Assessing Building Damage by Learning the Deep Feature Correspondence of Before and After Aerial Images

Maria Presa-Reyes and Shu-Ching Chen
School of Computing and Information Sciences
Florida International University, Miami, Florida 33199
Emails: {mpres029,chens}@cs.fiu.edu

***Abstract*—The damage caused by a natural disaster such as a hurricane, not only impacts human lives but can also be detrimental to the city's infrastructure and potentially cause the loss of historical buildings and essential records. Delivering an effective response requires quick and precise analyses concerning the impact of a disastrous event. With the current technological developments to acquire massive volumes of data and the recent advances in artificial intelligence and machine learning, now more than ever, disaster information integration and fusion have the potential to deliver enhanced situational awareness tools for humanitarian assistance and disaster relief efforts. Given the aerial images of a residential building taken before and after a natural disaster, recent applications of Convolutional Neural Networks (CNNs) work well when differentiating two types of damage (i.e., whether the structure is intact or destroyed) but underperform when trying to differentiate more damage levels. According to our findings: (1) including enough surrounding context provides essential visual clues that help the model better predict the building's level of damage and (2) learning the correspondence between the features extracted from pre- and post-imagery boosts the performance compared to a simple concatenation. We propose a two-stream CNN architecture that overcomes the difficulties of classifying the buildings at four damage levels and evaluate its performance on a curated, fully-labeled dataset assembled from open sources.**

***Keywords*-damage assessment; deep learning; convolutional neural networks**

I. Introduction

September 10, 2017 marks the date Hurricane Irma struck the Florida Keys as a category four storm with the maximum sustained wind speed of 132 mph and storm surge reaching up to 8 feet [1]. The eye of the storm made landfall over Cudjoe Key, and consequently, the Lower and Middle Keys received the highest impact. The moment such a catastrophic event strikes, quick and accurate situational information is critical in delivering an effective response. Emergency responders need to have complete situational awareness, including location, cause, and severity of the damage before they can act on an affected area. However, disasters often cause disruptions to local communication systems and transportation infrastructure, making the areas in most needs inaccessible.

Remote sensing data, which is the data collected by a high-flying aircraft or the satellite scanning of the earth, has become a crucial tool to survey the damage of the affected regions with limited accesses rapidly. Current efforts have started to rely on aerial photographs captured right after the tragedy to assess the resulting damage and make decisions [2]. However, the search and assessment of the damage concerning specific regions become a laborious and time-consuming job for emergency responders when large areas get affected. Hence, there is a need for tools and methods that can aid the rapid assessment of damaged homes in these time-sensitive situations.

Figure 1. Samples of aerial photographs depicting the different levels of damage caused by Hurricane Irma on the Florida Keys.

Machine learning and deep learning have shown tremendous success in various research areas [3], [4], [5], [6]. Convolutional Neural Networks (CNNs) are renowned for achieving state-of-the-art performance in image classification. Given the current technological developments to acquire massive volumes of data and the recent advances in artificial intelligence and machine learning, now more than ever, disaster information integration and fusion have the potential to deliver enhanced situational awareness tools for humanitarian assistance and disaster relief efforts [7], [8]. We propose a model trained on pre- and post-disaster satellite imagery. As a case study, we selected the event of Hurricanes Irma and the damage it has inflicted on the Florida Keys. Figure 1 shows a sample of the manually-labeled dataset depicting residential buildings under four varying levels of damage.

Our objective is to explore the optimum application of

CNNs to solve this real world problem. Thereupon, Section II, reviews related studies that apply deep learning methods and other statistical techniques to damage assessment. Section III introduces our proposed approach, the challenges, and how to overcome the challenges. Then, Section IV describes the dataset selected as our case study to demonstrate the effectiveness of our proposed work. In Section V, the effectiveness of our proposed model is shown through experimental results. Finally, Section VI concludes the paper with some potential future work.

II. Related Work

The technological advances in data collection and the growing availability of high-quality data have enabled substantial research in the field of automated and rapid damage assessment. Indeed some of our proposed techniques were inspired and improved upon Fujita *et al.*'s research previously published in 2017 [9]. The study proposed the application of a two-stream CNN model to classify whether a building survived or was washed away by a tsunami that followed the wake of the Great East Japan earthquake on March 11, 2011. Another source of inspiration and information was the survey conducted after Hurricane Irma by Xian *et al.* [10] assessing the damage of more than 1600 residential buildings in the Florida Keys. The study also provided a statistical analysis that reveals distinct factors potentially influencing the degree of damage in some of the most affected regions.

