TorontoCity: Seeing the World with a Million Eyes
Authors

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

* 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 km² region!
From all the views!
Dataset

Aerial

Ground Level Panorama

Data Source
Dataset

Ground Level Panorama

Aerial

LIDAR

Data Source
Dataset

Ground Level Panorama

Aerial  LIDAR  Stereo

Data Source
Dataset

Aerial

Ground Level Panorama

LIDAR

Drone

Stereo

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.

I’m not as rich as Jensen 😊
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

• 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!

I’m not as rich as Jensen 😊
Map as Annotations
Map as Annotations

HD Map

3D Building

Maps
Map as Annotations

HD Map

3D Building

Meta Data
Together, the rich sources of data enable a plethora of exciting tasks!
Building Footprint Extraction
Road Curb and Centerline Extraction
Building Instance Segmentation
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

<table>
<thead>
<tr>
<th>Method</th>
<th>Road</th>
<th>Building</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN</td>
<td>74.94%</td>
<td>73.88%</td>
<td>74.41%</td>
</tr>
<tr>
<td>ResNet-56</td>
<td>82.72%</td>
<td>78.80%</td>
<td>80.76%</td>
</tr>
</tbody>
</table>

Metric: Intersection-over-union (IOU), higher is better
Pilot Study with Neural Networks
Building Instance Segmentation
Pilot Study with Neural Networks
Building Instance Segmentation

<table>
<thead>
<tr>
<th>Method</th>
<th>Weighted Coverage</th>
<th>Average Precision</th>
<th>Recall-50%</th>
<th>Precision-50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN</td>
<td>41.92%</td>
<td>11.37%</td>
<td>21.50%</td>
<td>36.00%</td>
</tr>
<tr>
<td>ResNet-56</td>
<td>40.65%</td>
<td>12.13%</td>
<td>18.90%</td>
<td>45.36%</td>
</tr>
<tr>
<td>Deep Watershed Transform</td>
<td>56.22%</td>
<td>21.22%</td>
<td>67.16%</td>
<td>63.67%</td>
</tr>
</tbody>
</table>

Metric: Weighted Coverage, AP, Precision-50%, Recall-50%, higher is better
Pilot Study with Neural Networks
Building Instance Segmentation

<table>
<thead>
<tr>
<th>Method</th>
<th>Weighted Coverage</th>
<th>Average Precision</th>
<th>Recall-50%</th>
<th>Precision-50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN</td>
<td>41.92%</td>
<td>11.37%</td>
<td>21.50%</td>
<td>36.00%</td>
</tr>
<tr>
<td>ResNet-56</td>
<td>40.65%</td>
<td>12.13%</td>
<td>18.90%</td>
<td>45.36%</td>
</tr>
<tr>
<td>Deep Watershed Transform</td>
<td>56.22%</td>
<td>21.22%</td>
<td>67.16%</td>
<td>63.67%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Non-Road IOU</th>
<th>Road IOU</th>
<th>Mean IOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN</td>
<td>97.3%</td>
<td>95.8%</td>
<td>96.5%</td>
</tr>
<tr>
<td>ResNet-56</td>
<td>97.8%</td>
<td>96.6%</td>
<td>97.2%</td>
</tr>
</tbody>
</table>

Metric: Intersection-over-Union, higher is better
# Pilot Study with Neural Networks

**Ground-view Zoning Classification**

<table>
<thead>
<tr>
<th>Method</th>
<th>From-Scratch</th>
<th>Pre-trained from ImageNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td>66.48%</td>
<td>75.49%</td>
</tr>
<tr>
<td>GoogLeNet</td>
<td>75.08%</td>
<td>77.95%</td>
</tr>
<tr>
<td>ResNet</td>
<td>75.65%</td>
<td>79.33%</td>
</tr>
</tbody>
</table>

**Metric:** Top-1 Accuracy, higher is better
Statistics

- # of buildings: 397846
- Total area: 712.5 km²
- 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