VISUAL 3D UNDERSTANDING OF MIXED TRAFFIC IN BUSY INTERSECTIONS

James Elder

Centre for Vision Research, York University

TEAM

ITS Innovation Lab
Prof. Abdulhai, PI, Director

University of Toronto
Prof. Waslander, Associate Director, CV
Prof. Sanner, Associate Director, AI

Dr. Taha, ITS Program Manager

Dr. Carrara, RL Expert

Dr. Smirnov, ML Expert

SOW1	SOW2	SOW4	SOW5	SOW6	SOW7	SOW8
Prof. Waslander Prof. Abdulhai	Prof. Elder Prof. Abdulhai	Prof. Sanner Prof. Abdulhai	Prof. Abdulhai Prof. Sanner	Prof. Shalaby	Prof. Shalaby Prof. Abdulhai	Prof. Abdulhai
A. Mahmoud	Dr. Hou	T. Ting	X. Wang	Dr. Srikukenthiran	W. Hu	L. Elmorshedy
J. Willes	Dr. Claudio	X. Li	P. Jaggi	F. Elsaid	K. Othman	O. Elsamadisy
C Reading	P. Rao	X. Wang	Z. Yan	C. White		Y. Ai
J. Carrillo	K. Liu			Y. Liu		
Dataset Engineer	Y. Qian					

Markham CRC
Dr. Ma, Team Leader

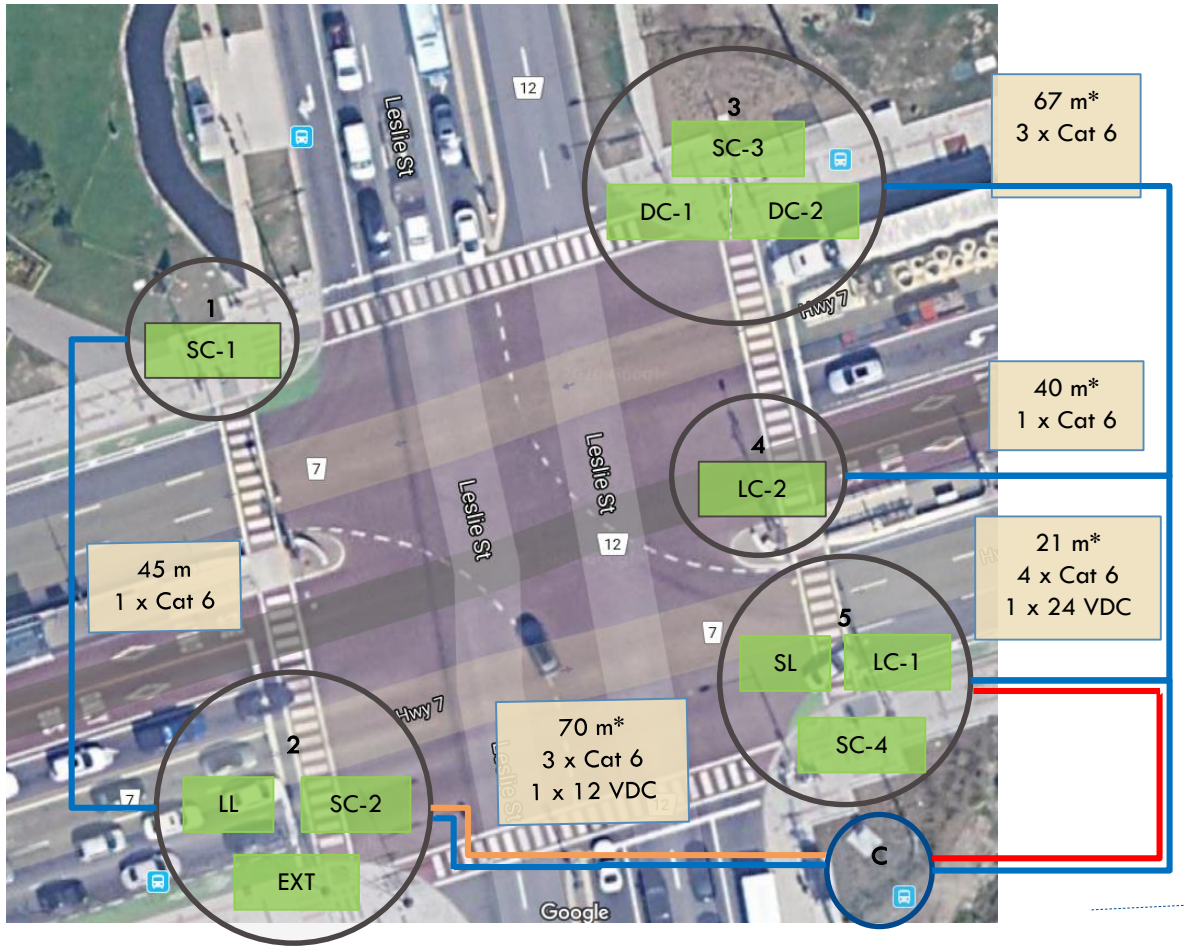
- Dr. Shabestary
- Dr. Fazlali
- Dr. Duanshun
- Dr. Chen
- Dr. Zhang
- A. Norouzi

Research Objectives

- Core Objective
  - Real-time 3D detection, classification, accurate geo-location and tracking of all road users
    - Cars, trucks, buses, pedestrians, bicycles,…
    - Diverse conditions: day/night, sun, cloud, rain, snow,…
- Derived Applications and Use Cases
  - Traffic counting and trajectory classification
  - Accurate speed measurement
  - Traffic anomaly and incident detection
  - Detection of near-misses
  - 3D visualization (digital intersection)
  - Automatic identification of and attention to vulnerable road users

