Collapsed Building Detection Using Residual Siamese Neural Network On LiDAR Data

Mgs M Luthfi Ramadhan[1], Grafika Jati[1], Machmud Roby Alhamidi[1], Riskyana Dewi Intan P[1], Muhammad Hafizhuddin Hilman[2], and Wisnu Jatmiko[1]

[1]Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
[2]Faculty of Engineering & IT, The University of Melbourne, Melbourne, Australia
mgs.m01, grafika.jati51, machmud.robi, riskyana.dewi{@ui.ac.id}, muhammad.hilman@unimelb.edu.au, wisnuj@cs.ui.ac.id

***Abstract*—Evaluation of buildings is crucial to aid emergency response but it costs a lot of resources to do it manually. Many approaches have been proposed to automate the process using artificial intelligence. Most of them, use handcrafted feature, difference calculation between pre-disaster and post-disaster feature, and a classifier model separately. In this study, the process from feature extraction, feature difference and classification are represented by a single model which is siamese neural network. Furthermore, we modify siamese neural network by implementing residual connection for feature concatenation purposes. We evaluate our model on Kumamoto Prefecture earthquake LiDAR data. The result shows the modified model is able to outperform the baseline model with Accuracy and F-measure of 90.91% and 79.28% respectively.**

Keywords— earthquake, siamese neural network, deep learning, collapsed building assessment

I. Introduction

Buildings are one of the most crucial things for humans to live. However, buildings are very vulnerable to disasters such as earthquakes which make them seriously damaged or even completely collapsed [1]. Thus, it is important to immediately observe building conditions after an earthquake to aid in emergency response and rescue operations [2, 3]. Much of the observing process is done manually by doing a field survey which is costly and time-consuming. Moreover, a field survey cannot be done in some cases due to broken or blocked roads caused by the earthquake [1, 4, 5, 6]. As an alternative, remote sensing can be used to observe the affected area. Although visual observation can assess building damage from remote sensing data, it's still cost a lot of resource and time to observe the affected area [7]. As a result, an automated damage detection system that is accurate and usable is required [8, 9].

Much research has been done to automate the process of assessing building damage through remote sensing data using artificial intelligence. For example, Miura et al. [10] conducted research to classify collapsed buildings in Kumamoto based on post-earthquake images using CNN. The challenge of assessing building damage in Kumamoto is that its citizens tend to cover their rooftop with a blue tarp as a coping mechanism. This habit makes it harder to assess the building. Miura et al. managed to get an accuracy of 95% and concluded that most of the blue tarp-covered buildings are moderately damaged. Min ji [1] using pre-earthquake and post-earthquake images, examined CNN feature and GLCM feature to classify collapsed buildings in Port-au-Prince. They concatenate features from both pre-earthquake and post-earthquake images. They then classify the obtained feature using Random forest classifier (RDF). They concluded that CNN-RDF performs better than GLCM-RDF with an accuracy of 87%. Hajeb et al. [4] use multimodal SAR and LiDAR data to classify building damage. Both SAR and LiDAR data are acquired before and after the mainshock of the Kumamoto earthquake occur. They rasterize the LiDAR data and export it as DSM. They then use DSM difference, GLCM difference, and coherence difference (SAR data) as features. With the resulting feature, they classify it using SVM and RDF. They concluded that SVM with DSM difference feature performs better with an accuracy of 87%. Moya et al. [3] use pre-earthquake and post-earthquake LiDAR data to classify building damage. Just like Hajeb et al. [4], they transform LiDAR data into DSM. They use DSM difference, standard deviation, and correlation coefficient as features and classify it using SVM. They got an accuracy of 93%. As an addition, Zhang [11] utilize siamese neural network to detect building structural change. The two identical sub-networks of siamese neural network receive two different inputs which are the difference between old DSM and new DSM and R, G, B channels from the orthoimage. As a comparison, they stacked the old DSM, new DSM, R, G, and B as a single tensor to train a CNN model. They concluded that the Siamese neural network performs better than CNN with an accuracy of 76%.

