Street-Level Disaster Location Detection Using Image Matching of Social Media Images

Pai-Hui Hsu [1], Manikandan Sathianarayanan [2, 3]

[1] Department of Civil engineering, National Taiwan University, Taiwan – hsuph@ntu.edu.tw
[2] Department of Civil engineering, National Taiwan University, Taiwan – d06521026@ntu.edu.tw
[3] Research Centre for Humanities and Social Sciences (RCHSS), Academia Sinica, Taiwan.

Keywords: Disaster Response, Geotagging, Social Media, GSV Images, Image Similarity, Image Inpainting.

Abstract

Rapid identification of disaster locations is essential for effective emergency response and situational awareness. However, a large proportion of images shared on social media during disasters lack geographic metadata, limiting their usefulness for operational decision-making. Existing approaches mainly rely on geotags or textual geoparsing, which often fail when metadata is missing or ambiguous. As a result, valuable visual information from social media images remains underutilized for disaster mapping. This study proposes a deep learning–based framework to estimate the geographic location of disaster images using visual scene matching. The approach compares query images from social media with georeferenced Google Street View imagery to infer their potential locations. The framework integrates image pre-processing, deep feature extraction, and similarity-based matching to identify the most likely geographic correspondence. Preliminary experiments demonstrate promising results in detecting key scene elements within disaster imagery, providing a foundation for reliable visual matching. By leveraging visual cues rather than textual metadata, the proposed framework aims to improve the usability of non-geotagged social media images for disaster response. The proposed approach has the potential to support rapid disaster mapping by transforming citizen-generated imagery into spatially actionable information for emergency management.

1. Introduction

1.1 Background Study

Taiwan frequently experiences natural hazards such as earthquakes, typhoons, floods, and landslides due to its geographic position within the Pacific seismic belt and along major typhoon pathways (Chang, 1996). These hazards are often intensified by the island's complex geological structure and steep mountainous terrain, which increase vulnerability to disaster impacts. One of the most catastrophic events in recent history was the 1999 Chi-Chi earthquake, which had a magnitude of 7.3 and caused severe destruction, resulting in more than 2,000 fatalities and the collapse of approximately 50,000 homes. Similarly, Typhoon Morakot in 2009 caused widespread devastation, leading to over 600 deaths, more than 1,050 injuries, and the destruction of around 600 houses. In August 2018, a tropical depression triggered severe flooding, which resulted in seven fatalities and forced nearly 8,500 residents to evacuate to temporary shelters. According to official disaster management statistics, natural hazards including floods, typhoons, earthquakes, and landslides caused more than 4,000 deaths in Taiwan between 1994 and 2020. In response to the severe impacts of the 1999 Chi-Chi earthquake, the Taiwanese government enacted the Disaster Prevention and Protection Act to strengthen disaster management and protect public safety and property (Sathianarayanan et al., 2024).

In recent years, social media platforms such as Twitter, Facebook, and Instagram have emerged as important channels for sharing real-time information during disaster events. Users frequently post textual updates, photographs, and videos that capture on-the-ground conditions during emergencies. The widespread use of smartphones equipped with GPS sensors has significantly increased the availability of location-related information within social media content, gradually transforming these platforms into location-aware communication networks (Sathianarayanan and Hsu, 2025). However, the enormous volume and heterogeneous nature of social media data present significant challenges for researchers. A large proportion of the information is noisy, redundant, or irrelevant, making it difficult to efficiently extract useful insights. Despite these challenges, previous studies have highlighted the growing importance of social media data for disaster management, providing valuable situational awareness for governmental agencies and non-governmental organizations involved in emergency response.

