International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Multi Disaster Building Damage Assessment with Deep Learning using Satellite Imagery Data

Edy Irwansyah*[1], Hansen Young[2], Alexander A. S. Gunawan[3]

Submitted: 23/10/2022 **Revised:** 11/12/2022 **Accepted:** 27/12/2022

Abstract*:* Massive disaster events that have an impact on infrastructure damage, especially buildings, require the latest tools and technology so that an assessment of the damage that occurs can be carried out quickly and efficiently. Artificial Intelligence (AI), especially with machine learning (ML) and deep learning (DL), is one approach that can be a solution. The study uses satellite imagery data from the xBD dataset repository with a proposed two-stage deep learning model to assess the level of damage to buildings consisting of a building segmentation approach using the PSPNet and UNet models with the ResNet-18 backbone and the ResNet50 model for classification. The flood fill algorithm is inserted between the segmentation and classification stages with the aim of producing better extraction of segmentation results. The deep learning model with PSPNet and ResNet-18 produces an evaluation value and an accuracy value of building damage classification which is slightly better than the previous building segmentation research, with F1 values of 0.8494 and 0.8338 for precision, respectively. Referring to the resulting evaluation value, future research is still very open to developing models to achieve better results.

Keywords: *Disaster, building damage, deep learning, satellite imagery*

1. Introduction

The age of the earth that is getting older is marked by the earth system that is getting more and more unstable. One indicator is the increasing number of natural disasters around the world. Based on the disaster typology, droughts, floods, and earthquakes are disasters with the largest death impact in this century. The drought hazard proved to be the deadliest hazard over the past 10 years causing 650,000 deaths, followed by storms which caused 577,232 deaths; the flood, which claimed 58,700 lives; and extreme temperature events, in which 55,736 died [1]

Each year, numerous natural disasters happen all around the world. Natural disasters happen due to the natural process of the earth or the results of human acts. In both cases, a large-scale natural disaster may threaten the lives of many people and disturb the economics of a country. The overall number of casualties in 2019 was 6,133,500 persons, including the dead, missing, injured, suffering, and displaced. The number of catastrophe victims climbed to 6,324,534 persons in 2020, as the number of disasters increased. The number of catastrophe victims climbed to 6,324,534 persons in 2020, as the number of disasters increased [2], Although pre-disaster system may provide warnings about the incoming disaster, the damage done may not always be completely avoidable. This is especially true for immovable assets such as buildings or houses. With this problem stated, damage assessment is needed to provide support and reduce the loss caused by the natural disasters.

Currently, the collection and evaluation of disaster data after the event especially for damage assessment was conducted manually so that it takes time and resources. In many disaster scenarios, data collection is obtained in the field by surveying the population, which is time-consuming and may not provide reliable results due to factors such as the time and location of the incident, or the use of outdated demographic data [3][4][5], The reason for the large amount of time and resources is the reason many disaster researchers tried to find alternative methods for solving these problems. On the other hand, developments in remote sensing (RS) technology and computer science, especially ML and DL, make it possible to become tools and method solutions for faster and more accurate disaster impact assessments.

According with [6], the use of DL with RS data only started in 2015 and increased significantly in the 2018s with the most applications in the field of land use and landcover (LULC) analysis. In particular, the use of DL in damage assessment is still very rare even though it has been acknowledged by the joint regarding the prospect of using those method. As for the implementation of DL

[1] *Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta –11530, Indonesia*
ORCID ID: 0000-0002-3876-1943
[2] *Department of Mathematic, School of Computer Science, Bina Nusantara University, Jakarta –11530, Indonesia*
[3] *Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta –11530, Indonesia*
ORCID ID: 0000-0002-1097-5173
** Corresponding Author Email: eirwansyah@binus.edu*

specifically in the case of disaster response, it is known that it only started in 2018 – 2020 [7][8]. Research conducted [9], proposes the use of two encoders to extract features from both pre-disaster and post-disaster image with a common decoder to do a semantic segmentation process for classifying burned buildings from unburned ones. Those research gives a model with 0.859 F1-score for burned buildings and 0.818 for unburned buildings. In addition to previous research, there has also been research by [10] about using various convolutional neural networks (CNN) to classify flooded and unflooded buildings. The result shows that data augmentation and drop out layer combined with adam optimizer can aid in the training process to give a better accuracy. The actual research by Bouchard [11] with their DamageNet architecture specifically for building damage assessment using the xBD dataset, the new F1-score reached 0.846 for building segmentation and 0.709 for damage classification. All researchers suggest that DL should be used for the assessment of buildings after natural disasters.

Based on the previous research achievement, this research attempts to take on the problem one step forward by trying to quantify the damage more by using three damage levels scale from the xBD dataset. It consists of very-high-resolution (VHR) pre- and post-disaster remote sensing images from 18 disaster events worldwide, containing a diversity of climate, building, and disaster types. The dataset is annotated with building polygons classified according to a joint damage scale with four ordinal classes: no damage, minor damage, major damage, and destroyed [12]. Furthermore, this research will focus only on the post-disaster images because pre-disaster images are not always available. The model that will be used for this research are PSPNet and UNet for the segmentation process, and ResNet for the classification.

The following is the structure of this paper include: (1) introducing about natural damage impact and related problems; (2) identifying state of the art DL algorithm that has been in development for damage assessment using RS data; (3) Dataset and development of DL models; (4) results; and the (5) conclusion.