Field Site: Highway 7 & Leslie

Complete Hardware Layout



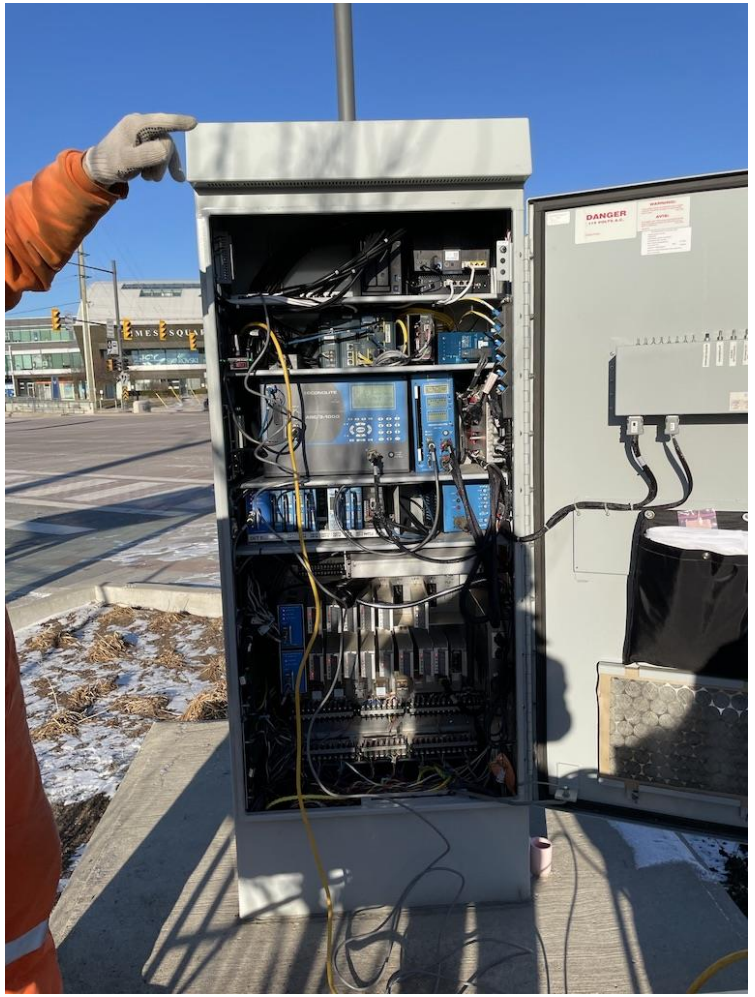
Cat 6

24 VDC

12 VDC

RT NAS

ES PC

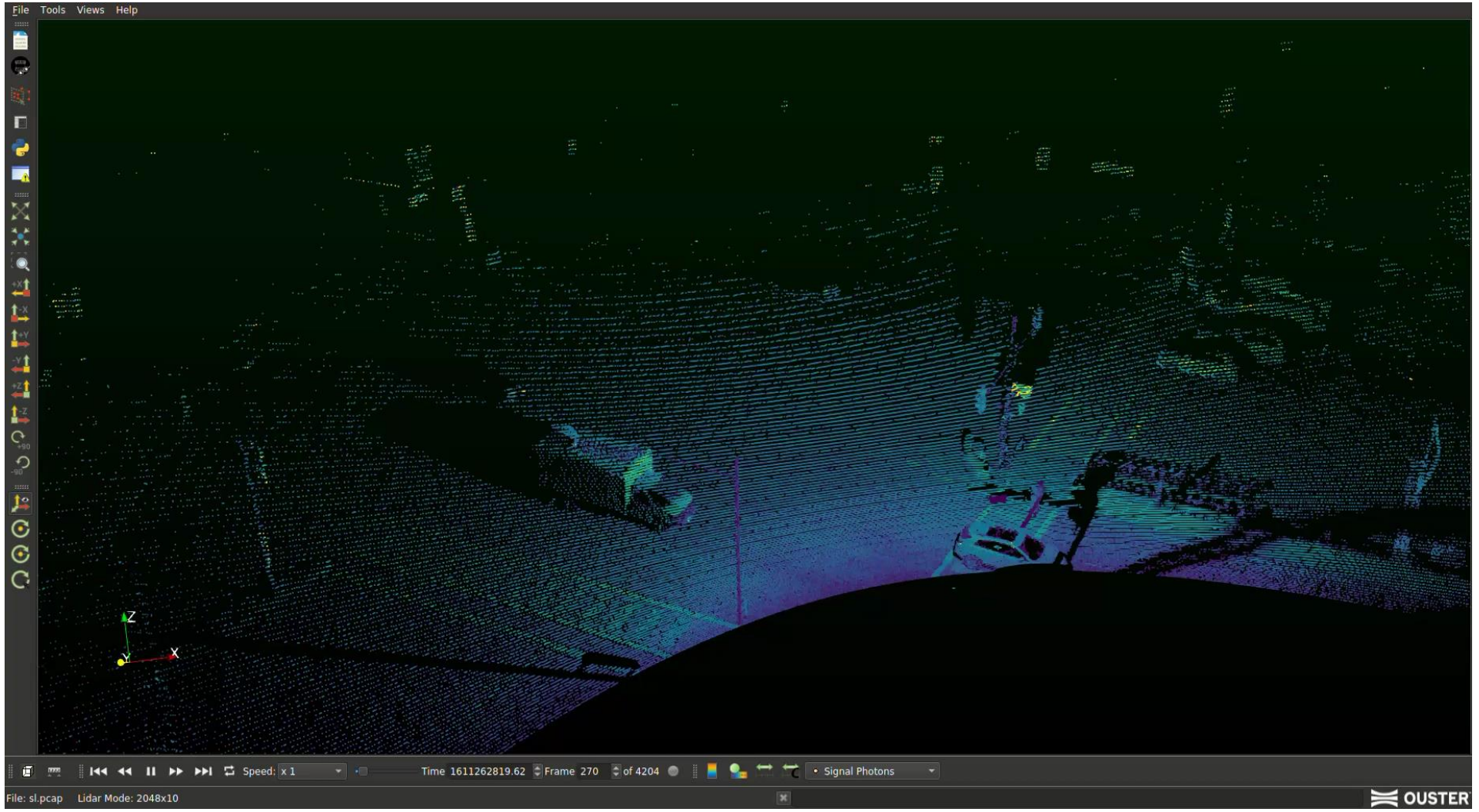


Sensors

Axis Q1798 Camera (SOW 2)

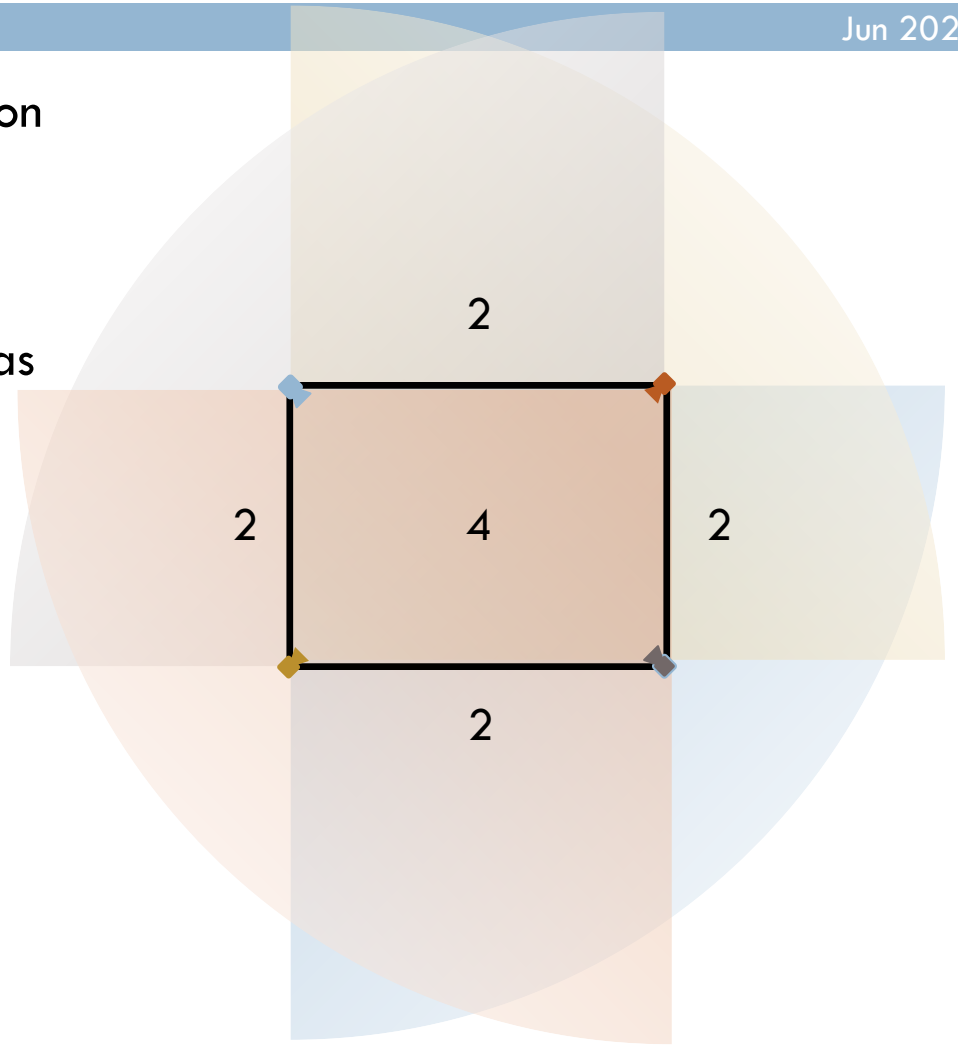
Cepton Vista X LiDAR (SOW 1)

Intersection LiDAR