Some of the most recent work includes Ci *et al.*, which introduced a new approach by combining CNNs for extracting features and a new loss function known as ordinal regression for optimizing classification results, aiming at assessing the degree of building damage caused by the 2010 Yushu earthquake and the 2014 Ludian earthquake [11]. Moreover, in the Summer of 2019, the Defense Innovation Unit (DIU) announced the xView2 Challenge [12] which aimed to stimulate applied research with a focus on automating the process of assessing building damage after a natural disaster. DIU publicly released a high-quality and large-scale dataset known as xBD, composed of satellite imagery annotated with building localization and levels of damage before and after natural disasters.

The previously introduced damage assessment studies work well when classifying damaged buildings into two categories (i.e., intact or destroyed). The goal of our proposed approach is to overcome the difficulties of classifying the building at different damage levels and greatly boost the performance of the model.

III. Approach

A two-stream CNN deep feature fusion network is designed to take a pair of image patches as input and predict the damage level of the building at the center of the patch. The concept of two-stream architecture is not new in the context of damage assessment from before and after images [9], [11]. However, we demonstrate the improvements made by our proposed work by first preprocessing the input data in a unique way that reduces the uncertainty and increases the performance of the model (Sec. III-A) and then applying a new network configuration focusing on a fusion technique more advanced than simply concatenating the deep features from each stream (Sec. III-B).

A. Data Preprocessing

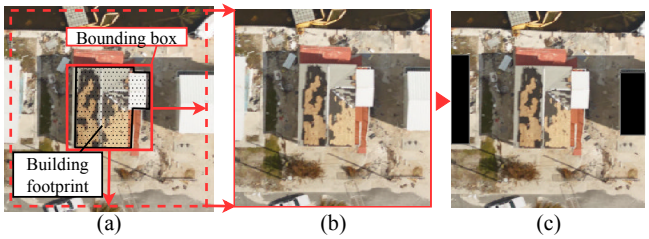


Figure 2. The input patch preprocessing steps start with (a) the bounding box surrounding the building's footprint is extended to cover enough surrounding area; (b) then the resized patch containing the building in the center is cropped; and (c) finally, nearby buildings are occluded to avoid confusing the model.

Our study finds that the surroundings of a building provide critical contextual information and visual cues that can help the model better predict the building's level of damage. However, when including too much of the surrounding area increases the risk of confusing the model by inadvertently feeding it image patches that contain multiple buildings of different damage levels. As shown in Figure 2, after resizing the bounding box to a size 80% larger than the footprint's geometric bounding box, and cropping the widened patch containing the building in the center, the nearby buildings are occluded by negating the pixels found inside the surrounding buildings' footprints. Hence, for each input pair, the model will only focus on the building in the center, and the clues found in the surrounding area will serve to help improve the performance. This method is very effective in reducing the uncertainty of the model.

B. Framework Configuration

Figure 3 demonstrates the proposed architecture with both CNN streams following the ResNet50 architecture pretrained on ImageNet [13]. The last classification layer is removed to fuse the outputs of the networks' average pooling layer and fine-tune the weights of the entire network. Both networks are trained on their own unshared weights, giving each of them the flexibility to learn specific features from their individual input data stream. Training both streams on unshared weights have shown to be much more effective rather than sharing the weights [9], especially when there is a significant gap between the time the images were taken

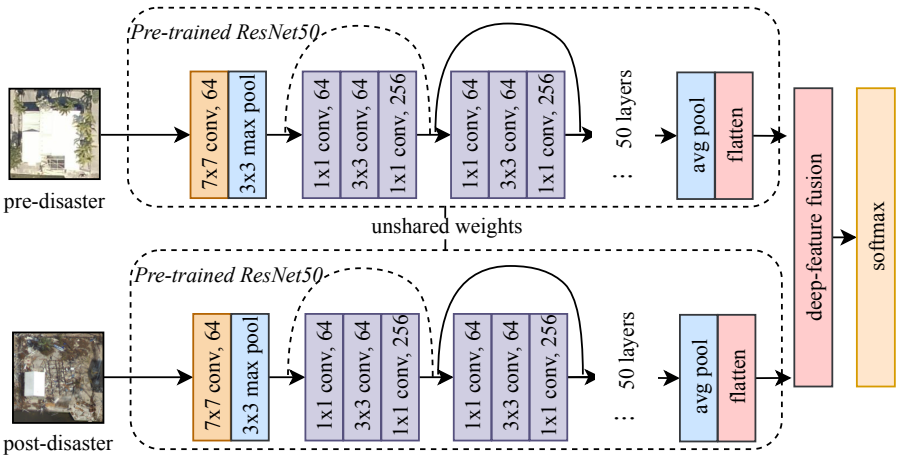


Figure 3. Proposed two-stream CNN architecture

(as described in Sec. IV), causing the image's appearance to be distinct.