Most studies on disaster management using social media have primarily focused on textual information such as posts, messages, and tweets for crisis analysis and situational awareness (Imran et al., 2015). Previous research has demonstrated the usefulness of textual data for assessing disaster impacts and supporting decision-making processes for governmental agencies and non-governmental organizations involved in emergency response (Alam et al., 2018). With recent advances in artificial intelligence and deep learning, researchers have increasingly explored the potential of visual content shared on social media platforms. Images posted during disasters often capture on-the-ground conditions and provide valuable contextual information that complements textual reports. Several studies have investigated the use of social media imagery for disaster analysis and detection. For instance, analyses of images shared on platforms such as Instagram and Flickr have shown strong correlations between visual content and the occurrence of disaster events, highlighting their potential for improving disaster awareness (Peters, 2015). Other research has explored the use of geotagged images and spatial–temporal metadata to help identify and locate disaster incidents. Recent developments in deep learning have further enabled automated extraction of information from images using convolutional neural networks (CNNs), which have been applied to tasks such as fire detection, flood identification, and damage assessment (Ahmad et al., 2017). These approaches often integrate multiple data sources, including satellite imagery, UAV imagery, and social media content, to improve disaster monitoring and prediction. Despite these advancements, many existing approaches rely heavily on geotagged images or metadata, which are often unavailable in real-world social media

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1277-2026

posts. Consequently, accurately identifying the geographic location of disaster images remains a significant challenge.

Rapid and precise identification of disaster locations remains critical for effective emergency response and relief coordination. During the immediate aftermath of disasters, responders often struggle with limited situational information, which can delay decision-making and resource deployment. While satellite imagery and ground-based sensing technologies provide valuable large-scale observations, their usefulness can be restricted by acquisition delays, high operational costs, and limited spatial resolution (Huang et al., 2020). In contrast, social media platforms such as X (formerly Twitter), Instagram, and Facebook have emerged as important sources of real-time, crowd-sourced information during disaster events. Users frequently share images that capture on-the-ground conditions with detailed street-level perspectives. Such visual content often provides richer contextual information than textual posts or conventional remote sensing data and can significantly enhance situational awareness when properly utilized (Gopal et al., 2026; Sathianarayanan et al., 2024).

However, a major challenge remains: most images shared on social media lack precise geographic metadata. The absence of geotags significantly limits their usability for spatial analysis and emergency response applications. Although some studies attempt to infer locations using textual geoparsing or auxiliary metadata, these approaches often fail when textual descriptions are incomplete, ambiguous, or unavailable (Kaur et al., 2022). Similarly, existing image-based approaches frequently rely on additional metadata such as phone numbers or provide only coarse location estimates at the city scale (Hsu et al., 2025). Therefore, developing methods capable of estimating the geographic location of disaster images directly from visual features remains an important research challenge.

1.2 Literature and Research Gap

Extensive research has examined the classification of disaster-related information from social media using machine learning techniques such as Naïve Bayes, Support Vector Machines (SVM), and Convolutional Neural Networks (CNN). Previous studies have demonstrated strong relationships between social media imagery and real disaster events, enabling the identification of disaster-related content from large volumes of online data (Firmansyah et al., 2024b). Although multimodal frameworks combining textual and visual information have been proposed for damage assessment, accurately determining the geographic location of disaster images remains a major challenge (Firmansyah et al., 2024a).

Current state-of-the-art approaches face three primary limitations: (i) a strong dependence on geotagged information, which is often missing in social media posts; (ii) reliance on textual metadata that may be incomplete or unreliable; and (iii) limited spatial precision, typically restricted to neighbourhood or city-level estimates. Recent deep learning–based approaches that utilize external APIs to extract information such as phone numbers from images represent notable progress, but their performance decreases significantly when such metadata is unavailable (Sathianarayanan et al., 2024). To address these limitations, this study proposes a deep learning–based framework that estimates the geographic location of disaster images using purely visual scene matching, with the goal of achieving street-level geolocation accuracy.

2. Materials and Methods

2.1 Study Area

The Taipei Basin (25°N, 121°E), encompassing parts of Taipei City and New Taipei City, is selected as the study area in this research. Situated approximately 10 km inland from the northern coast of Taiwan and surrounded by mountainous terrain, the basin has undergone rapid urban expansion and intensification over recent decades. Most urban development is concentrated within the relatively flat basin floor, resulting in a densely built urban landscape. The study boundary was defined based on the extent of the urbanized area in 2022, focusing mainly on built-up regions while excluding the surrounding mountainous areas (Fig. 1). Taipei City covers approximately 27,180 hectares and has a population of about 2.51 million. The region experiences a humid subtropical climate with an average annual temperature of approximately 23 °C. Previous studies have reported a significant increase in the frequency and intensity of hot summer days in the Taipei Basin over the past four decades.