However, in [11] study, the DSM difference is calculated outside of siamese neural network resulting in a lack of difference feature whereas this difference feature is the most essential factor in deciding whether a building is collapsed or not [3]. Thus, in this study we utilize siamese neural network containing subtraction layer inside the model to calculate meaningful difference feature on a higher dimension. Furthermore, we modify the model by implementing residual connections to concatenate features from different levels. By using residual connections, we get even richer features so that we can analyze the difference between the two data better, so the model will produce better performance. Therefore, the novelty of this study is that we develop a siamese neural network that takes into account features from different levels to calculate the similarity or difference between the two inputs. The remainder of this paper is organized as follows, section 2 explaining how the data is obtained and processed.

Fig. 1. Study area (inside the green polygon)

2021 6th International Workshop on Big Data and Information Security (IWBIS) | 978-1-6654-2451-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IWBIS53353.2021.9631844

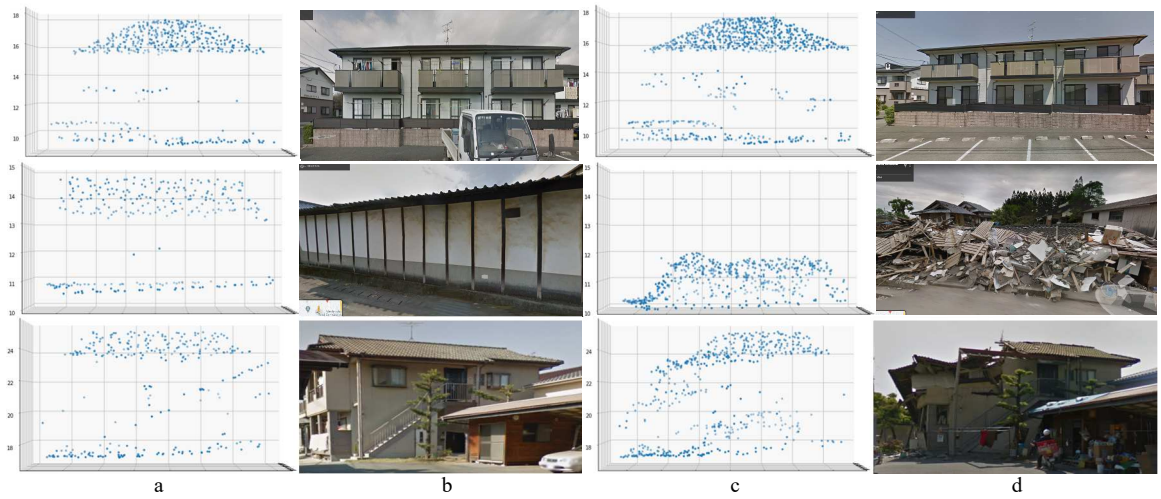

Fig. 2. Point cloud and google street view. A, b, c, and d representing pre-earthquake point cloud, pre-earthquake street view, post-earthquake point cloud, and post-earthquake street view respectively. The first row is an example of a non-collapsed building while the second and third are collapsed buildings.

Section 3 describes the siamese neural network method that we proposed. Section 4 is the results and discussion. Finally, section 5 concludes the paper.

II. DATA

In this study, we use Kumamoto Prefecture earthquake LiDAR data. Most of the heavily damaged buildings are located around Mashiki town [4, 12]. Thus, in this study we only use the data at Mashiki town. To be specific, it is located between Route 28 and Akizu river as shown in figure 1. This LiDAR data can be downloaded from Open Topography [13, 14]. It consists of two point clouds namely pre-earthquake and post-earthquake. The ground truth for each building is not included in the data, thus we annotate each building by ourselves. We considered 7 damage rates from the Architectural Institute of Japan (AIJ) namely D0, D1, D2, D3, D4, D5, and D6, which then reduced to be Non-collapsed and collapsed as shown in table I [4, 10].

