TorontoCity: Seeing the World with a Million Eyes

UNIVERSITY OF TORONTO

Authors









Shenlong Wang, Min Bai, Gellert Mattyus, Hang Chu, Wenjie Luo, Bin Yang
Justin Liang, Joel Cheverie, Sanja Fidler, Raquel Urtasun

UNIVERSITY OF TORONTO

* Project Completed by Summer 2016

Why Toronto?

The best place to live in the world*

- Toronto 4

*According to 2015 Global Liveability Ranking

Why Toronto?

The best place to live in the world*

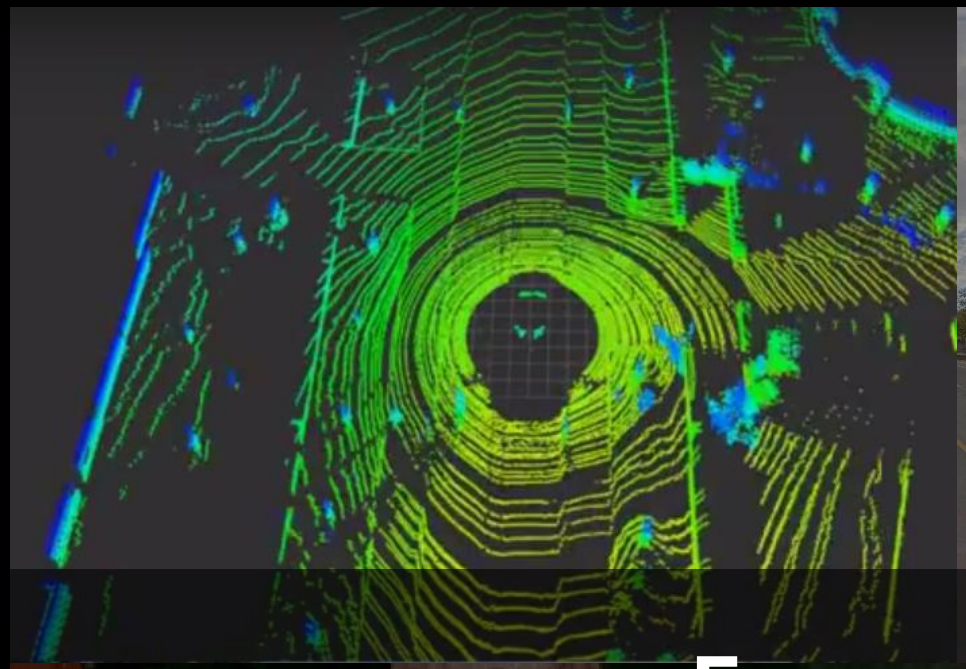
- Toronto 4

The places you are working at:

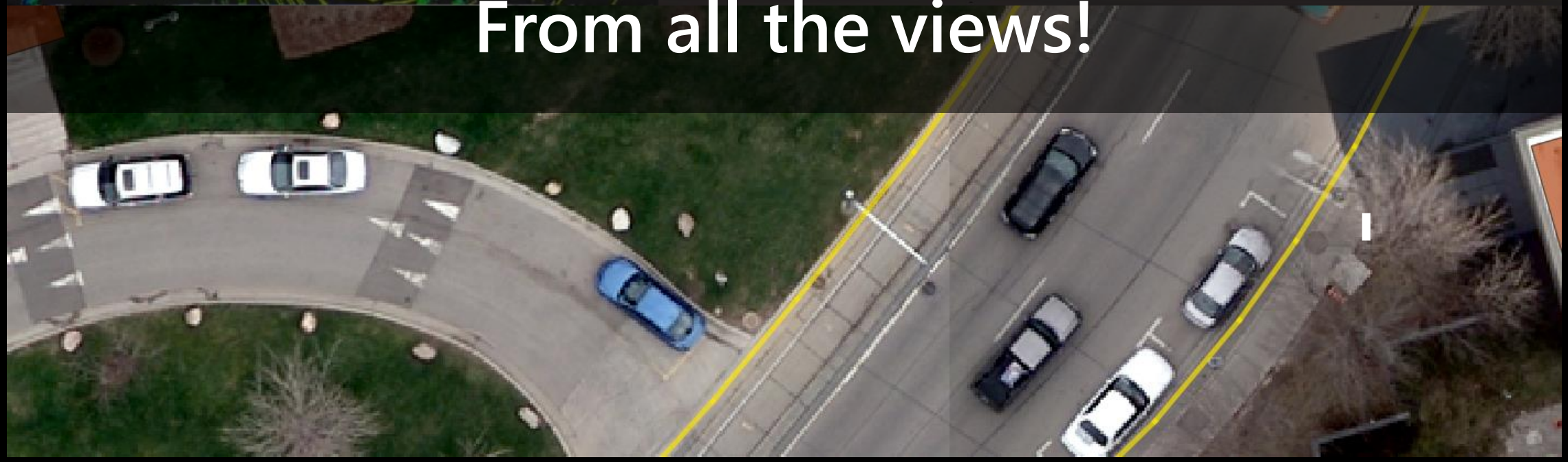
- Boston 36
- Pittsburgh 39
- San Francisco 49
- Los Angeles 51

