ImpactNet: A Robust Two-Stage Disaster Damage Mapping Model For Cross-Domain Satellite Imagery

Nishat Ara Nipa
Dept. of Modeling, Analysis
and Simulation
Old Dominion University,
Norfolk, Virginia
nnipa001@odu.edu

Mohammed Hamza Chao
Dept. of Computer Science
Old Dominion University,
Norfolk, Virginia
mchao002@odu.edu

Sachin Shetty
Dept. of Modeling, Analysis
and Simulation
Old Dominion University,
Norfolk, Virginia
sshetty@odu.edu

Abstract—**Rapid assessment of damage and situational information is essential for effective restoration and recovery following a natural disaster. Addressing the current limitation with manual interpretation, significant efforts have been made to automate the process by leveraging satellite imagery with an unbiased overhead view of pre- and post-disaster scenarios. Existing approaches in segmenting and classifying damaged buildings often fall short in scenarios where damaged and undamaged buildings coexist within close proximity. In this work, we propose a two-stage disaster damage framework that localizes buildings through segmentation and then classifies the associated damage level at an object level. The framework employs a Siamese Vision Transformer (ViT-Small) backbone to create a rich feature representation that is informed by the entire scene. From this map, building-specific features are extracted, enabling robust identification of subtle and complex damage patterns often missed by pixel-based models. Experimental results demonstrate a localization F1 score of 0.8485 and a damage classification F1 score of 0.71, yielding an overall F1 score of 0.7507. These results surpass the baseline by a significant margin and outperform existing state-of-the-art solutions, while ensuring generalizability under high class imbalance and domain shift.**

Index Terms—**Disaster damage assessment, Building segmentation, Vision Transformers, Siamese networks, Object-level classification, Remote sensing, xBD dataset**

I. INTRODUCTION

The devastating consequences of natural disasters on human life and infrastructure can be mitigated to a great extent if accurate situational information can be achieved in a timely manner. Satellite imagery is capable of capturing spatial patterns of affected areas such as flooded regions, wildfire spreads, or structural damage caused by earthquakes and storms. They can be used as a critical tool for Natural Disaster Management (NDM) by guiding emergency operations and prioritizing resource allocation to expedite the restoration of normalcy within the affected community [1]. A number of benchmark datasets, including Open-EarthMAp [2], INRIA [3], WHU-OHS [4] and xBD [5] have been introduced to accelerate progress in this domain. However, the bottleneck with the current analysis scheme has been the amount of time human subject matter experts take to identify the damaged areas, taking hours/days, where time is of the essence in this scene.

Fig. 1: Pre-(left) and corresponding Post-disaster satellite image from Hurricane Florence.

In the past few years, Deep Learning (DL) algorithms have emerged as a powerful tool applied to a wide range of applications in computer vision, including object detection and image segmentation. While their application has been predominantly centered around medical applications [6], their potential extends beyond the limits of healthcare.

The ability of DL models to learn hierarchical features automatically makes them highly suited for analyzing complex satellite imagery. DL algorithms highly rely on large-scale, high-quality pre-and post-disaster satellite imagery with fine-grained annotations. One major challenge in real-world disaster management is the adaptation of the DL models in diverse geographical contexts, disaster types, and environmental conditions. The practicality of the DL models is evaluated on their ability to generalize in cross-domain and disaster types [7]. The xBD dataset plays a vital role in this context as it provides pre- and post-disaster images across 16 geographical regions and 7 disaster types with more than 800,000 annotated buildings covering different levels of damage. The damage assessment task is divided into two sub-tasks: 1. Localization/Segmentation, and 2. Damage classification. The baseline model was provided by Gupta et al.[8], integrating both the task where the segmentation was done by a modified U-Net, achieving an Intersection over Union (IoU) score of 0.66 and an F-1 score of 0.80. The classification task, on the other hand, is a much complex one and utilizes a modified ResNet50, providing an overall weighted F1 score of 0.2564 [8]. Since then, it has been a central challenge in the field. Many of the works focused either on segmentation or classification. Relatively few of them demonstrate the ability to be generalizable across multiple categories of disasters and locations.

