CS230

GIN & TONIC: Graph Infused Networks with Topological Neurons for Inference & Classification

Robert Trevino, Vinay Sawal, and Kevin Yang
Department of Computer Science
Stanford University
rptrevin@stanford.edu, vsawal@stanford.edu, kyang14@stanford.edu

Abstract

In the wake of a natural disaster, the need for actionable information towards a damaged area is critical. Before rescue groups can take action, volunteers must manually identify the location and classify the damage of the affected buildings. The XView2 organization cites this as a main analytical bottle-neck in a post-disaster workflow. Inspired by their challenge, our project aims to automate the process of assessing building damage post-disaster and enable a speedy and resource-efficient response operation. Our hypothesis is that graph-based learning should be effective given the spatial properties of natural disaster damage. Trained on high resolution satellite imagery, location, and temporal data, we designed a hybrid GCN + CNN model to classify damage-lvl of each building effectively. Our results were conclusive - showing a drastic boost in results over XView's provided baseline.

1 Introduction

According to the UN's DRR office, the world has seen a dramatic increase in the number of natural disasters in the last 20 years [1, 2, 3]. The brunt of the burden is experienced by poorer countries both in economic cost and lives lost. One month ago, for example, Taal Volcano, the Philippine's most active volcano, erupted and caused an estimated $63M in damage to physical structures in nearby plantations, farms, and cities. Fast, automated post-disaster analysis is then crucial. We believe in the shift from a slow human-annotated building assessment process to a rapid deep learning system. In order to realise this, we propose using state-of-the-art CNN models to vastly enhance the identification and classification of these disaster-stricken areas and let human response teams focus on what they do best.

The input data to our proposed algorithm is a set of high resolution, RGB satellite overhead imagery spanning 19 natural disaster-stricken areas. Figure 1 shows the satellite imagery examples of pre and post disaster images. The baseline model is provided by XView2 [5]. Our goal is to improve upon this with graph-based techniques. The proposed model is hybrid deep learning model with satellite images trained on 1) a Graph Convolutional Network for spatial classification of damage based on neighbor buildings and 2) A Resnet-50 classifier from the XView2's baseline with its upper-half trained on satellite image data and its lower half frozen to preserve lower-lvl features captured by imagenet; its output consist of annotated-versions of the satellite imagery, specifically with masked colored polygons to identify specific buildings in the area and a rating number signifying the level of damage a building sustained post disaster.

2 Related work

The exploration of deep learning models on satellite imagery has been used for a variety of different important tasks ranging from counting cars and mega-city planning to classification of land cover and crop types

CS230: Deep Learning, Winter 2020, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

Figure 1: Training data sample from DigitalGlobe disaster image sources [4]
L-R: Hurricane Harvey; Joplin tornado; Lower Puna volcanic eruption; Sunda Strait tsunami.
Top Row: Pre-disaster, Bottom Row: Post-disaster

[6, 7, 8, 9, 10, 11]. These satellite deep learning models are generally multi-layered Convolutional Neural Networks, combining and building upon state-of-the-art CNN models.

In addition, deep learning has been implemented extensively for building damage assessment utilizing various types of both aerial overhead and ground-based imagery [12, 13, 14, 15, 16, 17, 18, 19]. For example, Duarte et al. [15] used multi-resolution satellite imagery and CNNs for damage building assessment. By combining satellite imagery with overhead imagery obtained from both manned and unmanned aerial vehicles the authors were able to implement a CNN to classify a building in damaged or undamaged categories. The model, however, requires additional images that may be difficult to obtain and limits classification to a binary category. The proposed method makes no requirement on additinal data and provides a more nuanced damage assesment.

3 Dataset and Features

The xView2 [5] challenge provided the xBD dataset [4], which is the largest and highest quality public data set of expertly annotated high-resolution satellite imagery available online. The data consists of $850,736$ buildings in $22,000$ images spanning $45,361$ square kilometers for 19 disaster events, namely the Guatemala Volcano, Hurricane Harvey, Mexico Earthquake, Midwest Flooding, Palu Tsunami, and many others. The image data is paired with a labeled data set of corresponding information such as image uid, coordinates (latitude, longitude), building label, damage classification (0- No Damage, 1- Minor Damage, 2- Major Damage, 3- Destroyed), and wkt polygon shape for the buildings.

