PRESSNet: Assessment of Building Damage Caused by the Earthquake

Dewa Ayu Defina Audrey Nathania[1], Alexander Agung Santoso Gunawan[2], Edy Irwansyah[3]
Computer Science Department, Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia[1]
Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia[2, 3]

***Abstract*—Loss of life and property often occur due to natural disasters and other significant occurrences like earthquakes, which make manual damage assessment a time-consuming and inefficient process. In an attempt to address this challenge, researchers have been investigating the field of automated damage assessment in Remote Sensing. With time, this area of research has transformed from conventional machine learning techniques to more sophisticated deep learning techniques. The study puts forward the PRESSNet model as a solution for assessing building damage. The effectiveness of the proposed PRESSNet model is compared to that of a baseline model, PSPNet, and ResNet 50, across different types of damage. This study contributes by introducing the spatial attention module to the baseline model. The xBD Dataset was used both before and after the Palu earthquake disaster. The results show that PRESSNet performs similarly or slightly better than the baseline model in all damage categories. This illustrates the impressive ability of the proposed PRESSNet architecture to accurately detect and classify building damage. This research sheds light on the development of effective models for assessing disaster damage and lays the foundation for future progress in this crucial area.**

***Keywords*—Remote sensing; deep learning; PSPNet; ResNet; spatial attention**

I. Introduction

In an era characterized by an increasing frequency of natural disasters, including earthquakes, floods, and hurricanes, which yield dire consequences, the importance of proficient crisis management becomes paramount. These catastrophic events result in not only the unfortunate loss of human lives but also substantial property damage [1]. Access to crucial information, both before and after a catastrophic event, proves to be of utmost importance in enhancing disaster response strategies and mitigating the impact on human lives and physical infrastructure [2]. In addition to enhancing the capacity for early detection and warning before the crisis begins, it is crucial to gather information about a disaster as soon as it occurs [3].

Assessing structural damage to buildings stands as a critical issue in the field of disaster response, as it has been identified as a prominent factor contributing to the loss of life during natural calamities [4]. The precise assessment of such harm is crucial to enhance emergency response efforts and ultimately preserve a greater number of lives. Recent advancements in remote sensing technology and the deployment of satellite constellations have greatly enhanced our ability to obtain high-resolution satellite (HRS) data [5]. Integrating this wealth of data with machine learning (ML) and deep learning (DL) methodologies offers a promising approach for evaluating structural damage in the aftermath of a calamitous event [5].

Assessing structural damage to buildings stands as a critical issue in the field of disaster response, as it has been identified as a prominent factor contributing to the loss of life during natural calamities [4]. The precise assessment of such harm is crucial to enhance emergency response efforts and ultimately preserve a greater number of lives. Recent advancements in remote sensing technology and the deployment of satellite constellations have greatly enhanced our ability to obtain high-resolution satellite (HRS) data. Integrating this wealth of data with machine learning (ML) and deep learning (DL) methodologies offers a promising approach for evaluating structural damage in the aftermath of a calamitous event [5].

However, recent evaluations of the PSPNet model, which employs ResNet 50 as its underlying framework, have identified shortcomings in effectively gathering and leveraging contextual information at various scales, particularly in the intricate realm of building degradation [8]. To address this challenge, the present study introduces the PRESSNet framework, which utilizes spatial attention mechanisms to enhance the identification and prioritization of critical regions inside a building that exhibit signs of structural deterioration. PRESSNet's performance in accurately evaluating the extent of building damage surpasses that of the PSPNet + ResNet-50 model, showcasing its capacity to augment disaster response efforts.

The primary objective of this study is to underscore the importance of spatial attention mechanisms within deep learning models for the purpose of disaster response. PRESSNet's contribution to the advancement of computer vision research in disaster management lies in its emphasis on the effective identification and classification of different levels of damage. This study highlights the significance of attention mechanisms in improving the performance of convolutional neural networks, thereby contributing to the development of more efficient disaster response systems.

