A Study on Solving the Data Imbalance Problem for Detecting Heunginjimun Roof Tilt Using Transfer Learning Algorithms

Sang-Yun Lee
Police Science & Public Safety ICT Research Center
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
syllee@etri.re.kr

Kang, Seok-Ju
Department of Radio Information and Communication Engineering
Chungnam National University
Daejeon, Republic of Korea
tnswnd35@cnu.ac.kr

***Abstract*—Cultural heritage with high historical value requires continuous management and protection. However, recognizing subtle changes with the naked eye has limitations and requires much time and personnel deployment. To solve this problem, we will automatically detect the tilt of Heunginjimun's roof using Transfer Learning algorithms. In a previous study, among single environments classified into nine types according to season and weather, the ratio of normal and abnormal images in the winter/night and winter/daytime datasets was unbalanced at 9:1 and 8:2. As a result, problems with poor prediction accuracy occurred in some experiments. In this paper, to solve this problem, we adjusted the composition ratio of the dataset and measured the prediction accuracy. When comparing the measurement results with previous studies, the dataset size was reduced by half, but the accuracy was higher. This showed that higher accuracy and performance can be expected by achieving the balance between classes rather than increasing the dataset size.**

Keywords— Roof tilt detection, Heungingimun, Deep Learning, CNN, Composition ratio of the dataset

I. Introduction

Cultural heritage has high historical value and, when located outdoors, is at high risk of damage from natural and artificial influences, so it requires continuous management and protection. However, the existing visual inspection method, in which a person directly or indirectly observes cultural heritage to detect minute changes, has limitations and requires much time and human resources.

In order to replace this existing method, we aim to automatically detect the tilt of the roof of Heunginjimun by using deep learning technology. The deep learning model uses a CNN(Convolutional Neural Network)[1], widely used for anomaly detection. Because CNN learns patterns at the pixel level of images, it can detect irregular patterns or movements well.

In previous studies [2-4], the ratio of the training set, validation set, and test set of the dataset classified into nine single environments according to season and weather was 50:25:25. However, among the nine single environments, the ratio of normal and abnormal images in the winter/night and winter/day datasets was 9:1 and 8:2, so there was a problem with the ratio between the two classes being different. Accordingly, in some experiments, the input accuracy of abnormal images was over 95%, but the input accuracy of normal images was lowered to below 95%.

In this study, the experiment was conducted after solving the data imbalance problem between normal and abnormal images by configuring the ratio of normal and abnormal images in the dataset to be the same. In addition, the total dataset was maintained at 20,000, and the ratio of training set/validation set/test set was adjusted to 50:30:20. The pre-trained models used for learning were the EfficientnetB0[5], Resnet18[6], EfficientnetB2[5], Shufflenet_v2[7], and Efficientnet_v2_s[8] models that showed high accuracy in previous studies

In this paper, we present and analyze the experimental results of this study that solved the data imbalance problem. In addition, we analyze the results of comparing the accuracy of this study and previous studies for each pre-learning model and suggest directions for future research.

II. Heunginjinum for Tilt Detection Methodology

A. *Preprocessing process and dataset composition*

Heunginjimun CCTV footage is classified into nine environments according to season and weather: winter/night, winter/day, spring/night, spring/day, sunny/night, cloudy/day, rainy/day, rainy/night, and snowy/night. Normal images were created by extracting images at regular frame intervals from the classified images, and abnormal images were created by expressing the tilt of the roof from the normal images using the warp function of the Photoshop program. Warp function is a function that distorts or transforms the shape of an image or graphic element. There were a total of 10 places where distortion was applied to the normal image, 5 at the top and 5 at the bottom of the roof, and an abnormal image was created by distorting 50 to 150 pixels in the downward direction for each area. The data imbalance problem between normal and abnormal images was solved by maintaining the same ratio of normal and abnormal images in the dataset at 5:5.