*According to 2015 Global Liveability Ranking

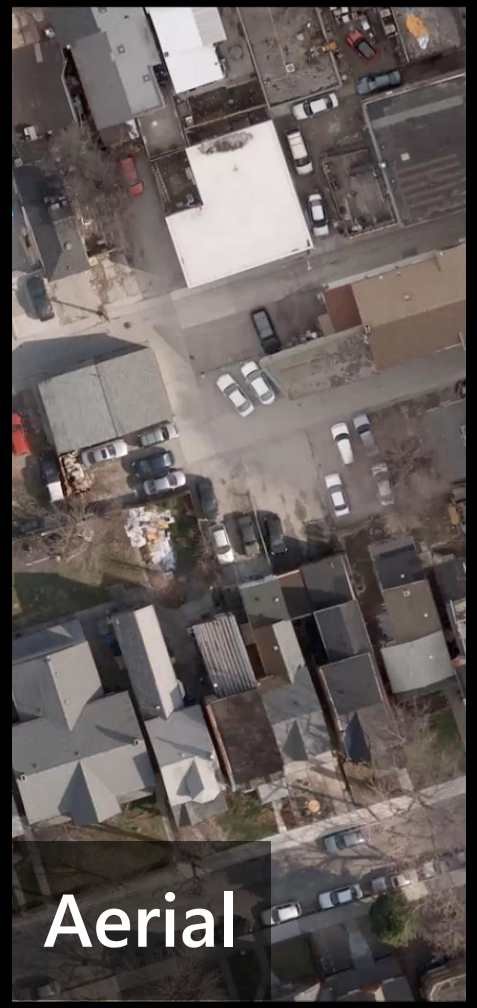
A dataset over 700 km^2 region!



From all the views!