Figure 1. Study area used for this study

2.2 Methodology Adopted for This Study

To address the limitations of existing geolocation approaches, this study proposes a novel deep learning–based image-matching framework for estimating the geographic location of disaster events from social media imagery. The proposed framework leverages visual scene information rather than relying on textual metadata or geotagged information, which are often unavailable in real-world social media posts. The main idea is to compare a query disaster image with a database of georeferenced Google Street View (GSV) images in order to identify visually similar scenes and infer the most likely geographic coordinates. By utilizing street-level imagery as a spatial reference, the framework aims to achieve accurate location estimation for disaster images.

The overall methodology consists of three main stages (Figure 2):

- Image pre-processing and feature enhancement
- Deep feature extraction and image matching
- Location inference and spatial verification

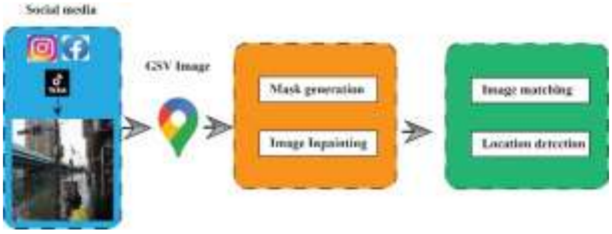

Figure 2. Methodology for street level accuracy of disaster location detection

2.3 Image Segmentation Using Mask R-CNN

To improve the reliability of image matching, transient objects such as vehicles and pedestrians must be removed from disaster images before comparing them with reference street-view imagery. In this study, instance segmentation is applied to identify and mask these dynamic objects. We employ the Mask R-CNN framework, a widely used deep learning model that performs object detection and instance segmentation simultaneously. The model is capable of generating pixel-level masks for individual objects while also providing bounding box localization.

Mask R-CNN utilizes a backbone convolutional network combined with a Feature Pyramid Network (FPN) to extract multi-scale features from input images. A Region Proposal Network (RPN) then identifies candidate object regions, which are further refined through classification and bounding-box regression (Figure 3). The ROIAlign operation is used to maintain precise spatial alignment during feature extraction. In this study, Mask R-CNN is applied to detect vehicles and people in disaster images and generate segmentation masks, enabling these transient objects to be removed prior to the image-matching stage with Google Street View imagery (He et al., 2017).

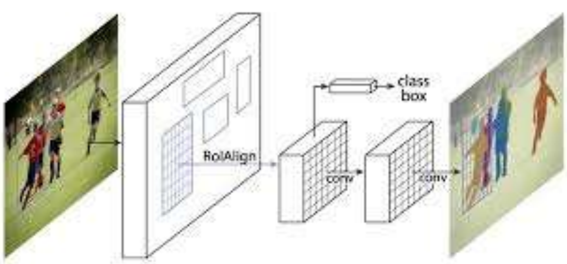

Figure 3. Mask RCNN architecture (He et al., 2017)

2.4 Image Inpainting

Image inpainting is a computer vision technique used to reconstruct missing or occluded regions in images by utilizing information from surrounding pixels. In this study, inpainting is applied after the segmentation stage to restore areas where dynamic objects such as vehicles and pedestrians have been removed. We employ Telea's inpainting algorithm based on the Fast-Marching Method (Telea, 2004), which propagates information from known regions toward missing areas to achieve visually consistent reconstruction. This step helps recover background structures in both disaster images and Google Street View imagery, improving the reliability of subsequent image-matching processes.

2.5 Image Similarity

To determine visual correspondence between disaster images and Google Street View (GSV) imagery, an image similarity measure is applied. In this study, the Learned Perceptual Image Patch Similarity (LPIPS) metric is used to evaluate perceptual similarity between pairs of images based on deep feature representations. Unlike conventional metrics such as RMSE and PSNR, which rely on pixel-level comparisons, LPIPS computes similarity using feature maps extracted from deep neural networks, enabling a more perceptually meaningful comparison. The metric is therefore used to match query disaster images with reference GSV images and identify the most probable geographic location.

2.6 Evaluation Metrics

To evaluate the performance of the proposed framework, standard object detection metrics were used to assess the accuracy of mask generation in disaster images. Detecting objects in complex disaster scenes requires accurate identification of both object presence and spatial location.