TABLE I. AIJ DAMAGE RATE

No	Picture	Label	Description	Used Label
1		D0	No damage	Non-Collapsed
2		D1	Hair-line cracks in walls	Non-Collapsed
3		D2	Falls of plaster of walls; Detachment of roof tiles	Non-Collapsed
4		D3	Significant incline of columns; Serious failure of beams and walls	Non-Collapsed
5		D4	Significant incline of columns; Serious failure of beams and walls	Non-Collapsed
6		D5	Partial collapsed	Collapsed
7		D6	Complete collapsed	Collapsed

The annotation is done by using CloudCompare. With the help of Google Earth, we extract each building by carefully creating polygons on both pre-earthquake and post-earthquake point clouds. We then annotate the data by visualizing the point clouds and observing them on Google Earth. Figure 2 shows the example of non-collapsed and collapsed building point clouds.

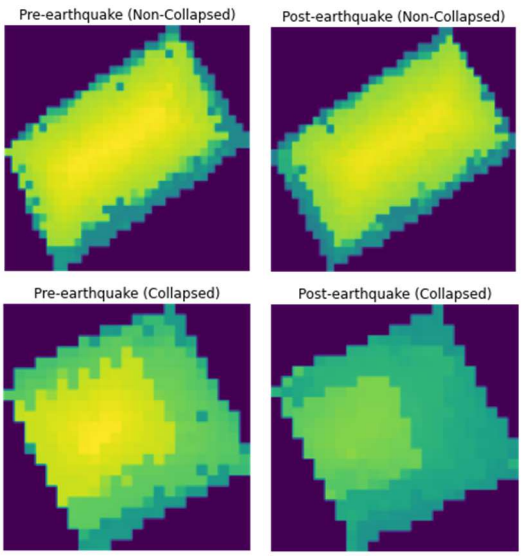

Fig. 3. Pre-earthquake and post-earthquake DSM of non-collapsed and collapsed buildings

Most machine learning and deep learning algorithms expect the input to be fixed while each building has a different number of point. In the case of disaster assessment, the point cloud should not be sampled because sampling a point cloud

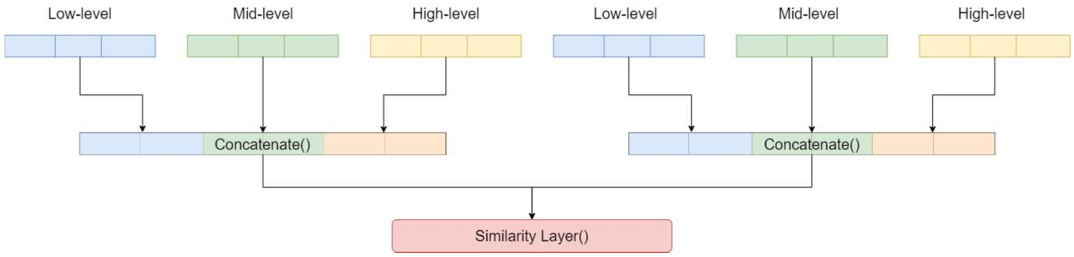

Fig. 4. Concatenation illustration

will cause several points to be removed, resulting in different point cloud patterns between the same non-collapsed building. Thus, in this study, we transform the point cloud into DSM, as shown in figure 3.

As many as 1.011 buildings have been annotated. 750 of which are non-collapsed while the remaining 261 are collapsed. Due to imbalanced data, we conducted a train-test-split using the stratified method with a ratio of 75:25 for training and testing respectively. The Stratified split will equalize the ratio of non-collapsed and collapsed classes in the training and testing data to reduce the potential bias towards the resulting model.

