# U-Net Based Disaster Damage Detection Through Semantic Segmentation

Nishat Ara Nipa
*Virginia Modeling, Analysis & Simulation Center*
*Old Dominion University*
Norfolk, VA, USA
nnipa001@odu.edu

Dr. Yan Lu
*Virginia Modeling, Analysis & Simulation Center*
*Old Dominion University*
Norfolk, VA, USA
y21u@odu.edu

Dr. Sachin Shetty
*Virginia Modeling, Analysis & Simulation Center*
*Old Dominion University*
Norfolk, VA, USA
sshetty@odu.edu

2023 International Symposium on Networks, Computers and Communications (ISNCC) | 979-8-3503-3559-0/23/$31.00 USG | DOI: 10.1109/ISNCC58260.2023.10323958

***Abstract*—Satellite image analysis of natural disasters is critical for effective emergency response, relief planning, and disaster prevention. They can provide a comprehensive view of the affected area allowing for quick and accurate assessments of the extent of damage by providing critical information on the disaster's development. Semantic segmentation is believed to be one of the best techniques to capture pixel-wise information in computer vision. In this work, we will be using U-Net-based architectures to address the building localization and damage detection of the complex Xview2 dataset. We propose a generic algorithm that is not constrained by the nature of the disaster and can be deployed to various classes of disaster types with slight modifications.**

***Keywords*—Natural disaster, Damage assessment, Semantic Segmentatio, U-Net, Attention Mechanism.**

## I. Introduction

Proper assessment of situational information during or after a disaster can be helpful in assisting the authorities in identifying the regions most vulnerable to damage and initiating appropriate relief efforts. Proper damage assessment can help the authorities determine the extent of damage, and what can be replaced, restored, or salvaged. Traditional damage assessment schemes involve on-site inspections, surveys, and data collection by human experts. They are all very valuable methods but it's high time we take leverage of advanced Remote Sensing, Artificial Intelligence technologies to complement the traditional methods by providing more accurate and time-sensitive information.

High-resolution satellite Images though providing a foundation for unbiased overhead views of post-disaster scenarios are not merely enough to recommend recovery initiations. A lot needs to be deduced from the satellite image to conclude [1]. The disaster data cover a large ground area with very few instances of major disaster damage instances. The data is inherently imbalanced and one must search through a huge pixel space to localize and mark the damaged area of interest.

(a) Pre-disaster (b) Post-Disaster

Analyzing pre and post-disaster images to detect damage has been a common practice in recent years. Researchers initially tried to apply ML methods such as Support Vector Machines [2], Random Forest [3], or Multilayer Feedforward Neural Networks (MFNN) [4]. For example, the authors in [2] used SVMs to analyze hyperspectral data both on a pixel and object level and obtained a better quantification of the object level. The work in [3] combines Linear Clustering with Random Forest to do super-pixel level segmentation and finally build a classifier based on the extracted features. The MFNN method in [4] provides the best accuracy score that assesses damage using high-resolution satellite data from the 2010 Haiti earthquake. The drawback with the ML algorithms is that each of the models is designed to address a specific disaster and is not robust to the real-world scenarios that demand a generic model able to work for several disaster types. The deep learning algorithm can provide us with a generic architecture that can overcome the case-specific ML approaches and can be deployed through multiple disaster scenarios. For example, Xu et al.(2019) [5] did a binary classification based on satellite images from three disaster types by deploying four different CNN architectures. Their result proves the Twin-tower Substract (TTS) Model Functioning is best for capturing the difference in the pre- and post-disaster images being a good indicator of building damage. The work in [6] uses multi-temporal fusion techniques by using Mask R-CNN with FPN architecture fused with the final segmentation layer and gets an improved result compared to the xDB baseline. [7] develops a two-stage U-Net-based network where the first stage focuses on building segmentation and is followed by a Siamese U-Net to perform the damage classification. The work in [8] compares viz.U-Net with PSPNEt for a single view semantic segmentation on the World-view 3 satellite dataset and concludes that U-Net outperforms with a better Mean-IOU value.