Furthermore, this study introduces the PRESSNet model as a solution for assessing building damage. It compares the effectiveness of PRESSNet to that of a baseline model, PSPNet, and ResNet 50, across different types of damage. The xBD Dataset is utilized both before and after the Palu earthquake disaster. The results demonstrate that PRESSNet performs similarly or slightly better than the baseline model in all damage categories. The baseline model exhibits strong performance with a macro-average F1 score of 89.5%, while

PRESSNet slightly outperforms it, achieving a macro-average F1 score of 90%. This illustrates the impressive ability of the proposed PRESSNet architecture to accurately detect and classify building damage. This research sheds light on the development of effective models for assessing disaster damage and lays the foundation for future progress in this crucial area.

II. Related Studies

A. Building Damage Assessment in High Resolution Satellite Imagery using Deep Learning

The assessment of building damage using deep learning models, particularly convolutional neural networks (CNNs), has been a prominent subject of investigation within the domain of remote sensing utilizing satellite and aerial photography. The authors, Garg et al. [8], introduced a deep convolutional neural network (CNN) that utilizes transfer learning to do building damage assessment using satellite photos. The model underwent training by utilizing a pre-existing VGG16 network, which was subsequently fine-tuned using a dataset specifically curated for earthquake-affected buildings. The results indicate that the suggested model had superior performance compared to standard machine learning techniques, achieving an accuracy of 87% on the test dataset.

The authors, Hu et al. [9], introduced an innovative methodology that uses deep learning techniques to evaluate structural harm in buildings through the analysis of aerial photography. The employed methodology involved the utilization of a Siamese network architecture to conduct a comparative analysis of aerial photos captured before and after a catastrophic event. The objective was to accurately detect and pinpoint areas where structural damage to buildings had occurred. The model achieved an overall accuracy of 92.65 percent on the test dataset.

Using satellite data, Shah et al. [10] created a deep convolutional neural network (CNN) method that makes use of transfer learning to assess the degree of building damage. The model under consideration utilized a pre-existing ResNet-50 network [11], which was subsequently refined through the process of fine-tuning. A dataset of buildings damaged by Hurricane Harvey served as the basis for this refinement. Waseem et al. [12] developed a system based on deep learning methods to assess the degree of building damage using satellite data. The proposed model integrates features derived from a pre-trained ResNet 50 network with manually engineered characteristics, including texture and color features.

B. PSPNet

The assessment of building damage holds significant importance in the aftermath of disasters, as it serves to guide relief operations, optimize resource allocation, and ascertain the overall consequences of both natural and human-induced calamities. The assessment of building damage has been significantly improved with the application of convolutional neural networks (CNNs), thanks to recent breakthroughs in computer vision and deep learning. The Pyramid Scene Parsing Network (PSPNet) is a convolutional neural network (CNN) architecture that has gained significant recognition due to its capacity to effectively capture contextual information at several scales.

The Pyramid Scene Parsing Network (PSPNet) is a convolutional neural network that uses a pyramid pooling module to capture contextual information at many scales. Additionally, it incorporates a spatial attention module to enhance important features and suppress non-informative ones [13]. The inclusion of the pyramid pooling module in PSPNet facilitates the efficient extraction of features at different scales. This capability is crucial in accurately identifying zones of building damage that exhibit diverse sizes. The architecture of PSPNet incorporates a pyramid pooling module that effectively combines contextual information from multiple scales. The module presented in this study is designed to gather contextual information at many scales. This enables the model to effectively capture both global and local contextual information that is relevant for building damage assessment. The inclusion of the pyramid pooling module significantly contributes to the model's ability to effectively differentiate between regions that are damaged and those that are unaffected.

Dilated convolutions, which are also called atrous convolutions, are used in the PSPNet architecture to increase the receptive field while keeping the spatial resolution the same. Dilated convolutions enable the neural network to record a wider contextual range by introducing gaps inside the convolutional kernels, thereby maintaining the integrity of spatial details. The PSPNet employs a fusion technique to effectively integrate contextual information by merging elements at many scales. The final segmentation map is generated by combining and refining the multi-scale feature maps from the pyramid pooling module using convolutional layers.

The PSPNet model comprises three primary components, each serving distinct roles. By using ResNet 50 as the underlying framework, it is possible to create a feature map from an input image. The pyramid parsing module utilizes the feature map of the initial module, ResNet 50, to extract representations of four separate sub-representations. These sub-representations are obtained using convolution operations of sizes 1×1, 2×2, 3×3, and 6×6. The result of the representation that was restored in the preceding layer is increased in size and combined with all of the increased representations, together with the characteristics that were mapped in the initial module. The convolutional layer utilizes a representation of the input image produced from the second module in order to obtain the final semantic segmentation results.