2. Deep Learning for Damage Assessment

2.1. Development of DL Algorithm

DL is a subset of what is known as ML, an AI algorithm was introduced in 2015 as a method of data learning with the aim to create representative data with layers using data preprocessing where the representative data is not explicitly made by humans but created by a learning algorithm [13]. The development of the DL algorithm is mentioned as the impact of a significant achievement from the use of the CNN algorithm with the AlexNet architecture in the ImageNet contest, one of which is supported by the efficiency in the use of GPU graphics processing units [14]. LeNet5 model is the first CNN model that utilizes structural information in the data through a few convolution and pooling operations inside the NN layer. This proves to be successful in handling data classification problems thus encouraging the development of more complex DL models for image analysis using RS data [15] . Deep attention regarding the use of DL in RS data analysis has started since 2014. Although it was small in numbers, it reached a significant number in 2017-2018 in various applications such as LULC classification and geographic object detection [6]. DL is specifically implemented for damage assessment using RS data with DL in 2017 at the Proceedings of the 11th international workshop on structural health monitoring -IWSHM [16]. The following discussion specifically introduces some of the popular DL models including the CNN model which was implemented in this research and the fully convolutional networks (FCN).

CNN is the most extensive and massive DL model for its use in various fields including remote sensing application. According to [6], the characteristics of the CNN model are very suitable for processing multiband RS data composed of a regular number of pixels. CNN has a fully connected network, which is a network structure of interconnected nodes. LeNet5 is the first CNN model that utilizes structure information in image data for character recognition through a few convolution and pooling operations [15] for example: adjacent pixels in an image or adjacent words in a text. Additionally, the CNN model has lower complexity, faster model training time, and requires fewer training data samples compared to traditional neural networks (NN) [16]. In general, a CNN model consists of several layer, namely (1) Input layer: functions to generate input data, (2) Convolutional layer: function for convolution operations on a number of nodes using several filters, (3) ReLU layer: is an activation function from output layer previously, (4) Pooling layer: this layer serves to sample the output of the previous layer to produce a structure with smaller dimensions, and (5) Fully connected layer: functions to calculate the output of the last layer [15]. Furthermore, some of the architectures of popular CNN models such as VGGNet [17] , UNet [18], ResNet [19] as well as DeepLab [20] will be discussed. The development of CNN models such as DeepCNN [21], RCNN [22] and FastRCNN [23] cannot be excluded from the discussion as these methods have been used for this research review.

FCN is the most used deep learning architecture, especially in semantic segmentation. It uses locally connected layers, such as convolution, concatenation, and up sampling. Avoiding the use of dense layers means fewer parameters, thus making the network faster in the training process. It also means that FCN can work for variable size images as all connections are local [24]. Taking advantage of this

FCN's rich class of models and special case of modern classification convnets, it is possible to extend the classification network for segmentation and improve the architecture with multi-resolution layers. This combination dramatically improves the state-of-the-art while simultaneously simplifying and speeding up the learning and inference [25]. One of FCN developments for damage assessment is by using skip connection to develop 3D point cloud instances. Traditionally, the development of this method requires many training iterations to achieve acceptable accuracy. However, the modification of skip connections allows the network to recover the most useful features during the training process at a faster rate. It also helps the recovery of finer details in precisions and reduce gradient vanishing issues [26]. In general, FCN model consists of several layers namely: (1) Input layer: functions to generate input data; (2) Convolution layers: this includes encoding and decoding layer, which is very common to find in semantic segmentation algorithm such as FCN; (3) Pooling: this functions to reduce the resolution of the convoluted features; (4) Un-pooling: sending the learned filters back to capture class-specific information; (5) Fully connected layer: this functions to fuse the output and giving the classification results. The development of FCN models such as 3D FCN [26] and FCN [27].

2.2. Deep Learning Algorithm for Damage Assessment

In damage assessment, there are variations and roles for scene and object. Even though this review only uses a few scene datasets, it is critical to understand the differences between the two because they are frequently misconstrued. Scene categorization is based on visual categories from many photographs as well as training samples from labeled photographs [6]. Coastal areas [28], forest [29], and other types of scenes are some instances of scene classification. Object detection, on the other hand, seeks to detect specific items in a single visual scene, such as buildings [30], bridges [31], roofs [32], and so on. Depending on the dataset utilized, object and scene detection for damage assessment has advantages and downsides.

For object detection using satellite imagery, [28] compared the use of the K-Means algorithm as change subtraction and as encoded concatenate images using deep embedding clustering algorithms (DEC) to the convolutional autoencoder (AE) DEC algorithm with initial clusters of 4 then reduced by the algorithms to 3. [33] uses data fusion to detect object degradation using intensity-hue-saturation (IHS) and wavelet transform. Because data fusion can generate higher-quality information, integrating Synthetic-aperture radar (SAR) and optical pictures necessitates the use of IHS fusion, which can produce substantial spectrum distortion to integrate the two. The image is then decomposed into groups of multi-resolution images with wavelet coefficients using wavelet transformation. A deep convolutional neural network (DCNN) is also used in satellite imagery object detection along with BDD-Net and using the backbone of EfficientNet-B0 as the feature-extractor and encoder [34]. BDD-Net is a variant of the U-Net architecture that was created to categorize every pixel of a post-disaster image into numerous categories of damaged buildings, non-damaged buildings, or background classes automatically. Other approaches for object detection and classification include a raid assessment framework based on geographical information system (GIS) technology [35], CNN with an adapted U-Net architecture [35], and CNN for image classification paired with fully convolutional network (FCN) architecture [36] to provide dense predictions.