The single environments classified into nine were designated as ENV-01 to ENV-09 in order, and the detailed composition of the single environment dataset is shown in Table I. The training set/validation set/test set ratio of the dataset was composed of 50:25:25, and the data composition of the single environment dataset was visualized as shown in Fig. 2.

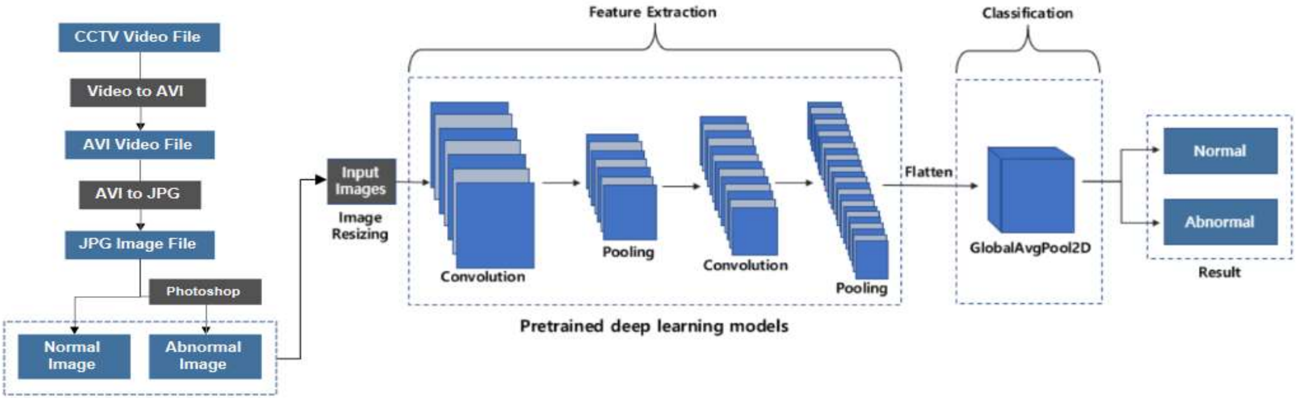


Fig. 1. Heunginjimun roof tilt detection preprocessing process and system framework

TABLE I. SINGLE ENVIRONMENT DATASET COMPOSITION

Dataset	Train		Validation		Test	
	Nor.	*Abnor.*	*Nor.*	*Abnor.*	*Nor.*	*Abnor.*
ENV-01 (winter /night)	2500	2500	1250	1250	1250	1250
ENV-02 (winter /day)	2500	2500	1250	1250	1250	1250
ENV-03 (spring /night)	2500	2500	1250	1250	1250	1250
ENV-04 (spring /day)	2500	2500	1250	1250	1250	1250
ENV-05 (sunny /night)	2500	2500	1250	1250	1250	1250
ENV-06 (cloudy /day)	2500	2500	1250	1250	1250	1250
ENV-07 (rainy /day)	2500	2500	1250	1250	1250	1250
ENV-08 (rainy /night)	2500	2500	1250	1250	1250	1250
ENV-09 (snowy /night)	2500	2500	1250	1250	1250	1250

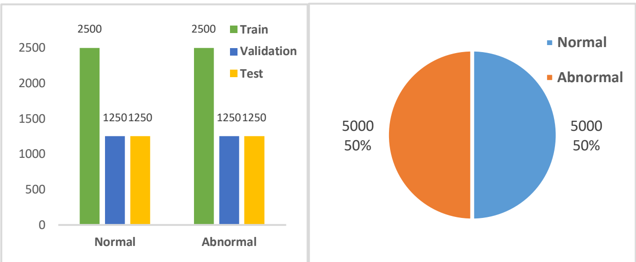


Fig. 2. Visualization of single environment data composition

The environments were created by adjusting the number, and in order, the ratio of integrated single environments was designated as ENV-10 to ENV-15. ENV-15 consists of 9 single environments in equal proportions, and the detailed composition of the integrated environment dataset is shown in Table II. While maintaining a total of 20,000 sheets, the training set/validation set/test set ratio was adjusted to 50:30:20, and the data composition of the integrated environment dataset was visualized, as shown in Fig. 3.