The point-wise attention blinders used for the research conducted by Hamdi et al. [14] exhibit dissimilarities when compared to the blinders employed in other studies. The PSA module specifically instructs on the implementation of location- and category-sensitive masks that possess self-adaptive capabilities. The Public Service Announcement (PSA) acquires the ability to gather contextual information for each unique point in a manner that is flexible and tailored to the specific needs of the user.

The method of pooling known as shift pooling was introduced by Yuan et al. [15] and was implemented to enhance the performance of PSPNet. The repositioning of the pooling grid allows for the comprehensive acquisition of local feature information by the pixels located at the edges and corners of the grid, resulting in enhanced segmentation outcomes.

C. Spatial Attention Mechanism Module

The spatial attention mechanism (SAM) has been identified as a highly successful approach for evaluating building damage and other computer vision tasks. The Spatial Attention Module (SAM) is employed to consolidate comparable picture information and augment the network's capacity to depict these characteristics. The utilization of SAM allows the model to effectively allocate importance to informative traits while simultaneously suppressing non-informative ones through the selective concentration on pertinent regions. This approach has been widely employed in several deep learning models to evaluate the extent of building damage.

Chen et al. [16] introduced a novel approach for enhancing the precision and consistency of building damage assessment. Their proposed method involves utilizing a change detection feature extractor, which incorporates a pyramid spatial temporal attention module. The experiments showed that this module, called SAM, makes it possible for the network to capture similar features and highlight the unique features of damaged regions.

The domain of remote sensing through satellite and aerial imaging has witnessed significant research activity in the realm of building damage assessment. Deep learning models, particularly convolutional neural networks (CNNs), have emerged as a prominent approach in this subject. The authors, Garg et al. [8], introduced a deep convolutional neural network (CNN) that utilizes transfer learning to do building damage assessment using satellite photos. The model underwent training by utilizing a pre-existing VGG16 network and subsequently underwent refinement through the utilization of a dataset consisting of buildings impacted by earthquakes. The results indicate that the suggested model exhibited superior performance compared to standard machine learning techniques, achieving an accuracy rate of 87% on the test dataset.

Attention mechanisms are widely used in deep learning. An attention model was created by Mnih et al. [17] that picks a number of regions or sites for processing in an adaptive manner. Multiple attention masks were found by Chen et al. [18] to combine feature maps or forecasts from many branches. Pre- and Post-Disaster Imagery were recovered from a single model with the same weight in investigations conducted by Weber et al. [19], and the output features were layered (concatenated) to derive features between pre and post.

A self-attention machine translation model was developed by Vaswani et al. [20]. The correlation matrix between each spatial location in the feature map was calculated by Wang et al. [21] to identify attention masks. Zhao et al.'s [22] point-wise spatial attention network (PSANet) was suggested as a way to relax the local neighborhood constraint. A self-adaptively learnt attention mask connects each location on the feature map to every other location. Additionally, scene understanding is allowed through bidirectional information propagation. The forecast of the current position can be aided by knowledge from previous positions, and the prediction of other positions can benefit from information from the current one.

III. Research Methodology

A. Dataset

The applied dataset is the 2018 Palu earthquake from the Tier 1 XBD dataset (https://xview2.org). Fig. 1. shows the map of Palu. No damage, minor damage, major damage, and destroyed are the four levels of damage. Pre- and post-disaster dataset images are distributed as follows: 80% to the training dataset and 20% to the test dataset.

B. Data Preprocessing

The original dimension of the Palu dataset was 128 x 128 pixels, which has been reduced to 64 x 64. No Damage and Destroyed are the two classes or levels within this research. Pre- and post-disaster images for the Train Dataset increased from 54 to 3264. Pre and Post Images increased from 15 Images (1024x1024) to 1024 Images (256x256) in the test dataset. The training dataset is composed of 70% Training data and 30% Validation data. The training dataset is then rotated by 30 degrees to generate a new patch. The Steps per Epoch used for this research are 16000 for Training and 5000 for validation.

