Detecting damaged buildings in post-Hurricane satellite imagery

Chetan Bademi | bademi@stanford.edu | Poster presentation: https://youtu.be/GSEkwDC27_s

Introduction

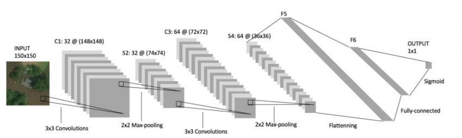
The goal of this work is to label buildings with the extent of damage from satellite imagery after a hurricane. This information can be useful for guiding rescue operations and for insurance purposes.

Data

Chen et al[1] 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’. 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 80 and 150 pixels.

Model

We used the same architecture as the best performing model by Cao et al[2], as described below.



We also experimented with adding/removing some convolutional and fully connected layers from this baseline architecture.

Results

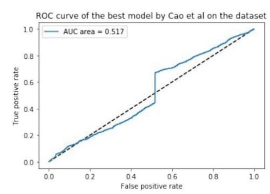
We tried building multiple models, but none of them performed better than random.

Model	Training Error	Test Error	Average Precision	Average Recall
Five class AFF, DES, MIN, MAJ, None	1.25	1.30	0.19	0.20
Three class DES, MAJ, None	0.925	0.975	0.33	0.33
Two class DES, None	0.5	0.55	0.51	0.52

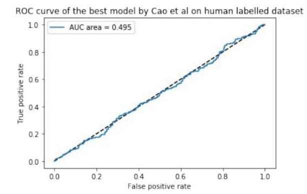
Training data had ~7000 samples per class and test and validation had ~1500 samples per class. We used Adam to optimize the cross entropy loss.

Discussion

We hypothesize among all buildings labelled “destroyed” and “major damage”, the damage was not visible from the satellite for some of the buildings. To verify this, we ran the best model by Cao et al[2] on a dataset with “destroyed” and “no damaged” images. It performed barely better than random.



To further verify if this hypothesis is correct, we manually divided images labelled “destroyed” into “damaged” (293 images) and “not damaged” (206 images) by looking for flooding and/or building destruction. We used the best model by Cao et al on this dataset and found that it performed as good as random. We hypothesize might be because the model was trained on data which includes many other kinds of damage



Conclusion and Future Work

We found that the dataset provided by Chen et al is not useful for building object detectors and classifiers as the damage assessed by FEMA is not necessarily visible in satellite imagery. Human labelling is needed to make these datasets more accurate.

References

1. 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
2. Cao, Quoc Dung, & Choe, Youngjun. 2018. Deep Learning Based Damage Detection on PostHurricane Satellite Imagery. arXiv preprint arXiv:1807.01688.