COMBINING SOCIAL MEDIA IMAGES AND BITEMPORAL SATELLITE IMAGES FOR AUTOMATED DETECTION OF DAMAGED AREAS AFTER FLOODING

JIAXI YU[1], TOMOHIRO FUKUDA[2] and NOBUYOSHI YABUKI[3]
[1,2,3] Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Osaka 5650871, Japan
[1]yu@it.see.eng.osaka-u.ac.jp, 0009-0009-4112-8178
[2]fukuda.tomohiro.see.eng@osaka-u.ac.jp, 0000-0002-4271-4445
[3]yabuki@see.eng.osaka-u.ac.jp, 0000-0002-2944-4540

Abstract. This paper addresses an urgent need for efficient and accurate flood damage assessment, a process currently hampered by labor-intensive and time-consuming methods. The study aims to harness the power of deep learning (DL) to create a model that integrates bitemporal satellite images and social media photos for automated flood-damaged building detection. Our original contribution lies in the novel combination of diverse data sources, which has shown the potential to enhance the generalization of damage detection models. The research question we tackle is: How can the integration of multi-source data improve the performance of flood damage detection? We deployed a bitemporal image transformer (BIT) incorporating a Convolutional Neural Network (CNN) as a feature extractor to merge features from satellite and social media images. Our model was tested on the Midwest-flooding dataset and yielded a 2% F1-score improvement over the baseline method while maintaining fewer parameters. This preliminary evidence suggests that social media disaster images contain crucial information for enhancing the performance of Disaster Detection Deep Learning (DDDL) methods. Integrating multi-source data proves beneficial in developing more sophisticated DDDL methods, which can promise fast and effective humanitarian relief in disaster scenarios.

Keywords. Deep learning, Change detection, Data fusion, Model distillation, Remote sensing, Social media image dataset.

1. Introduction

Change detection (CD) in remote sensing is essential for monitoring alterations in the Earth's surface over time. It involves identifying differences between multi-temporal remote sensing images taken in the same geographic area (Bruzzone and Bovolo, 2013). This process is pivotal for understanding land surface changes and has been extensively applied in diverse fields such as disaster damage assessment, agricultural

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 2, 59-68.

measurements, and ecosystem monitoring. The fusion of artificial intelligence techniques with traditional remote sensing practices represents a frontier in the evolution of CD technologies (Shi et al., 2020).

In disaster assessment, the study of flood events has garnered considerable attention. The Intergovernmental Panel on Climate Change (IPCC) underscores the growing importance of evaluating flood-induced damages, propelled by the increasing frequency and severity of such events. Traditional damage assessment methods are thorough yet suffer from being labor-intensive, time-consuming, and inherently uncertain during the initial phases of damage estimation (Shi et al., 2020). Fortunately, the advent of DL models in the change detection field promises a new avenue with considerable potential for disaster assessment tasks.

Deep Learning-based Change Detection (DLCD) models are generally architected around an encoder-decoder framework (Shi et al., 2020). Earlier works conveniently adapted image segmentation algorithms for CD issues with minor modifications (Alcantarilla et al., 2018). More recent approaches have shifted towards a dual-stream architecture (Caye et al., 2018), using Siamese encoders for dual temporal inputs.

Despite DLCD methods achieving promising results in various disaster scenarios, challenges with model robustness persist. A probable cause is using standard datasets for training, such as CCD (Lebedev et al., 2018) and WHU (Ji et al., 2019), which imply an abundance of positive change labels. However, when applied to disaster detection tasks, especially with the xBD dataset used for assessing building damages, changes can be subtle and often at a pixel-level resolution, making it difficult for existing DLCD encoders to extract sufficient features from limited training data. Acquiring large volumes of annotated pre- and post-disaster imagery is also highly costly.

The primary objective of this research is to enhance the generalization and efficiency of flood damage detection. Our approach incorporates supplementary data from social media images, which contain similar disaster-related information as satellite images but are more readily available. By integrating these two data sources, our model aims not only to improve the generalization of flood damage detection but also to augment the model's performance under conditions of limited positive labels availability.

The contributions of our work are threefold:

- We capitalize on low-cost social media data as supplementary information in scenarios where satellite images are sparse, enhancing the model's performance.
- The model encoder has been refined to effectively utilize features from diverse data types, including disaster information.
- In the Midwest-flood event, the proposed methods can detect more complete building outlines and a greater number of affected buildings with the same parameter count.

