In the Shadow of Disaster: Finding Shadows to Improve Damage Detection

Ilkin Bayramli , **Elizabeth Bondi** and **Milind Tambe**
Harvard John A. Paulson School of Engineering and Applied Sciences, Cambridge, MA
ibayramli@college.harvard.edu, ebondi@g.harvard.edu, milind_tambe@harvard.edu

Abstract

Rapid damage assessment after natural disasters is crucial for effective planning of relief efforts. Satellites with Very High Resolution (VHR) sensors can provide a detailed aerial image of the affected area, but current damage detection systems are fully- or semi-manual which can delay the delivery of emergency care. In this paper, we apply recent advancements in segmentation and change detection to detect damage given pre- and post-disaster VHR images of an affected area. Moreover, we demonstrate that segmentation models trained for this task rely on shadows by showing that (i) shadows influence false positive detections by the model, and (ii) removing shadows leads to poorer performance. Through this analysis, we aim to inspire future work to improve damage detection.

1 Introduction

According to the National Centers for Environmental Information (NCEI), the United States has sustained 241 natural disasters since 1980 with cumulative costs exceeding $1.6 trillion. During 2018 alone, the U.S. was impacted by 14 separate billion-dollar disaster events ranging from tropical cyclones to storms and wildfires. Moreover, the number and cost of disasters have seen a surge over time due to the increasingly pronounced impact of climate change on the frequency of the most extreme events. The past three years (2016-2018) have been especially active with the annual average number of billion-dollar disasters more than doubling the long-term average [Smith, 2019].

Although natural disasters have increased in volume and costliness, damage detection systems have not evolved as quickly. Current damage detection systems typically rely on manual segmentation of satellite maps by teams of volunteers (e.g., Tomnod), meaning they can be costly, slow, and error-prone. This necessitates development of an automated system that can provide government officials and emergency relief agencies with an accurate damage map of the afflicted areas.

2 Related Work

One of the most popular tasks in remote sensing is building footprint segmentation which has inspired competitions like DeepGlobe [Demir *et al.*, 2018] and SpaceNet [Etten *et al.*, 2018]. Encoder-decoder convolutional architectures such as SegNet [Badrinarayanan *et al.*, 2015], UNet [Ronneberger *et al.*, 2015] and modifications thereof with more powerful encoders such as ResNet [He *et al.*, 2015] have been consistently ranked as the best-performing solutions to these challenges [Rudner *et al.*, 2018]. Damage segmentation is similar to building footprint segmentation, but requires specifying damaged structures and building ruins as well. [Rudner *et al.*, 2018] introduced Multi3net for flood damage segmentation from multi-temporal medium- and very-high-resolution maps which relies on an encoder-decoder network with a ResNet backend and additional pyramidal context aggregation modules modeled after the PSPNet architecture [Zhao *et al.*, 2017]. The xView2 challenge and xBD dataset recently supported further development in this area [Gupta *et al.*, 2019]. In fact, the xBD dataset is the largest building damage assessment dataset prepared so far, containing around $700,000$ building annotations across over $5,000$ km^2 of imagery from 15 countries and 7 disaster types.

Model	**Data**	**F1 score**	**IoU**
Multi3Net VHR	post	0.54	0.37
Multi3Net VHR EF	pre-post	0.61	0.44
UNet	post	0.58	0.41
UNet EF	pre-post	0.77	0.63
FuseNet	pre-post	0.79	0.65
Siamese-Diff	pre-post	0.76	0.62
Siamese-Conc	**pre-post**	**0.81**	**0.68**

Table 1: Performance metrics for damage segmentation networks.

Change detection methods are also highly relevant for efficiently integrating pre- and post-disaster information into the model. [Hazırbaş *et al.*, 2016] integrates the additional depth channel into segmentation models in RGB-D imagery. [Daudt *et al.*, 2018] introduced fully convolutional Siamese-Difference networks for change detection in satellite imagery. This architecture uses a shared encoder through which both pre- and post-disaster images are passed through. The decoder continues with post-disaster input yet still receives pre-disaster data through skip connections which are generated by subtracting activations before downsampling pre-disaster images from post-disaster images. Also introduced in [Daudt *et al.*, 2018], Siamese-Concatenation networks are designed similarly to Siamese-Difference networks, except while creating skip connections, activations of convolutional blocks are concatenated.