4-Camera Geometry for Intersection

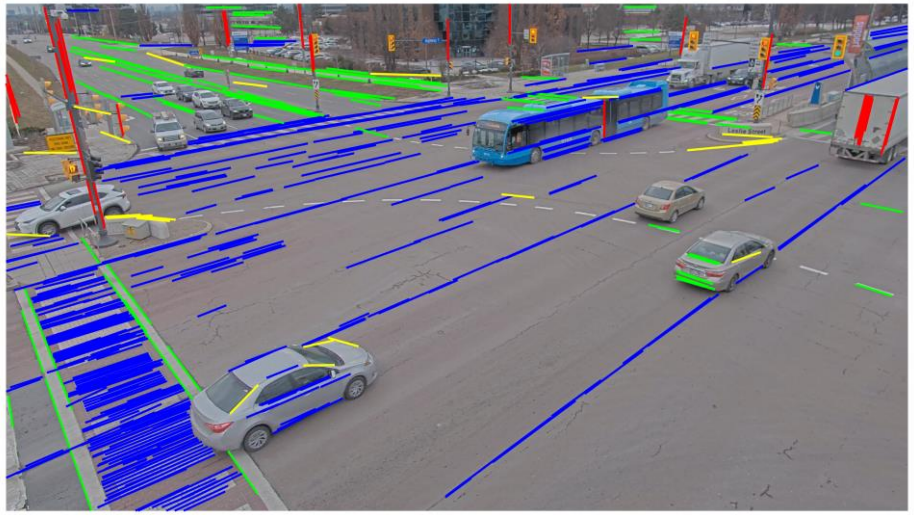
- Cameras mounted at 10m height at 4 corners of intersection
- Oblique angle expected to yield better recognition rates
- Objects interior to intersection seen by 4 cameras
- Objects entering and exiting intersection seen by 2 cameras