TABLE II. INTEGRATED ENVIRONMENT DATASET COMPOSITION

Data set	Comp.	Train		Validation		Test	
		Nor.	*Abnor.*	*Nor.*	*Abnor.*	*Nor.*	*Abnor.*
ENV -10	winter /night	1250	1250	750	750	500	500
	winter /day	1250	1250	750	750	500	500
	rainy /day	1250	1250	750	750	500	500
	rainy /night	1250	1250	750	750	500	500
	total	5000	5000	3000	3000	2000	2000
ENV -11	winter /night	1250	1250	750	750	500	500
	winter /day	1250	1250	750	750	500	500
	spring /night	1250	1250	750	750	500	500
	spring /day	1250	1250	750	750	500	500
	total	5000	5000	3000	3000	2000	2000
ENV -12	winter /night	500	500	300	300	200	200
	winter /day	2000	2000	1200	1200	800	800
	spring /night	500	500	300	300	200	200
	spring /day	2000	2000	1200	1200	800	800
	total	5000	5000	3000	3000	2000	2000
ENV -13	winter /night	2000	2000	1200	1200	800	800
	winter /day	500	500	300	300	200	200
	spring /night	2000	2000	1200	1200	800	800
	spring /day	500	500	300	300	200	200
	total	5000	5000	3000	3000	2000	2000
ENV -14	winter /night	1500	1500	900	900	600	600
	spring /day	2000	2000	1200	1200	800	800
	rainy /night	1500	1500	900	900	600	600
	total	5000	5000	3000	3000	2000	2000
ENV -15	all env.	5000	5000	3000	3000	2000	2000

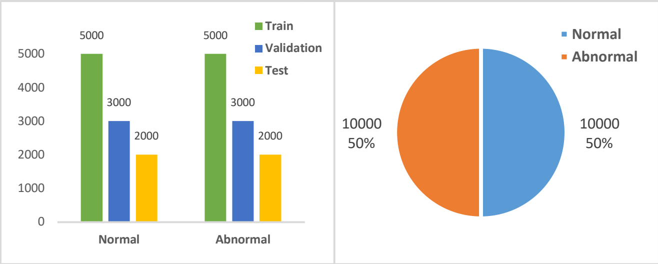


Fig. 3. Visualization of integrated environment data composition

B. *Algorithm and pre-trained model*

In this paper, the preprocessing process and system framework shown in Fig. 1 are used to detect the tilt of the roof of Heunginjimun. Transfer learning technique was used to learn data by importing a pre-trained model. Transfer learning is a learning technique that builds a model by reusing a model learned in a field with abundant training data in a field with insufficient data. By using this technique, we can solve the data shortage problem and shorten the training time.

A binary class dataset consisting of normal and abnormal images is used as the input image for the pre-trained model. The pre-learning models used in the experiment are the EfficientnetB0, Resnet18, EfficientnetB2, Shufflenet_v2, and Efficientnet_v2_s models that have shown high accuracy in previous studies, and are widely used models for image classification and object detection. After training the dataset using a pre-trained model, it classifies normal and abnormal images into two classes.

III. Experimental Results and Analysis

A. *Heungingimun roof tilt detection results*

Table III. shows the results of measuring accuracy in nine single environments after learning the ENV-10~ENV-15 dataset.

ENV-11~ENV-13 are datasets consisting of winter/night, winter/day, spring/night, and spring/day and are constructed so that the ratio of night and day environments varies. The highest accuracy was achieved at 97.52% when using the Efficientnet_v2_s model on the ENV-12 dataset, which consists of 80% of the daytime environment.