1) Deep Feature Fusion: Our goal is to fuse the features generated by the two streams in a manner that allows the overall model to learn the correspondence of the inputs from both networks [14]. Namely, a fusion function $f : \mathbf{x}_n^l, \mathbf{x}_n^r, \rightarrow \mathbf{y}_n$ fuses the feature pair at the output of both models' n-th layer given the feature maps $\mathbf{x}_n^l \in \mathbb{R}^{H \times W \times D}$ and $\mathbf{x}_n^r \in \mathbb{R}^{H' \times W' \times D'}$ to produce the output feature $\mathbf{y}_n \in \mathbb{R}^{H'' \times W'' \times D''}$, where H, W, and D constitute the height, width, and number of channels of the corresponding feature map. In our approach, f is applied as a late-fusion. The following assumptions are made, $H = H' = H''$, $W = W' = W''$, $D = D' = D''$, and the subscript n is dropped for simplicity purpose.

Concat Fusion. $\mathbf{y}^{\text{cat}} = f^{\text{cat}}(\mathbf{x}^l, \mathbf{x}^r)$ joins the feature sequence along an existing axis. In our case, both features are joined across their width:

$$\mathbf{y}^{\text{cat}} = \text{concat}\{x_{i,j,d}^l, x_{i,j,d}^r\},$$

where $1 \leq i \leq H$, $1 \leq j \leq W$, $1 \leq d \leq D$ and $\mathbf{y} \in \mathbb{R}^{H \times 2W \times D}$. Concatenation is one of the most common fusion techniques. This type of fusion often works well and is simple to implement. However, this technique by itself does not define a correspondence between the features.

Conv Fusion. $\mathbf{y}^{\text{conv}} = f^{\text{conv}}(\mathbf{x}^l, \mathbf{x}^r)$ first the two feature maps are stacked at the same spatial location (i, j) across channel d, namely:

$$y_{i,j,2d}^{\text{stack}} = x_{i,j,d}^l, \qquad y_{i,j,2d-1}^{\text{stack}} = x_{i,j,d}^r,$$

where $\mathbf{y} \in \mathbb{R}^{H \times W \times 2D}$. The layers that follow define the correspondence by learning the suitable filters when convolving the stacked data defined as follows.

$$\mathbf{y}^{\text{conv}} = \mathbf{y}^{\text{stack}} * \mathtt{f} + \mathtt{b},$$

where $\mathtt{b} \in \mathbb{R}^{D'''}$ is the term for bias and $\mathtt{f} \in \mathbb{R}^{1 \times 1 \times 2D \times D'''}$ is a bank of filters. In our approach, the total number of filters is set to $D''' = 100$. Thus, $\mathtt{f}$ is trained on the weighted combinations of $\mathbf{x}^l$ and $\mathbf{x}^r$ at the same feature location.

IV. Data

A. Aerial Photographs Taken Before and After Hurricane Irma

1) Pre-disaster images: The Florida Keys are part of Monroe County, which Geographic Information System (GIS) department [15] makes available high-quality satellite images throughout different years, with the latest three years being 2012, 2015, and 2018. Thus, the data available for the year 2015 was selected to serve as a reference for how the building structure looked like before the catastrophic event. Unlike the post-disaster images, the images provided by Monroe County have a ground sample distance (GSD) of 50 cm per pixel. They are also better orthorectified and depict a more consistent level of illumination throughout the region.

2) Post-disaster images: Right after the event of a natural disaster, the National Oceanic and Atmospheric Administration (NOAA)'s Remote Sensing Division often conducts aerial photography missions using a Trimble Digital Sensor System (DSS) from an altitude of 2,500 to 5,000 feet [2], [16]. The images are captured using specialized cameras aboard NOAA's aircraft, then posted online for the public to access [17]. The images captured right after Hurricane Irma are high-quality, with each pixel representing a ground sample distance (GSD) of 15 cm to 30 cm.