Object detection and classification utilizing aerial imagery typically yields a large amount of data and requires picture segmentation for small object detection. Traditional machine vision systems can detect objects, but they are still strongly reliant on the effect of feature extraction from post-event data, which is a difficult task [37]. To address this issue, [5] suggests using Inception V3 prediction to extract damage information from group buildings using a hybrid model of CNN and GIS data. Considering this, [38] proposed a model of single shot detector (SSD) algorithms for identifying and assessing post-disaster buildings. VGG-16 is used as a basic network to extract image feature information and to use the training weights to initialize the SSD weights. Another option is to employ the You Only Look Once (YOLO) algorithm [39], which has stronger generalization and detection capabilities. [39] proposed employing YOLOv3, a CNN-based object detection method, to locate collapsed buildings efficiently and accurately in post-earthquake occurrences. To minimize the number of parameters and enhance detection speed, the feature extractor was changed from Darknet53 to ShuffleNet v2. There are many other methods that improve on the CNN model, such as [40] who compared 2D CNN architecture with 3D FCN with 3D point clouds dataset and skip connection, [41] with a CNN model with dense connections and dilated convolutions that will allow it to capture the spatial context of the images patches using gapped kernel instead of contiguous one, and [42] with three multi-resolution CNN feature fusion approaches.

3. Dataset and Development of DL Models

The development of DL model begins by finding appropriate dataset that can be used for training the model. At this stage, the xBD dataset, which can be found in https://xview2.org/, is chosen. During the time that this dataset was used, it consists of 10 sub-datasets, each containing images from different natural disasters that occurred in the past few years. The dataset contains 5598 pairs of images, each pair having a raw satellite image, which will be referred to from now on as the frame, and its corresponding segmentation, which will be referred to as

the mask, and having a total of 7.7GB in size. Each pair of images further contains another pair which are the image before the event (pre-disaster) and the corresponding image after the event (post-disaster). As described on the previous part of this paper, pre-disaster images are discarded, reducing the effective number to only 2799 pairs of images. The images in this dataset are 24-bit in depth and have dimension of 1024 x 1024 pixels. The damage classes in this dataset are classified into 4 categories based on the Joint Damage Scale (JDS) convention, starting from the no damage class, which are buildings that do not get affected by the event, to the destroyed class, which are buildings that get completely wrecked. The rest of the class are minor damage and major damage which define the damage level between the no damage and destroyed class.

Next, the dataset is inspected to find any possible images that may affect the quality of the training process. These images are the ones which are mostly covered with clouds as can be seen in Fig.1a. Although there are some images that are partially blacked out (Fig.1b), these images will not hinder the model training since its corresponding segmentation is still accurate.

Fig. 1. Example of Special Datasets-cloud cover problem (a) and partially blacked out (b)

Before the images are ready to be used as a training set, the image is split up into four different quadrants, each having a dimension of 512 x 512. This is done to reduce the input size of the model without resizing the original image, hence reducing the number of times taken to finish a single iteration. During this phase, it was also noticed that buildings that are classified as destroyed can be barely seen on the image. Without the corresponding pre-disaster image, the authors figured that it would be really challenging for the model to be able to detect any building. As a result, the category destroyed is removed from the training set. Finally, some images which does not have any buildings are removed from the training set so that the number of pixels labeled as background will be lessened to reduce imbalances.

The DL model used in this research is divided into two stages. The first of the model is used for binary segmentation to identification the position of buildings on the image (localization) whereas the second part of the model is used to predict the damage level on each building by the first model (classification).

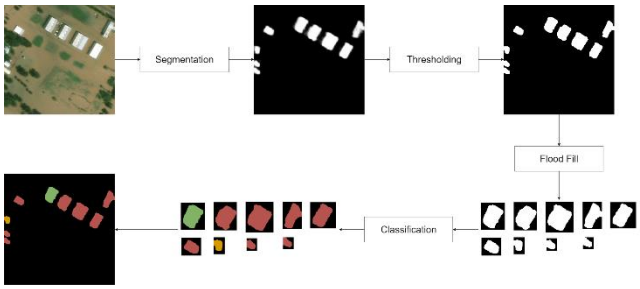

Fig. 2. Flow Process of Damage Assessment Model using Segmentation and Classification

From Fig. 2, there are two more intermediate stages have been conducted before the classification process. The first intermediate stage is the thresholding process which converts each pixel value from being a probability to a crisp value. This is conducted by choosing a value, β, so that every pixel, pi, such that pi ≤ β, will be set to 0, and the rest will be set to 1. By default, the value of β is arbitrarily chosen as 0.5. Now that the image contains only zeros and ones, it is easier to extract the buildings from the image. The algorithm below describes a series of flood fill algorithm is used to extract the buildings.

INPUT: Image I

OUTPUT: Array of coordinates defining the location of each building

let R be an empty array to store the result
let C be the set of all pixel coordinates in I
while C is not empty **do**
  let s be any arbitrary coordinate from C
  let c_i be an array of coordinates by flooding I from s
  if I[s] == 1 **then**
    append c_i to R
  end if
  C = C / c_i
end while

The model that is used for binary segmentation in this research is the PSPNet model with a ResNet18 backbone. PSPNet is a fully convolutional neural network that is excellent at capturing both local and global context information from the image than to the pyramid pooling module that uses different sizes of pooling kernel to extract different information from the image. The pooling kernel used in this research have sizes of 1, 2, 4, and 8. Architecture for the model as can be seen in the Fig. 3.

