International Cartographic Association
Association Cartographique Internationale

Perceiving Multidimensional Disaster Damages from Street-View Images Using Visual-Language Models

Yifan Yang [a], Lei Zou [a,*]

[a] *Department of Geography, Texas A&M University, Yifan Yang - yyang295@tamu.edu, Lei Zou - lzou@tamu.edu*

* Corresponding author

Keywords: Large Language Models, Disaster Perception, Street-view Imagery, Multimodal AI, Resilience

Abstract:

Timely and accurate disaster damage perception, including the assessment of impact types and intensities, is a critical component of effective disaster management. It plays a pivotal role in identifying victims in need, prioritizing emergency resource allocation, and enabling efficient and equitable recovery efforts. Traditional disaster damage perception efforts often rely on field surveys or satellite-based observations, which lack the capacity to provide extensive spatial coverage and hyperlocal detail. The increasing availability of street-view images offers a new avenue for ground-level disaster damage assessment. Pre-disaster street-view images capture rich information about the landscape and built environment, while post-disaster images provide visual evidence of disaster impacts, such as flooded streets, fallen trees, collapsed buildings, and damaged infrastructure. However, automatically perceiving precise, multidimensional disaster impacts from street-view images remains technically challenging (Yang et al., 2025).

The recent advances in large language models (LLMs) demonstrate their extensive contextual knowledge of natural disasters, powerful text generation capabilities, and emerging perceptual abilities, offering opportunities to address the challenge. This study investigates and extends the potential of LLMs in perceiving multidimensional disaster impacts through bi-temporal street-view imagery. Street-view images captured before and after Hurricane Milton in 2024 in Horseshoe Beach, Florida, United States, were collected for experimental analysis. We chose GPT-4o-mini as the LLM in this investigation. The objectives are: (1) to evaluate the capabilities of LLMs in disaster perception tasks, including structured damage scoring, post-disaster scene description, and change detection between pre- and post-disaster imagery; and (2) to synthesize pseudo post-disaster imagery by combining pre-disaster visuals with LLM-generated descriptions, facilitating comparative learning with real-world disaster scenes and supporting scenario simulation.

Hurricane Milton made landfall in Florida on October 9, 2024, and dissipated on October 13. It caused approximately $34.3 billion in damages and at least 35 fatalities. The first analysis explores the capacity of GPT-4o-mini in street-view image description, damage classification, and structured scoring of impacts by type and intensity. Figure 1 shows the study area, data coverage, and LLM-based damage perception framework. We collected 3246 street-view images captured on October 17, 2024, in Horseshoe Beach, Florida, U.S. to assess the post-disaster conditions. Another 2610 pre-disaster street view images of the same study area were obtained. We manually screened the two collections and identified 2556 pre- and post-disaster street view image pairs. The API of the GPT-4o-mini model was leveraged to explore disaster perception tasks.

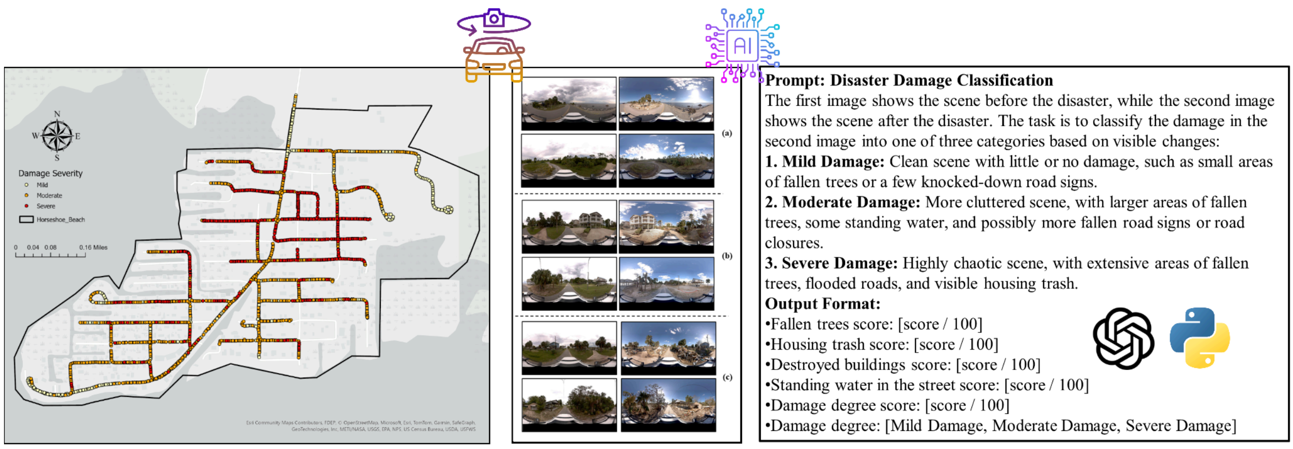

Figure 1. Disaster Perception Framework with GPT-4o-mini Using Street View Images in Horseshoe Beach, Florida