ENV-15 is a dataset composed of 9 single environments in equal proportions, and when using the EfficientnetB0 model, it showed the highest accuracy among all experiments at 98.49%.

TABLE III. Heungingimun Roof Tilt Detection Prediction Accuracy

Dataset	**Environment**	**Pre-trained model**	**Accuracy**	**Average**
ENV-10	winter/night 25% winter/day 25% rainy/day 25% rainy/night 25%	EfficientnetB0	95.26%	94.65%
		Resnet18	95.57%	
		EfficientnetB2	**97.77%**	
		Shufflenet_v2	88.08%	
		Efficientnet_v2_s	96.59%	
ENV-11	winter/night 25% winter/day 25% spring/night 25% spring/day 25%	EfficientnetB0	95.98%	94.35%
		Resnet18	93.93%	
		EfficientnetB2	96.55%	
		Shufflenet_v2	88.30%	
		Efficientnet_v2_s	**96.97%**	
ENV-12	winter/night 10% winter/day 40% spring/night 10% spring/day 40%	EfficientnetB0	96.90%	95.48%
		Resnet18	94.76%	
		EfficientnetB2	97.07%	
		Shufflenet_v2	91.16%	
		Efficientnet_v2_s	**97.52%**	
ENV-13	winter/night 40% winter/day 10% spring/night 40% spring/day 10%	EfficientnetB0	95.35%	94.19%
		Resnet18	94.58%	
		EfficientnetB2	95.05%	
		Shufflenet_v2	89.07%	
		Efficientnet_v2_s	**96.91%**	
ENV-14	winter/day 30% spring/day 40% rainy/night 30%	EfficientnetB0	96.36%	93.01%
		Resnet18	94.80%	
		EfficientnetB2	95.07%	
		Shufflenet_v2	80.44%	
		Efficientnet_v2_s	**98.37%**	
ENV-15	All environments	**EfficientnetB0**	**98.49%**	97.15%
		Resnet18	97.59%	
		EfficientnetB2	98.26%	
		Shufflenet_v2	92.93%	
		Efficientnet_v2_s	98.48%	

B. *Pre-trained model accuracy analysis*

Fig. 4 is a graph showing the accuracy of the pre-learning model. Efficientnet_v2_s had the highest accuracy at 97.47%, and Shufflenet_v2 had the lowest accuracy at 88.33%.

The Efficientnet model uses a compound scaling method to achieve high performance with a small amount of calculation through coordination between model depth, width, and image resolution. [5]

The Efficientnet_v2_s model is a successor to the Efficientnet model and is a model that focuses on fast learning. It uses a Progressive Learning method that gradually increases the size of the image during training. [8]

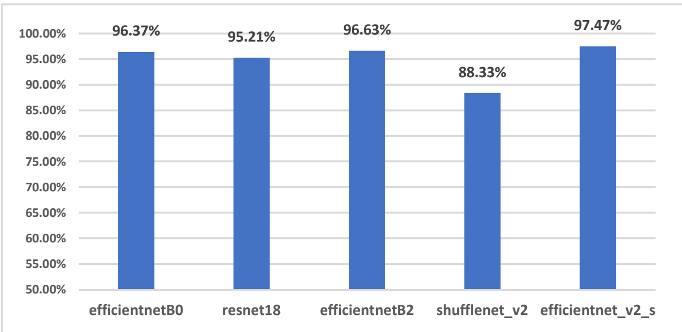


Fig. 4. Pre-training model accuracy graph.

Fig. 5 shows the accuracy graph and loss graph when prediction accuracy was measured using the Efficientnet_v2_s model, which showed the highest accuracy among the five pre-learning models used in this study on the ENV-15 dataset.

The accuracy graph slopes upward and converges to 100, and the loss graph slopes downward and converges to 0, showing that learning was carried out smoothly.