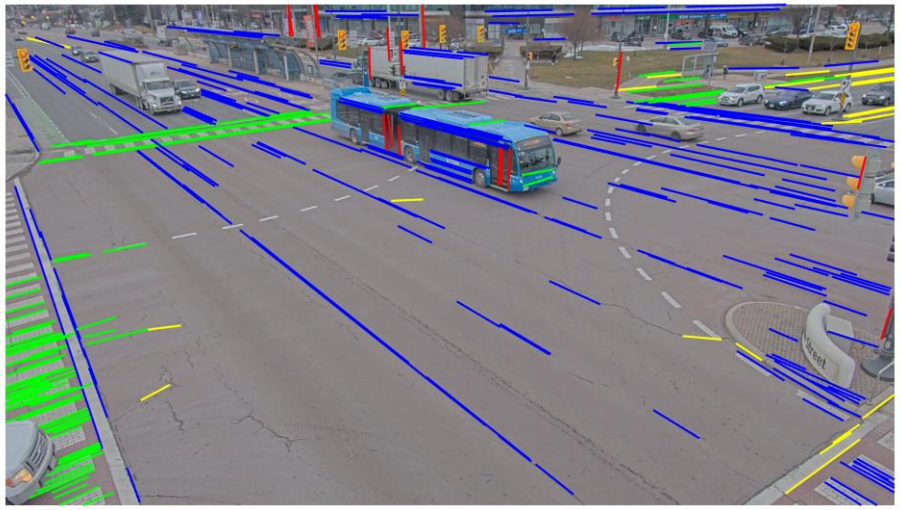


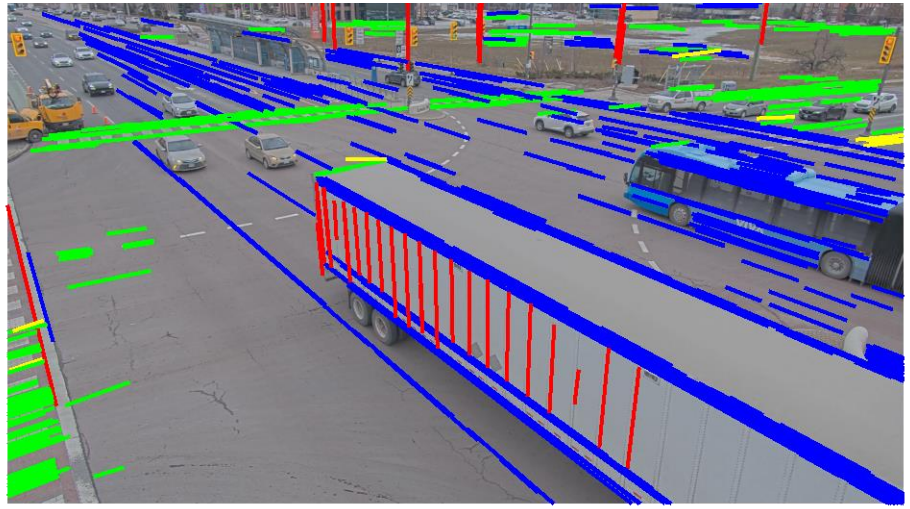
Current Work

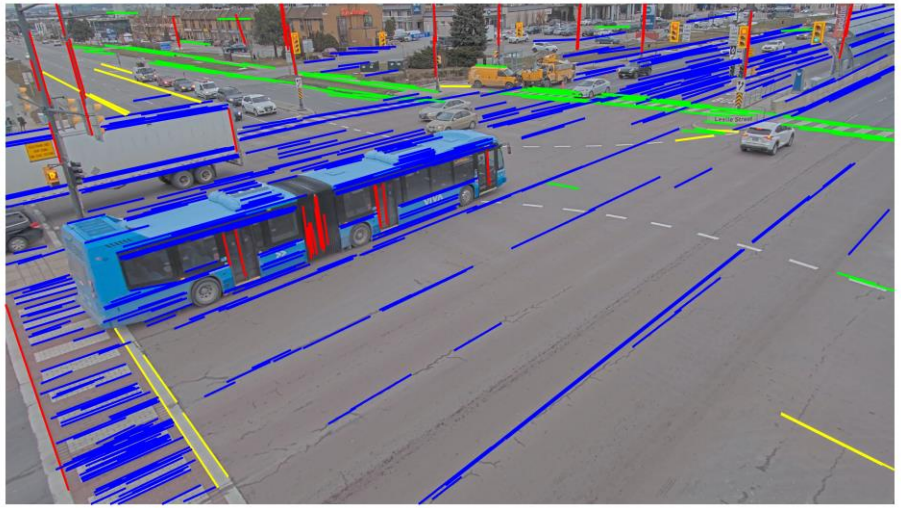
- Online camera calibration
- Object detection and segmentation
- Pipeline for trajectory classification
- Attending to vulnerable road users

