Rapid Computer Vision-aided Disaster Response via Fusion of Multiresolution, Multisensor, and Multitemporal Satellite Imagery

Tim G. J. Rudner* University of Oxford
Marc Rußwurm TU Munich
Jakub Fil University of Kent
Ramona Pelich LIST Luxembourg
Benjamin Bischke DFKI & TU Kaiserslautern
Veronika Kopačková Czech Geological Survey
Piotr Biliński University of Oxford & University of Warsaw

Abstract

Natural disasters can cause loss of life and substantial property damage. Moreover, the economic ramifications of disaster damage disproportionately impact the most vulnerable members of society. In this paper, we propose *Multi*^{3}*Net*, a novel approach for rapid and accurate disaster damage segmentation by fusing multiresolution, multisensor, and multitemporal satellite imagery in a convolutional neural network. In our method, segmentation maps can be produced as soon as at least a single satellite image acquisition has been successful and subsequently be improved upon once additional imagery becomes available. This way, we are able to reduce the amount of time needed to generate satellite imagery-based disaster damage maps, enabling first responders and local authorities to make swift and well-informed decisions when responding to disaster events. We demonstrate the performance and usefulness of our approach for earthquake and flood events. To encourage future research into image fusion for disaster relief, we release the first open-source dataset of fully preprocessed and labeled multiresolution, multispectral, and multitemporal satellite images of disaster sites along with our source code at `https://github.com/FrontierDevelopmentLab/multi3net`.

Introduction

In 2017, Houston, Texas, the fourth largest city in the United States, was hit by tropical storm Harvey, the worst storm to pass through the city in over 50 years. Harvey flooded large parts of Houston, inundating over 154,170 homes and leading to more than 80 deaths. According to the National Hurricane Center, the storm caused over 125 billion USD in damage, making it the second costliest storm ever recorded in the United States. Natural disasters can cause loss of life and substantial property damage. Moreover, the economic ramifications of disaster damage disproportionately impact the most vulnerable members of society.

When a region is hit by a natural disaster, authorized representatives of national civil protection, rescue, and security organizations can activate the International Charter 'Space and Major Disasters'. Once the Charter has been activated, commercial Earth observation companies and national space organizations task their satellites to acquire imagery of the affected region. As soon as images have been obtained, satellite imagery specialists visually or semi-automatically interpret them to create flood maps to be delivered to disaster relief organizations. However, Due to the semi-automated nature of the map generation process, delivery of flood maps to first responders can take several hours after the imagery was provided.

First workshop on AI for Social Good (NeurIPS 2018), Montréal, Canada.

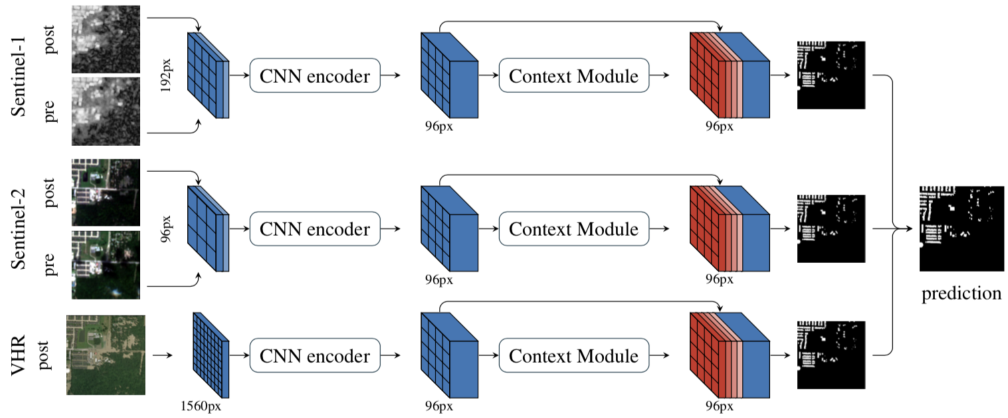


Figure 1: Overview of Multi^{3}Net's multi-stream architecture. Each satellite image is processed by a separate stream that extracts feature maps using a CNN-encoder and then augments them with contextual features. Features are mapped to the same spatial resolution, and the final prediction is obtained by fusing the predictions of individual streams using additional convolutions.