III. Proposed Method

In this study, we utilize Siamese neural network to classify collapsed or non-collapsed buildings. Siamese neural network was first introduced by [15] for handwritten signature verification. It consists of 2 sub-networks that are identical to each other. The two sub-networks have the same architecture and weight; this aims to extract features with the same weight for two different inputs. After extracting the features, both features are passed to the similarity layer to measure the similarity between the two. Overall this model accepts two inputs then produces the similarity between the two. In our study, the two inputs are pre-earthquake DSM and post-earthquake DSM. As Moya [3] stated in their study, the most important factor in deciding whether a building is collapsed or not is the height difference between pre-earthquake and post-earthquake; therefore, the Siamese neural network can be used for this problem. If the pre-earthquake DSM and post-earthquake DSM are similar, the building is likely non-collapsed, on the other hand, if the pre-earthquake DSM and post-earthquake DSM are not similar, the building is highly likely to collapsed. The similarity is measured using Euclidean distance.

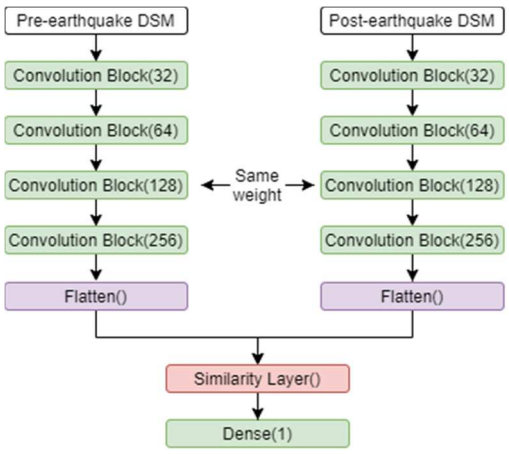

Fig. 5. Baseline siamese neural network architecture

$$d(a,b) = \sqrt{\sum_i^n (a_i - b_i)^2} \quad (1)$$

Equation 1 shows the formula for Euclidean distance with a and b representing pre-earthquake feature and post-earthquake feature respectively. As a comparison we also propose element-wise subtraction as a replacement to euclidean distance. The element wise-subtraction works by subtracting each element from each feature vector

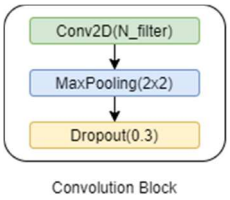

Fig. 6. Convolution block illustration

correspondingly. Unlike euclidean distance which produces a scalar, element wise subtraction produces an *n*-dimensional vector. Equation 2 shows the formula for element-wise subtraction with a, b, and c representing pre-earthquake feature, post-earthquake feature, and the resulting element-wise subtraction between both features respectively.

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} - \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} \quad (2)$$

The pre-earthquake feature and post-earthquake feature are extracted automatically by using two identical convolutional neural networks. The overall baseline architecture is illustrated in figure 5 with each convolutional block illustrated in figure 6. The similarity layer which we mentioned inside the illustrations in figure 4, 5, and 7 is either a Euclidean distance or subtraction layer.