B. Building Footprints

Other than high-quality orthorectified natural images, the Monroe County GIS department provides information on the geometric references of various building footprints. These geometries are then combined together with the geometries

of building footprints from OpenStreetMaps [18] to obtain a more cohesive set.

C. Preliminary Damage Assessment

For Hurricane Irma, Monroe County released a preliminary damage assessment report [19]. Although the assessment appears to be incomplete, only including references for some of the majorly damaged and destroyed residential homes, it still provides a good reference on how to get started with labeling the different types of damages. The assessment from this survey was combined with the information from Xian *et al.* [10] to create the dataset summarized in Table I. The following lists the definition of each classification label in accordance to FEMA's official guide to assess damage and impact [20].

Table I
THE STATISTICAL INFORMATION OF THE BUILDING DAMAGE DATASET.

No.	Concepts	# of Instances
1	No damage	17,482
2	Minor damage	2,633
3	Major damage	962
4	Destroyed	1,436
	Total	22,513

- *No damage*: the structure exhibits no damage or minimal change, such as some missing shingles in the building rooftop.
- *Minor damage*: it constitutes a wide range of damages that do not necessarily affect the integrity of the structure.
- *Major damage*: the building sustained significant damage to its structural elements and requires extensive repairs.
- *Destroyed*: the structure is rendered a total loss, and repair will not be feasible.

V. EXPERIMENTS AND ANALYSES

A. Experimental Setup

The fusion techniques introduced in Section III-B1 (i.e., `concat fusion` and `conv fusion`) are compared with a scenario (called `post only`) where one pre-trained ResNet50 model is trained on the images taken after the disaster event. Our approach is evaluated on our fully-labeled dataset depicting various damage levels resulting from Hurricane Irma's landfall on the Florida Keys. Our inputs are also divided into two sets (1) `crop center`, the size of the patch is only as large as the size of the bounding box surrounding the building footprint, the information in the patch will focus mostly on the building rooftop, and the final patch sizes are 112×112; and (2) `extended patch`, where the bounding box is extended to cover enough surrounding area as described in Sec. III-A, and the final patch sizes are 224×224.

To overcome the highly-imbalanced nature of the Hurricane Irma dataset (as depicted in Table I), data augmentation is applied to enhance the training set. Data augmentation is a common approach that serves to improve the performance of the CNN models as well as its generalization capability by applying random transformations to the input data. The augmentation techniques which are randomly applied to our dataset consist of horizontal and vertical flips, rotation, shear transformations, and zooming. Moreover, given the limited size of the dataset, 5-fold cross-validation is utilized to evaluate our model and to prevent over-fitting. In each fold, 80% of the images are randomly selected for training and the remaining 20% for testing. Each model configuration is trained for a total of 80 epochs on each fold.

The Adam solver is employed to optimize our model with an initial learning rate ($\eta = 0.0001$) that is small enough to update the transferred weights slowly when fine-tuning the model and to achieve a more optimal set of final weights [21]. During training, the learning rate will drop to 10% of its current learning rate after there have been no improvements to the testing loss for 10 epochs. Moreover, the model is also trained using small batches of size 16. The smaller batch sizes introduce noise to the training process, leading to a regularizing effect that improves the model's generalization capability for a given computational cost [22].