There are two important issues that need to be addressed in any damage assessment algorithms- 1. Localization: where the focus is proper building location segmentation. In satellite images, the buildings are congested in their placement and the background is also complex comprising different objects. So a proper building segmentation model is crucial. 2. Damage Classification where the level of damage is captured. In this work, we use a generic U-Net architecture for semantic segmentation of buildings for the Localization problem using the publicly available Xview2 dataset using both Pre and Post-Disaster sets. On the next level, we take advantage of the attention mechanism to incorporate into the U-Net architecture and finally pixel-based object detection to detect the damaged part of the image.

## II. Methods & Model

### A. Localization

The first stage of the framework focuses on Building a Segmentation network using U-Net architecture as a baseline. In the pre-disaster images, the buildings are intact and can be utilized to segment the location of the buildings. Satellite images are rich in feature information and Semantic Segmentation is believed to be one of the best techniques to capture pixel-wise information in computer vision. Though dominantly used in medical image segmentation, the U-Net architecture is well-proven to work for a diverse set of satellite imagery.

U-Net for Semantic Segmentation:

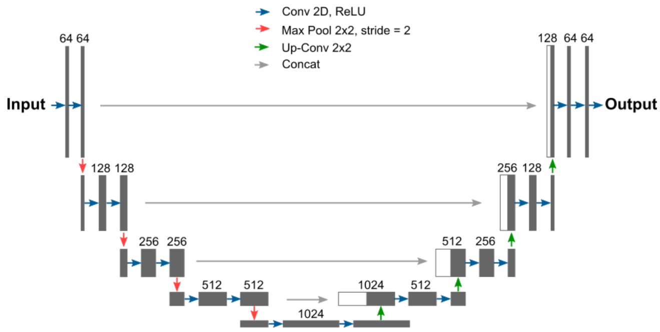


Fig 1. U-Net Framework for Building Segmentation.

The U-net architecture consists of a contracting or encoder that downsamples the images and an expanding or decoder path that upsamples the images. A sequence of convolutional and pooling layers is used in the contraction path to reduce the spatial dimensions of the input while increasing the number of channels. High-resolution characteristics from the contracting path are coupled with upsampled output for localization. A subsequent convolution layer can then learn to assemble information more precisely. The upsampling part eventually possesses a large number of feature channels which by default allows the network to propagate context information to higher resolution layers making the expansive path symmetric to the contracting path[9]. The expanding path is made up of convolutional and upsampling layers that enhance the spatial dimensions of the input while decreasing the number of channels.Skip connections connect the encoder and decoder by concatenating feature maps from the contracting path to the equivalent feature maps in the expanding path. The concatenation of the feature maps is what gives us the benefit of providing localized information and results in optimum segmentation accuracy.

### B. Damage Detection

The next part of the framework focuses on detecting damage from the post-disaster images. The dataset consists of 5 labels. The challenge with the Xview dataset is the high level of variability present in the data. The dataset consists of 5 classes (0-Background, 1-No Damage, 2-Minor Damage, 3-Major Damage, 4-Destroyed) being highly biased towards the "No Damage" class. As the damaged part consists of very few pixels of the entire image, it's not possible for the general U-Net architecture to segment those few pixels with damaged labels. Soft attention introduced to the network weights different parts of the image differently. So, the relevant parts of the image get larger weights compared to the less relevant ones. So, Attention U-Nets are able to highlight the relevant activations during training and provide a better generalization of the network.

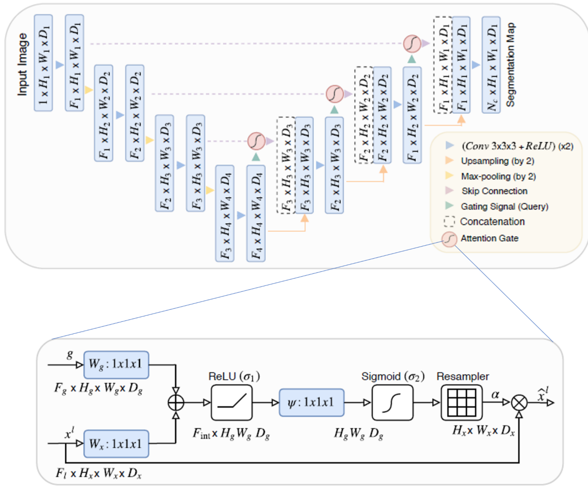


Fig 2. Attention gate introduced to U-Net Architecture [10]

The concatenation paths in U-Net combine the spatial information from the down-sampling path with the up-sampling path to retain good spatial information. The bottleneck here is that it brings along poor feature representation as it comes from a very early stage of the network. Soft attention implemented at the skip connections actively suppresses activations at relevant regions and can provide more weightage to the feature of interest[10]. As shown in Fig 2. $g$ is the gating signal that comes from the next lowest layer of the network. As it comes from a deeper part of the network, it has better feature representation. The input features, $x^l$ are scaled up with attention coefficients at the attention.