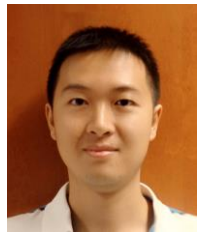
Online Camera Calibration







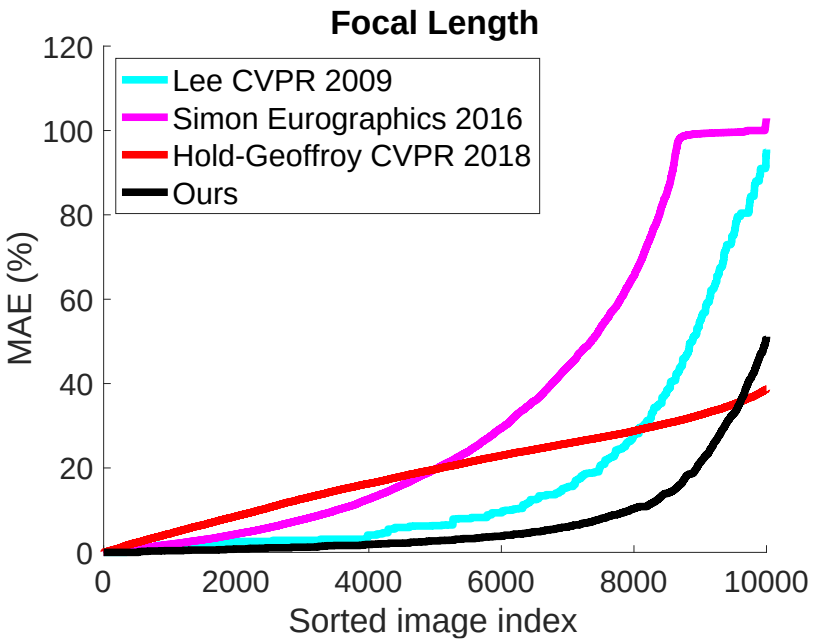


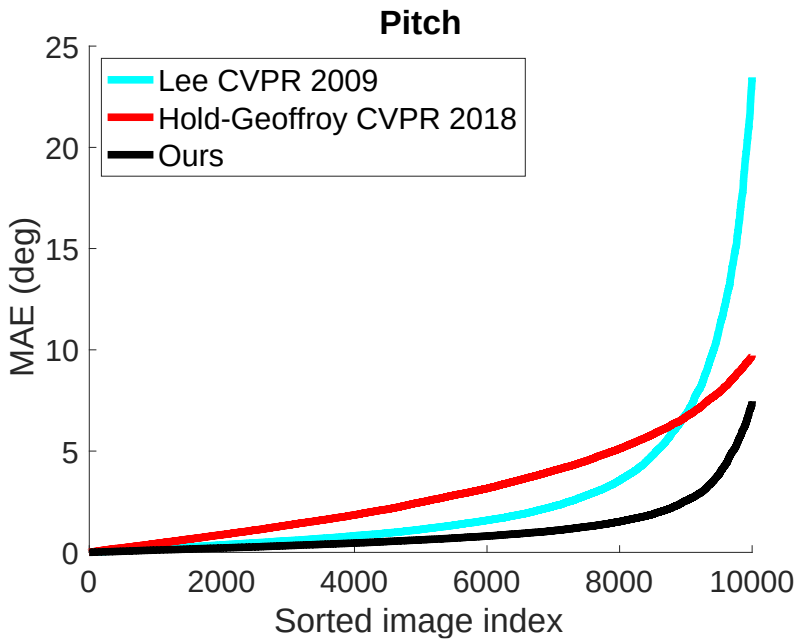


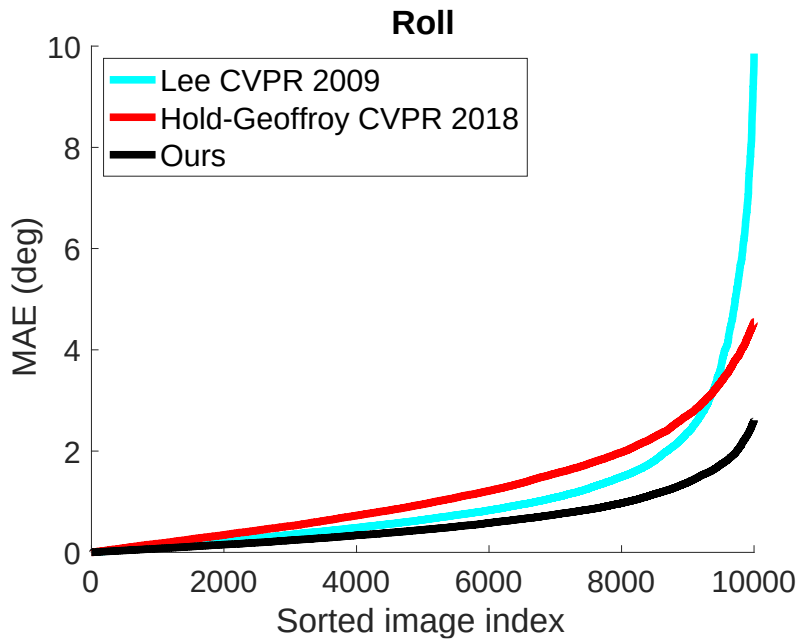
Dr. Yiming Qian
J. Elder

Improving State-of-the-Art

- Images sorted by MAE independently for each algorithm and parameter.