The training dataset was used to optimize the neural network, whereas the validation dataset was used to evaluate the network's training efficacy. Using the evaluation dataset, the effectiveness of the optimized neural network was determined. The validation and training images were read into (256, 256, 4) arrays of the form. The labels were interpreted as an array of shapes (256, 256, 1), with the values 1 or 0 indicating whether the item is damaged or not.

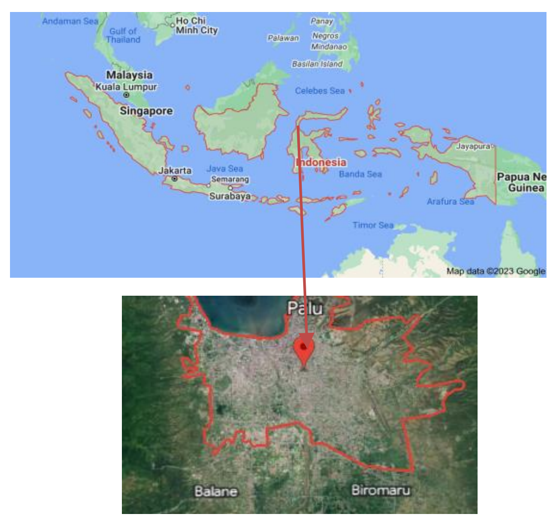


Fig. 1. Palu earthquake, 2018.

C. Proposed Method

This study proposed the model PRESS-Net, which consists of the single segmentation model PSPNet[10] using ResNet[12] as a with different backbone using ResNet 50 and ResNet 101 with Hyperparameter Tuning and Spatial Attention

Module [23] to learn which feature and where the location of feature is important to be selected.

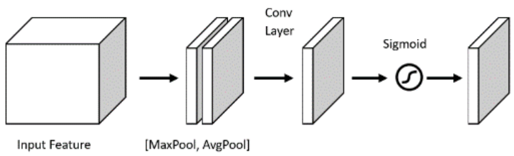


Fig. 2. Spatial attention module architecture.

The spatial attention map is generated by applying the interspatial relationship between features. When compared to channel attention, the spatial attention module architecture (see Fig. 2) emphasizes "where" as an additional information component. We first use average-pooling and max-pooling operations along the channel axis to compute spatial attention, and we then concatenate the results to provide an efficient feature descriptor. It has been proven that aggregating procedures applied along the channel axis can effectively highlight informative regions. We apply a convolution layer to the concatenated feature descriptor to generate a spatial attention map $M_s(F) \in R^{H\ X\ W}$ that encodes where to highlight or suppress.

The computation for spatial attention is as follows:

$$M_s(F) = \sigma\left(f^{7x7}\ ([AvgPool(F); MaxPool(F)])\right)$$
$$= \sigma\left(f^{7x7}\ \left(\left[F_{avg}^{s}\ ;\ F_{max}^{s}\right]\right)\right)$$

- Where σ denotes the sigmoid function and f^{7x7} represents a convolution operation with the filter size of 7 × 7 [23].

Algorithm 1: Pseudocode for building segmentation using CNNs

Inputs: *xBD Dataset*
Output: *Segmented building damage*

```
Start:
  Load and preprocess the xBD Dataset
  Apply noise removal to the Palu dataset
  Initialize training loop counter i = 1
  Set maximum training loops as MAX_LOOPS
  Do while i <= MAX_LOOPS:
  # Training Data Augmentation
      Augment training data and masks
  # Split dataset into training and validation sets
      Split Augmented Data into X_train, X_val and Augmented
Masks into y_train, y_val
  # Create the PSPNet model with ResNet-50 backbone and spatial
attention
      pspnet_model = create_pspnet(input_shape, num_classes)
   # Train the model
      Train pspnet_model on X_train and y_train for a specified
number of epochs
   # Evaluate the model's performance on validation set
      Evaluate pspnet_model on X_val and y_val to get validation
loss and accuracy
   # Print evaluation results (optional)
      Print "Validation Loss:", validation_loss
      Print "Validation Accuracy:", validation_accuracy
   # Save the trained model (optional)
      Save pspnet_model to disk with a suitable filename
   # Increment training loop counter
       Increment i
  End do while
End
```

- The process begins by loading a dataset called xBD, containing images of building damage. These images are prepared for analysis by removing any distracting noise from them, which helps the algorithm focus on the key information. To train an accurate model, a training loop is set up. A loop counter named 'i' starts at 1, and a maximum number of loops (MAX_LOOPS) is predetermined to manage the training process. Inside the loop, data augmentation is applied to the training images and their corresponding masks.