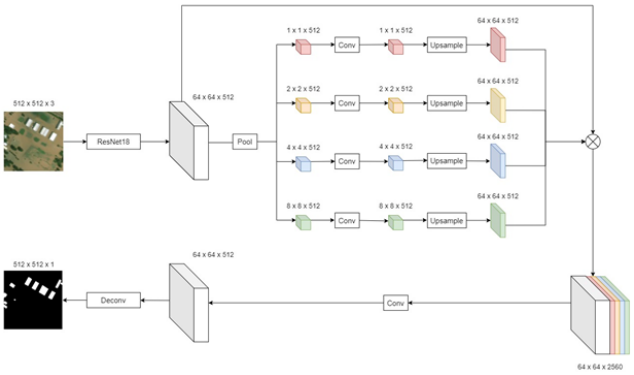

Fig. 3. Binary Segmentation for Building Using PSPNet and ResNet-18

As a comparison, the UNet model with Vanilla backbone and 4 down-sampling layers is trained in the same manner as the PSPNet model. UNet has an encoder-decoder architecture which provides its ability to predict whether is pixel is part of a building (encoder) and locate where a building is on the image (decoder). The feature output sizes of each down-sampling layers are 64, 128, 256, and 512. Architecture for the UNet model as can be seen in the Fig. 4.

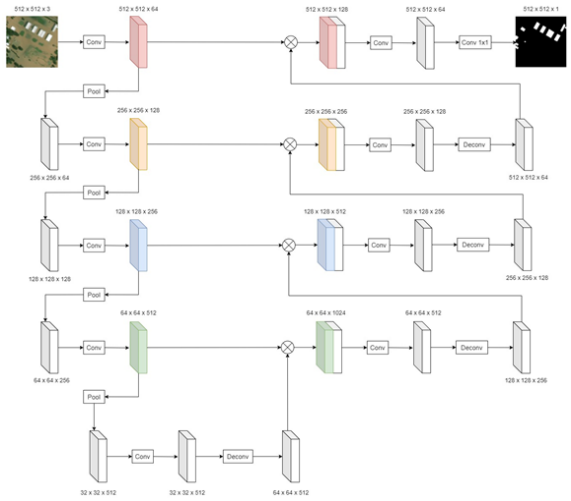

Fig. 4. Binary Segmentation for Building Using UNet and ResNet-18

For the classification stage, the ResNet50 is chosen (Fig. 5). The ResNet model consists of several layers of residual block that is used for feature extraction. The skip connection is a way of preventing the vanishing gradient problem. The original 7 x 7 convolution layer is replaced with 2 layers of 3 x 3 convolution kernel to provide more feature for the model.

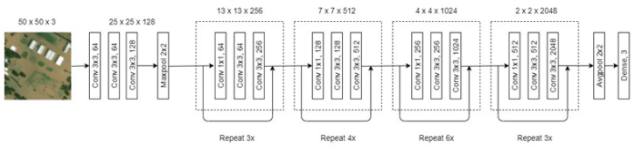

Fig. 5. ResNet50 Architecture used in the classification stage

Then implementation of the DL model is done using the Keras framework built on top of Tensorflow. The cyclical learning rates (CLR) was used to avoid the need to perform a lot of experiments for finding the best learning rate (LR) values. LR range test was first run to find the lower bound and upper bound of the LR. The triangular policy was chosen to update the LR values after each iteration. Along with it, the Adam optimizer was used to update the learnable parameters of the model. Finally, the dice coefficient is used as a metric to measure the performance of the segmentation model, and accuracy was used to measure the performance of the classification model. Once the model has been trained, it is tested on unused data to check its feasibility.

4. Result

In accordance with the model development methodology, the discussion of the results in this paper will be divided into two, the first development and implementation of the model for building and non-building segmentation and the second is the development and implementation of a classification model for the level of building damage due to natural disasters. The development of the segmentation model has been conduct using PSPNet with Resnet-18 backbone and using UNet with vanilla backbone. Because the models are trained on a machine without using any GPU. It takes roughly about 40 minutes to finish a single epoch for the binary segmentation model and 10 minutes for the classification model. Details of each training process can be found on Table 1.

Table 1. Training parameters for data in the segmentation process

		PSPNet	**UNet**
Image	Image Count	1003	
	Dimension	(Frame) 512 x 512 x 3 (Mask) 512 x 512 x 1	
	Batch Size	4	
	Augmentation	Horizontal Flip, Vertical Flip and Rotation	
	Validation Split	0.15	
Parameter	Output Activation	Sigmoid	Sigmoid
	Loss Function	Dice + CE	Dice + CE
	Optimizer	Adam	Adam
	Epochs	60	100

The results of the evaluation of the segmentation proposed in this paper are then compared with the results of evaluations that have been carried out by previous researchers using the same dataset as the conducted by [27], [43], [44], [45], [46]. The results of the evaluated F1 scores are based on publications that have been compiled by [47] . In Table 2. The F1 score of the PSPNet method and algorithm used in research is slightly better than the results of previous studies, except for the results of the study [46] with a value of 0.869. The results are different with a smaller value when compared to the results of previous studies with experiments conducted using UNet with vanilla backbone.