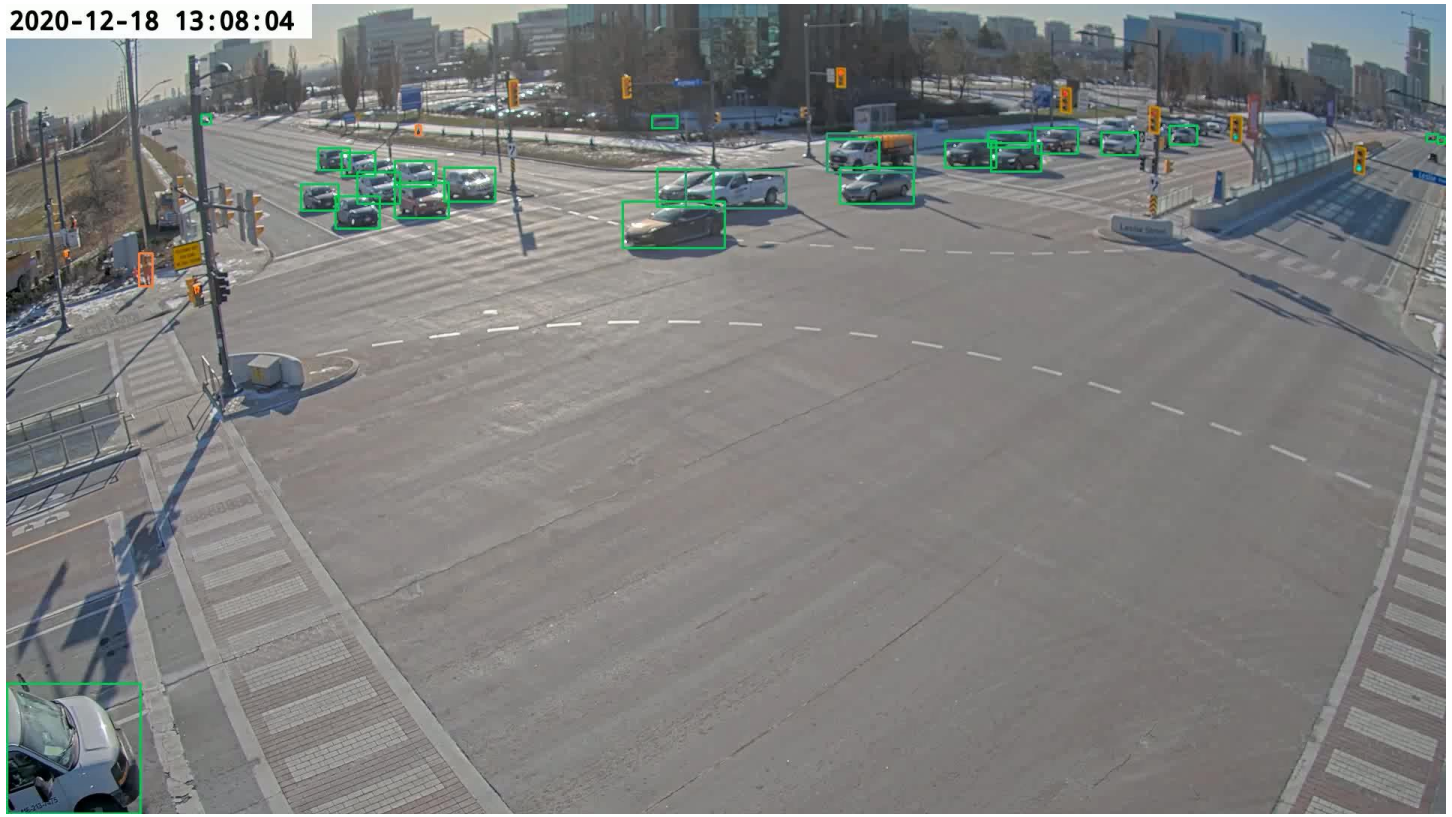






2D Vehicle Detection

- CenterNet with Resnet50 Backbone trained on MS COCO



Duan *et al*. Centernet: Keypoint triplets for object detection. ICCV 2019.

Keyi Liu

2D Vehicle Segmentation

2D Figure/Ground Completion

Improving State-of-the-Art

- DTM-based method with sparse coding outperforms both Contour-based and Mask-based top performers
- All methods are trained for 12 epochs on MS COCO with ResNet-50 backbone

Representation	Method	Input Scale	Encoding Dim.	Active Dim.	Mask AP
Contour	ESE-Seg[1]	416	20	20	0.216
	PolarMask[2]	400	72	72	0.229
Mask	MEInst[3]	400	60	60	0.239
	Ours	**400**	**128**	**57**	**0.249(+1.0%)**
DTM	**Ours**	**400**	**128**	**56**	**0.257(+1.8%)**

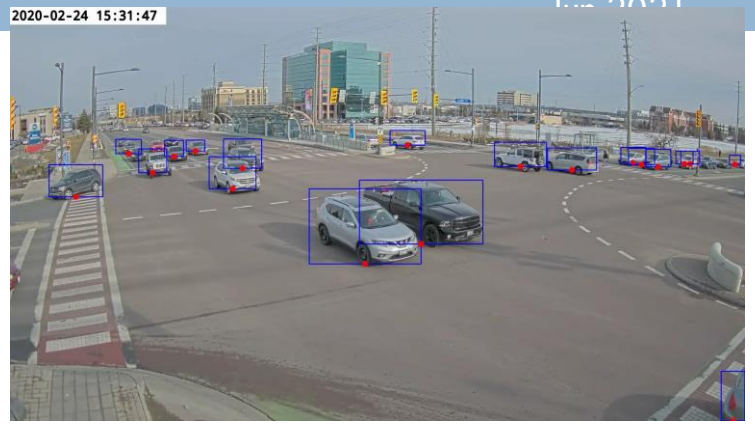
[1] RuWenqiang Xu, Haiyang Wang, Fubo Qi, and Cewu Lu. Explicit shape encoding for real-time instance segmentation. In ICCV, 2019.