In this paper, we propose *Multi^{3}Net*, a novel approach for rapid and accurate disaster damage segmentation by fusing multiresolution, multisensor, and multitemporal satellite imagery in a convolutional neural network. The network consists of multiple deep encoder-decoder streams, each of which produces an output map based on data from a single sensor. If data from multiple sensors is available, the streams are combined into a joint prediction map.

Our method aims to reduce the amount of time needed to generate satellite imagery-based flood maps by fusing images from multiple satellite sensors. Segmentation maps can be produced as soon as at least a single satellite image acquisition has been successful and subsequently be improved upon once additional imagery becomes available. This way, the amount of time needed to generate satellite imagery-based flood maps can be reduced significantly, helping first responders and local authorities make swift and well-informed decisions when responding to flood events. Additionally, by incorporating multitemporal satellite imagery, our method allows for a speedy and accurate post-disaster damage assessment, helping governments better coordinate medium- and long-term financial assistance programs for affected areas.

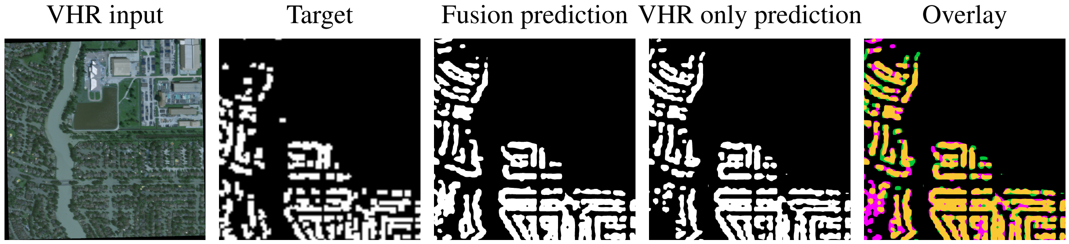
Related Work

Mapping disaster damage using high-resolution imagery has long been an area of research in the field of remote sensing (Barnes, Fritz, and Yoo, 2007a; Yamazaki, 2001), where methods are typically tailored to specific disaster types, such as floods (Scarsi et al., 2014; Goldberg et al., 2018), hurricanes (Cao and Choe, 2018; Ramlal, Davis, and De Bellott, 2018), or earthquakes (Brunner, Lemoine, and Bruzzone, 2010; Cooner, Shao, and Campbell, 2016). Damage caused by hurricanes and earthquakes is often idetified using high-resolution optical or radar imagery (Barnes, Fritz, and Yoo, 2007b), whereas floods (in non-urban areas) are usually identified using low-spatial resolution long-wavelength radar satellite images (Scarsi et al., 2014). Identifying flooding in urban areas, however, is more challenging for conventional remote sensing approaches (Soergel, 2010).