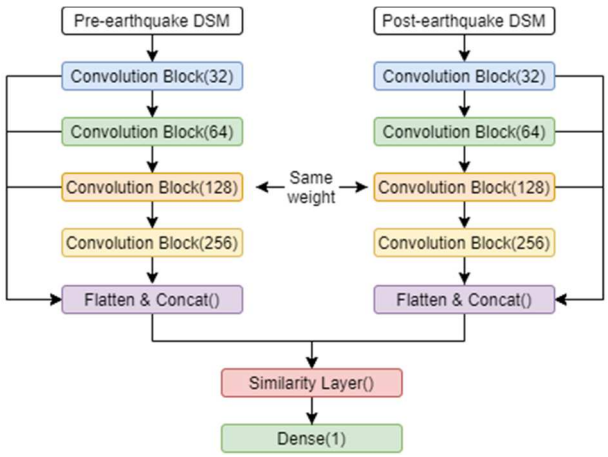

Fig. 7. Residual siamese neural network architecture

To enhance the result, we modify the model by implementing residual connections. The residual connection was first introduced by He et al. [16]. It is used to propagate a larger gradient for the front layer so that the front layer can learn as fast as the deeper layer do. Simply put, it is usually used to build a deeper neural network and prevent vanishing gradient at the same time. In this study, we use the residual connection for feature concatenation purposes as shown in figure 7. The idea behind this is to extend the number of features that are being fed to the similarity layer by concatenating the feature from the low level all the way through to the high-level as illustrated in figure 4. By doing so, the similarity layer is not only calculating the similarity of the high-level feature, instead, it calculating the similarity of the concatenation of feature from low-level to high-level. To ease the concatenation process, each feature map from each

$$\vec{u} = \{u_0, u_1, \dots, u_n\}$$
$$\vec{v} = \{v_0, v_1, \dots, v_n\} \quad (3)$$
$$\vec{w} = \{w_0, w_1, \dots, w_n\}$$
$$\vec{x} = [\vec{u}, \vec{v}, \vec{w}]$$

convolution block is flattened before concatenating. Equation 3 shows the formula for concatenation where u, v, and w representing the resulting feature vector from each convolution block while x representing the resulting concatenation vector. The x dimension is equal to the summation of u dimension, v dimension, and w dimension. When using subtraction as similarity layer, the subtraction is done by subtracting each level of feature from pre-earthquake DSM and post-earthquake DSM correspondingly. By doing so, this model is not only analyzing differences in high-level features, instead, it analyzes the difference from low-level feature all the way through to the high-level feature. The result of this subtraction will provide more features and information for the output layer to make the decision whether the input building is collapsed or not. Thus leading to better classification performance.

We implemented the model using Python programming language with the help of Tensorflow. The script was run in Google Colaboratory with NVIDIA Tesla K80 GPU and 12 GB of RAM. We lock random seed for Tensorflow and Numpy with the value of 1.234 and 0 respectively. By doing so, it is guaranteed that each trial will have the same random treatment. The model is trained using Adam optimizer as many as 100 iterations with a learning rate of 0.001 and batch size of 32.

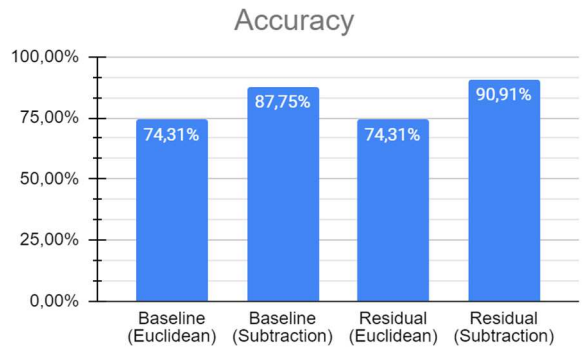

Fig. 8. Accuracy of proposed method compared with baseline

IV. Result and Discussion

Four trials have been done in this study. In the first trial, we evaluate the baseline model with Euclidean distance layer, the second trial we evaluate baseline model with subtraction layer, the third trial we evaluate the residual model with Euclidean distance, and the fourth trial we evaluate the residual model with subtraction layer. The results of these trials are shown in table II.

TABLE II. Experiment result of proposed method compared with baseline

Model	Precision	Recall	F-Measure
Baseline (Euclidean)	0%	0%	0%
Baseline (Subtraction)	**100%**	52.31%	68.69%
Residual (Euclidean)	0%	0%	0%
Residual (Subtraction)	95.65%	**67.69%**	**79.28%**