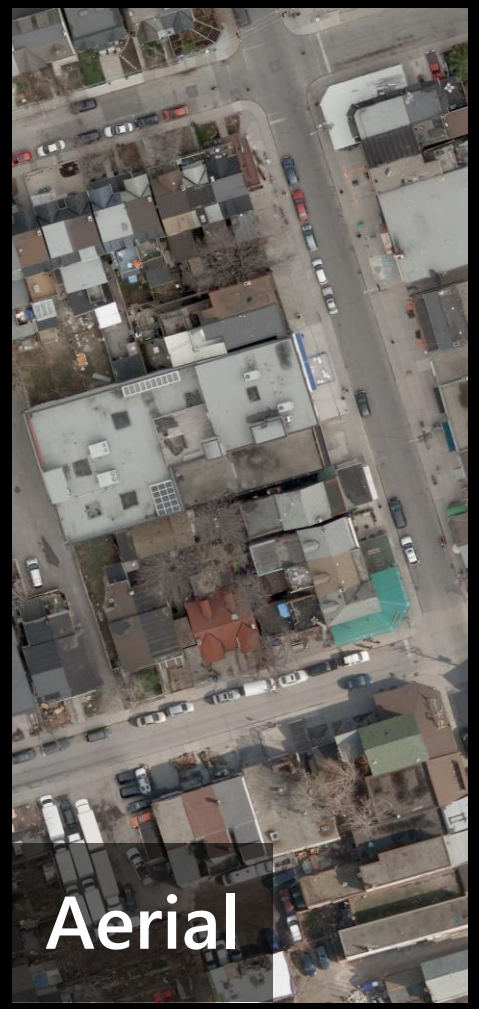


Dataset



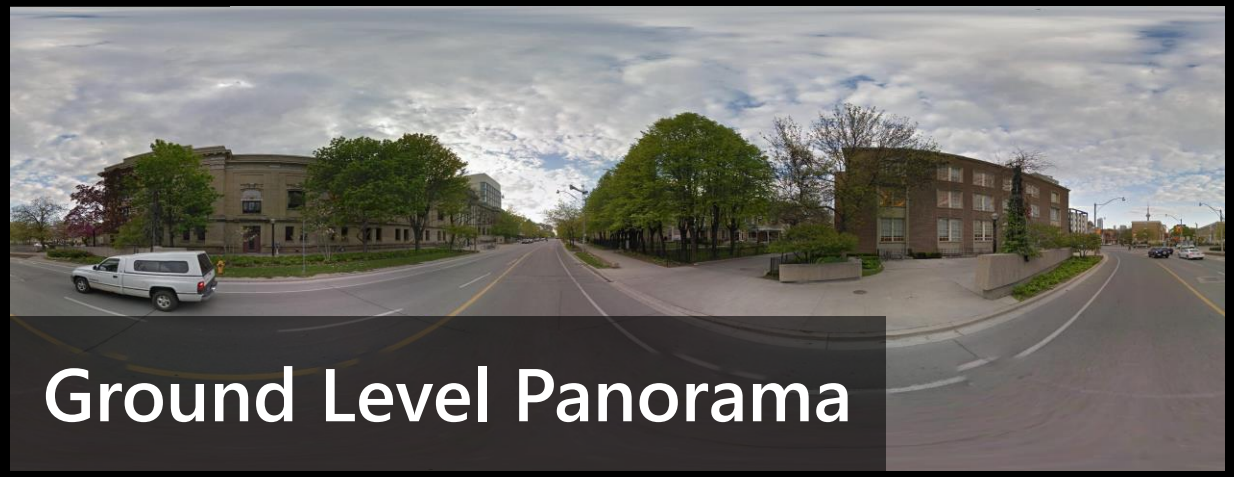
Data Source

Dataset



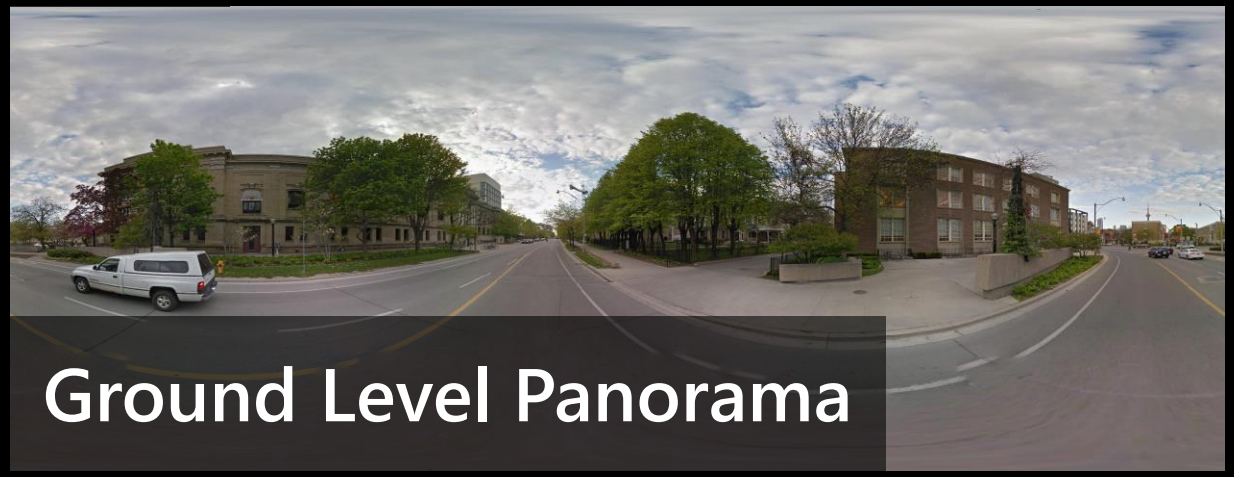
Data Source

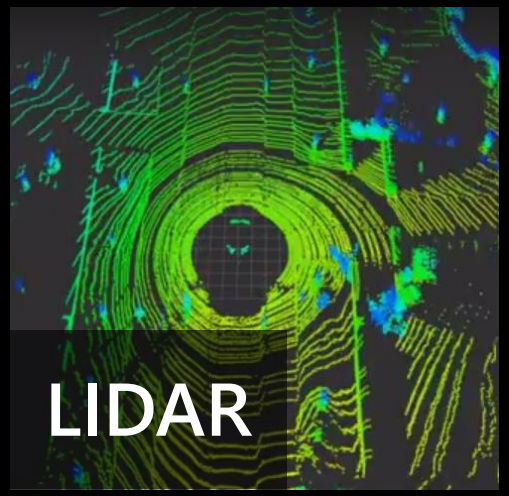
Dataset

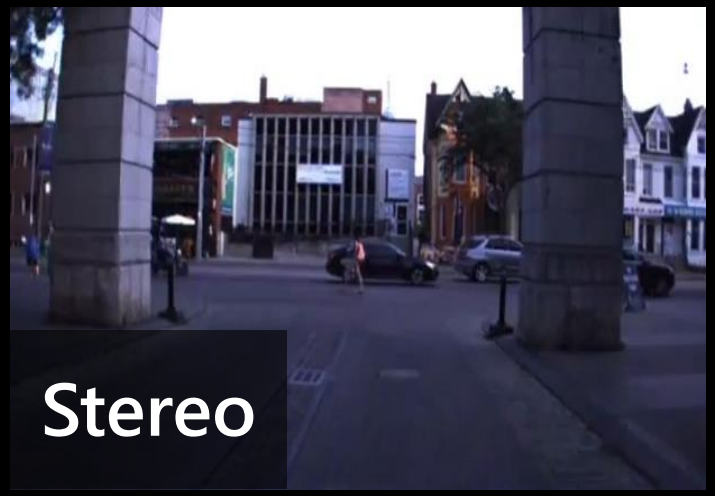


Data Source

Dataset