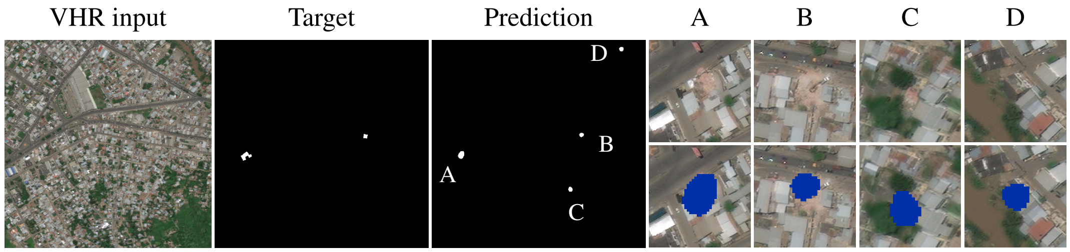
Recent advances in computer vision and the rapid increase of commercially and publicly available medium- and high-resolution satellite imagery have given rise to a new area of research at the interface of machine learning and remote sensing, as summarized by Zhu et al. (2017) and Zhang, Zhang, and Du (2016). Single-stream convolutional neural network approaches have demonstrated the benefits of deep feature learning in end-to-end architectures (Sun et al., 2017; Narazaki et al., 2018). For the segmentation of building footprints from satllite images, U-Net-based approaches that replace the original VGG architecture (Simonyan and Zisserman, 2014) with, for example, ResNet encoders (He et al., 2016) achieved the best results in the 2018 DeepGlobe challenge (Hamaguchi and Hikosaka, 2018). Recently developed computer vision models, such as DeepLab-v3 (Chen et al., 2017), PSPNet (Zhao et al., 2017), or DDSC (Bilinski and Prisacariu, 2018), however, use improved encoder architectures with a higher receptive field and additional context modules.

Multi^{3}Net

Multi^{3}Net uses an encoder-decoder architecture. In particular, we use a modified version of ResNet (He et al., 2016) with dilated convolutions as feature extractors (Yu, Koltun, and Funkhouser, 2017)



(a) Comparison of predictions for the segmentation of flooded buildings for fusion-based and VHR-only models. In the overlay image, predictions added by the fusion are marked in magenta, predictions that were removed by the fusion are marked in green, and predictions present in both are marked in yellow.



(b) Segmentation of collapsed buildings in the Ecuadorian town of Portoviejo after an earthquake in 2016.

Figure 2: Qualitative segmentation results for flooded and collapsed buildings, resctively.

that allows us to effectively downsample the multi-resolution input streams to a common spatial dimension. Motivated by the recent success of multi-scale features (Zhao et al., 2017; Chen et al., 2017), we enrich the feature maps with an additional context aggregation module as described in (Zhao et al., 2017). The decoder component of the network uses three blocks of bilinear upsampling functions with a factor of $\times 2$, followed by a 3×3 convolution, and a PReLU activation function to learn a mapping from latent space to label space. This way, *Multi*^{3}*Net* is able to fuse images obtained at multiple points in time from multiple sensors with different resolutions and capture different properties of the Earth's surface across time. The network is trained end-to-end using backpropagation.

Multisensor Fusion We used a late fusion approach where each image type is fed into a dedicated information processing stream as shown in the segmentation network architecture depicted in Figure 1. We first extract features separately from each satellite image. Next, we combine the class predictions from each individual stream by first concatenating them and then applying additional convolutions. We compared the performance of several network architectures, fusing the feature maps in the encoder (as was done in FuseNet (Hazirbas et al., 2016)) and using different late-fusion approaches, such as sum fusion or element-wise multiplication, and found that a late-fusion approach, in which the output of each stream is fused using additional convolutional layers, achieved the best performance. In this setup, the segmentation maps from the different streams are fused by concatenating the segmentation map tensors and applying two additional layers of 3×3 convolutions with PReLU activations and a 1×1 convolution.

Multiresolution Fusion In order to best incorporate the satellite images' different spatial resolutions, we follow two different approaches. When only medium-resolution images are available, we transform the feature maps into a common resolution of 96px $\times$ 96px at a 10m ground resolution by removing one upsampling layer in the Sentinel-2 encoder network. Whenever very high-resolution (VHR) optical imagery is available as well, we also remove the upsampling layer in the very high-resolution subnetwork to match the feature maps of the two Sentinel imagery streams.

Multitemporal Fusion To detect changes in an image scene over time, we use pre- and post-disaster images. We achieved the best segmentation results by concatenating pre- and post-disaster images into a single input tensor and processing them with the network described in Figure 1.