Table 2. Segmentation results of various models from several previous researchers

	Localization *F1*	
Weber [43]	0.835	
RescueNet [27]	0.840	
BDANet [44]	0.864	
DCFNet [45]	0.864	
DamFormer [46]	0.869	
BuildingNet [47]	0.846	
Weber [43]	0.835	
Our Model	PSPNet	UNet
	0.8494	0.8371

During the process of data segmentation, the research conducted was faced with obstacles where in certain areas, the building size was relatively small and quite dense. This condition causes difficulties in segmenting buildings individually. Referring to research [48], it is tried to implement a flood fill algorithm so that several different buildings can be extracted as buildings with a single identity, which then affects the classification process. The solution to the problem of building separation in a congested location is as an example that can be seen in Fig. 6.

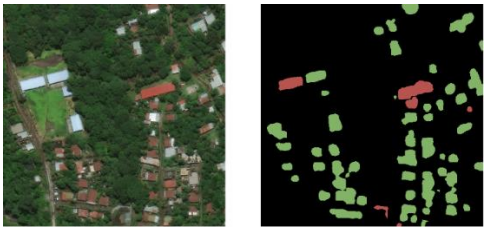

Fig. 6. Segmentation on building density area using flood fill algorithm. Example of imagery data with building density relatively high (a) and the result of segmentation for individual building (b)

In this study, three models of multiclass segmentation have also been conducted and the results of which are discussed as follows: The 1st model classifies all buildings into the no damage category. This is because the distribution of background and no damage pixels has up to 96.4% of the total dataset, so the model is able to get an accuracy value of 0.9407 only by classifying all buildings in the no damage category. Although the model has a high accuracy value, this model cannot be used because it does not provide any information about the level of damage to the building (see Fig.7).

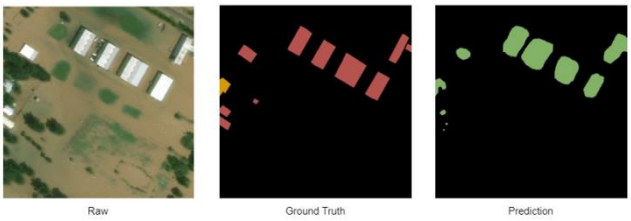

Fig. 7. The results of the multiclass segmentation for 1st model. Data (a), ground truth (b) and result of multiclass segmentation (c)

In the 2nd model, building with minor damage and major damage categories were extracted and pasted on the existing dataset to create a synthetic building dataset. In the 2nd dataset, the distribution of pixels is more even than the previous dataset. The model is given 20 epochs for the training process. From the accuracy graphic, the best model is found in the 11th epoch with a dice coefficient value of 0.8643 and an accuracy of 0.8225. A test is then carried out to ensure that the model is working properly. The results of the test concluded that the model would classify all buildings in the no damage category except for synthetic buildings. This happens because there is a high gradient value at the edge of the building that is made so that the model incorrectly learns the features that it should. An example of the results of the 2nd model can be seen in Fig. 8.

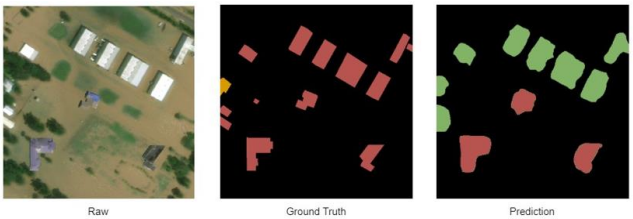

Fig. 8. The results of the multiclass segmentation for 2nd model. Data (a), ground truth (b) and result of classification (c)

The 3rd model combines both under sampling and oversampling techniques to balance the dataset. The distribution of pixels in percent for each background class, no damage, minor damage, and major damage is as follows 98.26; 0.59; 0.58; and 0.58. The 3rd model gets high dice coefficient and accuracy values, but this model also cannot

be used because the number of pixels in the background category makes the model predict many background categories. The results of the 3rd model can be seen in Fig. 9.

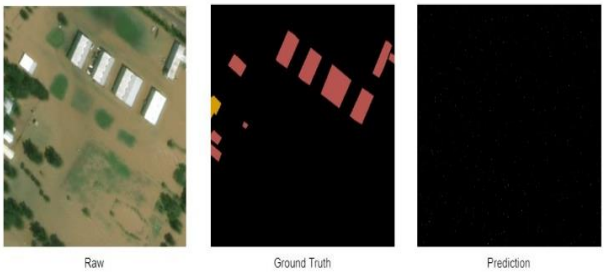

Fig. 9. The results of the multiclass segmentation for 3rd model. Data (a), ground truth (b) and result of classification (c)

After the segmentation process can produce individual buildings well, the next step is to classify the damage level of the building. In the research experiment, it has been tried to classify the level of damage to buildings with several different classes such as 2 classes (damage and no damage) and 3 classes (no damage, minor damage, and major damage). Parameters for classification stage, as well as segmentation, it uses dimensions of 50 x 50 x 3 with a batch size of 32. As for the augmentation and optimizer using the same method as the data segmentation stage, using horizontal, vertical and rotation with Adam optimizer and the number of epochs is 100. For further analysis of the classification model, a test set of consisting of 4309 images is created. In Fig.10, it can see a graph of the accuracy and loss of the ResNet50 model used in the classification stage. Accuracy results for training reached 0.6494 while data validation reached 0.6482.