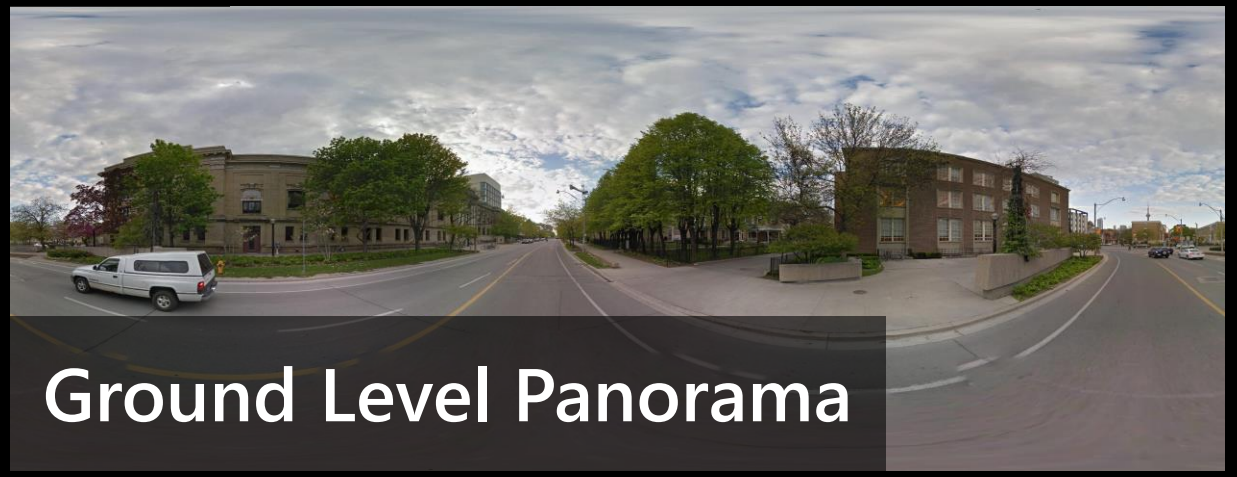


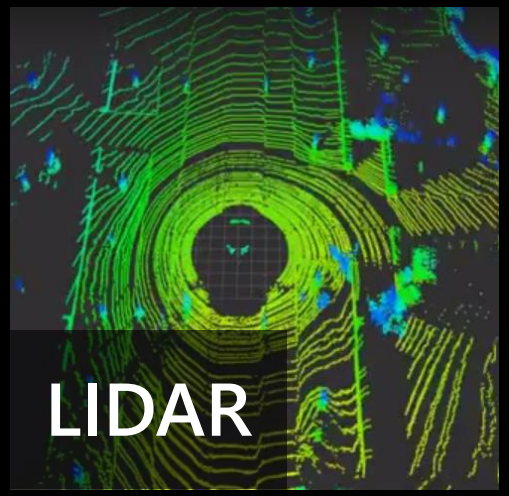




Data Source

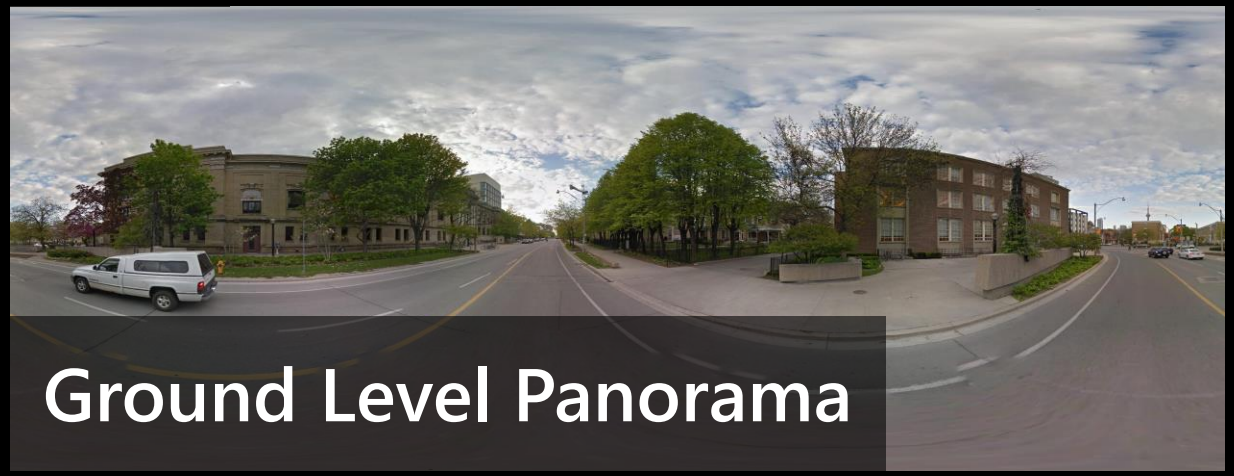
Dataset

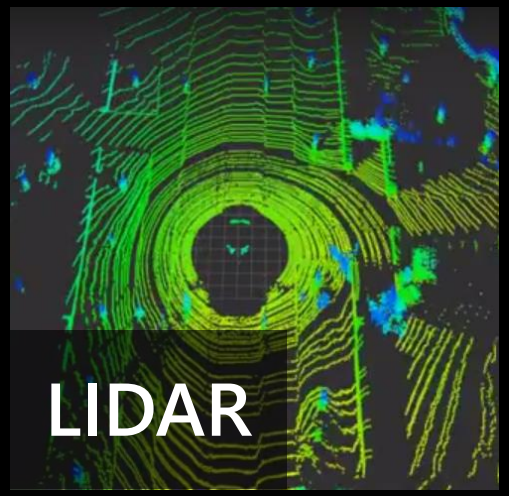


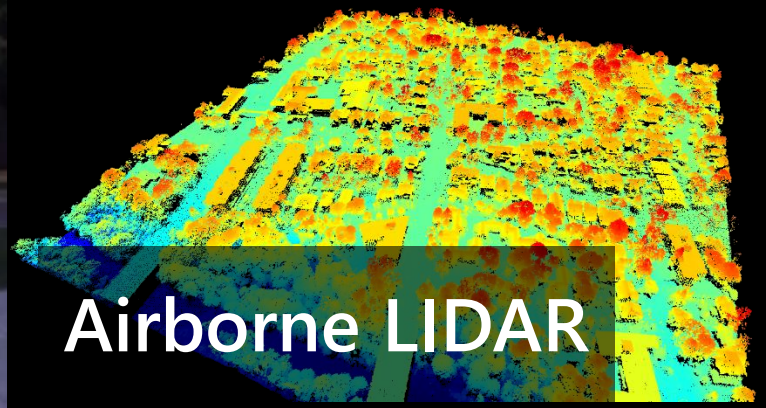


Data Source

Dataset







Data Source

Why we need this?