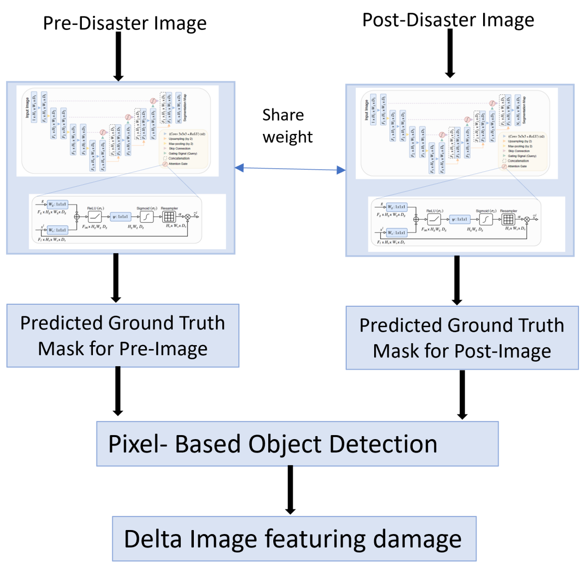


Fig 3. Framework for Building Damage Detection.

It comes from the early layers and has better spatial information. The attention gates can easily be plugged into the standard U-Net architecture to emphasize important features that are passed through the skip connections.

We use the Attention U-Net framework inspired by [10] which was originally trained for medical image segmentation. We train the model separately for Pre-Disaster and Post-Disaster images along with their ground truth mask and finally perform pixel-based object detection to extract the difference image from the predicted masks. We can see the summary of the model in Fig 3.

## III. Preparing the Dataset

We've used the publicly available xView2 dataset. The input RGB square images are paired in pre/post-identified by matching numerical IDs for each set along with their corresponding ground truth mask. The pixel label description of the dataset is shown in Table 1. Around the world, disasters, buildings, and land use patterns differ. For example, earthquake damage appears extremely different from damage caused by flood or a burnt building in Australia appears much different from one in Nepal. There was no publicly accessible dataset of labeled high-resolution satellite photos spanning a wide range of disaster kinds around the world before the release of this dataset. While it's a great collection of HRS images of various kinds, the quality of the dataset makes it extremely challenging to come up with a generic framework for this Dataset

| Mask Value | Description | Pre-Image | Post-Image |
|---|---|---|---|
| 0 | No Building | Yes | Yes |
| 1 | Undamaged Building | Yes | Yes |
| 2 | Building with minor damage | No | Yes |
| 3 | Building with major damage | No | Yes |
| 4 | Destroyed Building | No | Yes |

Table 1. Damage Description

For simplification, while training the attention U-Net model we considered labels 2,3 & 4 as one class (2) as instances for all these labels are very low compared to the background and the undamaged buildings.

## IV. Result & Evaluation

### A. *Building Segmentation*

The greatest challenge while doing the building segmentation is to differentiate the buildings from the complex background of the HRS image where a large number of pixels are occupied by various kinds of objects and are distributed in a very uneven way. We resized all our images to 256x256 size and fed them as input to the U-Net architecture. The encoder block consists of a repeated application of two unpadded convolutions. Each of them is followed by a Rectified Linear Unit which adds an extra pixel at the edges so the dimension of the output image is the same as the input image. We added a feature space of 16 and doubled the number at each downsampling step. The decoder part of the network consists of transposed convolutional layers followed by concatenation with the corresponding output from the encoder and convolution block. The dropout layer is utilized for regularization, and batch normalization is done after each convolutional layer to enhance model convergence.