- This augmentation involves making small changes to the images, like flipping or rotating, to create a more varied dataset. This diversity helps the model learn better. The augmented data is divided into two parts: one for training (X_train and y_train) and another for validation (X_val and y_val). This separation helps evaluate how well the model performs on new, unseen data. Within the loop, a specialized model called PSPNet is constructed. It's designed to understand and segment building damage.

- This model uses a ResNet-50 backbone, which helps identify important features, and a spatial attention module to focus on critical parts of the image. The PSPNet model is trained using augmented training data. It learns from the images and their associated masks that indicate where building damage is present. This training process continues for a set number of cycles, improving the model's accuracy with each iteration. After training, the model's performance is tested on the validation dataset. This helps measure how well the model has learned to identify building damage. The results, such as validation loss and accuracy, can be printed out to assess the model's progress. If desired, the trained PSPNet model can be saved to the computer's storage. This way, the model can be reused later without needing to go through the training process again. The loop repeats as long as the loop counter 'i' is within the set maximum number of loops (MAX_LOOPS). In each loop, the model's understanding of building damage is refined, leading to better performance.

- Once the desired number of loops is completed, the algorithm finishes its execution. This systematic approach helps create a reliable model for segmenting building damage from images, contributing to more accurate and efficient analyses.

- PRESSNet is a deep learning model that incorporates the PSPNet model, ResNet 50 architecture, and a spatial attention mechanism to improve image segmentation performance. The evaluation metrics for this research will be the macro-average F1 Score, which will be used to compare the baseline model and the proposed model. The architecture can be seen in Fig. 3.

- The ResNet 50 backbone functions as the network's foundation. Multiple residual blocks are layered to create a deep convolutional neural network. Each residual block is comprised of multiple convolutional layers, enabling the network to learn increasingly complex characteristics.

The skip connections in the residual blocks permit gradient flow during training, thereby resolving the vanishing gradient problem and enhancing the network's capacity to acquire meaningful representations. The ResNet 50 backbone analyzes the input image and extracts multiple levels of hierarchical features, capturing both low-level details and high-level semantic information. The PSPNet (Pyramid Scene Parsing Network) module is incorporated into the architecture to collect contextual data at multiple scales.

The module operates on the ResNet 50 backbone's generated feature maps. It utilizes a pyramid pooling mechanism to combine spatially distinct features. Pyramid pooling is accomplished by dividing the feature maps into multiple regions of differing sizes and then performing pooling operations (such as average pooling) within each region. By combining features at various dimensions, the PSPNet module captures both local details and global context, giving the network a comprehensive understanding of the image.

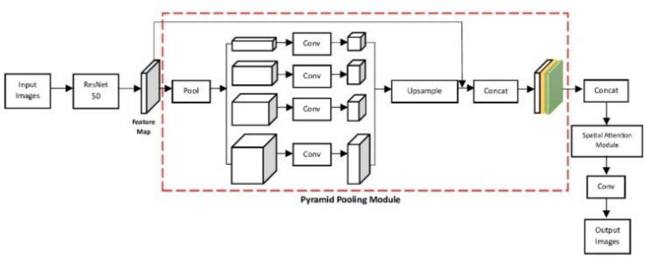


Fig. 3. The PRESSNet architecture.

The spatial attention mechanism is implemented to selectively highlight informative image regions. It improves the network's discriminative ability by focusing on pertinent features and suppressing irrelevant or chaotic regions. Indicating the significance of each location, the mechanism allocates attention weights to various spatial locations in the feature maps.

During the training process, these attention weights are learned, allowing the network to automatically attend to informative regions. The spatial attention mechanism helps the network make more accurate predictions and improves its overall performance by focusing on pertinent features. The ResNet-50 backbone in the PRESSNet architecture extracts complex image features. The PSPNet module then processes these features to capture contextual information at multiple scales. Lastly, the spatial attention mechanism selectively accentuates significant regions, thereby enhancing the network's capacity to concentrate on pertinent features.