2. Related Work

2.1. DEEP LEARNING CHANGE DETECTION

Traditional image segmentation models inspired initial DLCD methods. Alcantarilla et al. (2018) capitalized on the ability of DL model to generate change maps by feeding CNN with pre- and post-alteration images. Some studies have focused on expanding the model's reception field (RF) to attain global information in images. For instance, Chen et al. (2022) exploited attention mechanisms to boost the model's ability to recognize contextual tags.

Specifically in disaster assessment, Nguyen et al. (2017) were the first to use a pre-trained model to analyze disaster scenes, demonstrating the potential application of DL in this domain. To develop DL models for disaster scenes, Gupta et al. (2019) created the xBD dataset, which includes 19 disaster events of 5 types. Unlike multi-class detection problems, Kim et al. (2022) focused on improving rapid response capabilities during disasters. They used a simple Siamese-structured ResNet18 to process pre- and post-disaster images, establishing a lightweight network for detecting water-related building damage areas.

2.2. MODEL GENERALIZATION AND SPARSE DATA

Many researchers are devoted to enhancing model generalization performance in situations of data sparsity, with a common strategy involving modifications to the network architecture. Many networks in the DLCD domain have adopted U-Net as the backbone for further improvements (Fang et al., 2022). Apart from leveraging U-Net, pre-trained ResNet is also a prevalent choice as the backbone (Caye et al., 2018) (Kim et al., 2022). However, such small-scale networks tend to be data-oriented, which can limit their generalization capability (Zhu et al., 2017).

To train DL models with robust generalization skills, a large dataset containing critical feature information is essential (Zhu et al., 2017). Unfortunately, no public benchmark dataset for disaster detection is available from satellite images. The xBD dataset, covering the most diverse range of disaster events, includes 23,000 annotated images of building damage across six disaster types. However, it falls far short when compared to large-scale image classification datasets today, such as ImageNet (Deng et al., 2009), which has millions of images. Recently, the Incident1M dataset, released by Weber et al. (2023), includes 977,088 images sourced from social media, covering 43 disaster categories and 49 location category labels. It is the only large-scale dataset related to disasters.

Research by Yosinski et al. (2014) demonstrated that initializing DL models with transfer features can significantly enhance their generalization performance. Researchers frequently adopt this strategy. For example, Chen et al. (2022) used a ResNet model pre-trained on ImageNet for their bitemporal image transformer (BIT) model. Kim et al. (2022) used pre-trained models trained by standard satellite images to strengthen their encoder, effectively increasing the model's ability.

3. Proposed Method

3.1. METHOD PROCEDURE

We propose an innovative technique to enhance the disaster change detection model's capability by integrating multimodal data (Figure 1).

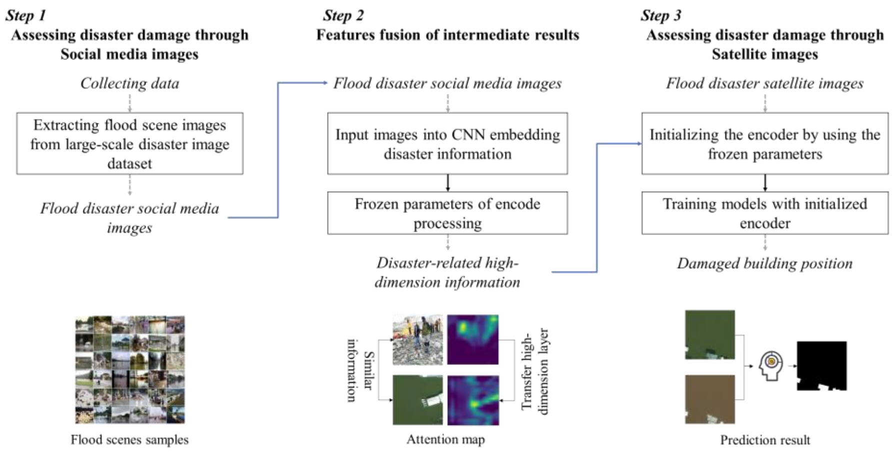

Figure 1. Overview of the proposed method.

The first step involves curating a social media image dataset encompassing flood disaster information by obtaining all images labeled as 'flood' from the comprehensive disaster dataset Incidents1M (Weber et al., 2023). The second phase initiates the training process within the ResNet18 model (He et al., 2016) using the established image dataset, extracting the embedded disaster information. In the third stage, we transfer the frozen parameters from the convolution layer obtained in step 2 into our formulated DDDL method by initializing the model's encoder.