4 Methods

Let $S = \{s_1, s_2, ...s_K\}$ be a set of K satellite images, where each image can have a varying number n of individual building sub-images $\{x_1, x_2, ...x_n\} \in s_i$ of varying sizes as well. Further, let $\mathcal{G}(V, E)$ be a graph structure, where each node $v_i \in V$ corresponds to an x_i building sub-image and each edge $e_{ij} \in E$ represents whether buildings v_i and v_j occur within a predefined threshold distance in satellite image s_k. Each node is defined by m features such that $v \in \Re^m$. A model is proposed such that $f(x_i, v_i) \rightarrow y_i$, where $y_i \in \{0, 1, 2, 3\}$ corresponding to no damage, minor damage, major damage and destroyed categories, respectively.

The proposed model $f(\cdot, \cdot)$ is defined as a combination of a Convolutional Neural Network and a Graph Convolution Network [20]. The proposed model is formerly defined as

$$f(v_i, x_i) = F(W_k \cdot \sigma([W_{k-1} \cdot AGG(h_u^{k-1}, \forall u \in N(v_i)), C^{k-1}(x_i)]))) \tag{1}$$

, where u represents the nearest neighbors of v and $AGG(h_u^{k-1}, \forall u \in N(v_i))$ is the GCN component of the model while C^{k-1} represents the CNN component.

The GCN component is defined as in [20], adding an additional mask function to ensure that only activations of v_i and neighbor(s) u are concatenated with CNN feature maps corresponding to v_i. The mask function is simply the row of the adjacency matrix, $A_{v_i,:}$, pertaining to v_i.

$$AGG(h_u^{k-1}, \forall u \in N(v_i)) = A_{v_i,:} \cdot RELU(\hat{A} \cdot RELU(\hat{A} \cdot XW_{k-1})W_k) \tag{2}$$

, where $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ is the normalized adjacency matrix. The $\tilde{A} = A + I_N$ represents an adjacecny matrix of an undirected graph G with added self connections. The $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ represents the degree of each node within graph G. The adjacency matrix is defined by the graph structure created by the buildings in the immediate proximity of the building of interest. The spatial information provides additional key information on disaster impact within a geographic location. A masking function ensures the specific information of a building of interest is isolated to be combined with a CNN model.

The CNN component C^{k-1} is comprised of the Resnet-50 model [21] previous trained on imagenet data [22] with the lower half of the layers frozen to preserver low level features such as edges and shapes that can be used to identify buildings in an image. Three additional convolutional layers are added before the Resnet-50 model and three dense layers are added after as in [5] to fit the image data set better.

As shown in Figure 2, the two components are concatenated together into a dense layer σ allowing for non-linear aggregation and activation. A softmax layer F is subsequently utilized to provide a probability distribution for each category in the building damage assessment.

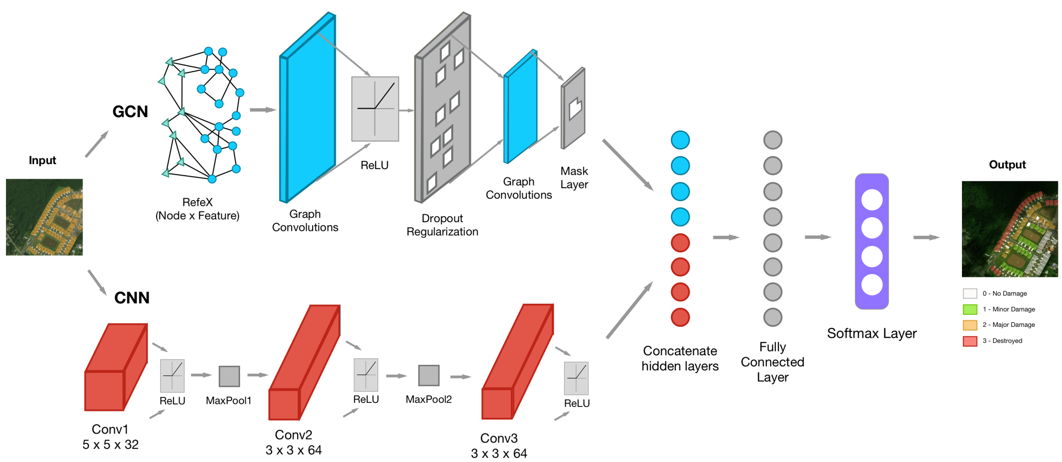

Figure 2: Our Hybrid GCN-CNN architecture

Allowing $f(u_i, x_i) = \hat{y}_i$, a categorical cross-entropy loss function defined as