Another crucial aspect was brought to attention in a very recent work by the authors in [9]. They analyzed the top-5 solutions of the x View2 challenge and indicated a common bias of over-reliance of the methods on correlative features. Almost all of the challenge winners focused on pixel-level (a region of damage) F1 scores, while it is more important to assess the damage at a building (object) level. In our work, we tried to address these gaps by -

- Developing a disaster damage detection framework for the xBD dataset evaluated across the ground truth of the pre- and post-disaster sets. Our framework is trained and tested for a global dataset, thus it is generalizable through cross-region and domain datasets.
- Our framework incorporates a two-stage robust model that takes raw high-resolution satellite imagery as input and performs segmentation and classification at an object level.
- We compared our best performing model with the state-of-the-art Deep Learning models used for the same dataset for both Localization and Damage Classification, and got a significant increase in the overall model performance.

II. LITERATURE REVIEW

Transfer Learning has been a popular choice among researchers in this domain for its learning ability from limited annotated data. Bouchard [17] and Calton [16] used established transfer learning techniques for damage assessment in natural disasters. In their work, Bouchard et al. argue about the generalizability of the xView2 challenge to capture damage in the context of global diversity, pointing to the false performance ceiling in the existing solutions. They also estimate the minimal number of annotated samples(as few as 1500 annotated buildings) to boost the performance of the fine-tuned model. In their attempt to enable transfer learning models in damage detection, Calton et al. demonstrate that the performance of MobileNet consistently outperforms the other models for their in-house dataset for both binary and multi-class flood damage detection. Wang et al.[12] proposed a Transformer-based semantic segmentation method for fine-resolution remote sensing images. They replaced the traditional FCN backbone with a Swin Transformer to better capture long-range contextual information. Additionally, a hybrid dual-encoder architecture combining Swin Transformers and CNNs was introduced and evaluated on the Vaihingen and Potsdam datasets, demonstrating improved segmentation performance. U-Net-based architecture, including attention mechanism, has also proved to be useful for the change detection task in several past works [14]. The work in [14] uses a unified framework that integrates building segmentation with damage classification in the context of the xView2 challenge. Their two-step model with Pre & Post concatenation and disaster label approach achieved a combined F1 score of 0.66, substantially surpassing the xView2 baseline of 0.28, and was among the first to explicitly aim for generalization across multiple disaster types. The authors explored both end-to-end architectures and a two-stage architecture where their best model combines a semantic segmentation CNN for building detection and a classification network for damage level prediction. One of the major shortcomings of their work lies in its persistent difficulty in distinguishing fine-grained damage categories, which creates an open challenge for further development and scalability.

III. DATASET

We've used the publicly available xView2 dataset. The total number of 2799 RGB square images is paired in pre/post- disaster sets identified

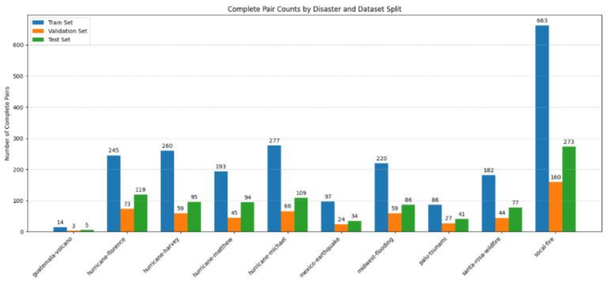

Fig. 2: Number of image pairs per disaster type and its split.