3.2. MODEL ARCHITECTURE

3.2.1. Overview of Model Architecture

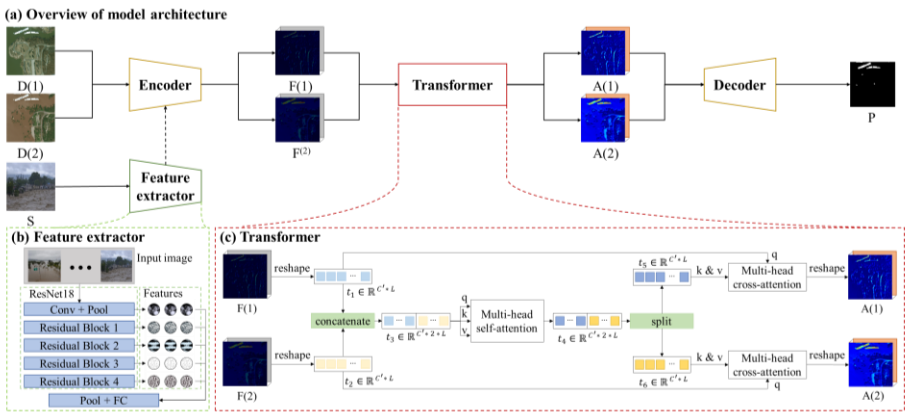

Figure 2. The proposed modified model architecture.

The overall process of our model is presented in Figure 2. Initially, we scrutinize the architecture of the encoder used for information extraction by the satellite image segmentation model (a). After defining the form of the feature extractor (b) based on this encoder. Using the social media images S as inputs, we train the features extractor to obtain fixated parameters for the convolutional layers. These parameters are then transferred into the encoder of the satellite image segmentation model. D(1) and D(2) are introduced into the enhanced satellite image segmentation model. Post-Encoder processing, we extract the features maps F(1) and F(2), which are then input into the Transformer module (c) to amplify the image feature information, producing the enhanced Attention feature maps A(1) and A(2). Finally, through our Decoder, we generate pixel-level predictions P.

3.2.2. Satellite Image Segmentation Model Based on Attention Mechanism

Our approach is designed to allow the new model to focus on disaster information within social media disaster images. Our intuition suggests that once a model possesses this pre-acquired disaster knowledge, it will enhance its understanding of disaster-stricken satellite images compared to the generic knowledge used in the original model.

We contemplated that disaster information gleaned from social media disaster images could also be preserved in feature maps F(1) and F(2). However, features that can be retained after satellite image training may be submerged during the downsampling process. Therefore, we hoped to enhance the disaster features hidden in the feature maps through the attention mechanism. Considering both computational resources and model structure, we chose to improve on the bitemporal image transformer (BIT) model (Chen et al., 2022) as the backbone for our experiments.

3.2.3. Process of Extracting Disaster Information from Social Media Images

The task of our feature extractor is to transform the flood information present in social media disaster images into high-level semantic concepts that align with the encoder paradigm of our primary model. Given the encoder structure of our main model, as depicted in Figure 3b, we built our feature extractor, utilized four convolutional layers to extract information from the images, and conducted layer-wise visualization. Ultimately, we retained the parameters within the first four convolutional modules and transferred these to the BIT model through parameter initialization.

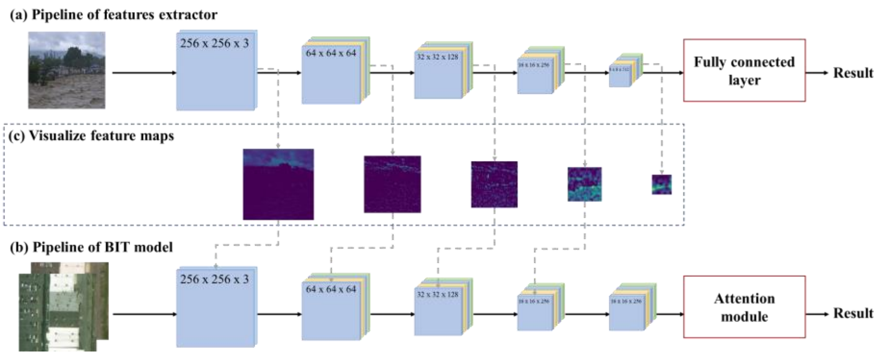

Figure 3. The processing of transfer disaster information to (b) The BIT model.