3 Baseline Results

We trained and tested [Hazırbaş *et al.*, 2016] and [Daudt *et al.*, 2018] in addition to Early Fusion (EF), in which we concatenate pre- and post-disaster RGB images in the color channel, effectively feeding in a 6-channel input to the model. While training the Siamese models above, we used the original UNet layers since our dataset was larger than that used by the authors. The network was initialized with standard weights sampled from the standard Normal distribution and trained using the Adam optimizer [Kingma and Ba, 2014] with a learning rate of 10^{-2}. The training set was augmented with random rotations and flips. Batch normalization [Ioffe and Szegedy, 2015] was used, with a batch size of 10. We used a *weighted cross-entropy* loss function with weights to account for class imbalance.

The results of these initial baselines are provided in Table 1. The F1 score of Siamese-Concatenation is comparable to the recent results using xBD data [Weber and Kané, 2020; Gupta and Shah, 2020], though we have focused on detecting damage or no damage only, rather than specific granularities of damage as in these recent papers. As we looked for ways of improving on these promising results, we found that there were several false positive examples involving shadows (such as in Fig. 2). We limit the following discussions to Siamese-Concatenation due to its superior performance, though qualitatively, architectures such as UNet, UNet EF, and Multi3Net VHR also reacted to shadows.

4 Shadows

Because satellite imagery is not always captured in perfect settings (e.g., at noon when the sun is high), shadows are often present [Dare, 2005]. The influence of shadows is even more pronounced in datasets such as xBD where satellite images in each pair are sometimes taken at different times and sun elevation angles [Gupta *et al.*, 2019]. In fact, in [Dare, 2005], the author states that, "From the earliest days of aerial photography, the effects of shadowing have been utilised to highlight ground features in applications such as archaeology and aerial reconnaissance. However, more often than not, shadows are considered a nuisance obscuring important object space detail." In the domain of disaster detection, we are dealing with both pre- and post-disaster images, and we have further noticed the presence of shadows may lead to false positive results, as shown in Fig. 2, so we may think of this as a nuisance and try to remove shadows as in [Dare, 2005]. There was also recent success in considering the removal of such features from water in Sea-thru [Akkaynak and Treibitz, 2019]. We therefore sought to improve our results through the use of shadow removal, consisting of several different preprocessing methods as an initial proof-of-concept.

To see the effect of removing shadows, we replaced shadows near some false positives manually with nearby pixels (e.g. colorpicking). In nearly all images, as seen in Fig. 2, we

	Tsunami	**Hurricane**	**Flood**	**Fire**	**Tornado**
Train	Orig	Orig	Orig	Orig	Orig
Orig	0.76	0.84	0.78	0.78	0.85
NS	0.75	0.81	0.66	0.75	0.81
	Quake	**Volcano**	**All**	**All**	
Train	Orig	Orig	Orig	NS	
Orig	0.77	0.74	0.81	0.10	
NS	0.74	0.64	0.75	0.90	

Table 2: F1 scores of Siamese-Conc for original (Orig) and no shadow (NS) test images, trained according to the train set rows. Quake is short for Earthquake.

found that preprocessed regions were no longer detected as a false positive. However, in some instances such as rows 3, 4 of Fig. 2, removal of shadows near certain structures led to false positives in other regions, even those not touched during manual editing. This led us to hypothesize that shadows must help convolutional models such as Siamese-Concatenation to detect buildings. Indeed, the new large false positives of our model do not look random at all: viewed perpendicularly from above, these brown-gray regions resemble real building roofs. What leads a human to distinguish these regions from buildings is perhaps an accurate sense of camera angle, which can be deduced from nearby buildings as well as shadows. The model may have learned to look for similar cues.

To test this hypothesis more rigorously, we automatically detected shadows and inpainted them using a combination of thresholding, Fast Marching Algorithm [Telea, 2004], and distance transformations [Borgefors, 1986]. Table 2 shows results both from training on the original or shadow-removed versions of the train set, then testing on the original or shadow-removed versions of the test set. On the shadow-removed test dataset (NS Test), the model trained on the original train set (Orig) performed worse both overall and across disaster types. In the model trained on the shadow-removed train set (Overall, NS), we suspect that the model learned to depend on the shadow inpaintings, leading to excellent performance on the shadow-removed test set and poor performance on the original test set. This problem could perhaps be alleviated by using a shadow detection algorithm such as [Kwatra *et al.*, 2012]. To further confirm the model's dependence on shadows, we observed changes in the activations of different encoder filters after shadow replacement (Fig. 1). We therefore hypothesize that although removing a shadow may help in some cases, overall, it seems that the shadows may actually provide an important cue to Siamese-Concatenation, similar to what [Dare, 2005] also alludes to.

