A BENCHMARK DATASET FOR DETECTING EARTHQUAKE-DAMAGED BUILDINGS FROM SINGLE POST-EVENT VHR SAR IMAGERY

Y. Sun[1], Y. Wang[1], M. Eineder[2]

[1] Data Science in Earth Observation, Technical University of Munich (TUM), (yao.sun, yi4.wang)@tum.de
[2] Remote Sensing Technology Institute, German Aerospace Center (DLR), michael.eineder@dlr.de

KEY WORDS: building damage detection, convolutional neural network (CNN), very high resolution (VHR), synthetic aperture radar (SAR), earthquake, OpenStreetMap (OSM), large-scale urban areas.

1. INTRODUCTION

Earthquakes cause significant damage, requiring fast and accurate detection of affected buildings in remote sensing imagery. Remote sensing technologies, such as Very High Resolution (VHR) optical images and Synthetic Aperture Radar (SAR), are crucial for disaster management (Dell'Acqua and Gamba, 2012, Dong and Shan, 2013, Contreras et al., 2021). While VHR optical images are easier to interpret, SAR imagery offers all-weather capabilities, and enhanced resolution of contemporary SAR satellite images enables the extraction of information at the individual building level (Sun et al., 2021, Chen et al., 2021, Sun et al., 2022), comparable to VHR optical data (Chen et al., 2021, Li et al., 2023). Existing research often relies on both pre- and post-earthquake SAR images; however, pre-event VHR SAR data is often unavailable in most regions. Therefore, the question of whether a single VHR SAR image acquired after an event can effectively identify damaged buildings remains to be addressed.

In this regard, benchmark datasets are crucial for developing and comparatively assessing various methodologies designed to address the following questions: *To what extent can a single VHR SAR image, acquired post-event, allow to identify damaged buildings and with which accuracy? Additionally, how do the outcomes derived from a single post-event SAR image compare with those of an optical image?* Challenges in creating such datasets include limited high-resolution SAR images, absence of damaged building labels, and lack of accurate terrain models. The presented work introduces the first dataset to assess the effectiveness of single post-event VHR SAR images in identifying damaged buildings, integrating publicly accessible satellite imagery and annotations from the 2023 Turkey-Syria earthquakes. The dataset includes over four thousand buildings with co-registered post-event SAR image patches, serving as a benchmark for machine learning and deep learning methods in damaged building detection.

This abstract briefly reports our work on a benchmark dataset for detecting earthquake-damaged buildings from single post-event VHR SAR imagery.

2. POST-EARTHQUAKE SAR-OPTICAL DATASET

The study area, Islahiye in southeastern Turkey near the Syrian border, was significantly affected by a magnitude 7.8 earthquake on February 6, 2023, followed by a 7.5 magnitude aftershock. The disaster led to widespread destruction, causing loss of life, injuries, and extensive damage to buildings and infrastructure. To create a dataset of post-earthquake buildings, we use a Spotlight SAR image from Capella Space acquired on February 9, 2023, with a pixel spacing of 0.35 m, and an optical image from Maxar WorldView-3, captured on February 7, 2023, with a ground sampling distance of 0.31 m, for comparison. Post-event building footprints and labels were obtained from Humanitarian OpenStreetMap Team. All data are projected to the UTM coordinate system for uniform processing, with the SAR image logarithmically scaled in dB for further analysis.

The dataset comprises 169 damaged and 3860 intact buildings in the study area, with each having four patches: SAR image, SAR footprint, optical image, and optical footprint. The dataset is generated in two main steps:

First, we co-register building footprints with the satellite images. Building-level analysis requires accurate registration of 2-D building footprints with satellite images. The ARD optical image aligns well with building footprints, requiring no additional registration. For the GEO SAR image, inspection shows that building polygons are not well-matched with the SAR image and further registration is needed. We apply the algorithm developed in (Sun et al., 2020) to improve the alignment of building polygons and the SAR image. The algorithm relies on the corresponding building features representing the bottom of sensor-visible walls in both the two data, i.e., double bounce lines in the SAR image and near-range boundaries of 2-D building polygons. As the majority of buildings remain upright, with expected double bounce line signatures on the SAR image, the algorithm is applicable.

Then, we crop SAR and optical image patches for each building, considering the target building area (footprint, wall, and roof) and possible ruins, excluding surrounding buildings. For side-looking SAR images, a 10-pixel buffer (around 3.5 m) is added for far-range sides, and an additional buffer is added for near-range sides to include layover areas. The optical image, with a 6.3° off-nadir angle, requires a 16-pixel buffer to compensate for roof offset, ensuring the image patches include the entire roof. Footprint masks are generated for each building corresponding to SAR and optical patches. For side-looking SAR, the mask helps locate the target building amidst signals from surrounding buildings. Nadir-looking optical data also include footprint masks for fair comparison with SAR.

3. EXPERIMENTAL RESULTS

3.1 Baseline Approaches and Implementing Details

Four methods are introduced for benchmarking: support vector machine (SVM), random forests (RF), a 3-layer convolutional neural network (CNN), and ResNet-18 (He et al., 2016).