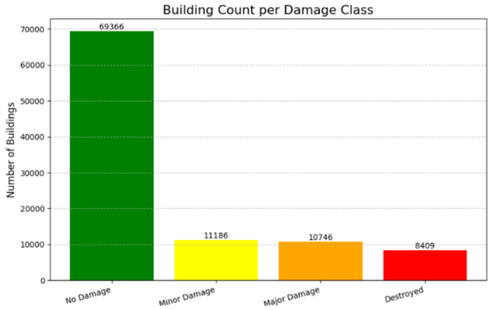

Fig. 3: Number of damaged buildings for each damage class.

by matching numerical IDs for each set along with their corresponding ground truth mask, whereas 933 pairs of pre-and post-disaster satellite images are available in the test set. The size of the images is 1024 X 1024 pixels. The damage level associated with the dataset is a joint damage scale ranging from 0 (no damage) to 3 (destroyed). The ground truth masks are only available for the training set. An example of the pre- and post-disaster image can be seen in Fig.1. A notable characteristic of the dataset is its inherent imbalance towards one class, which can be depicted in Fig.2 through the building count per damage class. A complete pair counts by specific disaster type for the 10 different disasters present in the dataset is shown in Fig.3.

IV. METHODS

Our framework consists of a two-stage model optimized to treat the building segmentation and damage classification as two distinct tasks (Fig.4).

A. *Model Architecture*

Building Segmentation Model: We tried and tested on several architectures and finally used the DCSwin Building Localizer as our segmentation model in stage 1. The DCSwin transformer is a state-of-the-art architecture specifically designed for semantic segmentation. The module was introduced by Wang et al. in [10] and was implemented on the ISPRS (Vaihingen and Potsdam datasets). It combines the power of a Swin Transformer backbone with a specialized decoder. The DC in the name comes from this novel decoder, the Densely Connected Feature Aggregation Module (DCFAM). This hierarchical approach was well-suited for our xDB dataset with some modest variation to precisely identify building footprints from a single image. We can visualize the results from the segmentation model in Fig.5.

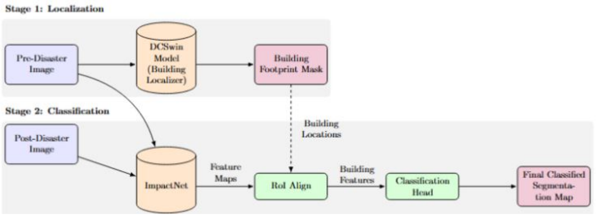

Fig. 4: Two-stage model architecture. Stage 1: Building Segmentation, Stage 2: Damage Classification

Building Classification Model: The baseline model proposed by Gupta et al. [8] uses a ResNet-50 backbone pre-trained on ImageNet and additional features from a Shallow CNN for damage classification. The correlation of the building's damage score to that of its nearest neighbor is 0.763, and they tend to be very close to each other. Including the baseline, all the top-scoring solutions of the xView2 challenge are indifferent to damaged buildings in undamaged surroundings, as they focus on a pixel-level (damaged region-based) detection. Our proposed ImpactNet does this damage classification by finding buildings in a mask first, and classifies the damage level of each building individually on an object level, contrary to the popular RescuNet architecture that predicts a class for every single pixel in the entire image. The model utilizes a Siamese network with a ViT-small backbone to classify the damage level of individual buildings by comparing the pre-and post-disaster images. The siamese structure is fundamentally suited for change detection, whereas the Vision Transformer architecture allowed the model to capture global relationships across the entire image patch, which is critical for subtle, large-scale damage patterns. The ViT-Small variant provides a strong balance between performance and computational efficiency of our model.

It processes pre- and post-disaster image patches

Fig. 5: Semantic Segmentation results on validation Set

of each building through the Siamese ViT-Small backbone with shared weights, extracting rich visual features independently for each temporal input. Region-specific features are obtained via ROI Align and concatenated to form a unified representation, which is passed through an MLP classification head to predict damage levels. The model is trained using Cross-Entropy Loss, with gradients backpropagated through the entire network and weights optimized via AdamW, enabling the Siamese backbones and classification head to learn features relevant for accurate building damage assessment.

V. EXPERIMENT

A. Loss Function

Our segmentation module uses a joint loss function combining the overlap maximization of dice loss and the hard example mining property of focal loss, making it suitable for performing under high class imbalance and complex segmentation.

$$\mathcal{L}_{\text{joint}} = \alpha\,\mathcal{L}_{\text{Dice}} + \beta\,\mathcal{L}_{\text{Focal}} \tag{1}$$

where α and β are weighting coefficients that balance the contribution of each component.