- Mapping for Autonomous Driving
- Smart City
- Benchmarking:
  - Large-Scale Machine Learning / Deep Learning
  - 3D Vision
  - Remote Sensing
  - Robotics

Source: Here 360

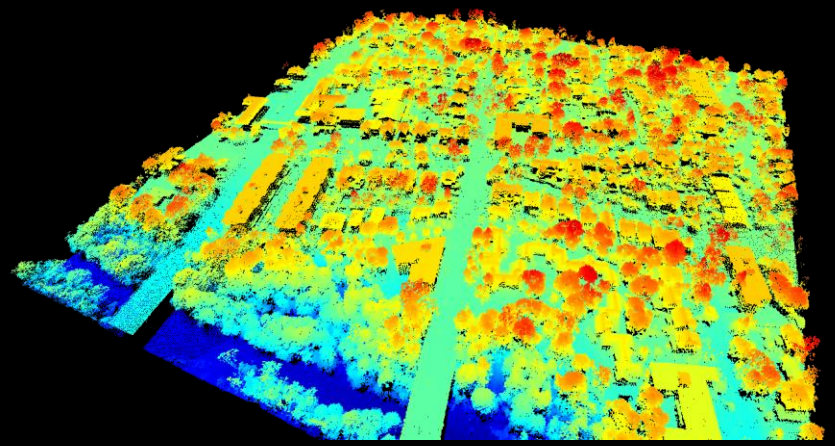
Why we need this?

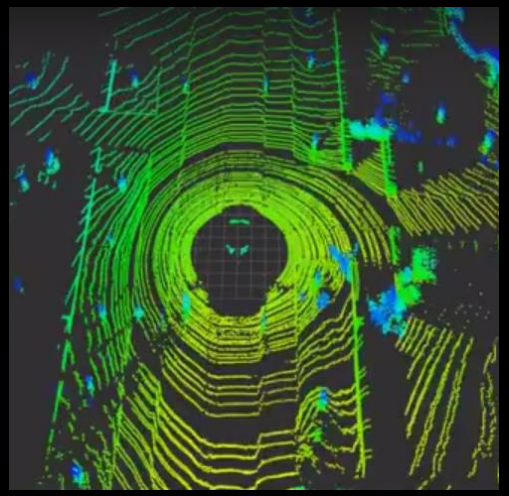
- Mapping for Autonomous Driving
- Smart City
- Benchmarking:
  - Large-Scale Machine Learning / Deep Learning
  - 3D Vision
  - Remote Sensing
  - Robotics

Source: Toronto SmartCity Summit

Why we need this?

- Mapping for Autonomous Driving
- Smart City
- Benchmarking:
  - Large-Scale Machine Learning / Deep Learning
  - 3D Vision
  - Remote Sensing
  - Robotics





Annotations

- Manual annotation? Impossible!
  - Suppose each 500x500 image costs $1 to annotate pixel-wise labels, we need to pay $11M to create ground-truth only for the aerial images.

Annotations

- Manual annotation? Impossible!
  - Suppose each 500x500 image costs $1 to annotate pixel-wise labels, we need to pay $11M to create ground-truth only for the aerial images.

Annotations

- Manual annotation? Impossible!
  - Suppose each 500x500 image costs $1 to annotate pixel-wise labels, we need to pay $11M to create ground-truth only for the aerial images.

- However, humans already collect rich knowledge about the world!

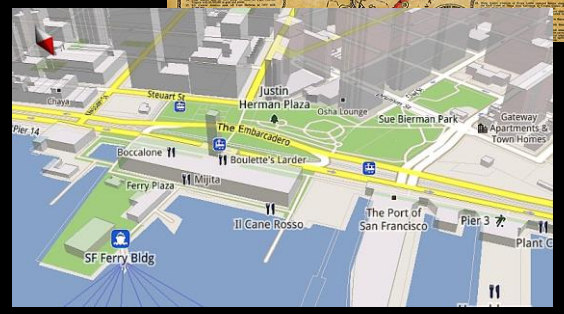
Annotations

I'm not as rich as Jensen ☹

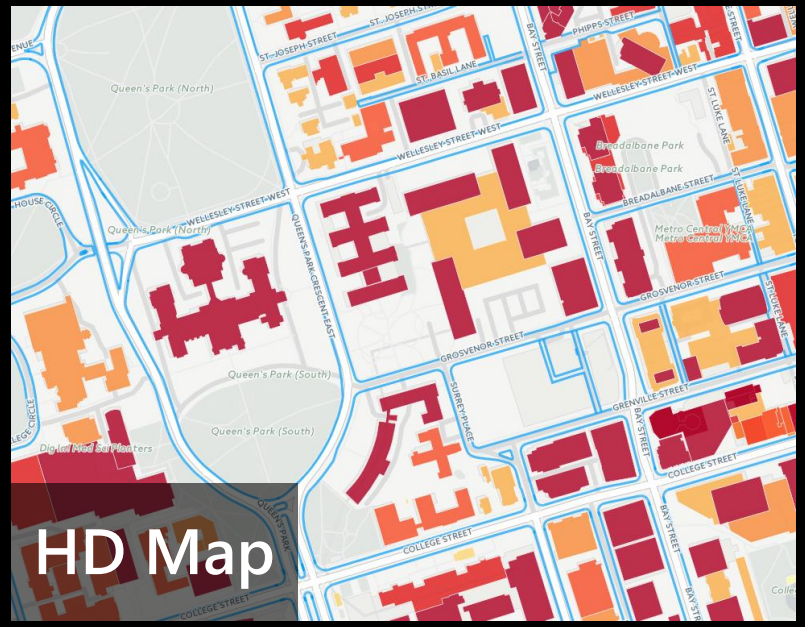
- Manual annotation? Impossible!
  - Suppose each 500x500 image costs $1 to annotate pixel-wise labels, we need to pay $1139200 to create ground-truth only for the aerial images.

- Humans already collect rich knowledge about the world!

Use maps!