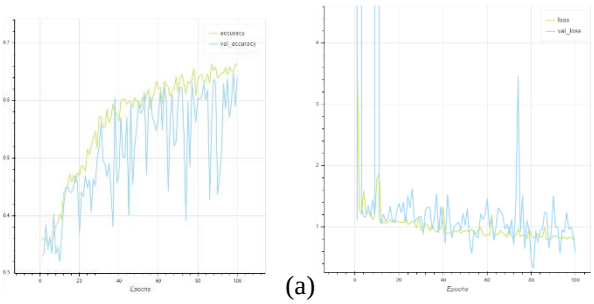

Fig. 10. Accuracy (a) and Loss graphic (b) ResNet50 model for classification

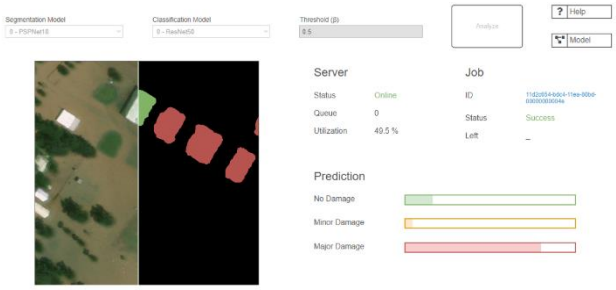

Fig. 11. Classification result using ResNet50 model for three class of buildings Damage (No damage, minor and major damage

To assess the accuracy of the model, a confusion matrix is then used with the actual distribution and predictions of each level of building damage as can be seen in Table 3.

Table 3. Confusion Matrix of three level of building damage

Actual/ Prediction	No Damage	Minor Damage	Major Damage
No Damage	929	135	431
Minor Damage	175	753	441
Major Damage	130	201	1114

From Table 4, it can be seen that the model is still quite difficult to determine the dividing line between buildings in the major damage category and buildings that are not. This conclusion is drawn based on the fact that the number of false positives in the major damage category is almost the same as the true positive numbers. This means, from 77.09% of the buildings that are predicted to be major damage, only about 56.09% of the buildings are actually relevant.

Table 4. Precision, Recall, and Accuracy of building damage

Actual/ Prediction	Precision	Recall	Accuracy
No Damage	0.7528	0.6214	0.7979
Minor Damage	0.6915	0.5500	0.7791
Major Damage	0.5609	0.7709	0.7208

From these results, it is deduced that the classification model is not able to properly differentiate minor damage and major damage class. If both damaged class is combined as a single class, then the result would give an average accuracy of 0.8338 (Table 5).

Table 5. Precision and recall on 2 Classes

Actual/ Prediction	Precision	Recall
No Damage	0.8594	0.7665
Damaged	0.8159	0.8919

5. Conclusion

The results achieved from previous experiments show that deep learning can be used as a new, efficient, and fast way to assess damage to buildings. Based on the research problems, the research proposed using a combined model of building segmentation followed by a classification model to detect building objects and classify the level of damage caused by various natural disasters using aerial image data.

The proposed model, using PSPNet with a Resnet-18 backbone, can improve the accuracy of building segmentation from several models in previous studies which achieved an F1 score of 0.8494 although this value is slightly lower than the F1 score of the DamFormer model proposed previously

The factor of the number of classes of building damage is very influential on the accuracy of the assessment where the more classes, the lower the accuracy obtained. Evaluation with a confusion matrix using two classes of building damage levels (No damage and Damage) can produce a precision model that reaches 0.8338. The results achieved are slightly better than previous studies.

Furthermore, there is an opportunity to improve relevant research with further development on making a better binary segmentation model for extracting buildings in dense areas from aerial images and classifying the level of damage to buildings for a greater number of damage classes.

5.1. Abbreviations

AI – Artificial Intelligence

ML – Machine Learning

DL – Deep Learning

RS – Remote Sensing

LULC - Land Use and Land Cover

CNN - Cconvolutional Neural Networks

VHR - very-high-resolution

FCN - Fully Convolutional Networks

NN - Neural Networks

DEC - Deep Embedding Clustering

IHS - Intensity-Hue-Saturation

SAR - Synthetic-aperture radar

DCNN - Deep Convolutional Neural Network

GIS - Geographical Information System

SSD - Single Shot Detector

YOLO - You Only Look Once

JDS - Joint Damage Scale

LR - Learning Rate

CLR - Cyclical Learning Rates

5.2. Availability of Data and Material

Data and material supporting the findings of this study are available from xBD dataset, which can be found in https://xview2.org/.

5.3. Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

5.4. Funding

This work was funded by Kementerian Pendidikan dan Kebudayaan Republik Indonesia, as a part of "Penelitian Terapan Unggulan Perguruan Tinggi" Research Grant to BINUS University titled "Sistem Penilaian Kerusakan Bangunan Pasca Bencana Dengan Segmentasi Citra Udara Menggunakan Deep Learning" (Applied Research grant to BINUS University with titled "Earthquake Building Damage Assessment with Aerial Imagery Segmentation Using Deep Learning") with contract number: 309/E4.1/AK.04.PT/2021 and 3507/LL3/KR/2021.