By integrating these elements, the architecture is able to effectively capture both local and global context, extract high-level characteristics, and selectively focus on informative regions. This improves the performance of various image comprehension tasks, such as semantic segmentation, where precise object boundary delineation and accurate pixel-level predictions are essential.

IV. Result and Discussion

PRESSNet made use of the Palu earthquake disaster dataset, which it acquired from the xBD dataset. The proposed method, which is called "PRESSNet," combines the effectiveness of the PSPNet model with the architecture of "ResNet-50" and adds a "spatial attention" mechanism to get great results in tasks that have to do with understanding pictures. The ResNet-50 architecture is part of the PRESSNet framework. It uses deep residual blocks to collect high-level features and get discriminative representations in a way that uses few resources. The PSPNet model improves network performance by incorporating a pyramid pooling module that effectively combines contextual information from several dimensions. This enables the model to capture both local and global context, leading to enhanced performance. Furthermore, the integration of a spatial attention mechanism into PRESS-Net enables the network to prioritize informative regions within the input image. This capability allows the network to concentrate on pertinent characteristics while effectively reducing the impact of noise and distractions. Through the integration of several components, the suggested PRESS-Net exhibits remarkable performance in the field of image analysis, showcasing the efficacy of including the PSPNet model alongside ResNet-50 and spatial attention mechanisms for the purpose of enhancing image processing tasks. The test results are displayed in Table I.

The assessment of various damage classes was conducted using three models: the baseline model (PSPNet+ResNet50), PRESSNet, and the PSPNet + ResNet 101 + Spatial Attention model. The evaluation revealed that the baseline model achieved an F1 score of 98.6% for the "Background" class. PRESSNet demonstrated a little superior performance compared to the base model, as seen by its F1 score of 98.62%. The F1 score achieved by the enhanced model PSPNet+ResNet101+spatial attention was 94.3%. The baseline model attained an F1 score of 88.6% for the "No Damage" category. The performance of PRESSNet was somewhat inferior to that of F1, as evidenced by its score of 88.2%.

The combination of PSPNet, ResNet 101, and spatial attention achieved a notable F1 score of 75.1%. The baseline model obtained an F1 score of 81.2% in the "Destroyed" class. The performance of PRESSNet showed a slight superiority compared to F1, as substantiated by its score of 82.4%. The PSPNet+ResNet101+spatial attention model had a significantly diminished performance, as evidenced by an F1 score of 28.3%. The observed substantial decrease in the F1 score pertaining to the "Destroyed" class, resulting from the utilization of an alternative backbone (PSPNet+ResNet101+spatial attention) in comparison to the baseline model (PSPNet+ResNet50) and PRESSNet, raises the inquiry regarding the underlying reasons for the drop in performance.

Several factors may have contributed to this performance loss that the PSPNet+ResNet101+Spatial Attention model is more complicated than the base model. It includes a more advanced ResNet architecture (ResNet 101) and spatial attention. It may have been more difficult to learn effective representations for the "Destroyed" class due to the increased model complexity, resulting in a lower F1 score.

If the "Destroyed" class is not adequately represented in the training data, the model may struggle to acquire meaningful patterns for this class. Insufficient data can result in inaccurate generalizations and diminished performance, especially when dealing with uncommon or underrepresented groups.

Imbalance in Class Distribution: If the dataset is imbalanced, i.e., there is a significant difference in the number of samples between classes, the model's performance may be negatively impacted. If the "Destroyed" class is underrepresented relative to other classes, the model may not have had sufficient exposure to acquire its unique characteristics, resulting in a lower F1 score.

To resolve the low F1 score for the "Destroyed" class in the improved model, it may be advantageous to investigate and analyze the specific difficulties and constraints posed by this class in greater depth. Obtaining more representative training data, meticulously balancing the class distribution, and refining the model architecture and hyperparameters could be potential solutions for improving the model's performance on the "Destroyed" class.