It can be inferred that the model with Euclidean layer both baseline and residual fail to converge. Both models have an f-measure, precision, and recall of 0% which means that the model cannot differentiate which building is collapsed and which building is non-collapsed. We assume this happens due to the lack of feature and information provided by Euclidean distance for the output layer since Euclidean distance only produces a scalar. On the other hand, the subtraction layer is able to make better predictions compared to Euclidean distance for both baseline and residual model. The non-zero score of f-measure, precision, and recall are indicators that the model is able to differentiate which building is collapsed and which building is non-collapsed. This happens because the subtraction layer produces an *n*-dimensional vector which leads to a richer feature. The precision score of baseline model with subtraction layer is 100% which indicates that the model never mistaken a non-collapsed building as collapsed. This happens due to the imbalance of data that favor the non-collapsed class, thus the model is more likely to predicting non-collapsed instead of collapsed. The precision score of residual model with subtraction layer is lower than the baseline model but as an exchange, it has significant improvements on recall which worth the tradeoff. This means that the model is less biased toward the non-collapsed class that the model is predicting collapsed class more frequently than the baseline model does.

Figure 8 shows the accuracy of these trials. The models with Euclidean layer have an accuracy of 74.31% which is a very strong indicator that the model is just straight predicting all testing data as non-collapsed since 74.31% is a ratio of non-collapsed class in testing data. On the other hand, models with subtraction layer have greater accuracy, it indicates that there are several collapsed building that is predicted correctly. When using subtraction layer, the residual connections provide more feature for subtraction layer which result on even richer difference feature that calculated from several levels of feature. This makes output layer easier to do classification, thus it has better classification performance. However, the residual model with subtraction layer still mistaken a collapsed building as a non-collapsed building a lot as shown in figure 9 which leads to a low recall score. Just like what happened to our baseline model, it might be caused

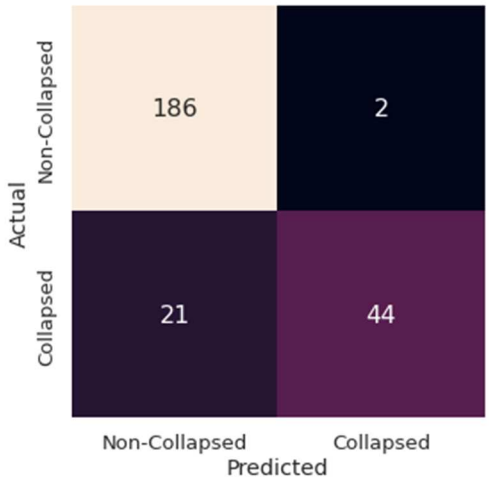

Fig. 9. Confusion matrix of residual siamese neural network with subtraction layer

by the imbalanced class in the training set that favors the non-collapsed class. Thus, further research needs to be done to improve the classification performance such as oversampling the minority class so that the model less biased towards the majority class. Another probable improvement is by implementing cost-sensitive learning that will allow the model to pay more attention to the minority class.

V. Conclusion

This study conducted a collapsed building detection using residual siamese neural network on LiDAR data. The Euclidean layer on siamese neural network is not optimally used in the case of classification. We assume it happens due to the lack of features provided by Euclidean distance since it only produces a scalar, as an alternative this research replaces the layer into an element-wise subtraction layer. We also study the impact of implementing residual connections on the siamese neural network. The experimental results show that the implementation of the residual connection is able to improve the classification results. However, the residual model with subtraction layer still mistaken a collapsed building as a non-collapsed building a lot as shown in figure 9. Further improvement needs to be done such as oversampling the minority class or implementing cost-sensitive learning.

Acknowledgment

We would like to thank Asia Air Survey Co. Ltd for providing the pre-earthquake and post-earthquake LiDAR data. This work was also supported by Indonesia Ministry of Research and Technology/National Research and Innovation Agency 2021, No: NKB-232/UN2.RST/HKP.05.00/2021.

References

[1] Ji, M., Liu, L., Du, R., & Buchroithner, M. F. (2019). A comparative study of texture and convolutional neural network features for detecting collapsed buildings after earthquakes using pre- and post-event satellite imagery. Remote Sensing, 11(10). https://doi.org/10.3390/rs11101202.

[2] Das, R., & Hanaoka, S. (2014). An agent-based model for resource allocation during relief distribution. Journal of Humanitarian Logistics and Supply Chain Management, 4(2), 265–285. https://doi.org/10.1108/JHLSCM-07-2013-0023.