4. Experiment

4.1. EXPERIMENTAL SETUP

This section of the experiment is divided into three subsections. Section 4.1.1 the dataset utilized in the experiment. Section 4.1.2 implementation details of the experiment. Finally, Section 4.1.3 describes the metrics used for evaluation.

4.1.1. Dataset

This study involves two phases of experimentation, each utilizing a different subset of datasets. Initially, We trained the feature extractor on a selected portion of the Incidents1M dataset. Subsequently, we carried out the modified BIT model experiments using a flood event of the xBD dataset.

The Incidents1M dataset (Weber et al., 2023) is a large multi-label dataset comprising 977,088 images. It encompasses 43 disaster incidents and 49 place categories. We utilized a subset of 18,513 images labeled with the 'flood' for feature extractor training.

The Midwest-flooding dataset is a subset of the xBD dataset, consisting of 445 pairs of pre- and post-disaster satellite images related to flood events. We adhered to the standard dataset partitioning method (training/validation/test) and segmented the images into smaller 256x256 pixel tiles. This process yielded a final dataset comprising 13,671 pairs of images for training, 1,280 pairs for validation, and 1,376 pairs for testing.

4.1.2. Implementation Details

We implemented our model using PyTorch and conducted the training on a single NVIDIA RTX A6000 GPU.

For the feature extractor model, we utilized Adam as the optimizer. We set the learning rate at 0.0001 and conducted the training over 15 epochs.

The modified BIT model's training strategy is the Step Learning Rate Scheduler (StepLR). We initiated the learning rate at 0.00003, with provisions to decay this rate to 20% of its initial value after 60 epochs. The training process, using Adam as the optimizer, spanned 200 epochs.

4.1.3. Evaluation Metrics

This study's primary metrics for evaluating the model's performance are the parameter counts and the F1-score for various classes. The F1-score is derived from the precision and recall values calculated on the test set, defined as follows:

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{1}$$

Additionally, both precision and recall metrics are reported, with their definitions being:

$$Precision = \frac{TP}{TP + FP} \quad (2) \qquad Recall = \frac{TP}{TP + FN} \quad (3)$$

In these formulas, TP (True Positive) represents the number of positive samples correctly predicted by the model, FP (False Positive) signifies the number of negative samples incorrectly predicted as positive, and FN (False Negative) indicates the number of positive samples incorrectly predicted as negative.

4.2. BENCHMARK DLCD METHODS COMPARISON

To evaluate the effectiveness of our proposed method, we compared it with three state-of-the-art (SOTA) DLCD benchmark methods for disaster detection:

- SNUNet (Fang et al., 2022): SNUNet is currently recognized as the most advanced DLCD method utilizing UNet as its backbone.
- P2V （Lin et al., 2023): P2V offers comprehensive temporal modeling to depict the change process and is considered at the forefront of temporal modeling for DLCD.
- BIT_base (Chen et al., 2022): Since our model is a modification of the BIT, it is imperative to compare it with the BIT_base for a comprehensive assessment.

In this experiment phase, we employed the same hyperparameters and code to train the three benchmark DLCD methods described in Section 4.2 and the modified BIT model mentioned in Section 4.1.2. Table 1 compares the performance of these four networks on the Midwest-flooding dataset. We evaluated method performances based on two dimensions: the model's parameter counts and its F1-score.

Table 1. Comparison results on Midwest-flooding datasets.

Model name	Precision	Recall	F1 score	Parameters
SNUNet	0.7649	0.2908	0.4213	10.29M
P2V-CD	0.7777	0.2755	0.4048	5.42M
BIT-base	0.7580	0.2663	0.3941	3.04M
BIT-modified	0.7451	0.2865	0.4141	3.04M

From the F1-score comparison, our BIT-modified method, as opposed to the original BIT-base, showed a decrease in precision by 1.29% but an increase in recall by 2.02% and an overall F1-score improvement of 2%. This indicates a more advantageous model performance from a quantitative perspective. In practical scenarios, especially in disaster response, recall is more critical than precision. Recall represents the rate of detecting all affected buildings, while precision indicates the accuracy of identified damaged buildings. Therefore, our approach is more suitable for disaster relief scenarios, emphasizing rapid identification of potential damage sites.

When comparing the parameter counts while maintaining the lowest parameter counts, our model outperformed the P2V-CD and was slightly below the SNUNet. This evidence suggests that our modifications made the model more applicable to disaster scenarios, with parameters more efficiently used for storing critical information.