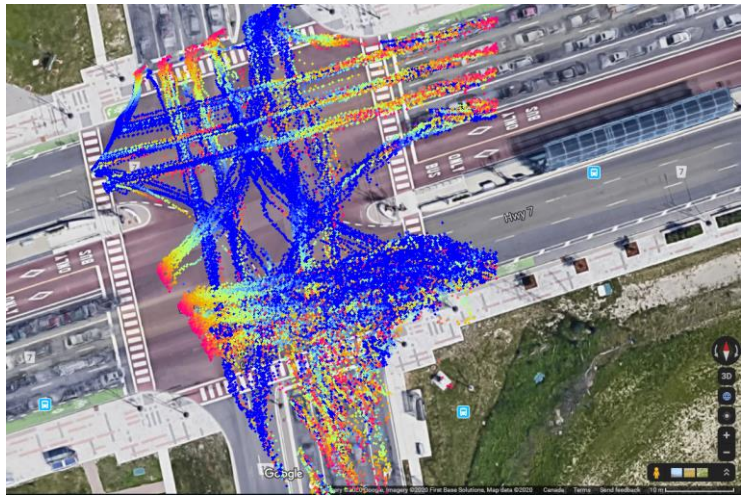
[2] Enze Xie, Peize Sun, Xiaogang Song, Wenhai Wang, Xuebo Liu, Ding Liang, Chun-hua Shen, and Ping Luo. PolarMask: Single shot instance segmentation with polar representation. In CVPR, 2020..

[3] Rufeng Zhang, Zhi Tian, Chunhua Shen, Mingyu You, and Youliang Yan. Mask Encoding for Single Shot Instance Segmentation. In CVPR, 2020.

Base Pipeline for Trajectory Counting

- Implemented complete pipeline for trajectory classification of motor vehicles at intersection.
- Evaluated on small provisional dataset derived from field site.
- Results provided as extra deliverable for Y1

