# Detecting damaged buildings in post-Hurricane satellite imagery

Chetan Bademi | bademi@stanford.edu | chetanbademi@gmail.com

## Introduction

The goal of this work is to label buildings with the extent of damage from satellite imagery after a hurricane. This information can be useful for guiding rescue operations and for insurance purposes. After every disaster in the United States, Federal Emergency Management Agency(FEMA) provides labels for every building in the affected region, with different categories of damage[1]. This is done by a field survey and disaster reports. The goal of this work is to study if this process can be automated.

## Related Work

Cao et al [2] and Duarte et al[3] have used CNNs to detect damage by using damaged and non-damaged buildings as two classes with good . Doshi et al[4] have used change in semantic segmentation of satellite images as a proxy for damage at a grid-level. To the best of our knowledge, no one has studied the problem of labelling individual buildings based on degree of damage.

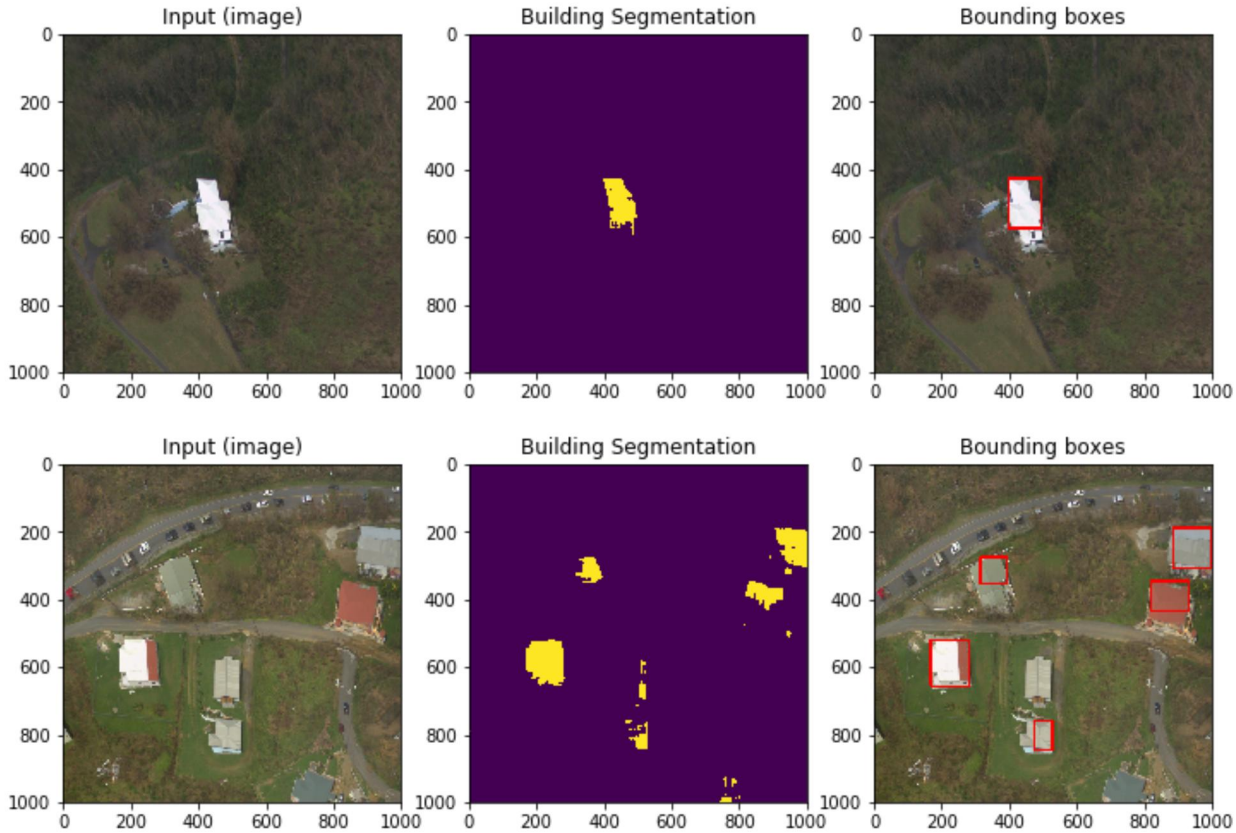
## Methods

The proposed algorithm to label individual buildings in a satellite image is as follows:

1. Run object detection/building footprint detection algorithm on the image
2. Crop out individual buildings from the image
3. Run the classifier on the cropped image to label the building

### Building Detection

To detect buildings in a satellite image, we run a building footprint detector and then crop the buildings from this semantically segmented image. Since training a model to damage segmentation task is a non-trivial task, we will be using the library provided by Github user motokimuara[6]. The semantic segmentation was then post processed to obtain bounding boxes for buildings in satellite images.