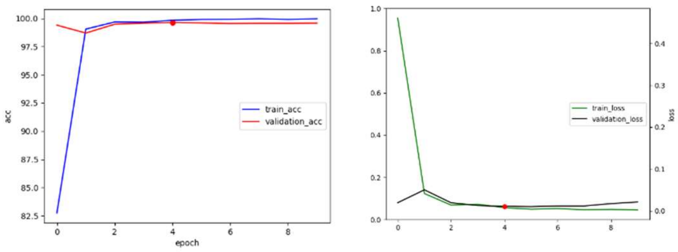


Fig. 5. Efficientnet_v2_s model accuracy/loss graph

Fig. 6 is a picture showing the classification results visualized through a confusion matrix when classifying normal and abnormal images of the test set after learning using the Efficientnet_v2_s model in the ENV-15 dataset.

The test set of ENV-15 is a total of 4,000 images, of which 2,000 are normal images and 2,000 are abnormal images. Among the normal images, 1,994 were determined as normal images, and among the abnormal images, 2,000 were determined as abnormal images.

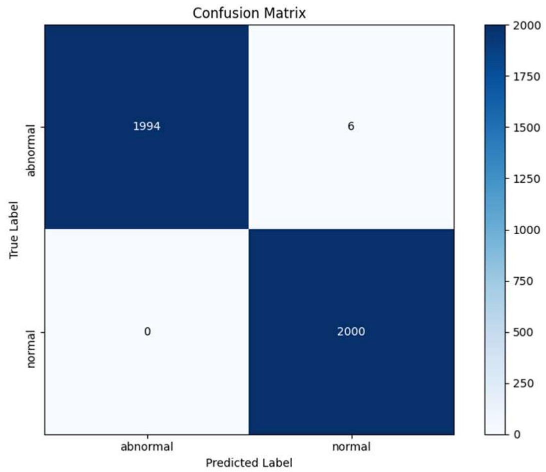


Fig. 6. Efficientnet_v2_s model confusion matrix result

C. Solving data imbalance problem

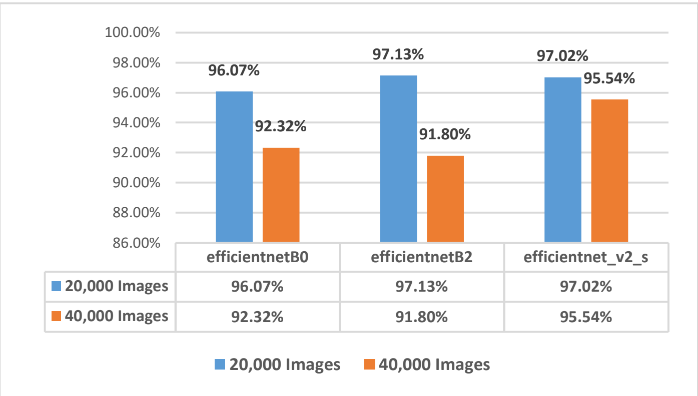


Fig. 7. Accuracy comparison graph before and after solving the data imbalance problem

In the row study, the dataset consisting of four single environments was 10,000 images for each single environment, for a total of 40,000 images. On the other hand, in this study, the number and ratio of integrated environments were adjusted while maintaining a total of 20,000 datasets.

In a previous study, the ratio of normal images to abnormal images in the winter/night and winter/daytime datasets among the nine single environments was 9:1 and 8:2, so there was a problem with the ratio between the two classes being different. Accordingly, in some experiments, the input accuracy of abnormal images was over 95%, but the input accuracy of normal images was lowered to below 95%.

As a result of comparing the accuracy of each pre-learning model between this study and previous studies, which solved the imbalance problem by setting the ratio of both normal and abnormal images in the dataset to 5:5, the dataset size was reduced by half, but higher accuracy was achieved. The detailed comparison results are shown in Fig. 7.