$$\mathcal{L}_f = -\frac{1}{n}\sum_c \sum_i (y_{(i,c)} log(\hat{y}_{(i,c)}))$$

is then minimized accross the different categories c to train the corresponding weights of the previously described GCN and CNN models.

5 Experiments/Results/Discussion

The proposed model was trained and evaluated using $\approx 10,000$ satellite images with over $200,000$ pre-defined building sub-images. The localization of these building sub-images is outside the scope of the current project with the intent on further investigating in future works. Spatial information was used to define a graph sturcture $\mathcal{G}(U, E)$ in the following manner

1. For each satellite image s_k, define every building/structure sub-image x_i as a node $u_i \in U$ in $\mathcal{G}$.
2. Compute the centroid c of each u_i within each satellite image s_k.

Table 1: Model Performance for Damage Classification

Damage Type	Model	Precision	Recall	F1
No Damage	CNN Baseline	0.857	0.451	0.591
	ReFeX + XGBoost	0.664	0.866	0.752
	GCN + CNN	**0.931**	**0.943**	**0.937**
Minor Damage	CNN Baseline	0.073	**0.484**	0.127
	ReFeX + XGBoost	**0.485**	0.290	0.363
	GCN + CNN	0.447	0.411	**0.428**
Major Damage	CNN Baseline	0.242	0.093	0.134
	ReFeX + XGBoost	**0.642**	0.388	0.484
	GCN + CNN	0.549	**0.466**	**0.504**
Destroyed	CNN Baseline	0.420	0.623	0.502
	ReFeX + XGBoost	0.612	0.446	0.516
	GCN + CNN	**0.744**	**0.777**	**0.760**

3. $d = Euclid(c_i, c_j)$ of each u_i and u_j in satellite image s_k.
4. if $d < \theta$ for some pre-defined threshold θ then $e_{ij} = 1, \forall e_{i,j} \in E$.

After $\mathcal{G}$ was generated, spatial recursive feature extraction algorithm (ReFeX) [23] was ran yielding 38 features per node. A feature matrix was constructed from the node features and was used in the training of the GCN algorithm in conjunction with the adjacency matrix from $\mathcal{G}$.

A CNN was concurrently trained using pixel data from building sub-images x_i. Transfer learning was leveraged using a Resnet-50 model, trained on the imagenet [22] data set. The lower half of the Resnet-50 model weights were frozen to take advantage of the low level feature maps including edges that distinguish between different objects. The upper portion was trained to provide feature maps specific to damaged buildings. The proposed algorithm's parameter count was

- Total params: 124,284,320
- Trainable params: 121,795,616
- Non-trainable params: 2,488,704

To mitigate the unbalanced data sets, data points were weighted as an inverse of the number of data points available per class in the cost function to provided more significance to sparse classes. In addition, data augmentation was performed with each data point having a corresponding augmented data point. The augmentation focused on image properties that are inherent within satellite imagery data: 1) horizontal flip, 2) vertical flip, 3) width shift, 4) height shift. The model was ran for 100 epochs with hyper-parameter tuning of dropout rates of $\{0.3, 0.5, 0.7\}$ and batch size $\{32, 64, 300, 400\}$.

Batch Normalization was implemented on convolutional layers in order to improve training time while dropout was used on dense layers to mitigate over-fitting issue.

5.1 Baseline CNN Architecture

XView2's baseline classification architecture is built upon a Resnet-50 [21] CNN model, trained on imagenet [22] data with all layers frozen. Three trainable convolutional layers are concatenated with the Renet-50 results and passed to three trainable dense layers with a RELU output function to classify the building damage into one of the 4 classes:No Damage, Minor Damage, Major Damage, Destroyed.