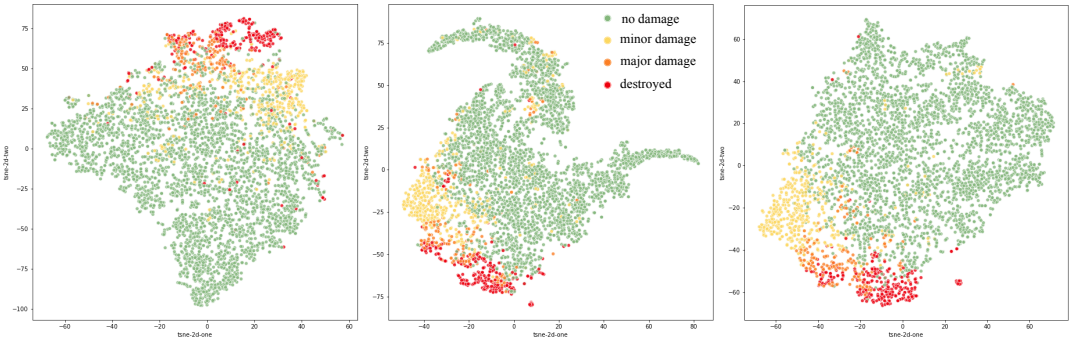
B. Results and Discussion

Table II demonstrates the performance summary for the class-specific F1-scores under the three configurations previously described (Sec. V-A) as well as considering the `crop center` and `extended patch` input scales. The results report the average of five runs from the cross-validation setup, with a standard deviation ranging from $\sigma = 9.6\text{e}{-04}$ to $\sigma = 3.8\text{e}{-02}$. The experimental results demonstrate the effectiveness of our proposed approach which (1) `extended patch`, pre-processing the data by extending the building patch to include enough surrounding context while occluding surrounding buildings of potentially different damage levels; and (2) `conv fusion`, applying an advanced fusion technique by learning the correspondence between two feature maps. The proposed approach consistently achieves the best results throughout the four damage categories.

Various aspects are emerged when analyzing the results. First, the fusion of deep features from the pre-post image pairs performs better than the `post only` configuration, even though there is a significant gap of 2 years between the times both images were captured. Other than the time difference, there are also significant changes in angle, illumination, and resolution. Moreover, because the imagery must be captured in a timely manner after the event, it is common for the post-disaster photos to have clouds present and sometimes present blurriness. The performance improvements demonstrate the need to include a reference of how the building looked

Table II
RESULTS SUMMARY OF EACH CLASS'S F1-SCORE AVERAGED FROM THE RESULTS OF 5-FOLD CROSS-VALIDATION

config	crop center				extended patch			
	no-dmg	minor-dmg	major-dmg	destroyed	no-dmg	minor-dmg	major-dmg	destroyed
post only	0.938	0.673	0.530	0.747	0.951	0.739	0.582	0.815
concat fusion	0.956	0.745	0.587	0.819	0.969	0.819	0.701	0.875
conv fusion	0.957	0.749	0.616	0.834	**0.971**	**0.831**	**0.711**	**0.901**



(a) (b) (c)

Figure 4. t-SNE visualization of the feature layer right before the softmax layer for the first fold in cross-validation trained on the input data of the extended patch for (a) post only; (b) concat fusion; and (c) conv fusion

like before possibly being affected by the event in order to make a better assessment. Second, there is also a boost in performance by extending the image patch to include the building's surroundings (i.e., `extended patch`) compared to using the image patches by including mostly the rooftop information (i.e., `crop center`).

The configuration `concat only` is a very common approach of fusing the features and performs well when compared with the `post only` option. However, it is demonstrated that by considering the correspondence between the features, the model's performance can be further improved. Figure 4 shows the capability of each model configuration to separate the four categories using t-Distributed Stochastic Neighbor Embedding (t-SNE) [23] to visualize the features from the layer right before the softmax classifier. t-SNE is a powerful tool that is able to map high-dimensional data into a lower dimension and has an insight into how the data is arranged. While Figures 4(a) and 4(b) demonstrate some overlaps between different damage levels, Figure 4(c) shows a better separation of the classes.

Our proposed model obtains a weighted F1-score of 0.94, calculated by first computing the F1-score for each damage class, then finding their average weighted by the number of true instances in each class, taking into consideration the imbalanced nature of the data. Table III summarizes the Precision, Recall, and F1-scores on a class-by-class basis for the proposed approach. The model demonstrates the weakest performance when classifying major damaged instances. This is due to the limited number of samples available (as shown in Table I) and the overlaps between major and minor damaged buildings.

Table III
PERFORMANCE MEASURES OF THE PROPOSED NETWORK CONFIGURATION

Damage Type	Precision	Recall	F1-Score
No Damage	0.96	0.98	0.97
Minor Damage	0.87	0.79	0.83
Major Damage	0.73	0.69	0.71
Destroyed	0.90	0.90	0.90

VI. CONCLUSION

Due to the technological advances in artificial intelligence and machine learning, now more than ever, disaster information integration and fusion have the potential to deliver enhanced situational awareness tools. In this study, CNNs are explored to assess the damages of residential homes by consolidating the aerial photographs taken before and after a disaster and finding the correspondence between the features of the image pair. The effectiveness of our model is demonstrated using a curated dataset collected from open data sources. As future directions for this work, we plan to 1) extend the problem to involve both the localization [24] of the building and classification of the damage level, and 2) integrate the image data with other remote sensing information, such as Light Detection and Ranging (LIDAR), to further improve the performance.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. HRD-1547798, which was awarded to Florida International University as part of the Centers of Research Excellence in Science and Technology (CREST) Program. This is contribution number 947 from the Southeast Environmental Research Center in the Institute of Water and Environment at Florida International University.

