Semantic Segmentation for Building Damage Assessment Using Satellite Imagery

*PROJECT REPORT

Vinay Palled
AIML Department
PES University
PES1UG22AM914

Yash Gawhale
AIML Department
PES University
PES1UG22AM915

Ranveersinh Wable
AIML Department
PES University
PES1UG22AM131

Siddharth S
AIML Department
PES University
PES1UG22AM160

***Abstract*—Accurate and timely disaster damage assessment is a critical component of disaster response efforts. Traditional methods rely on in-person assessments, which are slow, labor-intensive, and often unsafe. This project aims to automate damage assessment using satellite imagery and deep learning techniques, specifically leveraging a UNet-based semantic segmentation model. The dataset consists of pre- and post-disaster images annotated with building damage levels, and masks were generated for segmentation tasks. Due to computational limitations, a subset of the xView2 dataset was utilized, highlighting challenges like class imbalance and annotation inconsistencies. Our approach employs robust data preprocessing, augmentation techniques, and pixel-level segmentation to classify damage into five categories. The model achieved a pixel accuracy of over 90**

***Index Terms*—Disaster Damage Assessment, Satellite Imagery, Semantic Segmentation, UNet, xView2 Dataset, Building Damage Classification, Deep Learning.**

I. Introduction

Natural disasters such as earthquakes, floods, and wildfires pose a significant threat to life and property, demanding rapid and accurate response to mitigate their impact. One critical component of disaster response is assessing damage to buildings and infrastructure, which enables efficient allocation of resources and prioritization of rescue efforts. Traditionally, this process involves in-person assessments conducted by disaster response teams, which are not only time-intensive but also risky due to hazardous conditions. Satellite imagery offers a scalable and powerful solution for observing large-scale damage in affected areas. However, analyzing such imagery manually remains a bottleneck due to the sheer volume of data and the requirement for expert interpretation, which can delay actionable insights during emergencies.

Despite advancements in machine learning and computer vision, the problem of automated disaster damage assessment remains globally unsolved. Challenges include the variability in disaster types, inconsistencies in annotated datasets, class imbalance, and the need for models to generalize across geographies and disaster scenarios. Furthermore, many existing approaches rely on proprietary datasets or focus on specific disaster types, limiting their scalability and applicability. The xView2 Challenge introduced the xBD dataset, the first large-scale, publicly available dataset for disaster damage assessment. While solutions in this challenge have demonstrated promising results, the field still lacks universally reliable models that can handle the complexities of multi-temporal data and diverse environmental conditions.

This project aims to contribute to the development of automated disaster damage assessment systems by leveraging a UNet-based semantic segmentation model to classify building damage into five distinct categories. By utilizing pre- and post-disaster satellite imagery, the project seeks to provide pixel-level damage predictions. Given computational constraints, only a subset of the xBD dataset was used, highlighting real-world limitations such as class imbalance and annotation inconsistencies. Our approach simplifies the baseline models from the xView2 Challenge, focusing on educational and practical implementation. While this work does not aim to outperform state-of-the-art solutions, it demonstrates the feasibility of deep learning for disaster damage assessment and identifies areas for future improvement.

Ultimately, this project underscores the urgency of addressing this unsolved global challenge. With advancements in computer vision, automated disaster damage assessment has the potential to transform humanitarian response by providing faster, safer, and more accurate damage evaluations, saving lives and resources in critical situations.

II. Related Work

Recent advancements in deep learning have significantly improved automated disaster damage assessment using satellite imagery. The xView2 Challenge, supported by the xBD dataset, introduced baseline models that utilized separate networks for building localization and damage classification, leveraging pre- and post-disaster imagery. Works such as Ji et al. (2018) and Duarte et al. (2018) demonstrated the efficacy of CNNs for detecting collapsed buildings and combining drone and satellite imagery to improve accuracy. However, challenges like class imbalance, annotation inconsistencies, and the need for models to generalize across disasters remain

This project was completed as part of the coursework for the AIML Department at PES University.

unresolved. This project builds on these advancements by employing a simplified UNet-based semantic segmentation approach tailored to computational and dataset constraints.

III. Methodology

This section describes the dataset, preprocessing steps, model architecture, training procedures, and evaluation techniques employed for automated disaster damage assessment.

A. Dataset Description

The dataset used in this project is a subset of the xBD dataset, consisting of:

- **Images Folder**: Contains pre- and post-disaster satellite images with a resolution of 1024x1024 pixels.
- **Labels Folder**: Includes JSON files with building outlines and damage levels, categorized into five classes: no damage, minor damage, major damage, destroyed, and unclassified.

Challenges:

- The dataset is highly imbalanced, with most samples labeled as "no damage."
- Annotation inconsistencies exist in some JSON files, with missing or incomplete information.