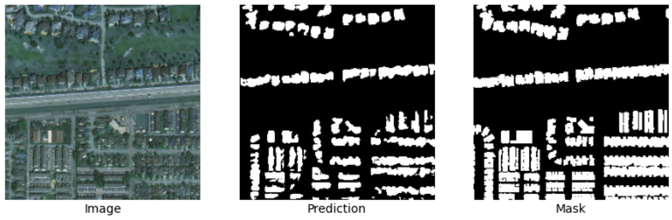


Fig 4. Visualization of Building Segmentation Result

We have evaluated the model based on F1 scores which are a summed average of precision (P) and recall (R) values and can be used when there is a class balance in the dataset. As per the pre-images consisting of a balanced number of building and background pixels F1 score is a reliable evaluation metric for this case to see how many times the model made a correct prediction over the entire dataset. The evaluation results are summarized in Table 2. We used True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) to calculate the scores. Where-

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1); \quad \text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

The F1 score is calculated as the harmonic mean of precision and Recall as follows:

$$\text{F1 Score} = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} = \frac{TP}{TP+1/2(FP+FN)} \quad (3)$$

| Model | Accuracy | Precision | Recall | F1 score |
|---|---|---|---|---|
| Building Segmentation | 89.13% | 0.94 | 0.67 | 0.79 |

Table 2. Evaluation scores for Building Segmentation

### B. *Damage Detection*

As mentioned earlier segmenting damaged pixels from the post-disaster set is difficult due to the imbalanced nature of the dataset. So, we took a different approach other than only segmenting the images. The predicted mask for the pre-disaster masks has all the buildings damaged or non-damaged (Mask Value: 0,1) before the occurrence of the disaster while the post-disaster disaster masks contain separate values for damaged classes (Mask Value : 0, 1, 2). We need to be more specific to differentiate the pixels marked as buildings to be damaged or non-damaged. So, we introduced attention mechanism to the U-Net architecture to get the inference images for the pre and post-disaster pair.

Instead of just segmenting the images our framework does a pixel-wise comparison between the pre and post-inference mask and comes up with a resultant delta image where the post-mask is different from the pre-mask which is by default the damage class. So,
delta_image = Pre_Inference_mask–Post_Inference_ Mask.

Because of the dominant class imbalance present in the data, the F1 score is no longer a reliable metric to evaluate the model. Rather we'll use IOU or Intersection-merge ratio to evaluate the comprehensive performance of the model. We can visualize the damage detection in Fig 5 with their corresponding IOU values for each image. Fig 6 superimposes an example post-image on top of a pre-image with different transparency and the comparable delta image (for damage class is placed on the right.

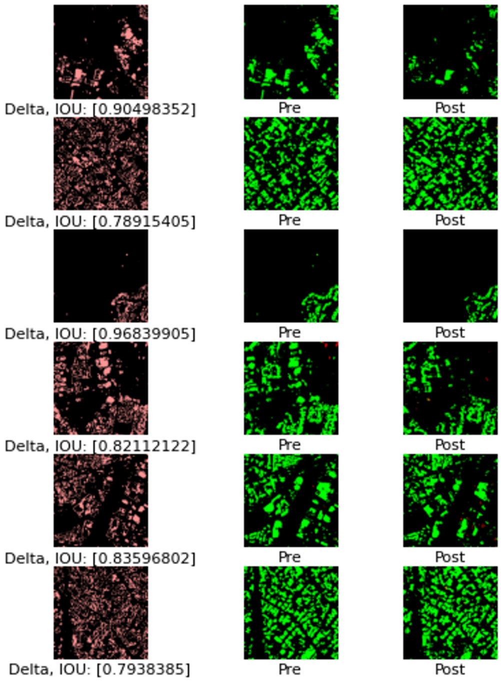


Fig 5. Comparable Pre, Post, and Delta images with Corresponding IOU scores.

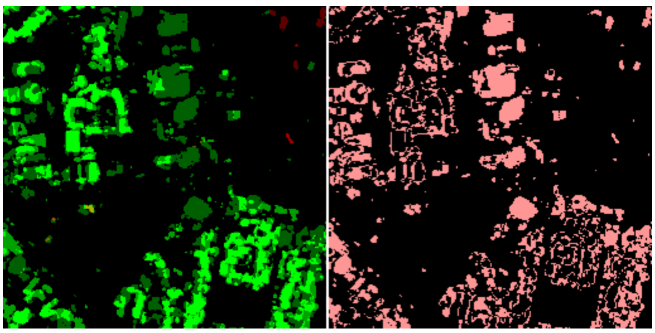

Fig 6: Post-image superimposed on the pre-image (Left side) and the delta image (Right side)