5 Vision

We believe that being able to help first responders to respond more quickly to disasters is an important area of future work, and that the use of satellite imagery and computer vision could play a role in this. We further believe that shadows may have a role to play in improving existing algorithms and deserve further investigation going forward. Two ideas for this are as follows: (i) modify the model itself to emphasize and refine shadow detection, and (ii) augment shadows in images.

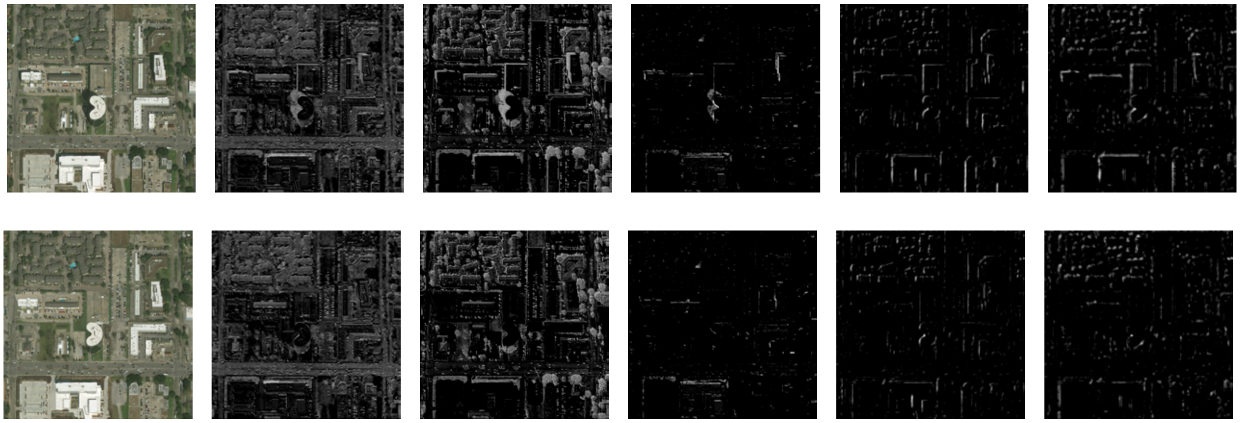

Figure 1: **Activations of the network.**
The first row consists of the activations of the original image while the second one correspond to the shadow-removed version.