Results and Discussion

To train our model, we use medium-resolution satellite imagery with a ground resolution of 5m–10m, acquired before and after disaster events, along with very high-resolution post-event images with

Table 1: Quantitative results from two experiments reporting building intersection over union (bIoU), mean IoU (mIoU), and pixel accuracy. Table 1a compares our method to state-of-the-art approaches for segmentation of building footprints. Table 1b compares different fusion inputs for segmentation of flooded buildings using *Multi*^{3}*Net*.

Model	bIoU	Accuracy
Maggiori et al. (2017b)	61.2%	94.2%
Ohleyer (2018)	65.6%	94.1%
This work	**73.4%**	**95.7%**

(a) Segmentation of building footprints using VHR imagery of Austin in the INRIA Aerial Labels Dataset.

Data	mIoU	bIoU	Accuracy
S-1 + S-2	59.7%	34.1%	86.4%
VHR	74.2%	56.0%	93.1%
S-1 + S-2 + VHR	**75.3%**	**57.5%**	**93.7%**

(b) Segmentation of flooded buildings in Houston, TX, following Hurricane Harvey, 2017.

a ground resolution of 0.5m. Medium-resolution satellite imagery is publicly available for any location globally and acquired weekly by the European Space Agency's Sentinel-1 and Sentinel-2 satellite constellations. To obtain finer image details, such as building delineations, we use very high-resolution post-event images obtained through the DigitalGlobe Open Data Program. For radar data, we construct a three-band image consisting of the intensity, multitemporal filtered intensity, and interferometric coherence. Details about the data acquisition process and remote sensing terminology can be found in the supplementary material.

Building footprint segmentation We demonstrated the competitive performance of our model for the segmentation of building footprints. We assessed our model vis-à-vis other approaches using pixel accuracy and the intersection over union (IoU) metric. Our method outperformed state-of-the-art approaches for building footprint segmentation, reaching a building IoU of 73.4% (see Table 1a) on the Austin partition of the INRIA aerial labels dataset (Maggiori et al., 2017a).

Segmentation of disaster damage To segment footprints of flooded buildings, we used pre- and post-event images obtained by Sentinel-1 and Sentinel-2 along with post-event VHR imagery. Table 1b shows that fusing images from all sensors across time yielded the best results (75.3% mIoU). Fusing only medium-resolution Sentinel-1 and Sentinel-2 images without high-resolution imagery yielded a good segmentation accuracy (59.7% mIoU) as well. Figure 2a shows predictions for the segmentation of flooded buildings obtained from the very high-resolution-only and full-fusion models. The overlay image shows the differences between the two predictions. Fusing images obtained at multiple points in time from multiple sensors with different resolutions eliminates the majority of false positives and helps delineate the shape of detected structures more accurately.

We also used our method to segment collapsed buildings in the Ecuadorian town of Portoviejo following an earthquake in 2016. This task is much more challenging than segmenting flooded buildings due to the relative sparsity of collapsed buildings in our sample images. To achieve high predictive accuracy, we first pre-trained the network to perform standard building footprint segmentation before training the model on the footprints of collapsed buildings. This way, the model first learns to identify the set of 'buildings', before learning to segment the subset of collapsed buildings. We also modified the loss function to assign penalties ($\times 100$) for incorrectly identifying pixels that are labeled as belonging to the footprint of a collapsed building to discourage the network from over-predicting non-collapsed buildings (which make up over 90% of the pixels). Figure 2b shows that our model was able to correctly identify collapsed buildings (points A and B) as well as two buildings that were labeled as severely damaged (points C and D).

Conclusion

In disaster response, fast information extraction is crucial for first responders to coordinate disaster relief efforts, and satellite imagery can be a valuable asset for rapid mapping of affected areas. In this work, we introduced a novel end-to-end trainable convolutional neural network architecture for image segmentation via fusion of multiresolution, multisensor, and multitemporal satellite images. Our network outperformed state-of-the-art approaches on building footprint segmentation and achieved high accuracy in the segmentation of flooded buildings. We demonstrated that publicly and globally available medium-resolution imagery alone can be used for efficient segmentation of flooded buildings, making our method massively scalable. The source code as well as a dataset containing fully preprocessed and labeled multiresolution, multispectral, and multitemporal satellite imagery of disaster sites will be made publicly available.