## Conclusion & Future Work

In this work, we constructed a framework for building damage detection which can segmentize the buildings from the Pre-Disaster images with a binary accuracy score of 89.13% and F1 score of 0.79 compared to the baseline F1 score of 0.77 for the dataset. We also proposed a framework for damage detection through inference masks obtained from semantic segmentation through Attention U-Net. We acquired an overall IOU score of 0.823 for the change detection task of the pre and post-image. The work is a specimen of advanced deep learning methods with Remote Sensing technologies and can be deployed as a general framework throughout a good many numbers of disasters. Future work opportunity lies in developing a better model at the inference level to address the class imbalance of the data more accurately.

## Data Availability

The data is publicly available through registration in this archive: https://xview2.org/dataset

## Acknowledgement


This work is supported in part by DoD Center of Excellence in AI and Machine Learning (CoE-AIML) under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory, National Science Foundation under Grant No. 2219742 and Grant No. 2131001.


## References


[1] Zhao, X. et al. Advances of satellite remote sensing technology in earthquake prediction. Nat. Hazard. Rev. 22, 03120001 (2021).

[2] Li, P., Xu, H. & Guo, J. Urban building damage detection from very high-resolution imagery using OCSVM and spatial features.

[3] Joshi, A. R., Tarte, I., Suresh, S. & Koolagudi, S. G. Damage identification and assessment using image processing on post-disaster satellite imagery. In Proceedings of the 2017 IEEE Global Humanitarian Technology Conference (GHTC). 1–7 (2017).

[4] Cooner, A. J., Shao, Y. & Campbell, J. B. Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake. Remote Sens. 8, 868 (2016).

[5] Xu, J. Z., Lu, W., Li, Z., Khaitan, P., & Zaytseva, V. (2019). Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks. arXiv preprint arXiv:1910.06444.

[6] Weber, E., & Kan, H. (2020). Building disaster damage assessment in satellite imagery with multi-temporal fusion. arXiv preprint arXiv:2004.05525.

[7] Deng, L., Wang, Y. Post-disaster building damage assessment based on improved U-Net. Sci Rep 12, 15862 (2022). https://doi.org/10.1038/s41598-022-20114-w

[8] Chaurasia, K., Nandy, R., Pawar, O. et al. Semantic segmentation of high-resolution satellite images using deep learning. Earth Sci Inform 14, 2161–2170 (2021). https://doi.org/10.1007/s12145-021-00674-

[9] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597

[10] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert. "Attention U-Net: Learning Where to Look for the Pancreas." 2018. arXiv:1804.03999 [cs.CV].

## References


[1] Zhao, X. et al. Advances of satellite remote sensing technology in earthquake prediction. Nat. Hazard. Rev. 22, 03120001 (2021).

[2] Li, P., Xu, H. & Guo, J. Urban building damage detection from very high-resolution imagery using OCSVM and spatial features.

[3] Joshi, A. R., Tarte, I., Suresh, S. & Koolagudi, S. G. Damage identification and assessment using image processing on post-disaster satellite imagery. In Proceedings of the 2017 IEEE Global Humanitarian Technology Conference (GHTC). 1–7 (2017).

[4] Cooner, A. J., Shao, Y. & Campbell, J. B. Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake. Remote Sens. 8, 868 (2016).

[5] Xu, J. Z., Lu, W., Li, Z., Khaitan, P., & Zaytseva, V. (2019). Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks. arXiv preprint arXiv:1910.06444.

[6] Weber, E., & Kan, H. (2020). Building disaster damage assessment in satellite imagery with multi-temporal fusion. arXiv preprint arXiv:2004.05525.

[7] Deng, L., Wang, Y. Post-disaster building damage assessment based on improved U-Net. Sci Rep 12, 15862 (2022). https://doi.org/10.1038/s41598-022-20114-w

[8] Chaurasia, K., Nandy, R., Pawar, O. et al. Semantic segmentation of high-resolution satellite images using deep learning. Earth Sci Inform 14, 2161–2170 (2021). https://doi.org/10.1007/s12145-021-00674-

[9] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597

[10] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert. "Attention U-Net: Learning Where to Look for the Pancreas." 2018. arXiv:1804.03999 [cs.CV].