5.5. Authors' Contribution

Conceptualization, E.I., and A.A.S.G; data collection: E.I and H.Y; formal analysis: E.I.; investigation: E.I; algorithm development: E.I., A.A.S.G., and HY; training and testing algorithm: H.Y., validation and supervision: A.A.S.G.; writing—original draft: E.I and HY.; writing—review and editing: E.I. and A.A.S.G. All authors have read and approved the final manuscript.

5.6. Acknowledgments

The authors would like to thank to the anonymous reviewers and editor for their insightful comments.

References

[1] D. Pavlinovic, "Climate and weather-related disasters surge five-fold over 50 years, but early warnings save lives-WMO report," *UN News*, 2021. https://news.un.org/en/story/2021/09/1098662

[2] H. Ritchie, P. Rosado, and M. Roser, "Natural Disasters Data Explorer," *Natural Disasters*, 2018. https://ourworldindata.org/natural-disasters

[3] S. Li and H. Tang, "Classification of Building Damage Triggered by Earthquakes Using Decision Tree," *Math Probl Eng*, vol. 2020, pp. 1–15, 2020, doi: 10.1155/2020/2930515.

[4] H. Ma, Y. Liu, Y. Ren, D. Wang, L. Yu, and J. Yu, "Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images," *Remote Sens (Basel)*, vol. 12, no. 2, pp. 1–16, Jan. 2020, doi: 10.3390/rs12020260.

[5] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019, doi: 10.1016/j.isprsjprs.2019.04.015.

[6] Y. Hu, S. Gao, D. Lunga, W. Li, S. Newsam, and B. Bhaduri, "GeoAI at ACM SIGSPATIAL," *SIGSPATIAL Special*, vol. 11, no. 2, pp. 5–15, Dec. 2019, doi: 10.1145/3377000.3377002.

[7] H. Ghanbari, M. Mahdianpari, S. Homayouni, and F. Mohammadimanesh, "A Meta-Analysis of Convolutional Neural Networks for Remote Sensing Applications," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 14, pp. 3602–3613, 2021, doi: 10.1109/JSTARS.2021.3065569.

[8] A. Trekin, G. Novikov, G. Potapov, V. Ignatiev, and E. Burnaev, "Satellite imagery analysis for operational damage assessment in Emergency situations," in *International Conference on Business Information Systems*, Feb. 2018, pp. 347–358. [Online]. Available: http://arxiv.org/abs/1803.00397

[9] Q. D. Cao and Y. Choe, "Building Damage Annotation on Post-Hurricane Satellite Imagery Based on Convolutional Neural Networks," *Natural Hazards*, vol. 103, no. 3, pp. 3357–3376, Jul. 2018, doi: 10.1007/s11069-020-04133-2.

[10] I. Bouchard, M. È. Rancourt, D. Aloise, and F. Kalaitzis, "On Transfer Learning for Building Damage Assessment from Satellite Imagery in Emergency Contexts," *Remote Sens (Basel)*, vol. 14, no. 11, pp. 1–29, Jun. 2022, doi: 10.3390/rs14112532.

[11] R. Gupta *et al.*, "Creating xBD: A dataset for assessing building damage from satellite imagery," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Nov. 2019, pp. 10–17. [Online]. Available: http://arxiv.org/abs/1911.09296

[12] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[13] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *Foundations and Trends in Machine Learning*, vol. 11, no. 3–4, pp. 219–354, Dec. 2018, doi: 10.1561/2200000071.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, vol. 86, no. 11, pp. 2278–2323. doi: 10.1109/5.726791.

[15] Y. Heryadi and E. Irwansyah, *Deep Learning: Aplikasinya di Bidang Geospasial*. Depok - Jawa Barat: AWI Technology Press, 2020. Accessed: Dec. 25, 2022. [Online]. Available: https://books.google.com/books/about/Deep_Learning_Aplikasinya_di_Bidang_Geos.html?id=UorwDwAAQBAJ

[16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2015, pp. 1–14. [Online]. Available: http://arxiv.org/abs/1409.1556

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, vol. 9351, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[18] B. Leibe, J. Matas, N. Sebe, and M. Welling, "Identity Mappings in Deep Residual Networks," *European conference on computer vision*, vol. 9906 LNCS. Springer Verlag, pp. 630-645-, 2016. doi: 10.1007/978-3-319-46493-0.

[19] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," *arXiv preprint arXiv:1706.05587*, Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.05587

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.

[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587. [Online]. Available: http://arxiv.

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Trans Pattern Anal Mach Intell*, vol. 38, no. 1, pp. 142–158, Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.

[23] E. Shelhamer, K. Rakelly, J. Hoffman, and T. Darrell, "Clockwork Convnets for Video Semantic Segmentation," in *European Conference on Computer Vision*, Aug. 2016, pp. 852–868. [Online]. Available: http://arxiv.org/abs/1608.03609

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, vol. 7, pp. 43369–43382.

[25] Y. Liao, M. E. Mohammadi, and R. L. Wood, "Deep learning classification of 2D orthomosaic images and 3D point clouds for post-event structural damage assessment," *Drones*, vol. 4, no. 2, pp. 1–19, Jun. 2020, doi: 10.3390/drones4020024.

[26] R. Gupta and M. Shah, "RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery," in *Proceedings - International Conference on Pattern Recognition*, Apr. 2020, pp. 4405–4411. [Online]. Available: http://arxiv.org/abs/2004.07312

[27] J. Sublime and E. Kalinicheva, "Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami,"

Remote Sens (Basel), vol. 11, no. 9, pp. 1–20, May 2019, doi: 10.3390/rs11091123.