- **Dice Loss:** In segmentation tasks, the extreme scarcity of foreground examples often biases the network toward predicting the background. To address this issue, Dice Loss was proposed [11] to balance the contribution of foreground and background classes. It is defined as:

  $$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon} \tag{2}$$

  where $p_i \in [0,1]$ is the predicted probability for pixel i, $g_i \in \{0,1\}$ is the ground-truth label, N is the total number of pixels, and ϵ is a small constant to prevent division by zero.
- **Focal Loss:** Focal Loss, proposed by Lin et al [12], addresses class imbalance by refining cross-entropy loss with a modulating factor that reduces the loss for well-classified examples and emphasizes harder ones:

  $$\mathcal{L}_{\text{Focal}}(\hat{y}) = -(1-\hat{y})^{\gamma}\log(\hat{y}) \tag{3}$$

  where $\hat{y}$ is the predicted probability of the true class and γ is a focusing parameter. In this work, Focal Loss is applied to mitigate the imbalance in the xBD dataset.
- For our primary contribution, the damage classification module (ImpactNet), we employed a **Weighted Cross-Entropy Loss** to address the significant class imbalance inherent in the xBD dataset. The standard Cross-Entropy Loss is defined as:

  $$\mathcal{L}_{CE} = -\sum_{c=1}^{C} y_c \log \hat{y}_c \tag{4}$$

  where C is the number of classes, y_c is a binary indicator (0 or 1) of whether class c is the correct class for a given sample, and $\hat{y}_c$ is the predicted probability for class c. However, this formulation can be biased towards the majority class.

  To mitigate this, our method introduces a coefficient w_c for each class, yielding the *Weighted Cross-Entropy Loss*:

  $$\mathcal{L}_{WCE} = -\sum_{c=1}^{C} w_c\, y_c \log \hat{y}_c \tag{5}$$

  The weights w_c were calculated to be inversely proportional to the class frequencies in the training set. This strategy compels the model to assign a higher penalty to errors made on the under-represented damage classes (minor, major, and destroyed), preventing it from simply defaulting to the majority "no-damage" class and leading to a more robust and effective model.

B. Evaluation Metrics

Mean Intersection-Over-Union (mIoU) and a Weighted F1 score (30% segmentation F1 score + 70% classification F-1 score to comply with the xView2 challenge) were used to evaluate the performance of the model. mIoU was used as the primary metric to evaluate segmentation quality. We also used Overall Accuracy (OA) in this part. The weighted F1 score balances precision and recall in a harmonic mean, which is well-suited for imbalanced datasets. The formula to obtain the F1 score is mathematically expressed as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

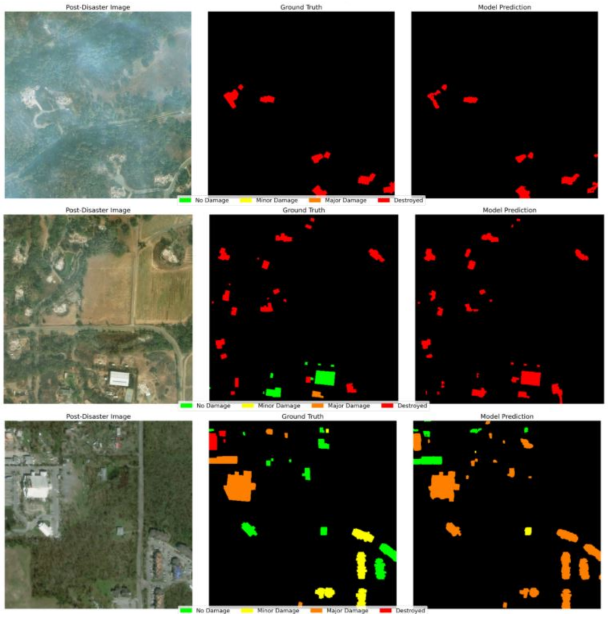

Fig. 6: Visualization for Building Damage Classification

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

where TP is the number of true positives (rate of positive instances correctly classified), FP is the number of false positives (rate of negative instances misclassified), and FN is the number of false negatives (rate of positive instances misclassified).