REFERENCES

[1] H. Wetherington, "Hurricane Irma recovery." [Online]. Available: https://www.monroecounty-fl.gov/726/Hurricane-Irma-Recovery

[2] "Assessing Irma's destruction from the air: Aerial images available." [Online]. Available: https://www.noaa.gov/news/assessing-irma-s-destruction-from-air-aerial-images-available

[3] S.-C. Chen, M.-L. Shyu, and C. Zhang, "Innovative shot boundary detection for video indexing," in *Video data management and information retrieval*. IGI Global, 2005, pp. 217–236.

[4] S.-C. Chen, S. H. Rubin, M.-L. Shyu, and C. Zhang, "A dynamic user concept pattern learning framework for content-based image retrieval," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 36, no. 6, pp. 772–783, 2006.

[5] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, "Effective supervised discretization for classification based on correlation maximization," in *2011 IEEE International Conference on Information Reuse & Integration*. IEEE, 2011, pp. 390–395.

[6] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, p. 92, 2019.

[7] L. Zheng, C. Shen, L. Tang, C. Zeng, T. Li, S. Luis, and S.-C. Chen, "Data mining meets the needs of disaster information management," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 5, pp. 451–464, 2013.

[8] M. E. P. Reyes, S. Pouyanfar, H. C. Zheng, H.-Y. Ha, and S.-C. Chen, "Multimedia data management for disaster situation awareness," in *International Symposium on Sensor Networks, Systems and Security*. Springer, 2017, pp. 137–146.

[9] A. Fujita, K. Sakurada, T. Imaizumi, R. Ito, S. Hikosaka, and R. Nakamura, "Damage detection from aerial images via convolutional neural networks," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. IEEE, 2017, pp. 5–8.

[10] S. Xian, K. Feng, N. Lin, R. Marsooli, D. Chavas, J. Chen, and A. Hatzikyriakou, "Brief communication: Rapid assessment of damaged residential buildings in the Florida Keys after Hurricane Irma," *Natural Hazards and Earth System Sciences*, vol. 18, no. 7, 2018.

[11] T. Ci, Z. Liu, and Y. Wang, "Assessment of the degree of building damage caused by disaster using convolutional neural networks in combination with ordinal regression," *Remote Sensing*, vol. 11, no. 23, p. 2858, 2019.

[12] R. Gupta, B. Goodman, N. Patel, R. Hosfelt, S. Sajeev, E. Heim, J. Doshi, K. Lucas, H. Choset, and M. 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*, 2019, pp. 10–17.

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

[14] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1933–1941.

[15] Monroe County - GIS, "Orthophotography viewer," http://monroecounty-fl.maps.arcgis.com/apps/webappviewer/index.html?id=d60ffd1357504b4b83319445b488b334.

[16] National Oceanic and Atmospheric Administration (NOAA), "Hurricane IRMA: Emergency response imagery of the surrounding regions." [Online]. Available: https://storms.ngs.noaa.gov/storms/irma/download/metadata.html

[17] US Department of Commerce and NOAA, "Emergency response imagery." [Online]. Available: https://storms.ngs.noaa.gov/

[18] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

[19] Monroe County - GIS, "Hurricane Irma - damage assessment." [Online]. Available: http://monroecounty-fl.maps.arcgis.com/apps/webappviewer/index.html?id=87fc14264bdf43fba1669413381dfe3a

[20] Federal Emergency Management Agency (FEMA), "Damage assessment operations manual," 2016. [Online]. Available: https://www.fema.gov/media-library-data/1558541566358-30e29cac50605aae39af77f7e25a3ff0/Damage_Assessment_Manual_4-5-2016.pdf

[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT press, 2016.

[22] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," *arXiv preprint arXiv:1804.07612*, 2018.

[23] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[24] S.-C. Chen, M.-L. Shyu, C. Zhang, and R. L. Kashyap, "Identifying overlapped objects for video indexing and modeling in multimedia database systems," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 04, pp. 715–734, 2001.