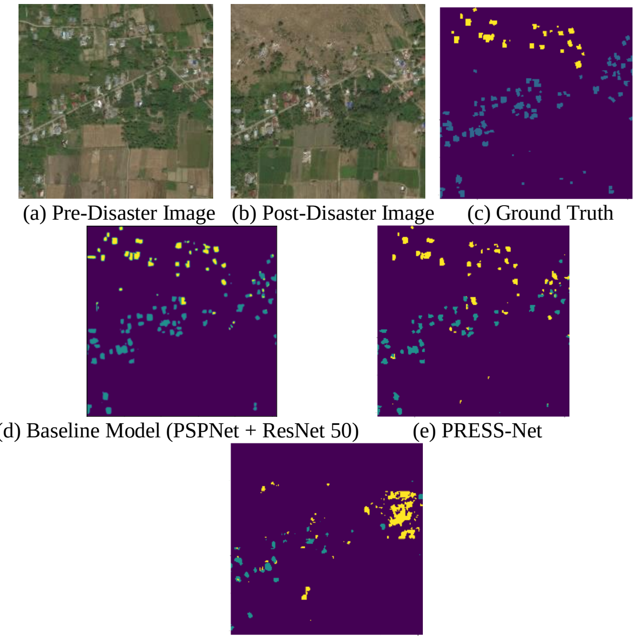
The baseline model (PSPNet+ResNet50) received an F1 score of 89.5% based on the macro-average F1 scores, which provide a comprehensive performance measurement. With an F1 score of 89.7%, PRESSNet outperformed the baseline model by a small margin. The performance of the improved model, PSPNet+ResNet101+spatial attention, was 66% on the F1 scale.

Fig. 4 shows the result between the baseline model, PRESS-Net, and PSPNet + ResNet 101 + Spatial Attention. In general, PRESSNet performs comparably or slightly better than the baseline model (PSPNet+ResNet50) across all damage classes. However, the model that used ResNet 101 backbone did not improve performance across all classes, with the "Destroyed" class experiencing a significant performance decline.

It is essential to observe that these results are dependent on the evaluation metrics and data set employed. To comprehend the factors contributing to the performance disparities between models, additional analysis and investigation are required.

TABLE I. THE TEST RESULT

Class	**Deep Learning Model**		
	Baseline Model (PSPNet + ResNet 50)	***PSPNet + Resnet 50 + Spatial Attention***	***PSPNet + Resnet 101 + Spatial Attention***
F1 Background (Class 0)	0.986	0.986	0.943
F1 No Damage (Class 1)	0.886	0.882	0.751
F1 Destroyed (Class 2)	0.812	0.824	0.280
Macro Average F1	0.895	0.900	0.659



(a) Pre-Disaster Image (b) Post-Disaster Image (c) Ground Truth

(d) Baseline Model (PSPNet + ResNet 50) (e) PRESS-Net

(f) PSP-Net +ResNet 101 +Spatial Attention

Fig. 4. The test result.

V. CONCLUSION

This paper presents the PRESS-Net methodology for developing semantic segmentation models, with a particular focus on leveraging the Palu earthquake disaster dataset obtained from the xBD dataset. The PRESS-Net model achieved remarkable results in the field of image understanding by combining the beneficial aspects of the PSPNet model with the ResNet-50 architecture while also incorporating a spatial attention mechanism. The examined technique has demonstrated enhanced performance in comparison to the baseline model (PSPNet+ResNet50), particularly in its capacity to capture both local and global context information.

Upon conducting a comparative analysis of different versions of the model, this research has come across surprising findings. The baseline model (PSPNet+ResNet50) demonstrated commendable performance, achieving an F1 Score of 89.5%. The percentage was elevated to 89.7% as a result of the intervention of PRESS-Net. The F1 Score of the more intricate model, which combines PSPNet, ResNet101, and spatial attention, experienced a decrease to 66.0%. This decrease was particularly notable in the "Destroyed" category. The potential cause for this reduction could be attributed to the model's excessive complexity or the inadequate availability of suitable instances for learning. In order to tackle this matter, the research proposes the acquisition of more photos that portray scenes of destruction, ensuring a balanced distribution of image categories, and making modifications to the structure of the model.