D. Accuracy verification using Grad-CAM

TABLE IV. GRAD-CAM VISUALIZATION RESULTS BY DISTORTION AREA

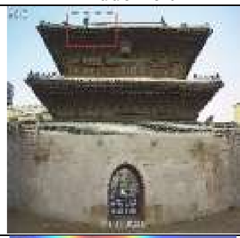
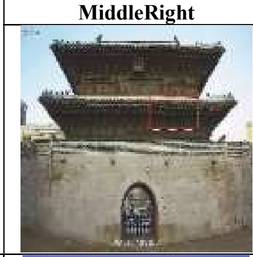
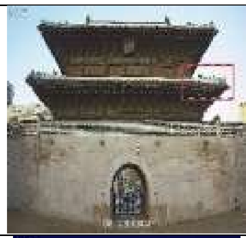
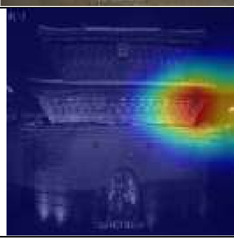
	Upperleft	**Upper MiddelLeft**
Distortion area		
Grad-CAM		
	Lower MiddleRight	**LowerRight**
Distortion area		
Grad-CAM		

TABLE IV. compares the distortion area of the classified image and Grad-CAM(Class Activation Map) [9] results by distortion area when the image was classified after learning using the Efficientnet model in the ENV-15 dataset. Since the heatmap is distributed around the distorted area, it can be seen that the model classified the image using the distorted area.

IV. CONCLUSION

Continuous management and observation are necessary to protect cultural assets. Still, the existing visual inspection method, in which humans directly or indirectly detect minute changes, requires a lot of time and personnel deployment.

To overcome this problem, research was conducted to automatically detect the tilt of the roof of Heunginjimun using a Transfer Learning algorithm.

In a previous study, the ratio of normal images to abnormal images in the winter/night dataset and winter/day dataset among nine single environments was unbalanced at 9:1 and 8:2, causing problems with classification accuracy in some experiments.

In this paper, the data imbalance problem was solved by adjusting the ratio of normal and abnormal images in the dataset to 5:5, and then the prediction accuracy was measured. As a result of comparing the accuracy of each pre-learning model before and after solving the data imbalance problem, when the data imbalance problem was solved, the entire dataset was reduced by half from 40,000 to 20,000 pieces, but the accuracy was higher.

This showed that higher accuracy and performance can be expected by achieving the balance between classes rather than increasing the size of the dataset.

In future research, it is expected that a more accurate and reliable prediction model can be built by balancing the ratio between classes by considering the imbalance problem of the dataset.

ACKNOWLEDGMENT

This paper was funded by the government (Cultural Heritage Administration) in 2024 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] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, pp. 541-551, 1989.

[2] S. Y. Lee, H. H. Cho, and D. G. Lee, "Heunginjimun Displacement Determination Using Artificial Intelligence Algorithm," Workshop on Image Processing and Image Understanding, pp. 1-26, 2023.

[3] S. J. Kang, S. Y. Lee, "A Study on Performance Lmprovement of Heunginjimun Roof Displacement Determination Model Using Artificial Intelligence Algorithm," Summer Annual Conference of IEIE, pp. 1050-1054, 2023.

[4] S. Y. Lee, S. J. Kang, "Heunginjimun roof tilt prediction experiment and analysis based on deep learning based on environment integration," J. The Society of Culthral Heritage Disaster Prevention., in press

[5] M. Tan, and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning," J. Proc. Mach. Learn. Res, pp. 6105-6114, May 2019.

[6] K. He, X. Zhang, S. Ren, and J Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[7] 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, pp. 116-131, 2018.

[8] M. Tan, and Q. Le, "Efficientnetv2: Smaller models and faster training," J. Proc. Mach. Learn. Res, pp. 10096-10106, July 2021

[9] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," In Proceedings of the IEEE international conference on computer vision, pp. 618-626, 2017