# Building Labelling

## Dataset

Chen et al[8] provide a database of damage buildings from Hurricane Harvey annotated with the degree of damage provided by FEMA. There are a total of 600k buildings and each building is labelled one of ‘no damage’, ‘affected’, ‘minor damage’, ‘major damage’ and ‘destroyed’.

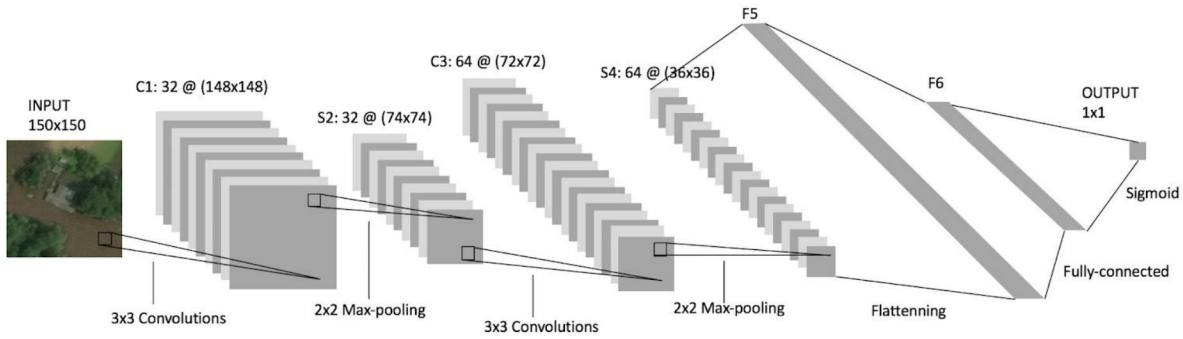
The database lists the coordinates of the bounding box for each building. Satellite images for Hurricane Harvey were downloaded from the National Oceanic and Atmospheric Administration’s website[9]. Using the bounding boxes from Chen et al[8], each building was cropped from NOAA’s images. Since the label ‘Destroyed’ had the fewest number of buildings (approximately 9,000), images for others were randomly sampled to create a balanced dataset. The dataset of 45,000 was further split into train, dev and test splits with 70, 15 and 15 percent of examples each. Each image has 3 channels (RGB) and the height and width are between 50 and 150 pixels.

## Methods

The first step in building the classifier was to build a baseline model which performed better than random. However, none of the models trained using the dataset performed much better than random. This section lists all the ideas that were tried.

## Model

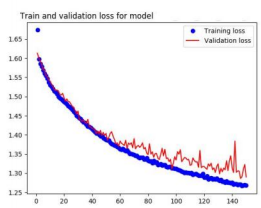
Since the dataset was similar to the one used by Cao et al[7], we used the architecture of their best performing model.




Model architecture as described in Cao 2018.

## Training a five class model with CNN architecture from Cao et al

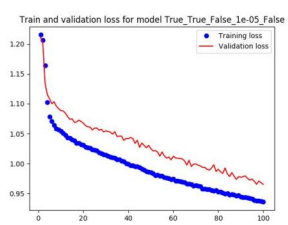
The CNN architecture from Cao et al was used with softmax layer with five units. We experimented with adding more and removing some convolutional and/or fully connected layers. None of the models performed better than random.




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| AFF | 0.17 | 0.04 | 0.07 | 1453 |
| DES | 0.19 | 0.23 | 0.21 | 1381 |
| MAJ | 0.20 | 0.28 | 0.23 | 1652 |
| MIN | 0.20 | 0.12 | 0.15 | 1583 |
| none | 0.22 | 0.31 | 0.25 | 1806 |
| avg / total | 0.19 | 0.20 | 0.19 | 7875 |

## Training a three class model with CNN architecture from Cao et al

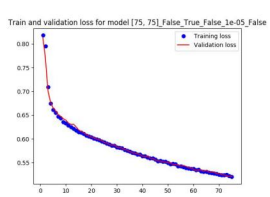
Since the five class model didn't perform well, we simplified the task to predict only the following labels: 'destroyed', 'major damage' and 'no damage'.  None of the models performed much better than random.




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| DES | 0.28 | 0.39 | 0.33 | 1381 |
| MAJ | 0.33 | 0.26 | 0.29 | 1652 |
| none | 0.38 | 0.34 | 0.36 | 1806 |
| avg / total | 0.33 | 0.33 | 0.33 | 4839 |

## Training a binary classifier with CNN architecture from Cao et al

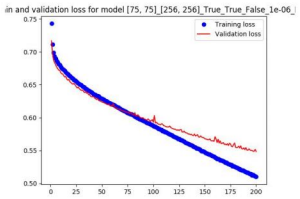
We further simplified the task to distinguish between 'destroyed' and 'no damage' labels.