5.2 Classical ML approach

In order to compare performance of various techniques, we also experimented with graph-based feature extraction and decision-tree based damage classification. As seen in figure 3, our ReFeX-XGBoost method combines graph-based recursive feature extraction [23, 24] and XGBoost [25] based node classification technique. XGBoost algorithm was used on node features & labels derived from graph $\mathcal{G}(V, E)$ to train a multi-class classifier. Our results show a modest improvement over the baseline and are presented in 1.

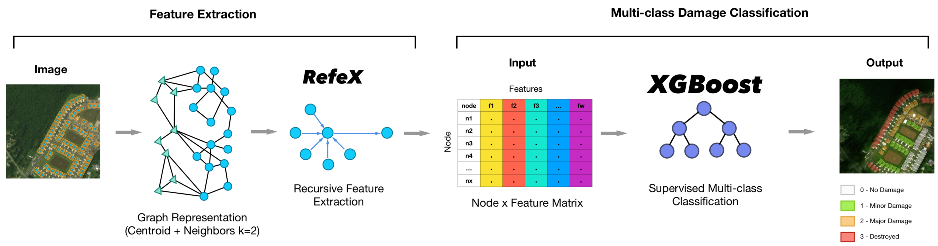

Figure 3: Our ReFeX-XGBoost architecture

5.3 Error Analysis

The proposed model performed quite well on identifying no damage buildings. Although the model performed better than the other models, it continues to lag in major, minor, and destroyed categories. This was especially true for major and minor damage, where it performed with an F1 score no better than 50%. The most plausible reason for this discrepancy is the unbalanced data, with the "no damage" category having at least a 10 to 1 ratio as compared to the other categories. Although data augmentation and class weighting mitigated some of this imbalance, it did not complete remove it. The collection of more data for the under represented categories should assist in overcoming this short fall.

Additionally, the localization process may not be ideally suited using image segmentation, since if there is no border readily visible, buildings will be grouped together impacting the pixel distribution patterns from one sub-image to the next. Rescaling is also done on the image to bring it to a size of $(128, 128, 3)$. Since each segmentation may be significantly different in size, some pixel information of a building may be removed while others are expanded, again disrupting pixel distribution. Using a object detection algorithm such as YOLO may reduce this issue and provide better sub-image quality.

6 Conclusion/Future Work

Building damage assessment is critical for appropriately allocating limited resources post natural disasters when time is of the essence for rescue efforts. Utilizing graph theory in conjunction with a powerful convolutional neural network (CNN), a model was proposed that out-performed baseline values previously reported using a baseline Resnet-50 architecture as well as a powerful classical machine learning algorithm called XGBoost. The proposed model improved upon recall, precision and F1 scores in each of the different damage categories by statistically significant margins.

Although there were significant improvements across the different categories, F1 scores remain relatively low for major and minor damage categories assessment. More research should be done to enhance localization of buildings by further exploring object detection as opposed to image segmentation models. This may also reduce the amount of preprocessing of segmented objects within an image, providing a much cleaner object to assess.

In addition, since it was demonstrated that graph structures provide useful information on building damage assessment, future research will focus on different methods of defining edges between building images that may prove to be useful in better assessing building damage for minor and major damage categories. Improving upon building damage assessment across each category can undoubtedly assist in rescue efforts when natural disasters occur.

7 Contributions & Acknowledgements

Throughout the course of this project, all three members of the team contributed equally in all aspects of the project including project planning & execution, code development, model training & debugging, result analysis, documentation, poster & video creation. We would like to thank CS230 Instructors and TA team for their help and valuable insights during the course of this project.

References

[1] Anthony Oliver-Smith and Anthony Oliver-Smith. *Disasters and Large-Scale Population Dislocations: International and National Responses*. Number February. 2018.

[2] Matteo Coronese, Francesco Lamperti, Klaus Keller, Francesca Chiaromonte, and Andrea Roventini. Evidence for sharp increase in the economic damages of extreme natural disasters. *Proceedings of the National Academy of Sciences of the United States of America*, 116(43):21450–21455, 2019.

