Laboratory in ETRI. He has been working at ETRI since 1999. In 2008, he received Ph.D. in Electronics and Telecommunications Engineering at the University of Hanyang(Rep. of Korea). He has been developing technologies in the fields of Broadcasting

Communication, System Software, Embedded Software, Artificial Intelligence, and 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.



**Ho-Hyun Cho** (B'20) was Born in Bucheon, Gyeonggi-do, South Korea. graduated from high school in Daejeon and majored in Business Administration at Korea University. After graduation, he got interested in both blockchain and artificial intelligence. He is currently working for Season, Inc. located in Sejong. His main studies are as follows. First, load cell sensor data analysis was performed by using t-SNE, isolation forest clustering algorithm. Second, antibody analysis modeling for

corona kit was performed by using OpenCV and MaskR-CNN. Next, plasma sensor data analysis and visualization were performed. At the moment, he is contributing to the discovery of new preservation methods for cultural properties by combining artificial intelligence with elements of cultural heritage management. In addition, in his recent studies, he showed interest in the application of artificial intelligence algorithms related to image processing and hopes to continue further research on these in the future.