[28] D. E. Kislov and K. A. Korznikov, "Automatic windthrow detection using very-high-resolution satellite imagery and deep learning," *Remote Sens (Basel)*, vol. 12, no. 7, pp. 1–17, Apr. 2020, doi: 10.3390/rs12071145.

[29] A. J. Cooner, Y. Shao, and J. B. Campbell, "Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake," *Remote Sens (Basel)*, vol. 8, no. 10, pp. 1–17, Oct. 2016, doi: 10.3390/rs8100868.

[30] W. Liu and F. Yamazaki, "Extraction of collapsed bridges due to the 2011 Tohoku-oki earthquake from post-event SAR images," *Journal of Disaster Research*, vol. 13, no. 2, pp. 281–290, Mar. 2018, doi: 10.20965/jdr.2018.p0281.

[31] F. Wang, J. P. Kerekes, Z. Xu, and Y. Wang, "Residential roof condition assessment system using deep learning," *J Appl Remote Sens*, vol. 12, no. 01, pp. 1–20, Mar. 2018, doi: 10.1117/1.jrs.12.016040.

[32] W. Zhang and M. Xu, "Translate SAR Data into Optical Image Using IHS and Wavelet Transform Integrated Fusion," *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 1, pp. 125–137, Jan. 2019, doi: 10.1007/s12524-018-0879-7.

[33] J. Shao, L. Tang, M. Liu, G. Shao, L. Sun, and Q. Qiu, "BDD-net: A general protocol for mapping buildings damaged by a wide range of disasters based on satellite imagery," *Remote Sensing*, vol. 12, no. 10. MDPI AG, pp. 1–11, May 01, 2020. doi: 10.3390/rs12101670.

[34] J. Zhao *et al.*, "A rapid public health needs assessment framework for after major earthquakes using high-resolution satellite imagery," *Int J Environ Res Public Health*, vol. 15, no. 6, pp. 1–18, Jun. 2018, doi: 10.3390/ijerph15061111.

[35] B. J. Wheeler and H. A. Karimi, "Deep learning-enabled semantic inference of individual building damage magnitude from satellite images," *Algorithms*, vol. 13, no. 8, pp. 1–15, Aug. 2020, doi: 10.3390/A13080195.

[36] S. Cotrufo, C. Sandu, F. Giulio Tonolo, and P. Boccardo, "Building damage assessment scale tailored to remote sensing vertical imagery," *Eur J Remote Sens*, vol. 51, no. 1, pp. 991–1005, Jan. 2018, doi: 10.1080/22797254.2018.1527662.

[37] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector Wei," in *European Conference on Computer Vision*, 2016, vol. 9905, pp. 21–37. doi: 10.1007/978-3-319-46448-0.

[38] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788. [Online]. Available: http://pjreddie.com/yolo/

[39] H. Ma, Y. Liu, Y. Ren, and J. Yu, "Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3," *Remote Sens (Basel)*, vol. 12, no. 1, pp. 1–19, Jan. 2020, doi: 10.3390/RS12010044.

[40] F. Nex, D. Duarte, F. G. Tonolo, and N. Kerle, "Structural building damage detection with deep learning: Assessment of a state-of-the-art CNN in operational conditions," *Remote Sens (Basel)*, vol. 11, no. 23, pp. 1–17, Dec. 2019, doi: 10.3390/rs11232765.

[41] D. Duarte, F. Nex, N. Kerle, and G. Vosselman, "Multi-resolution feature fusion for image classification of building damages with convolutional neural networks," *Remote Sens (Basel)*, vol. 10, no. 10, pp. 1–26, Oct. 2018, doi: 10.3390/rs10101636.

[42] E. Weber and H. Kané, "Building Disaster Damage Assessment in Satellite Imagery with Multi-Temporal Fusion," *arXiv preprint arXiv:2004.05525*, pp. 1–7, Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.05525

[43] Y. Shen *et al.*, "BDANet: Multiscale Convolutional Neural Network with Cross-Directional Attention for Building Damage Assessment from Satellite Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3080580.

[44] H. Xiao, Y. Peng, H. Tan, and P. Li, "Dynamic Cross Fusion Network for Building-Based Damage Assessment," in *Proceedings - IEEE International Conference on Multimedia and Expo*, 2021, pp. 1–6. doi: 10.1109/ICME51207.2021.9428414.

[45] H. Chen, E. Nemni, S. Vallecorsa, X. Li, C. Wu, and L. Bromley, "Dual-Tasks Siamese Transformer Framework for Building Damage Assessment," *arXiv preprint arXiv:2201.10953*, pp. 1600–1603, Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.10953

[46] S. Shrestha and L. Vanneschi, "Improved fully convolutional network with conditional random fields for building extraction," *Remote Sens (Basel)*, vol. 10, no. 7, Jul. 2018, doi: 10.3390/rs10071135.

[47] E. Irwansyah, A. A. S. Gunawan, and Nurhasanah, "Deep Learning Model Comparison for Dense Building Segmentation in the City Using Aerial Imagery Data," in *2022 International Conference on Science and Technology, ICOSTECH 2022*, 2022, pp. 25–27. doi: 10.1109/ICOSTECH54296.2022.9828812.