Abstracts of the International Cartographic Association, 10, 310, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-310-2025

Figure 2 illustrates the workflow and results of leveraging GPT-4o-mini for the automatic annotation of disaster-related damage descriptions and assessments from post-disaster street-view imagery. We guided GPT-4o-mini to generate detailed textual narratives and structured damage evaluations based on pre- and post-disaster images. The prompt design targeted specific disaster elements, including fallen trees, debris, building collapse, and water inundation, enabling consistent and interpretable annotation outputs. These structured annotations, as shown in Figure 2 (center panel), facilitate large-scale dataset construction and enhance the interpretability of disaster impacts. Additionally, we explored the generative capabilities of GPT-4o by synthesizing pseudo post-disaster street-view images. By combining pre-disaster visuals with GPT-4o-mini-generated damage descriptions, the model produced synthetic disaster scenes that mimic real-world post-event conditions. As depicted in the bottom-right panel of Figure 2, these synthetic images closely resemble actual post-disaster imagery, offering a novel pathway for scenario simulation, data augmentation, and contrastive learning between real and synthetic disaster scenes.

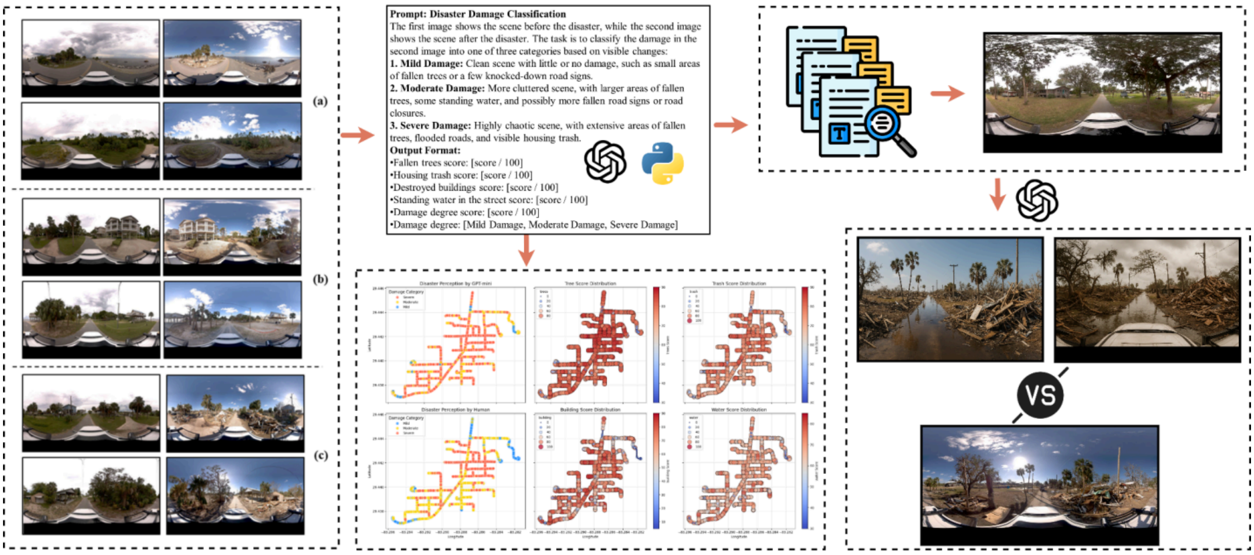

Figure 2. Exploring pre-disaster and post-disaster imagery using ChatGPT.

To systematically assess the reliability and effectiveness of LLM-generated annotations for disaster perception, we conducted a series of classification experiments across different input modalities: image-only, text-only, and multimodal (image + text) approaches. In the text-based and multimodal experiments, both human-generated and LLM-generated descriptions were incorporated, allowing for a direct and controlled comparison between human and machine-generated inputs. This design enables a deeper understanding of the relative contributions of visual and textual information, as well as the quality gap between human and LLM annotations. Table 1 shows that image-only models (e.g., ViT-B/16) outperform text-only and multimodal models, as the complexity of post-disaster street-view scenes and the redundancy of generated descriptions limit CLIP's effectiveness.

Model Name	ResNet-50	ViT-B/16	ViT-B/32	RoBERTa (LLM)	RoBERTa (Human)	CLIP (LLM)	CLIP (Human)
Modality	Image Only	Image Only	Image Only	Text Only	Text Only	Image+Text	Image+Text
Accuracy	0.7398	0.7567	0.7488	0.6506	0.65706	0.7453	0.7453

Table 1. Performance Comparison of Image-Only, Text-Only, and Multimodal Models on Disaster Severity Classification.

References

Yang, Y., Zou, L., Zhou, B., Li, D., Lin, B., Abedin, J., & Yang, M. (2025). *Hyperlocal disaster damage assessment using bi-temporal street-view imagery and pre-trained vision models* (No. arXiv:2504.09066). arXiv. https://doi.org/10.48550/arXiv.2504.09066

https://doi.org/10.5194/ica-abs-10-310-2025