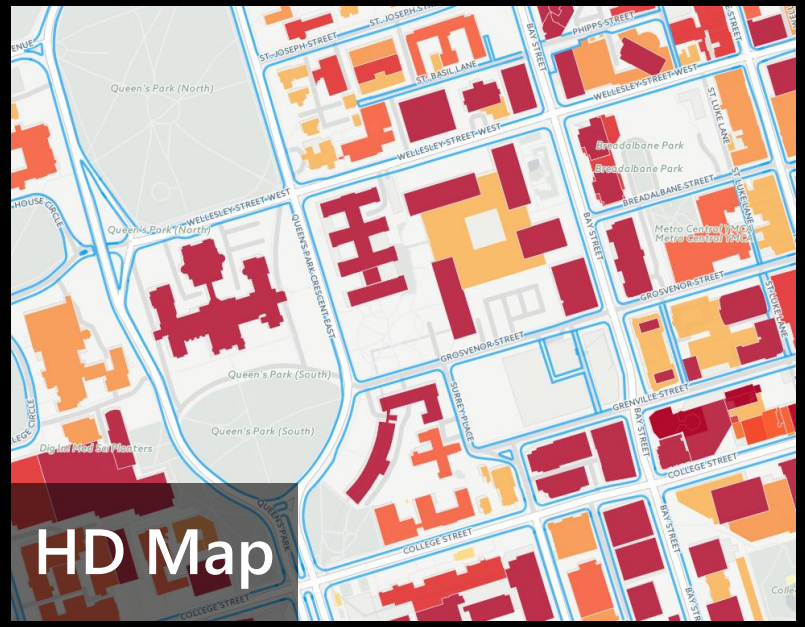


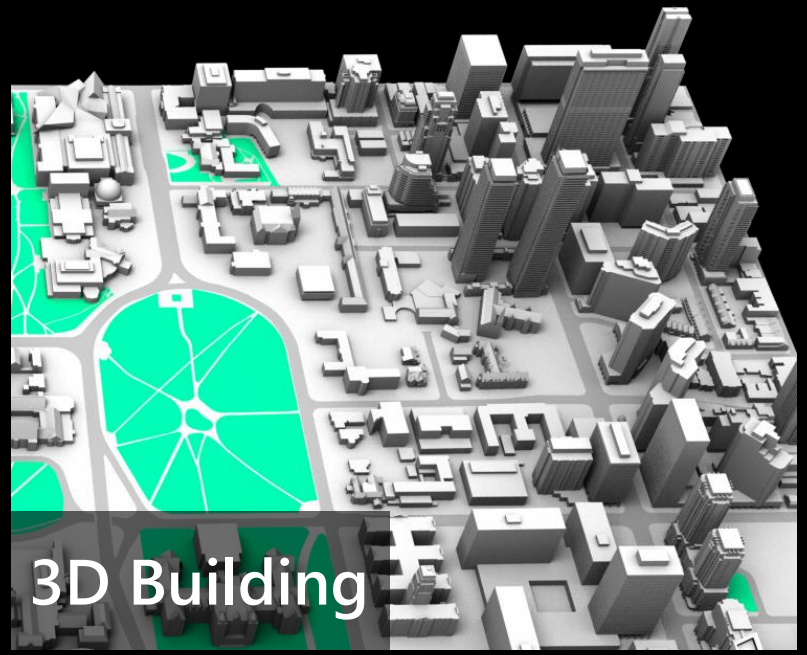
Map as Annotations



Maps

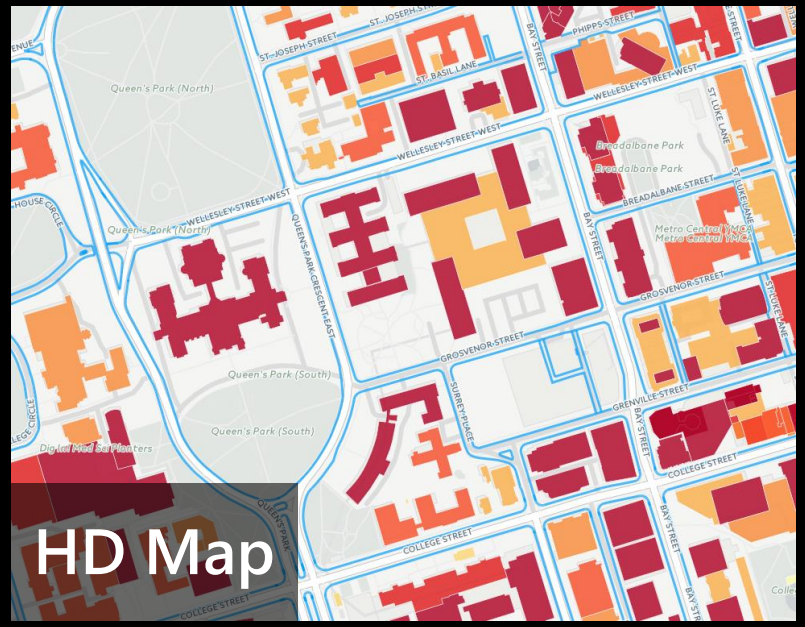
Map as Annotations

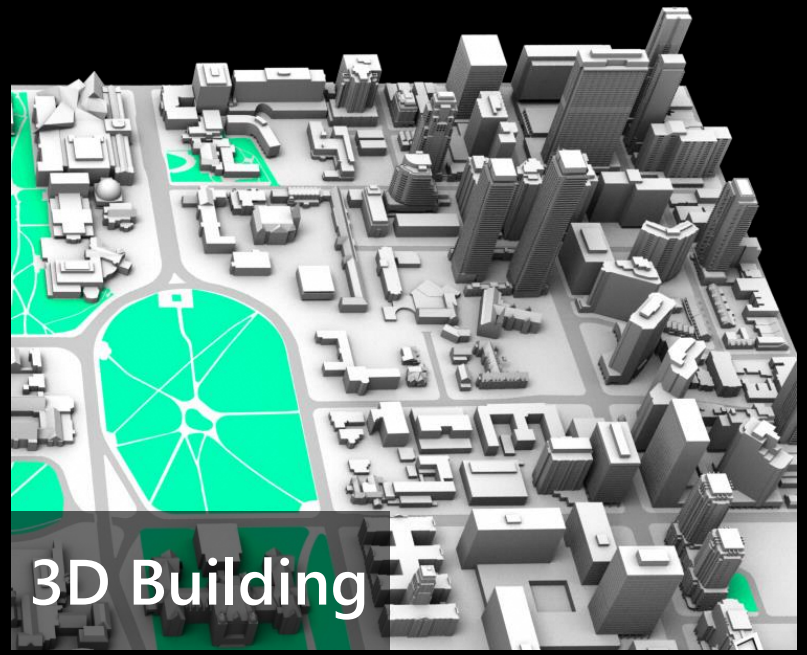




Maps

Map as Annotations





Maps


Together, the rich sources of data enable a plethora of exciting tasks!

Building Footprint Extraction

Road Curb and Centerline Extraction

Building Instance Segmentation

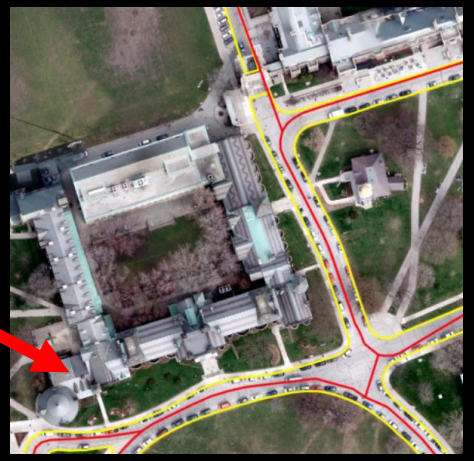

Zoning Prediction
Commercial
Residential
Institutional

Technical Difficulties

Mis-alignment and Data Noise

Aerial-ground images mis-alignment from raw GPS location data

Road centerline is shifted



Building's shape/location is not accurate