In the future, there are several exciting research possibilities for improving building damage assessment. One option is to explore different types of deep learning models that might be better at recognizing structural damage. This could involve trying out various model designs or new methods in deep learning. Using larger and more diverse datasets can also help the model work better in different situations. These datasets should cover a wide range of disasters and locations. To truly understand how well the model works in the real world, we need to test it with actual aerial and satellite images. Additionally, ongoing efforts to improve the model, like transfer learning or fine-tuning, can make a big difference when working with larger datasets. We should also consider adding more types of data, like weather or location information, to make the predictions more accurate. Lastly, we should validate the model's results in real-life situations to ensure it works effectively in disaster management. These suggestions are intended to guide future research in making deep learning models for building damage assessment more effective and reliable.

REFERENCES

[1] Y. Wang, S. Ruan, T. Wang, and M. Qiao, "Rapid estimation of an earthquake impact area using a spatial logistic growth model based on social media data," International Journal of Digital Earth, vol. 12, pp. 1-20, 2018. doi:10.1080/17538947.2018.1549108.

[2] Y. Du, L. Gong, Q. Li, and F. Wu, "Earthquake-induced building damage assessment on SAR multi-texture feature fusion," in Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 6608-6610, 2020. doi:10.1109/IGARSS39084.2020.9323276.

[3] C. Lin, Y. Li, Y. Liu, X. Wang, and S. Geng, "Building damage assessment from post-hurricane imageries using unsupervised domain adaptation with enhanced feature discrimination," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-10, 2021. doi:10.1109/TGRS.2021.3060460.

[4] R. Zhang, H. Li, K. Duan, S. You, K. Liu, F. Wang, and Y. Hu, "Automatic detection of earthquake-damaged buildings by integrating UAV oblique photography and infrared thermal imaging," Remote Sensing, vol. 12, pp. 2621, 2020. doi:10.3390/rs12162621.

[5] S. Koshimura, L. Moya, E. Mas, and Y. Bai, "Tsunami damage detection with remote sensing: A review," Geosciences, vol. 10, pp. 177, 2020. doi:10.3390/geosciences10050177.

[6] 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 Sensing, vol. 11, pp. 2765, 2019. doi:10.3390/rs1123276.

[7] Y. Endo, B. Adriano, E. Mas, and S. Koshimura, "New insights into multiclass damage classification of tsunami-induced building damage from SAR images," Remote Sensing, vol. 10, pp. 2059, 2018. doi:10.3390/rs10122059.

[8] S. Garg et al., "Building damage assessment using deep learning from satellite imagery," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 133-140, 2017.

[9] H. Hu et al., "Building damage assessment from aerial imagery using Siamese convolutional neural network," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, pp. 3836-3849, 2018.

[10] S. K. Shah et al., "Building damage assessment from satellite imagery using deep learning," Remote Sensing, vol. 11, pp. 2953, 2019.

[11] K. He et al., "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[12] M. Waseem et al., "Building damage assessment using a hybrid deep learning approach from post-disaster satellite imagery," Remote Sensing, vol. 12, pp. 3674, 2020.

[13] H. Zhao et al., "Pyramid scene parsing network," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881-2890, 2017.

[14] Z. M. Hamdi, M. Brandmeier, and C. Straub, "Forest damage assessment using deep learning on high-resolution remote sensing data," Remote Sensing, vol. 11, pp. 1976, 2019.

[15] W. Yuan, J. Wang, and W. Xu, "Shift pooling PSPNet: rethinking PSPNet for building extraction in remote sensing images from entire local feature pooling," Remote Sensing, vol. 14, pp. 4889, 2022.

[16] J. Chen et al., "DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 1194-1206, 2021.

[17] V. Mnih et al., "Recurrent models of visual attention," in NIPS.

[18] L. Chen et al., "Attention to scale: Scale-aware semantic image segmentation," in CVPR.

[19] E. Weber and H. Kané, "Building disaster damage assessment in satellite imagery with multi-temporal fusion," 2020. doi: 10.48550/arxiv.2004.05525.

[20] A. Vaswani et al., "Attention is all you need," in NIPS.

[21] X. Wang et al., "Non-local neural networks," in CVPR.

[22] H. Zhao et al., "Psanet: Point-wise spatial attention network for scene parsing," in Proceedings of the European conference on computer vision (ECCV),pp. 267-283, 2018.

[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proceedings of the European Conference on Computer Vision (ECCV), pp. 267-283, 2018.