C. Training Details

The segmentation model was trained for 70 epochs with a batch size of 8 (training) and 4 (validation) using a Lookahead-AdamW optimizer. The learning rate was set to 5×10^{-4} for the decoder and 5×10^{-5} for the pre-trained backbone, with a LinearLR warmup for 10 epochs followed by CosineAnnealingLR. To improve model generalization and prevent overfitting, we utilized a robust data augmentation pipeline during training, which included random horizontal flips, vertical flips, and color jittering to alter the brightness, contrast, and saturation of the input images.was strong regularization (weight decay 0.01, Lookahead, joint loss) were used to mitigate overfitting. Training was implemented in PyTorch Lightning, with training images randomly cropped to 512×512. Performance metrics for segmentation classes are summarized in TABLE I. The damage classification model was trained for up to 400 epochs with a batch size of 12 using the AdamW optimizer with a weight decay of 1×10^{-4}. The initial learning rate was set to 1×10^{-4} and scheduled using CosineAnnealingLR. Input images were resized to 512×512 pixels, with data augmentation including random horizontal and vertical flips and ColorJitter (brightness, contrast, saturation, hue). Overfitting was mitigated through a dropout layer ($p = 0.5$) in the classification head, weight decay, and strong data augmentation. The Siamese architecture processed pre- and post-disaster images in parallel, and the resulting features were fused for final damage classification. The classification report is summarized in TABLE II.

D. Test Results

With our two-stage model, we obtained a localization F1 score of 0.8454 for building segmentation and 0.71 for damage classification. These metrics are combined to obtain an overall F1 score of 0.7507, which exceeds the xBD baseline by a large margin. Our overall F1 score is also competitive with other works leading the xView2 leaderboard. We got a 8.66% increase in the overall F1 score compared to DeepDamageNet [14] and a 1.27% increase compared with the ranked place 2 in Track 3 of the challenge [15].

TABLE I: Performance metrics for segmentation classes

Class	**Precision**	**Recall**	**F1-Score**	**IoU**
Background	0.9904	0.9898	0.9901	0.9804
Building	0.8457	0.8533	0.8494	0.7383
Average	0.9180	0.9215	0.9198	0.8593

TABLE II: Comparison of Baseline and Our Results

Damage Type	Baseline			Ours		
	F1	Prec.	Rec.	F1	Prec.	Rec.
No Damage	0.6631	0.8770	0.5330	0.91	0.90	0.93
Minor Damage	0.1435	0.1971	0.1128	0.51	0.59	0.45
Major Damage	0.0094	0.7259	0.0047	0.65	0.62	0.67
Destroyed	0.4657	0.5050	0.4321	0.72	0.76	0.68
Macro F1		0.7066			0.70	
Accuracy		0.8266			0.82	

VI. CONCLUSION

In this work, we presented a two-stage disaster damage detection framework that addresses key limitations in existing approaches. Our framework leverages a robust object-level strategy, first localizing buildings and then classifying damage for each structure individually. This is powered by a Siamese ViT-Small backbone, whose architecture creates a rich feature representation that is informed by the entire scene, from which building-specific features are then extracted for a final, robust classification. By operating on individual buildings, our method excels in complex environments where damaged and intact structures coexist—a common failure case for pixel-level models. Our experiments demonstrate the success of this approach, achieving a damage classification macro F1 score of 0.71, which significantly surpasses the xBD baseline. This object-level reasoning provides a more reliable and interpretable tool for first responders, enabling more efficient and targeted resource allocation in the critical hours after a disaster. Future work can be pursued along several promising avenues. To directly boost classification metrics, further experimentation with alternative backbone architectures and advanced data augmentation techniques is a clear next step. Additionally, a significant research opportunity lies in combining uncertainty-aware methods with our framework to provide confidence scores for each damage classification, which would further enhance its reliability and utility in real-world scenarios.