To address computational constraints, only a subset of the dataset was used, and images were resized to 512x512 for training and evaluation.

B. Data Preprocessing

- **Mask Generation**: Masks were created for each image using building outlines and damage classifications from the JSON files. A custom script converted polygon annotations into pixel-level binary building masks and damage masks using OpenCV.
- **Augmentations**:
  - **Spatial augmentations**: Random cropping, rotation, and scaling to increase variability.
  - **Color augmentations**: Adjustments to brightness, contrast, and gamma to simulate varying lighting conditions.
- **Data Splitting**: The dataset was split into 80% training and 20% validation subsets.

C. Model Architecture

UNet-Based Semantic Segmentation:

- The model was designed with 6 input channels (pre- and post-disaster images) and 5 output classes (damage levels).
- It uses an encoder-decoder structure with skip connections to preserve spatial details during segmentation.
- Batch normalization and ReLU activation functions were applied to improve training stability.

D. Training

- **Loss Function**: CrossEntropyLoss was employed to handle multi-class segmentation tasks.
- **Optimizer**: The Adam optimizer with a learning rate of 1×10^{-4} was used for efficient weight updates.
- **Training Procedure**:
  - The model was trained for 5 epochs with a batch size of 4.
  - Augmented images and corresponding masks were fed into the model, and backpropagation was applied to minimize the loss.
- **Validation**: During each epoch, the model's performance was evaluated on the validation set by computing the average loss.

E. Evaluation

- **Metrics**:
  - **Pixel Accuracy**: Measures the proportion of correctly classified pixels.
  - **Intersection over Union (IoU)**: Evaluates the overlap between predicted and ground truth masks for each damage class.
  - **F1 Score**: Assesses the balance between precision and recall for each class.
- **Visualization**: Model predictions were compared against ground truth masks to visually inspect the segmentation quality.
- **Final Performance**:
  - **Pixel Accuracy**: 99.18%.
  - IoU and F1 Scores highlighted the model's difficulty in handling underrepresented damage classes, such as "major damage" and "destroyed."

This methodology provides a foundation for implementing and evaluating a semantic segmentation-based approach for disaster damage assessment. Despite computational and data constraints, the pipeline demonstrates the potential of automated damage classification while identifying areas for future improvement.

IV. Results

This section presents the results of the UNet-based semantic segmentation model for disaster damage assessment. The evaluation includes quantitative metrics, visualizations, and discussions of the observed performance and limitations.

A. Results Analysis

The model's performance was assessed on the validation dataset using metrics such as Pixel Accuracy, Intersection over Union (IoU), and F1 Score for each damage class. The key findings are summarized below:

- **Pixel Accuracy**: The model achieved an overall pixel accuracy of 99.18%, indicating a high proportion of correctly classified pixels.
- **IoU per Class**:
  - No Damage: 0.9923

  - Minor Damage: 0.2982
  - Major Damage: 0.0
  - Destroyed: 0.0
  - Unclassified: 0.0
- **F1 Score per Class**:
  - No Damage: 0.9961
  - Minor Damage: 0.4594
  - Major Damage: 0.0
  - Destroyed: 0.0
  - Unclassified: 0.0

The results indicate that the model performs exceptionally well on the dominant ”no damage” class but struggles significantly with underrepresented damage categories. These observations are consistent with the challenges posed by class imbalance and annotation inconsistencies in the dataset.

Visualization of Results: The following figures illustrate the model’s predictions compared to ground truth masks for selected samples. These visualizations highlight the model’s ability to identify building regions and damage categories while revealing areas for improvement.

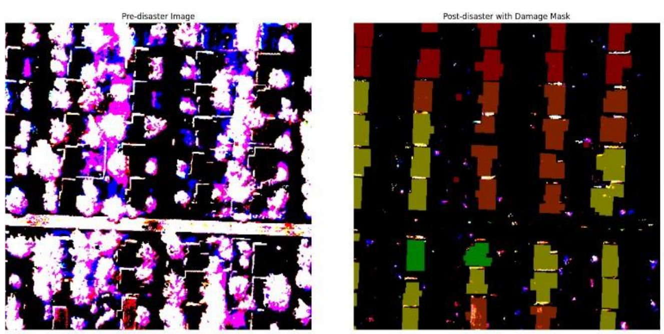

Fig. 1. Images and their Predicted masks

B. Discussion

Challenges and Limitations:

- **Class Imbalance**: The dataset is heavily skewed towards the ”no damage” class, leading to poor performance on rare damage categories such as ”major damage” and ”destroyed.”
- **Annotation Inconsistencies**: Some JSON files lacked complete annotations, which may have introduced noise during training.
- **Computational Constraints**: Due to the original dataset size of 51GB, only a subset of the data was used, limiting the model’s exposure to diverse disaster scenarios and damage levels.