In this study, Mean Average Precision (mAP), Precision, and Recall were used as evaluation metrics. The mAP metric measures detection accuracy based on the Intersection over Union (IoU), which quantifies the overlap between predicted bounding boxes and ground-truth annotations. Precision indicates the proportion of correctly detected objects among all predicted detections, while Recall measures the proportion of correctly detected objects relative to the total number of ground-truth instances. True Positives (TP) represent correct detections, whereas False Positives (FP) and False Negatives (FN) correspond to incorrect predictions.

$$Precision = \frac{TP}{(TP + FP)} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{2}$$

To evaluate the accuracy of the predicted geographic coordinates generated by the proposed framework, the estimated locations were compared with corresponding ground-truth coordinates. The prediction error was quantified using the Haversine formula, which calculates the great-circle distance between two points on the Earth's surface based on their latitude and longitude values. This metric is widely used for geographic distance estimation because it accounts for the curvature of the Earth and provides a reliable measure of spatial deviation between predicted and actual locations (Sathianarayanan et al., 2024).

3. Results and Discussion

3.1 Image Segmentation Using Mask RCNN

To evaluate the effectiveness of the proposed framework, the Mask R-CNN model was trained to detect dynamic objects such as vehicles and pedestrians in disaster images. The trained model was then tested on external disaster imagery that was not included in the training dataset. Performance was evaluated using standard object detection metrics, including Intersection over Union (IoU), Precision, Recall, and Average Precision (AP).

The model achieved an IoU score of 75%, with a precision of 69.7%, a recall of 79.1%, and an average precision (AP) of 63.8%. These results indicate that the model can reliably identify relevant objects within complex disaster scenes. Figure 4 illustrates example outputs of the detection results, where Mask R-CNN successfully identifies vehicles and pedestrians by generating bounding boxes and pixel-level segmentation masks.

The results shown in Figure 4 demonstrate that the model is capable of detecting multiple objects within a single image and distinguishing foreground objects from the surrounding environment. In several cases, the model was able to detect partially occluded objects, such as vehicles submerged in floodwater or pedestrians located near damaged structures. Although variations in lighting conditions may affect detection performance, the model still provides reliable segmentation results across different disaster scenarios.

These preliminary results confirm the effectiveness of using Mask R-CNN for identifying transient objects in disaster imagery, which is an essential preprocessing step for improving the subsequent image-matching and geolocation process.

Figure 4. Image segmentation results using Mask RCNN.

3.2 Image Mask Generation

Figure 5 illustrates examples of the mask generation and image inpainting results obtained from disaster images. The Mask R-CNN model successfully identifies dynamic objects such as pedestrians and vehicles and generates segmentation masks for these objects. After masking, the inpainting process reconstructs the removed regions to restore the background scene. The results demonstrate that transient objects can be effectively removed while preserving the structural features of the surrounding environment. This preprocessing step helps reduce visual noise in disaster images and improves the reliability of subsequent image-matching operations with Google Street View imagery for location estimation.

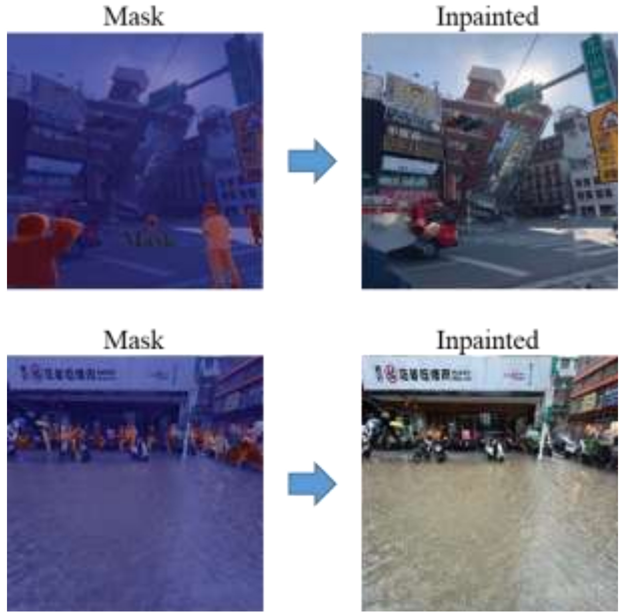

Figure 5. Example results of object masking and image inpainting. Dynamic objects such as vehicles and pedestrians are detected using Mask R-CNN (left) and removed through image inpainting to restore the background scene (right), improving image consistency for subsequent geolocation matching.