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| DES | 0.41 | 0.34 | 0.37 | 1381 |
| none | 0.55 | 0.62 | 0.58 | 1806 |
| avg / total | 0.49 | 0.50 | 0.49 | 3187 |

## Training a binary classifier using dense neural networks

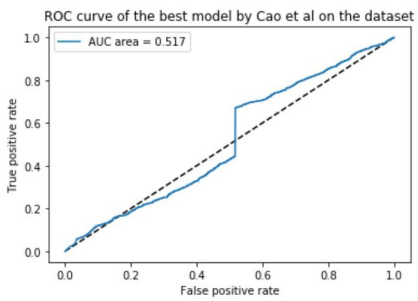
We experiment with one, two and three layer DNNs with a few different number of units (256, 512, 1024) in each layer to establish a baseline for the task.




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| DES | 0.44 | 0.40 | 0.42 | 1381 |
| none | 0.57 | 0.60 | 0.59 | 1806 |
| avg / total | 0.51 | 0.52 | 0.51 | 3187 |

## Evaluating the model from Cao et al on the dataset for binary classification

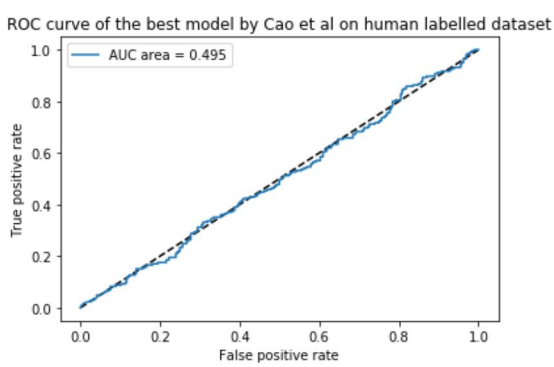
To evaluate if the dataset is somehow systematically biased, we used the best trained model from Cao et al to classify 'destroyed' and 'no damage' labels. Even though their model performed very well on their dataset (AUC = 0.96), it performs barely better than random (AUC = 0.517)




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| DES | 0.51 | 0.37 | 0.43 | 1381 |
| none | 0.60 | 0.74 | 0.66 | 1806 |
| avg / total | 0.56 | 0.58 | 0.56 | 3187 |

## Manually separating buildings labelled 'destroyed' into 'damaged' and 'not damaged'

To further test our hypothesis that the dataset contains many images where the damage is not visible, we divided images labelled 'destroyed' into 'damaged' (293 images) and 'not damaged' (206 images)by looking for flooding and building destruction. We found the model performs as good as random on this dataset.




This could be because of multiple reasons:

- The test set was small (499 images)
- The model by Cao et al was used incorrectly
- The model by Cao et al was trained on images from a different distribution
- The hypothesis that the dataset is noisy is incorrect and the reason for bad performance is bad modelling techniques

## Conclusion

We tried to build a classifier based on the dataset provided by Chen et al[8]. When the accuracy of the classifier was not better than random, we tried to prove that the dataset is inherently noisy due to the damage not being visible from the satellite images.

## Future Work

The dataset can be made less noisy by having each image labelled by multiple humans.

## Acknowledgements

We would like to thank Ishan Patil for his constant support and guidance throughout the duration of this project.

## Code

Building detection: https://github.com/BademiChetan/spacenet_building_detection
Damage classifier: https://github.com/BademiChetan/CS230Project

## References

1. Federal Emergency Management Agency, “Damage assessment operations manual,” Available: https://www.fema.gov/media-library/assets/documents/109040
2. Cao, Quoc Dung, & Choe, Youngjun. 2018. Deep Learning Based Damage Detection on PostHurricane Satellite Imagery. arXiv preprint arXiv:1807.01688.
3. Duarte, D, Nex, F, Kerle, N, & Vosselman, G. 2018. Satellite Image Classification of building damages using airborne and satellite image samples in a deep learning approach. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4(2).
4. Doshi, Jigar; Basu, Saikat; Pang, Gua 2018. From Satellite Imagery to Disaster Insights. eprint arXiv:1812.07033
5. Federal Emergency Management Agency, National Disasters, https://data.femadata.com/NationalDisasters/HurricaneMaria/
6. Motokimura, SpaceNet Building Detection, https://github.com/motokimura/spacenet_building_detection
7. SpaceNet on Amazon Web Services (AWS). “Datasets.” The SpaceNet Catalog. Last modified April 30, 2018.Accessed on [Insert Date]. https://spacenetchallenge.github.io/datasets/datasetHomePage.html.
8. Sean Andrew Chen, Andrew Escay, Christopher Haberland, Tessa Schneider, Valentina Staneva, Youngjun Choe. 2018. Benchmark Dataset for Automatic Damaged Building Detection from Post-Hurricane Remotely Sensed Imagery
9. N. Oceanic and A. Administration. (2017, Sep.) Hurricane harvey imagery. [Online].Available:https://storms.ngs.noaa.gov/storms/harvey/index.html#7/28.400/-96.690