References

Barnes, C. F.; Fritz, H.; and Yoo, J. 2007a. Hurricane disaster assessments with image-driven data mining in high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* 45(6):1631–1640.

Barnes, C. F.; Fritz, H. M.; and Yoo, J. 2007b. Hurricane disaster assessments with image-driven data mining in high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* 45:1631–1640.

Bilinski, P., and Prisacariu, V. 2018. Dense decoder shortcut connections for single-pass semantic segmentation. In *CVPR*.

Brunner, D.; Lemoine, G.; and Bruzzone, L. 2010. Earthquake damage assessment of buildings using vhr optical and sar imagery. *IEEE Transactions on Geoscience and Remote Sensing* 48:2403–2420.

Cao, Q. D., and Choe, Y. 2018. Deep learning based damage detection on post-hurricane satellite imagery. *CoRR* abs/1807.01688.

Chen, L.-C.; Papandreou, G.; Schroff, F.; and Adam, H. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Cooner, A. J.; Shao, Y.; and Campbell, J. B. 2016. Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 haiti earthquake. *Remote Sensing* 8:868.

Goldberg, M.; Li, S.; Goodman, S.; Lindsey, D.; Sjoberg, B.; and Sun, D. 2018. Contributions of operational satellites in monitoring the catastrophic floodwaters due to hurricane harvey. *Remote Sensing* 10(8):1256.

Hamaguchi, R., and Hikosaka, S. 2018. Building detection from satellite imagery using ensemble of size-specific detectors. In *CVPR Workshop*.

Hazirbas, C.; Ma, L.; Domokos, C.; and Cremers, D. 2016. Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture. In *ACCV*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*.

Maggiori, E.; Tarabalka, Y.; Charpiat, G.; and Alliez, P. 2017a. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *IGARSS*. IEEE.

Maggiori, E.; Tarabalka, Y.; Charpiat, G.; and Alliez, P. 2017b. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 55(2):645–657.

Narazaki, Y.; Hoskere, V.; Hoang, T. A.; and Spencer Jr, B. F. 2018. Automated vision-based bridge component extraction using multiscale convolutional neural networks. *arXiv preprint arXiv:1805.06042*.

Ohleyer, S. 2018. Building segmentation on satellite images. `https://project.inria.fr/aerialimagelabeling/files/2018/01/fp_ohleyer_compressed.pdf`. Accessed: 2018-08-26.

Ramlal, B.; Davis, D.; and De Bellott, K. 2018. A rapid post-hurricane building damage assessment methodology using satellite imagery. *West Indian Journal of Engineering* 41(1).

Scarsi, A.; Emery, W. J.; Serpico, S. B.; and Pacifici, F. 2014. An automated flood detection framework for very high spatial resolution imagery. *IEEE Geoscience and Remote Sensing Symposium* 4954–4957.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Soergel, U. 2010. *Radar Remote Sensing of Urban Areas*, volume 15. Springer.

Sun, G.; Hao, Y.; Rong, J.; Shi, S.; and Ren, J. 2017. Combined deep learning and multiscale segmentation for rapid high resolution damage mapping. In *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 1101–1105. IEEE.

Yamazaki, F. 2001. Applications of remote sensing and gis for damage assessment. *Structural Safety and Reliability* 1–12.

Yu, F.; Koltun, V.; and Funkhouser, T. A. 2017. Dilated residual networks. In *CVPR*.

Zhang, L.; Zhang, L.; and Du, B. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine* 4:22–40.

Zhao, H.; Shi, J.; Qi, X.; Wang, X.; and Jia, J. 2017. Pyramid scene parsing network. In *CVPR*.

Zhu, X. X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; and 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.