SVM and RF are chosen for their reported effectiveness in classifying collapsed and standing buildings from post-event SAR imagery (Gong et al., 2016). Feature extraction follows (Gong et al., 2016), using four first-order statistics and eight second-order image statistical measures. A 3-layer CNN and ResNet-18 are selected as deep learning methods. The CNN includes convolution-ReLU-maxpool blocks, followed by average pooling, a linear layer with dropout, and a final classification layer. ResNet-18 follows the standard design, consisting of 18 convolutional layers with residual connections. Similar to the simple CNN, a linear layer with dropout is added before the final classification layer. Cross-fold experiments are conducted on the dataset, split into 5 folds with a balanced number of damaged and intact buildings. Each experiment is run 5 times, with 4 folds for training and 1 for testing. Preprocessing involves removing 2% pixel outliers and normalizing pixel values to the range [0,1].

For deep learning methods, the CNN is randomly initialized, while ResNet-18 uses ImageNet pretrained weights. Data augmentations include random resized crop and random horizontal and vertical flips. To address class imbalance, weights are assigned during data sampling. Binary cross entropy loss is optimized with the AdamW optimizer for 30 epochs, following a cosine-decay schedule starting from 0.0001, with a batch size of 32.

3.2 Evaluation Metrics and Performance Comparison

Precision, recall, and F_1 scores, along with the area under the receiver-operator curve (AUROC), are reported for evaluating baseline methods. The metrics provide a comprehensive understanding of model performance, considering true positives, false positives, true negatives, and false negatives.

Table 1 presents model performance variations on the dataset. For SAR images, SVM shows lower precision but excels in recall (0.495). RF outperforms in precision but has lower recall. CNN produces a relatively high F_1 score and the highest AUROC (0.739), demonstrating proficiency in handling complex features directly from raw SAR data. ResNet-18 underperforms compared to CNN in all metrics.

For optical images, ResNet-18 stands out with the highest precision, recall, F_1 score, and AUROC. CNN demonstrates a balance between precision and recall (0.432), resulting in a high F_1 score and an impressive AUROC (0.853). RF attains high precision but lower recall. SVM yields less favorable outcomes, though its AUROC (0.723) surpasses three other models in the SAR image category. Comparing SAR and optical images, SAR images appear more challenging for all models, generally exhibiting lower performance. For deep learning models, ResNet-18 excels in optical images, while CNN outperforms ResNet-18 in SAR images. CNN's better performance on SAR data may be attributed to its ability to efficiently capture essential patterns in SAR's unique characteristics, reducing the risk of overfitting and leading to better generalization, especially with limited training data. In contrast, complex deep neural networks may not be optimized for SAR-specific traits.

4. CONCLUSION

Detecting earthquake-damaged buildings in post-event satellite imagery is challenging due to limited data. This study introduced a dataset, combining post-event SAR and optical images

Table 1. Benchmark results on the dataset: 5-fold mean (std). The highest values of different metrics are highlighted in **bold**.

image	model	Precision	Recall	F_1	AUROC
SAR	SVM	0.096 (0.021)	**0.495 (0.202)**	0.154 (0.037)	0.653 (0.068)
	RF	**0.223 (0.090)**	0.344 (0.208)	**0.231 (0.071)**	0.670 (0.079)
	CNN	0.167 (0.028)	0.371 (0.073)	0.228 (0.032)	**0.739 (0.026)**
	ResNet-18	0.155 (0.067)	0.276 (0.080)	0.184(0.050)	0.653 (0.043)
Optical	SVM	0.163 (0.044)	0.352 (0.148)	0.207 (0.057)	0.723 (0.047)
	RF	0.611 (0.215)	0.354 (0.098)	0.421 (0.106)	0.810 (0.062)
	CNN	0.489 (0.154)	0.432 (0.078)	0.449 (0.096)	0.853 (0.045)
	ResNet-18	**0.666 (0.121)**	**0.539 (0.147)**	**0.581 (0.100)**	**0.938 (0.022)**

with labeled damaged and intact buildings for image classification. Baseline methods revealed that detecting damage in SAR images is valuable but more challenging than in optical images, emphasizing the need for improved methods. While focusing on individual SAR image performance, the potential fusion of multi-modal data could enhance overall effectiveness when available simultaneously.

REFERENCES

Chen, J., Qiu, X., Ding, C., Wu, Y., 2021. CVCMFF Net: Complex-valued convolutional and multifeature fusion network for building semantic segmentation of InSAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–14.

Contreras, D., Wilkinson, S., James, P., 2021. Earthquake reconnaissance data sources, a literature review. *Earth*, 2(4), 1006–1037.

Dell'Acqua, F., Gamba, P., 2012. Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE*, 100(10), 2876–2890.

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84, 85–99.

Gong, L., Wang, C., Wu, F., Zhang, J., Zhang, H., Li, Q., 2016. Earthquake-induced building damage detection with post-event sub-meter VHR TerraSAR-X staring spotlight imagery. *Remote Sensing*, 8(11), 887.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Li, Q., Mou, L., Hua, Y., Shi, Y., Chen, S., Sun, Y., Zhu, X. X., 2023. 3DCentripetalNet: Building height retrieval from monocular remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, 120, 103311.

Sun, Y., Hua, Y., Mou, L., Zhu, X. X., 2021. CG-Net: Conditional GIS-Aware Network for Individual Building Segmentation in VHR SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 1-15.

Sun, Y., Montazeri, S., Wang, Y., Zhu, X. X., 2020. Automatic registration of a single SAR image and GIS building footprints in a large-scale urban area. *ISPRS Journal of Photogrammetry and Remote Sensing*, 170, 1-14.

Sun, Y., Mou, L., Wang, Y., Montazeri, S., Zhu, X. X., 2022. Large-scale building height retrieval from single SAR imagery based on bounding box regression networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 79-95.