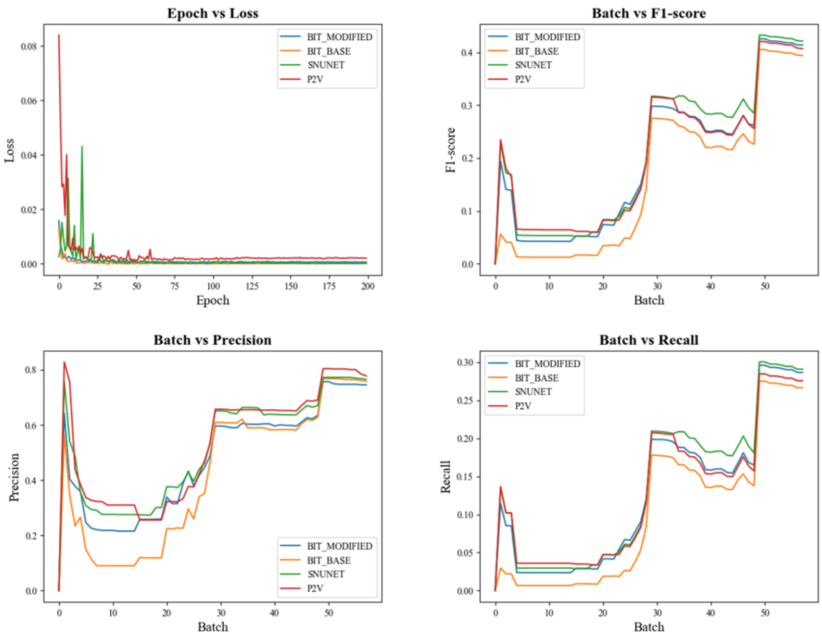

Figure 4. Loss on the training set and Precision, Recall, and F1-score on the test set.

Figure 4 illustrates the loss variation during the training process and the metrics change on the test dataset for the models. Our approach maintained the precision performance of BIT while significantly enhancing both Recall and F1-score on the test dataset.

5. Discussion

In the experimental findings presented in Chapter 4, we numerically demonstrated the superiority of our enhanced model. This chapter delves into a visual comparison on the dataset (Figure 5) to elucidate the advantages of our model over other methods. The visual representation employs green to indicate ground truth and white for the predictions of each model, with red denoting the portions where the model predictions are incorrect.

We categorize the advantages of our model into three types, as illustrated in Figure 5. The (a) row represents the proposed method can detect damaged building outlines that are undetectable by all other models. Rows (b) and (c) highlight the enhancement of our model's recognition capabilities compared to the BIT_base model. Rows (d) and (e) demonstrate our model retaining the advantageous performance of the BIT_base model. Upon observing the predictions of the BIT-base model and two other SOTA methods, it is evident that the BIT-base model can predict some areas that other methods cannot. This advantage is also preserved in our BIT-modified model. This visual analysis complements the numerical results and provides a qualitative understanding of the strengths of our modified model in disaster detection scenarios. The ability to identify previously undetected areas and improve upon the original model's performance positions our approach as a promising advancement in the field.

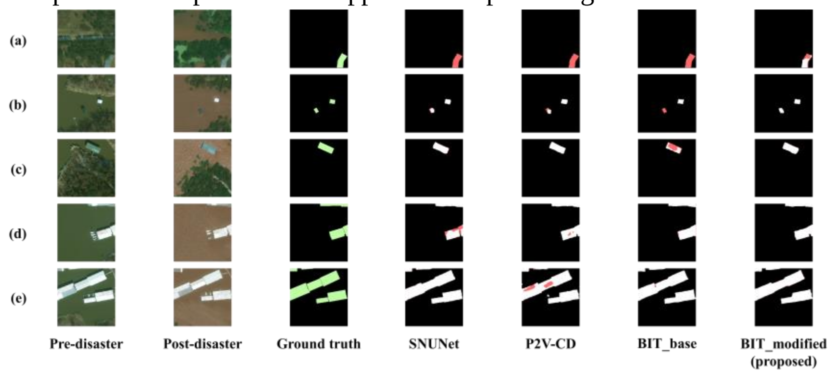

Figure 5. Visualization results of different methods on the test sets. Different colors are used for better view, i.e., We marked the unsuccessful prediction parts in red to show models limitations.