Data Pre-processing and Alignment

Appearance based Ground-aerial Alignment

Before Alignment

After Alignment

Data Pre-processing and Alignment

Instance-wise Aerial-map Alignment

Before alignment

Data Pre-processing and Alignment

Instance-wise Aerial-map Alignment

After alignment

Data Pre-processing and Alignment

Robust Road Surface Generation

Input Road Curb and Centreline (Noisy)

Polygonized Road Surface

Pilot Study with Neural Networks

Building Contour and Road Curb/Centerline Extraction

GT

ResNet

Pilot Study with Neural Networks

Semantic Segmentation

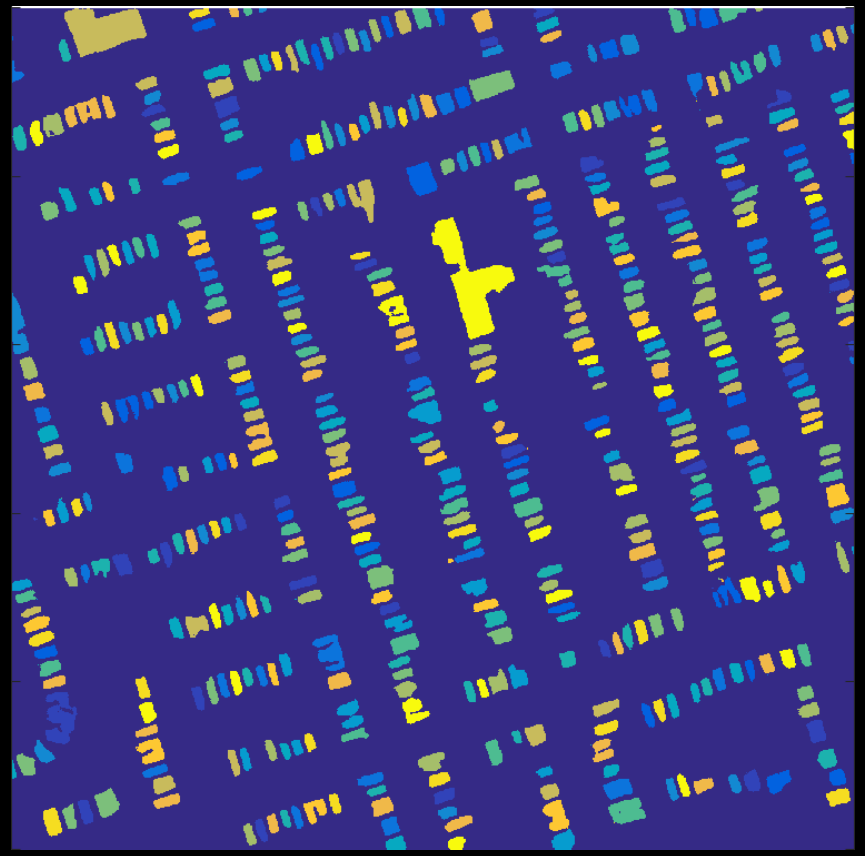
Method	Road	Building	Mean
FCN	74.94%	73.88%	74.41%
ResNet-56	82.72%	78.80%	80.76%