[3] Moya, L., Yamazaki, F., Liu, W., & Yamada, M. (2018). Detection of collapsed buildings from lidar data due to the 2016 Kumamoto earthquake in Japan. Natural Hazards and Earth System Sciences, 18(1), 65–78. https://doi.org/10.5194/nhess-18-65-2018.

[4] Hajeb, M., Karimzadeh, S., & Matsuoka, M. (2020). SAR and LIDAR datasets for building damage evaluation based on support vector machine and random forest algorithms—A case study of Kumamoto earthquake, Japan. Applied Sciences (Switzerland), 10(24), 1–18. https://doi.org/10.3390/app10248932.

[5] Zhai, W., Shen, H., Huang, C., & Pei, W. (2016). Building earthquake damage information extraction from a single post-earthquake PolSAR image. Remote Sensing, 8(3). https://doi.org/10.3390/rs8030171.

[6] Wie, S., Wang, X., & Sato, M (2016). Urban Damage Level Mapping Based on Scattering Mechanism Investigation Using Fully Polarimetric SAR Data for the 3.11 East Japan Earthquake. 54(12), 6919–6929. hhtps://doi.org/ 10.1109/TGRS.2016.2588325.

[7] Xie, S., Duan, J., Liu, S., Dai, Q., Liu, W., Ma, Y., Guo, R., & Ma, C. (2016). Crowdsourcing rapid assessment of collapsed buildings early after the earthquake based on aerial remote sensing image: A case study of Yushu earthquake. Remote Sensing, 8(9). https://doi.org/10.3390/rs8090759.

[8] Fujita, A., Sakurada, K., Imaizumi, T., Ito, R., Hikosaka, S., & Nakamura, R. (2017). Damage detection from aerial images via convolutional neural networks. Proceedings of the 15th IAPR International Conference on Machine Vision Applications, MVA 2017, 5–8. https://doi.org/10.23919/MVA.2017.7986759.

[9] Valentijn, T., Margutti, J., van den Homberg, M., & Laaksonen, J. (2020). Multi-hazard and spatial transferability of a CNN for automated building damage assessment. Remote Sensing, 12(17), 1–29. https://doi.org/10.3390/rs12172839.

[10] Miura, H., Aridome, T., & Matsuoka, M. (2020). Deep learning-based identification of collapsed, non-collapsed and blue tarp-covered buildings from post-disaster aerial images. Remote Sensing, 12(12). https://doi.org/10.3390/rs12121924.

[11] Zhang, Z., Vosselman, G., Gerke, M., Persello, C., Tuia, D., & Yang, M. Y. (2019). Detecting building changes between airborne laser scanning and photogrammetric data. Remote Sensing, 11(20). https://doi.org/10.3390/rs11202417.

[12] Yamada, M., Ohmura, J., & Goto, H. (2017). Wooden building damage analysis in mashiki town for the 2016 kumamoto earthquakes on April 14 and 16. Earthquake Spectra, 33(4), 1555–1572. https://doi.org/10.1193/090816EQS144M.

[13] Pre-Kumamoto Earthquake (16 April 2016) Rupture Lidar Scan. OpenTopography, High-Resolution Topography Data and Tools. Available online: https://portal.opentopography.org/lidarDataset?opentopoID=OTLAS.052018.2444.2.

[14] Post-Kumamoto Earthquake (16 April 2016) Rupture Lidar Scan. OpenTopography, High-Resolution Topography Data and Tools. Available online: https://portal.opentopography.org/lidarDataset?opentopoID=OTLAS.052018.2444.1.

[15] Bromley, J., Guyon, I., Lecun, Y., Sckinger, E., & Shah R. (1994). Signature Verification using a "Siamese" Time Delay Neural Network.

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem, 770–778. https://doi.org/10.1109/CVPR.2016.90.

978-1-6654-2451-6/21//$31.00