Ryerson University | Data Science & Analytics Graduate Studies

2018-2019 M.Sc. in Data Science and Analytics

Using Deep Learning and Satellite Imagery to Predict Road Safety

Claudia Choi

Supervisor: Dr. Tony Hernandez

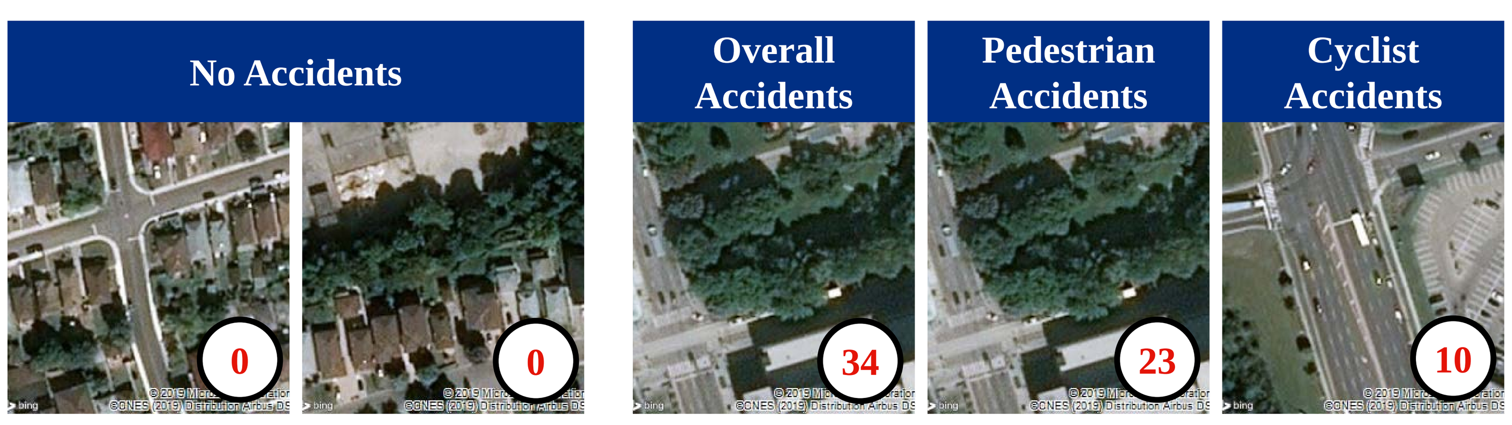
Objectives

1. A demonstration of the efficacy of using deep learning and satellite imagery to predict city-scale road safety
2. Constructing a deep model for predicting overall road safety for the City of Toronto
3. Fine-tuning deep models specifically for predicting pedestrian and cyclist safety in the City of Toronto

Background

In 2018…

Road Accidents in Toronto



Sample satellite images of geographic regions in the City of Toronto. The regions on the left have had no road accidents, while the regions on the right represent the areas with the highest number of accidents in each accident group between 2008 and 2018.

Previous Work

Exploring City-Scale Issues Using Deep Learning and Satellite Imagery

- Education
- Physical & Mental Health
- Neighborhood Crime

Combining Satellite Imagery and Open Data to Map Road Safety

A. Najjar, S. Kaneko and Y. Miyanaga, 2017

78% New York City

73% City of Denver

Data Sources

KSI
Toronto Police Services

Satellite Images
Bing Maps Imagery API

Deep Learning Model

Convolution Neural Network

ResNet50 Framework

Pre-trained on ImageNet

Implementation

Data Preprocessing

- Divide map of Toronto into 150m x 150m square regions
- Crawl satellite images from Bing Maps Imagery API
- Determine number of safety levels using k-means
- Label satellite images for each category:
  - *Accident Groups:* Overall road accidents, pedestrian accidents, cyclist accidents
  - *Labels:* 'highly safe' - 0, 1, 2, 3 - 'highly risky'

Balancing Imbalanced Data

Models trained on the Overall Road Accidents dataset

Class Balancing Approach	Accuracy Score	Macro F1 Score
Class Weights Approach	94%	32%
Undersampling Approach	69%	25%
Oversampling (SMOTE) Approach	95%	26%
SMOTE and Undersampling Approach	93%	35%

Experiments

- CNN with ResNet50 framework pre-trained on ImageNet
- SMOTE and Undersampling class balancing approach
- Cross-validated on three 80%/20% data splits

Results

Overall Road Accidents

92% Accuracy Score

36% Macro F1 Score

Individual Classes

	0	1	2	3
Distribution	54937	2049	732	136
Accuracy Score	95%	26%	18%	5%

Pedestrian Accidents

95% Accuracy Score

31% Macro F1 Score

Individual Classes

	0	1	2	3
Distribution	56313	1089	391	61
Accuracy Score	97%	19%	6%	3%

Cyclist Accidents

99% Accuracy Score

26% Macro F1 Score

Individual Classes

	0	1	2	3
Distribution	57353	326	138	37
Accuracy Score	99%	4%	0%	0%

Summary of Performance Results

Model	Accuracy Score	Macro F1 Score	Individual Class Accuracy Scores			
			0	1	2	3
Overall Road Accidents	92%	36%	95%	26%	18%	5%
Pedestrian Accidents	95%	31%	97%	19%	6%	3%
Cyclist Accidents	99%	26%	99%	4%	0%	0%

Conclusions

Observable Built Environment as a Predictor for Highly Safe Regions

Able to identify 'highly safe' regions for overall, pedestrian and cyclist accidents
Poor results for 'highly risky' regions

Limitations

Majority Class Bias
Low macro F1 scores & high accuracy scores

Highly Imbalanced and Limited Instances for Minority Classes
Model may improve with more data

Future Work

Improving on the prediction accuracy of 'highly risky' areas
Models that combine both satellite imagery and extraneous variables

Explore methods to increase sample of 'highly risky' class instances
Blending data from similar cities to train models