[3] Anthony D. Redmond, Timothy J. O'Dempsey, and Bertrand Taithe. Disasters and a register for foreign medical teams. *The Lancet*, 377(9771):1054–1055, 2011.

[4] Ritwik Gupta, Richard Hosfelt, Sandra Sajeev, Nirav Patel, Bryce Goodman, Jigar Doshi, Eric Heim, Howie Choset, and Matthew Gaston. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. 2019.

[5] xView2: Assess Building Damage, 2019.

[6] Barak Oshri, Xiao Chen, Marshall Burke, Annie Hu, Pascaline Dupas, David Lobell, Peter Adelson, Jeremy Weinstein, and Stefano Ermon. Infrastructure quality assessment in Africa using satellite imagery and deep learning. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 616–625, 2018.

[7] Benjamin Bischke, Patrick Helber, Joachim Folz, Damian Borth, and Andreas Dengel. Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks. *Proceedings - International Conference on Image Processing, ICIP*, 2019-Septe:1480–1484, 2019.

[8] Saikat Basu, Sangram Ganguly, Supratik Mukhopadhyay, Robert DiBiano, Manohar Karki, and Ramakrishna Nemani. DeepSat - A learning framework for satellite imagery. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 03-06-Nove:1–22, 2015.

[9] Lu Zhang, Zhenwei Shi, and Jun Wu. Surrounding Features for High-Resolution Optical Satellite Imagery. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10):1–15, 2015.

[10] Fan Zhang, Bo Du, and Liangpei Zhang. A multi-task convolutional neural network for mega-city analysis using very high resolution satellite imagery and geospatial data. pages 1–37, 2017.

[11] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 14(5):778–782, 2017.

[12] Anand Vetrivel, Markus Gerke, Norman Kerle, Francesco Nex, and George Vosselman. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140:45–59, 2018.

[13] M Matsuoka, Y Ishii, N Maki, K Horie, S Tanaka, R Nakamura, S Hikosaka, T Imaizumi, A Fujita, and R Ito. Damaged Building Recognition Using Deep Learning With Photos Taken After the Kobe Earthquake. *11th U.S. National Conference on Earthquake Engineering*, 2018.

[14] Kasthurirangan Gopalakrishnan, Hoda Gholami, Akash Vidyadharan, and Ankit Agrawal. Crack Damage Detection in Unmanned Aerial Vehicle Images of Civil Infrastructure Using Pre-Trained Deep Learning Model. *International Journal for Traffic and Transport Engineering*, 8(1):1–14, 2018.

[15] D. Duarte, F. Nex, N. Kerle, and G. Vosselman. SATELLITE IMAGE CLASSIFICATION of BUILDING DAMAGES USING AIRBORNE and SATELLITE IMAGE SAMPLES in A DEEP LEARNING APPROACH. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(2):89–96, 2018.

[16] Yundong Li, Shi Ye, and Ivan Bartoli. Semisupervised classification of hurricane damage from postevent aerial imagery using deep learning. *Journal of Applied Remote Sensing*, 12(04):1, 2018.

[17] Yanbing Bai, Chang Gao, Sameer Singh, Magaly Koch, Bruno Adriano, Erick Mas, and Shunichi Koshimura. A framework of rapid regional tsunami damage recognition from post-event terraSAR-X imagery using deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15(1):43–47, 2018.

[18] Yang Yu, Chaoyue Wang, Xiaoyu Gu, and Jianchun Li. A novel deep learning-based method for damage identification of smart building structures. *Structural Health Monitoring*, 18(1):143–163, 2019.

[19] Karoon Rashedi Nia and Greg Mori. Building damage assessment using deep learning and ground-level image data. *Proceedings - 2017 14th Conference on Computer and Robot Vision, CRV 2017*, 2018-Janua:95–102, 2018.

[20] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, pages 1–14, 2017.

[21] Songtao Wu, Shenghua Zhong, and Yan Liu. Deep residual learning for image Recognition. *Multimedia Tools and Applications*, pages 1–17, 2017.

[22] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

[23] It's who you know: Graph mining using recursive structural features, 2012.

[24] RolX: Structural Role Extraction & Mining in Large Graphs, 2012.

[25] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.