REFERENCES

[1] M. Krichen, M. S. Abdalzaher, M. Elwekeil, and M. M. Fouda, ”Managing natural disasters: An analysis of technological advancements, opportunities, and challenges,” Internet of Things and Cyber-Physical Systems, vol. 4, pp. 99–109, 2024, doi: 10.1016/j.iotcps.2023.09.002.

[2] J. Xia, N. Yokoya, B. Adriano, and C. Broni-Bediako, “Openearthmap benchmark dataset for global high-resolution land cover mapping,” in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.,2023, pp. 6254–6264.

[3] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, “Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2017, pp. 3226–3229.

[4] J. Li, X. Huang, and L. Tu, “WHU-OHS: A benchmark dataset for large-scale hersepctral image classification,” Int. J. Appl. Earth Observ. Geoinf., vol. 113, 2022, Art. no. 103022.

[5] R. Gupta et al., “XBD: A dataset for assessing building damage from satellite imagery,” 2019, arXiv:1911.09296.

[6] O. Ronneberger, P. Fischer, and T. Brox, ”U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[7] V. Benson and A. Ecker, “Assessing out-of-domain generalization for robust building damage detection,” arXiv preprint arXiv:2011.10328, 2020. [Online]. Available: https://arxiv.org/abs/2011.10328.

[8] Ritwik Gupta, Bryce Goodman, Nirav Patel, Ricky Hosfelt, Sandra Sajeev, Eric Heim, Jigar Doshi, Keane Lucas, Howie Choset, and Matthew Gaston. “Creating xbd: A dataset for assessing building damage from satellite imagery,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 10–17, 2019.

[9] D. Melamed, C. Johnson, I. D. Gerg, C. Zhao, R. Blue, A. Hoogs, B. Clipp, and P. Morrone, “Uncovering Bias in Building Damage Assessment from Satellite Imagery,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 2024, pp. 8095–8099. doi: 10.1109/IGARSS53475.2024.10642347

[10] N. A. Nipa, Y. Lu and S. Shetty, ”U-Net Based Disaster Damage Detection Through Semantic Segmentation,” 2023 International Symposium on Networks, Computers and Communications (ISNCC), Doha, Qatar, 2023, pp. 1-5, doi: 10.1109/ISNCC58260.2023.10323958.

[11] R. Zhao et al., ”Rethinking Dice Loss for Medical Image Segmentation,” 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 2020, pp. 851-860, doi: 10.1109/ICDM50108.2020.00094

[12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” arXiv preprint arXiv:1708.02002, 2018. [Online]. Available: https://arxiv.org/abs/1708.02002

[13] L. Wang, R. Li, C. Duan, C. Zhang, X. Meng and S. Fang, ”A Novel Transformer Based Semantic Segmentation Scheme for Fine-Resolution Remote Sensing Images,” in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6506105, doi: 10.1109/LGRS.2022.3143368.

[14] I. Alisjahbana, J. Li, B. Strong, and Y. Zhang, “DeepDamageNet: A two-step deep-learning model for multi-disaster building damage segmentation and classification using satellite imagery,” arXiv preprint arXiv:2405.04800, 2024. [Online]. Available: https://arxiv.org/abs/2405.04800 .

[15] E. Weber and H. Kané, ”Building Disaster Damage Assessment in Satellite Imagery with Multi-Temporal Fusion,” *arXiv preprint arXiv:2004.05525*, 2020. [Online]. Available: https://arxiv.org/abs/2004.05525

[16] Calton, Landon, and Zhangping Wei. 2022. ”Using Artificial Neural Network Models to Assess Hurricane Damage through Transfer Learning” Applied Sciences 12, no. 3: 1466. https://doi.org/10.3390/app12031466

[17] Bouchard, Isabelle, Marie-Ève Rancourt, Daniel Aloise, and Freddie Kalaitzis. 2022. ”On Transfer Learning for Building Damage Assessment from Satellite Imagery in Emergency Contexts” Remote Sensing 14, no. 11: 2532. https://doi.org/10.3390/rs14112532