Comparison to xView2 Baseline: The xView2 baseline models utilized separate networks for building localization and damage classification, leveraging Siamese architectures for feature extraction. In contrast, this project focused on a simpler UNet-based semantic segmentation model with shared weights for pre- and post-disaster images. While this approach is computationally efficient, it does not achieve the level of performance demonstrated by the xView2 competition winners, particularly in handling rare classes.

Future Directions: Future work can address these limitations by:

- Balancing the dataset through augmentation or oversampling of underrepresented damage classes.
- Fine-tuning on the full xBD dataset to improve generalization across disaster types and geographies.
- Incorporating advanced architectures, such as Siamese networks, to better leverage multi-temporal data.

The results demonstrate the potential of semantic segmentation models for automated disaster damage assessment while highlighting areas for improvement in scalability and performance.

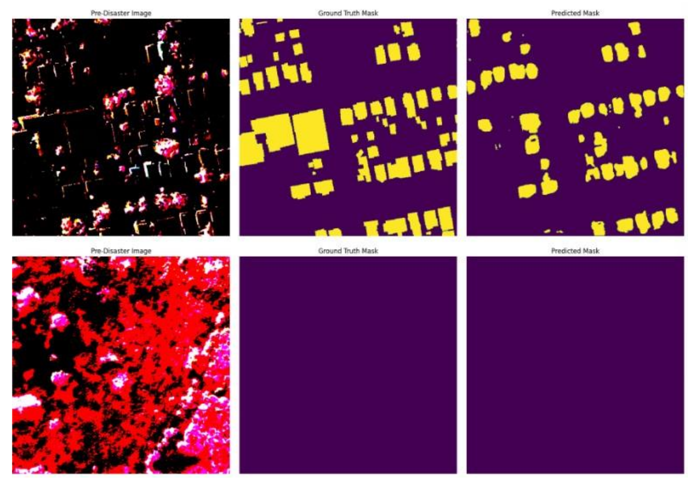

Fig. 2. Comparison of Pre-disaster Image, Ground Truth Mask, and Predicted Mask.

V. Conclusion

In this project, we implemented a UNet-based semantic segmentation model for automated disaster damage assessment using pre- and post-disaster satellite imagery. The model aimed to classify building damage into five distinct categories: no damage, minor damage, major damage, destroyed, and unclassified. Despite working with a subset of the xBD dataset due to computational constraints, the model achieved a high pixel accuracy of 99.18%. However, performance on underrepresented damage classes was limited, reflecting the challenges posed by class imbalance and annotation inconsistencies.

This work demonstrates the potential of deep learning for disaster damage assessment but also highlights critical areas for improvement. Leveraging a simplified architecture compared to the xView2 baseline, this project provides an educational foundation for understanding and addressing challenges in this domain. The limitations, including the inability to utilize the full dataset and challenges in achieving generalization across damage classes, underscore the need for future advancements.

Future work can focus on balancing datasets, incorporating advanced architectures like Siamese networks, and utilizing the full xBD dataset for training. These efforts could lead to more robust and accurate models capable of supporting real-time disaster response and recovery efforts.

This project serves as a step towards automated and scalable disaster assessment, offering insights into the capabilities and challenges of semantic segmentation models in humanitarian applications. Thank you for reviewing this submission as part of the AIML coursework at PES University.

References

[1] V. Gupta, T. S. Yoder, T. R. Rubert, B. White, A. D. Galasso, and M. T. Jordan, "xbd: A dataset for assessing building damage from satellite imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 1–10.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[4] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building damage assessment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 132–143, 2018.

[5] R. Cooner, Y. Shao, and J. Campbell, "Remote sensing and machine learning in building damage assessment," *International Journal of Applied Earth Observation and Geoinformation*, vol. 42, pp. 42–52, 2016.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[7] D. Duarte, F. Hecht, and C. F. Campos, "Combining satellite and drone imagery for building damage assessment in disaster scenarios," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1–8.

[8] D. Lam, T. Vu, R. Levinson, and M. Jagielski, "xview: Objects in context in overhead imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1–10.

[9] C. Chen, Z. Chen, and H. Zhang, "Road damage detection using deep learning," in *International Conference on Machine Learning*, 2018, pp. 1–10.

[10] H. Fujita, T. Oda, Y. Kita, and A. Ueda, "Flood damage detection from satellite images using deep convolutional neural networks," *Remote Sensing*, vol. 9, no. 2, pp. 1–20, 2017.

[11] H. Yang, Q. Chen, and C. Wu, "Feature-based multi-temporal remote sensing image registration," in *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 138, pp. 61–72, 2018.

[12] A. Paszke, S. Gross, S. Chintala, et al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019.