4. Conclusion

This study presents a deep learning–based framework for estimating the geographic location of disaster images shared on social media. The proposed approach addresses the limitations of existing methods that rely heavily on geotags or textual metadata, which are often unavailable in real-world social media posts. The framework integrates instance segmentation, image inpainting, and visual similarity analysis to improve the reliability of image matching with georeferenced Google Street View imagery.

Experimental results demonstrate that the Mask R-CNN model can effectively detect and segment dynamic objects such as vehicles and pedestrians in disaster scenes, achieving an IoU of 75%, a precision of 69.7%, and a recall of 79.1%. The preprocessing steps help reduce visual noise and enhance the quality of images used for location inference. These findings highlight the potential of using visual scene matching to support the geolocation of disaster imagery.

Future work will focus on improving the robustness of the framework under varying lighting and weather conditions and extending the approach to larger datasets for more accurate disaster location estimation.

References

Ahmad, K., Pogorelov, K., Riegler, M., Conci, N., Halvorsen, P., 2017. CNN and GAN Based Satellite and Social Media Data Fusion for Disaster Detection, MediaEval Benchmarking Initiative for Multimedia Evaluation, pp. 13–15.

Alam, F., Ofli, F., Imran, M., 2018. Processing Social Media Images by Combining Human and Machine Computing during Crises. International Journal of Human–Computer Interaction 34, 311–327.

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1277-2026

Chang, J.-C., 1996. Natural hazards in Taiwan. GeoJournal 38, 251–257.

Firmansyah, H.B., Fernandez-Marquez, J.L., Mulayim, M.O., Gomes, J., Ribeiro, J., Lorini, V., 2024a. Empowering Crisis Response Efforts: A Novel Approach to Geolocating Social Media Images for Enhanced Situational Awareness. Proceedings of the International ISCRAM Conference.

Firmansyah, H.B., Lorini, V., Mulayim, M.O., Gomes, J., Fernandez-Marquez, J.L., 2024b. Improving Social Media Geolocation for Disaster Response by Using Text From Images and ChatGPT, Proceedings of the 2024 11th Multidisciplinary International Social Networks Conference. Association for Computing Machinery, Bali, Indonesia, pp. 67–72.

Gopal, L.S., Prabha, R., Thirugnanam, H., Ramesh, M.V., Malamud, B.D., 2026. Review article: Social media for managing disasters triggered by natural hazards: a critical review of data collection strategies and actionable insights. Nat. Hazards Earth Syst. Sci. 26, 215–250.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.

Hsu, P.-H., Sathianarayanan, M., Gianoli, A., 2025. Maximizing cooling benefits through urban green and blue spaces in Taipei city. Discover Cities 2, 115.

Huang, X., Li, Z., Wang, C., Ning, H., 2020. Identifying disaster related social media for rapid response: a visual-textual fused CNN architecture. Int. J. Digit. Earth 13, 1017–1039.

Imran, M., Castillo, C., Diaz, F., Vieweg, S., 2015. Processing Social Media Messages in Mass Emergency: A Survey. ACM Comput. Surv. 47, Article 67.

Kaur, S., Gupta, S., Singh, S., Arora, T., 2022. A Review on Natural Disaster Detection in Social Media and Satellite Imagery Using Machine Learning and Deep Learning. International Journal of Image and Graphics 22, 2250040.

Peters, R., 2015. Porto de Albuquerque, J. Investigating images as indicators for relevant social media messages in disaster management, Procceedings of 12th International Conference on Information Systems for Crisis Response and Management, Palen, Kristiansand, Norway, pp. 24–27.

Sathianarayanan, M., Hsu, P.-H., 2025. Integrating RS and Social Media for Enhanced Disaster Monitoring, Assessment, and Response: Challenges and Opportunities, Global Land-Ocean Geospatial Applications. CRC Press, pp. 114–137.

Sathianarayanan, M., Hsu, P.-H., Chang, C.-C., 2024. Extracting disaster location identification from social media images using deep learning. Int. J. Disaster Risk Reduct. 104, 104352.

Telea, A., 2004. An Image Inpainting Technique Based on the Fast Marching Method. Journal of Graphics Tools 9, 23–34.