6. Conclusion

This paper presented an approach to flood damage assessment by integrating bitemporal satellite images with social media images using DL techniques. Our method demonstrates a significant improvement in identifying flood-damaged areas, evidenced by a 2% increase in F1-score compared to the original models. With its efficient use of parameters, the modified BIT model balances performance and computational resource requirements, making it suitable for rapid disaster response scenarios. We have made certain progress in this study but also encountered several limitations and challenges. Given the insufficiently annotated data, our improved model was constrained to experimental validation solely on the Midwest-flood dataset. Regrettably, this limitation hampers our ability to extend our analysis to a broader range of cases for thoroughly testing the efficiency of our modified module. Considering these limitations, our future research will employ more sophisticated and detailed methods to explore the effectiveness of other disaster scenarios. We believe that these efforts will enhance our understanding and utilization of social media image in disaster-related contexts, leading to more informed and practical approaches in this field.

Acknowledgements

During the preparation of this work, the authors used ChatGPT to improve the readability and language of the text. After using this service, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

Alcantarilla, P. F., Stent, S., Ros, G., Arroyo, R., & Gherardi, R. (2018). Street-view change detection with deconvolutional networks. Autonomous Robots, 42(7), 1301–1322. https://doi.org/10.1007/s10514-018-9734-5

Bruzzone, L., & Bovolo, F. (2013). A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. Proceedings of the IEEE, 101(3), 609–630. Scopus. https://doi.org/10.1109/JPROC.2012.2197169

Caye Daudt, R., Le Saux, B., & Boulch, A. (2018). Fully convolutional siamese networks for change detection. 2018 25th IEEE International Conference on Image Processing (ICIP), 4063–4067. https://doi.org/10.1109/ICIP.2018.8451652

Chen, H., Qi, Z., & Shi, Z. (2022). Remote sensing image change detection with transformers. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–14. https://doi.org/10.1109/TGRS.2021.3095166

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Fang, S., Li, K., Shao, J., & Li, Z. (2022). SNUNet-CD: A densely connected siamese network for change detection of VHR images. IEEE Geoscience and Remote Sensing Letters, 19, 1–5. https://doi.org/10.1109/LGRS.2021.3056416

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., ... & Gaston, M. (2019). xbd: A dataset for assessing building damage from satellite imagery. arXiv preprint arXiv:1911.09296.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90

Ji, S., Wei, S., & Lu, M. (2019). Fully Convolutional networks for multi-source building extraction from an open aerial and satellite imagery data Set. IEEE Transactions on Geoscience and Remote Sensing, 57(1), 574–586. Scopus. https://doi.org/10.1109/TGRS.2018.2858817on 3D Vision, 3DV 2018, 728-737.

Kim, D., Won, J., Lee, E., Park, K. R., Kim, J., Park, S., Yang, H., & Cha, M. (2022). Disaster assessment using computer vision and satellite imagery: Applications in detecting water-related building damages. Frontiers in Environmental Science, 10. https://www.frontiersin.org/articles/10.3389/fenvs.2022.969758

Lebedev, M. A., Vizilter, Yu. V., Vygolov, O. V., Knyaz, V. A., & Rubis, A. Yu. (2018). Change detection in remote sensing images using conditional adversarial networks. 42(2), 565–571. Scopus. https://doi.org/10.5194/isprs-archives-XLII-2-565-2018

Lin, M., Yang, G., & Zhang, H. (2023). Transition Is a Process: Pair-to-Video change detection networks for very high resolution remote sensing images. IEEE Transactions on Image Processing, 32, 57–71. https://doi.org/10.1109/TIP.2022.3226418

Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage assessment from social media imagery data during disasters. Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, 569–576. https://doi.org/10.1145/3110025.3110109

Shi, W., Zhang, M., Zhang, R., Chen, S., & Zhan, Z. (2020). Change detection based on artificial intelligence: State-of-the-art and challenges. Remote Sensing, 12(10). Scopus. https://doi.org/10.3390/rs12101688

Weber, E., Papadopoulos, D. P., Lapedriza, A., Ofli, F., Imran, M., & Torralba, A. (2023). Incidents1M: A large-scale dataset of images with natural disasters, damage, and incidents. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4), 4768–4781. https://doi.org/10.1109/TPAMI.2022.3191996

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Advances in Neural Information Processing Systems, 27. https://proceedings.neurips.cc/paper_files/paper/2014/hash/375c71349b295fbe2dcdca9206f20a06-Abstract.html

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307