Metric: Intersection-over-union (IOU), higher is better

Pilot Study with Neural Networks

Building Instance Segmentation

Input



DWT

Pilot Study with Neural Networks

Building Instance Segmentation

Method	Weighted Coverage	Average Precision	Recall-50%	Precision-50%
FCN	41.92%	11.37%	21.50%	36.00%
ResNet-56	40.65%	12.13%	18.90%	45.36%
Deep Watershed Transform	56.22%	21.22%	67.16%	63.67%

Metric: Weighted Coverage, AP, Precision-50%, Recall-50%, higher is better

Pilot Study with Neural Networks

Building Instance Segmentation

Method	Weighted Coverage	Average Precision	Recall-50%	Precision-50%
FCN	41.92%	11.37%	21.50%	36.00%
ResNet-56	40.65%	12.13%	18.90%	45.36%
Deep Watershed Transform	56.22%	21.22%	67.16%	63.67%

Join the other talk today to know more about the deep watershed instance segmentation:
Wednesday, May 10, 4:00 PM - 4:25 PM – Room 210G

Pilot Study with Neural Networks

Ground-view Road Segmentation

True Positive: Yellow; False Negative: Green; False Positive: Red

Pilot Study with Neural Networks

Ground-view Road Segmentation

Method	Non-Road IOU	Road IOU	Mean IOU
FCN	97.3%	95.8%	96.5%
ResNet-56	97.8%	96.6%	97.2%

Metric: Intersection-over-Union, higher is better

Pilot Study with Neural Networks

Ground-view Zoning Classification

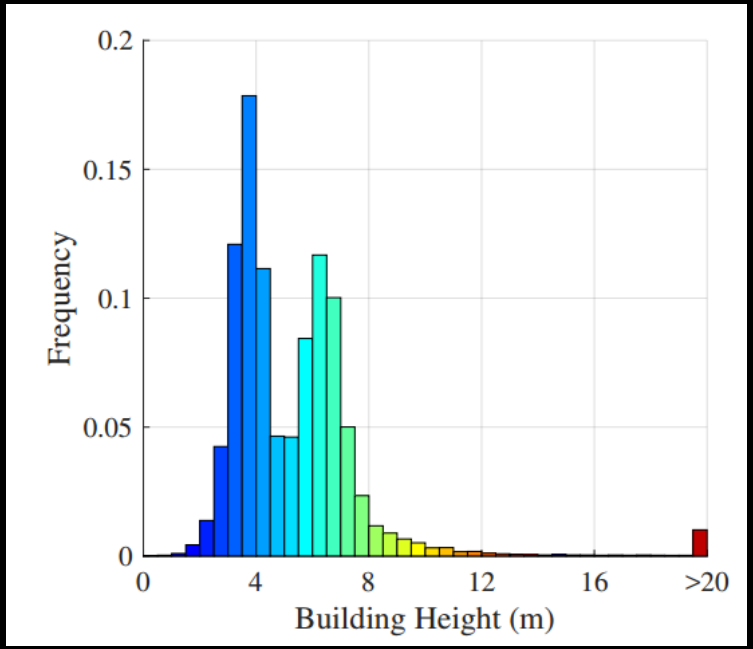
	Top-1 Accuracy	
Method	From-Scratch	Pre-trained from ImageNet
AlexNet	66.48%	75.49
GoogLeNet	75.08%	77.95%
ResNet	75.65%	79.33%

Metric: Top-1 Accuracy, higher is better

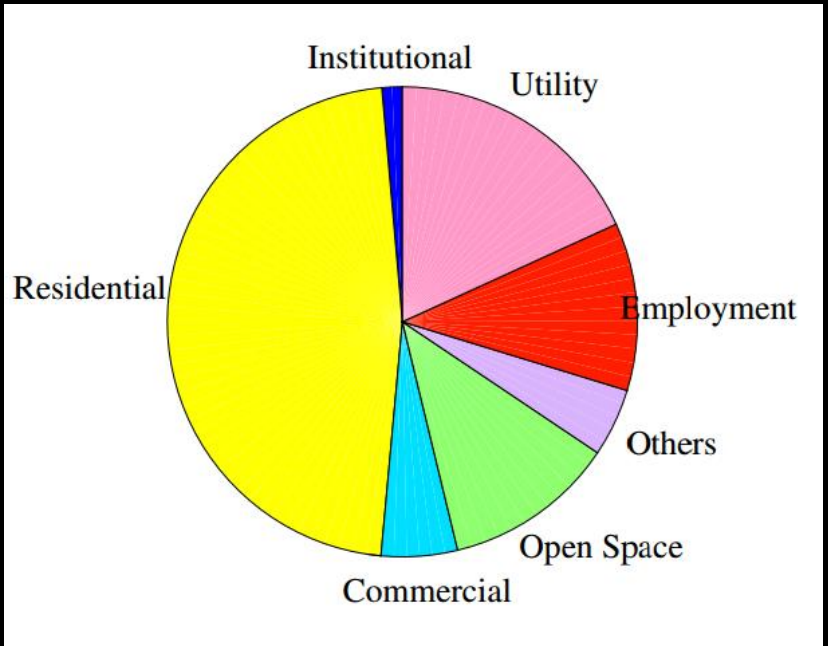
Statistics

- # of buildings: 397846
- Total area: 712.5 km^2
- Total length of road: 8439 km

Statistics



Building height distribution



Zoning type distribution

Conclusion

- We propose a large dataset with from different views and sensors
- Maps are used to create GT annotations
- In future we have many more exciting tasks to come
- Check our paper for more details: https://arxiv.org/abs/1612.00423
- Data available soon. Stay tuned and welcome to over-fit

Join the other talk today to know more about the deep watershed instance segmentation:
Wednesday, May 10, 4:00 PM - 4:25 PM – Room 210G