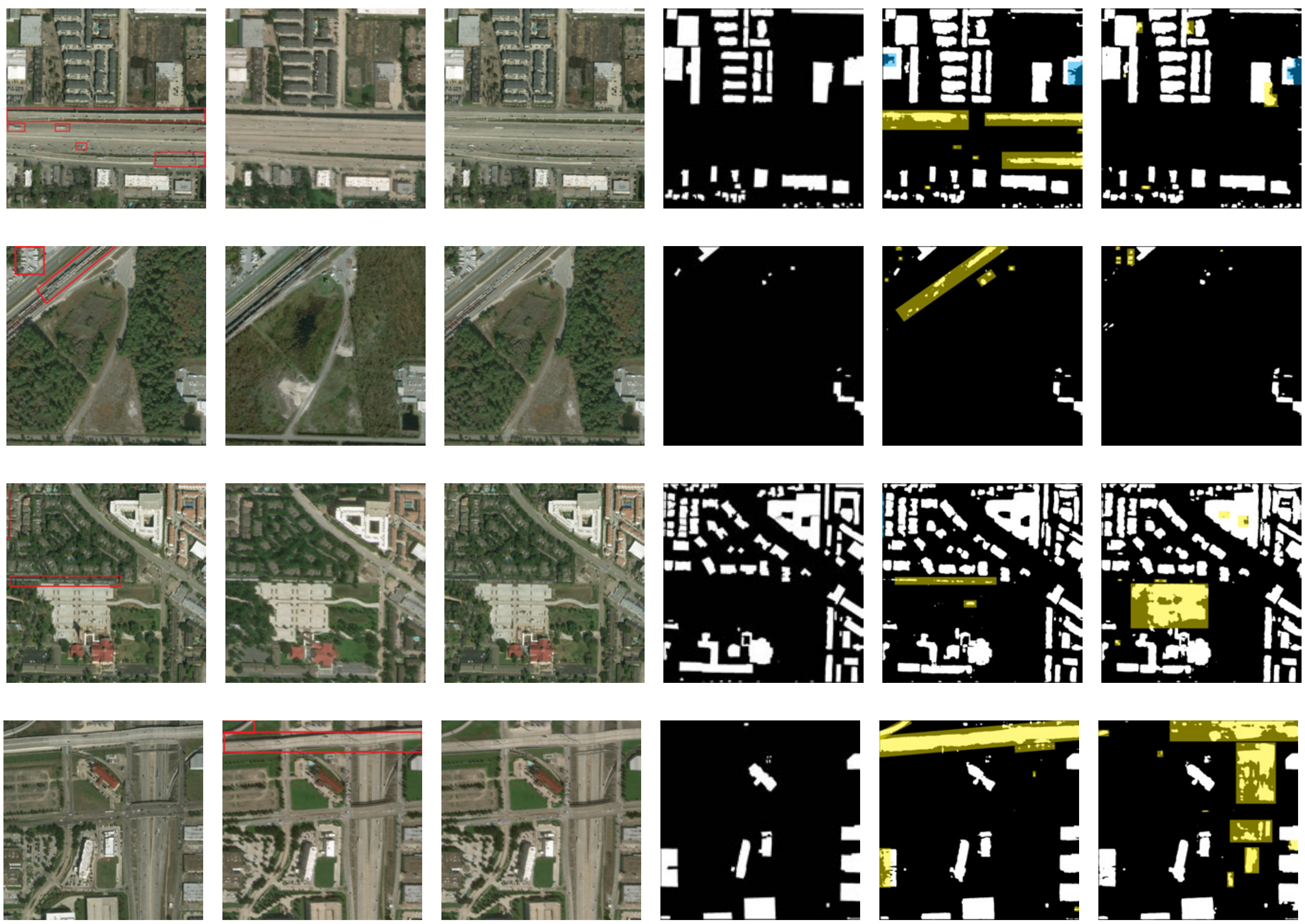

Figure 2: **Response of class predictions to replacement of shadows.**
a) pre-disaster image *b)* post-disaster image *c)* shadow replaced (pre or post) *d)* target *e)* original prediction *f)* shadow replaced prediction
Yellow highlight denotes false positives and cyan false negatives.

References

[Akkaynak and Treibitz, 2019] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1682–1691, 2019.

[Badrinarayanan *et al.*, 2015] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation, 2015.

[Borgefors, 1986] Gunilla Borgefors. Distance transformations in digital images. *Comput. Vis. Graph. Image Process.*, 34:344–371, 1986.

[Dare, 2005] Paul M Dare. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering & Remote Sensing*, 71(2):169–177, 2005.

[Daudt *et al.*, 2018] Rodrigo Caye Daudt, Bertrand Le Saux, and Alexandre Boulch. Fully convolutional siamese networks for change detection, 2018.

[Demir *et al.*, 2018] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raska. Deepglobe 2018: A challenge to parse the earth through satellite images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun 2018.

[Etten *et al.*, 2018] Adam Van Etten, Dave Lindenbaum, and Todd M. Bacastow. Spacenet: A remote sensing dataset and challenge series, 2018.

[Gupta and Shah, 2020] Rohit Gupta and Mubarak Shah. Rescuenet: Joint building segmentation and damage assessment from satellite imagery. *arXiv preprint arXiv:2004.07312*, 2020.

[Gupta *et al.*, 2019] Ritwik Gupta, Bryce Goodman, Nirav Patel, Richard 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 *International Conference on Man–Machine Interactions*. IEEE, 2019.

[Hazırbaş *et al.*, 2016] Caner Hazırbaş, Lingni Ma, Csaba Domokos, and Daniel Cremers. Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture. 11 2016.

[He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[Ioffe and Szegedy, 2015] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[Kingma and Ba, 2014] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[Kwatra *et al.*, 2012] Vivek Kwatra, Mei Han, and Shengyang Dai. Shadow removal for aerial imagery by information theoretic intrinsic image analysis. In *2012 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2012.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, page 234–241, 2015.

[Rudner *et al.*, 2018] Tim G. J. Rudner, Marc Rußwurm, Jakub Fil, Ramona Pelich, Benjamin Bischke, Veronika Kopackova, and Piotr Bilinski. Multi^{3}net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2018.

[Smith, 2019] Adam B. Smith. 2018's billion dollar disasters in context. 2019.

[Telea, 2004] Alexandru Telea. An image inpainting technique based on the fast marching method. *J. Graphics, GPU, Game Tools*, 9:23–34, 2004.

[Weber and Kané, 2020] Ethan Weber and Hassan Kané. Building disaster damage assessment in satellite imagery with multi-temporal fusion. *arXiv preprint arXiv:2004.05525*, 2020.

[Zhao *et al.*, 2017] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. CVPR 2017*, 2017.