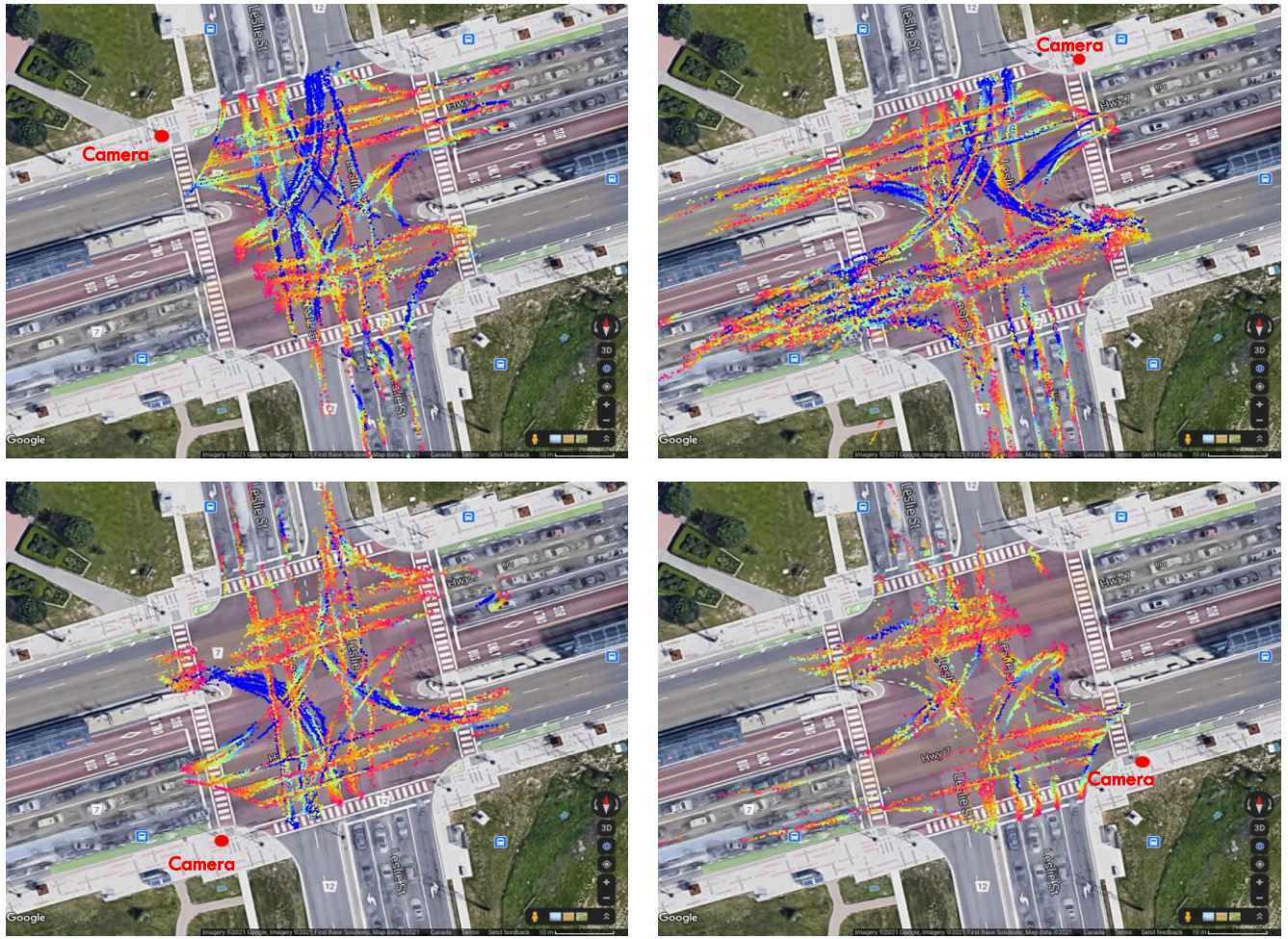




	Approaching from East	Approaching from West	Approaching from South	Approaching from North
Ground Truth	145	156	58	59
FRCNN	148	133	79	74
FRCNN error	+2.1%	-14.7%	+36.2%	+25.4%
SSD512	141	70	82	81
SSD512 error	-2.8%	-55.1%	+41.4%	+37.3%

Poorna Rao

Tracks from 4 Cameras



Digital Intersection

- ✓ 1. Accurate road shape
- ✓ 2. Traffic islands

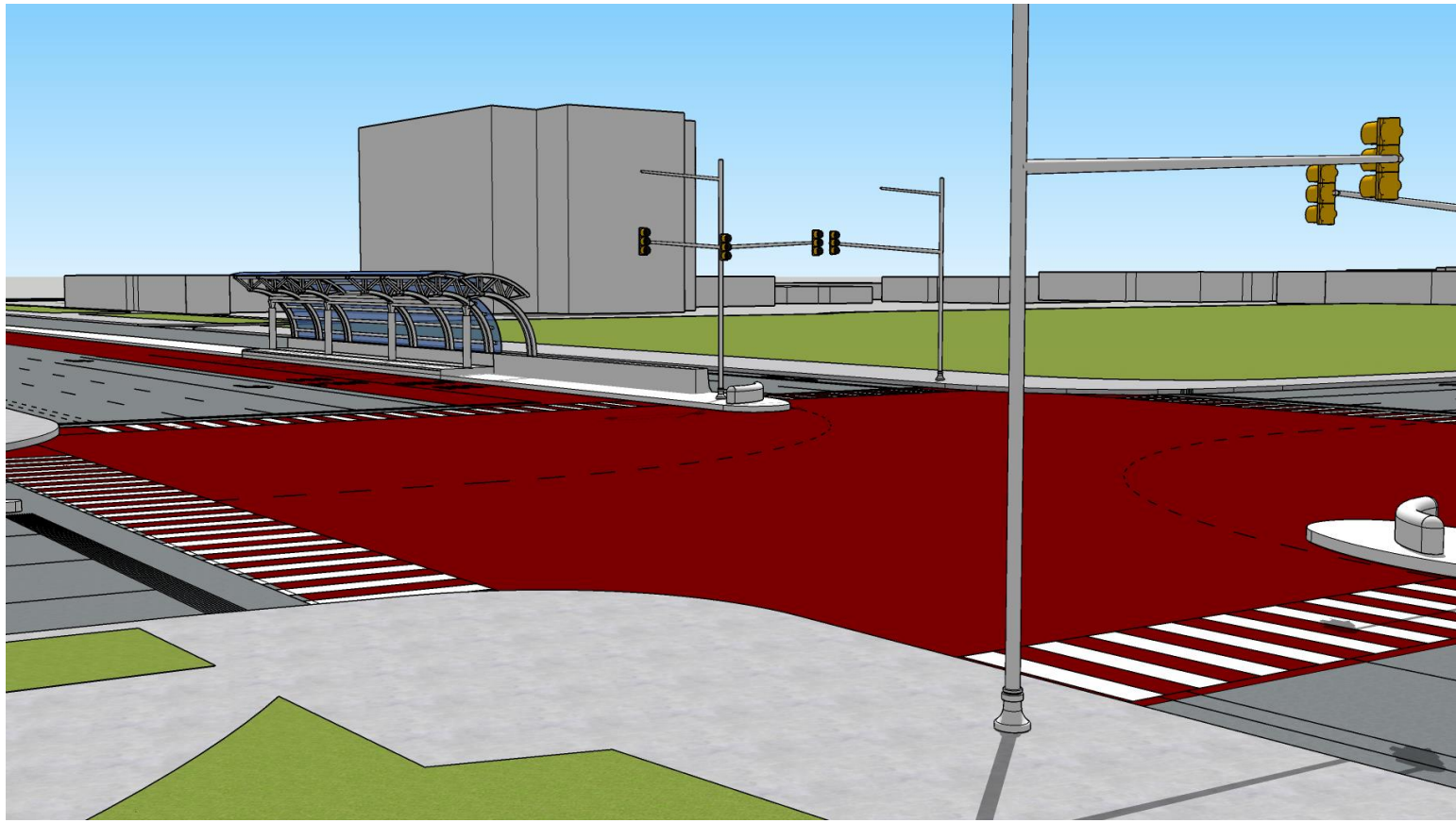
LOD 1

- ✓ 3. Textures: road markings
- ✓ 4. Grass
- ✓ 5. Building massing

LOD 2

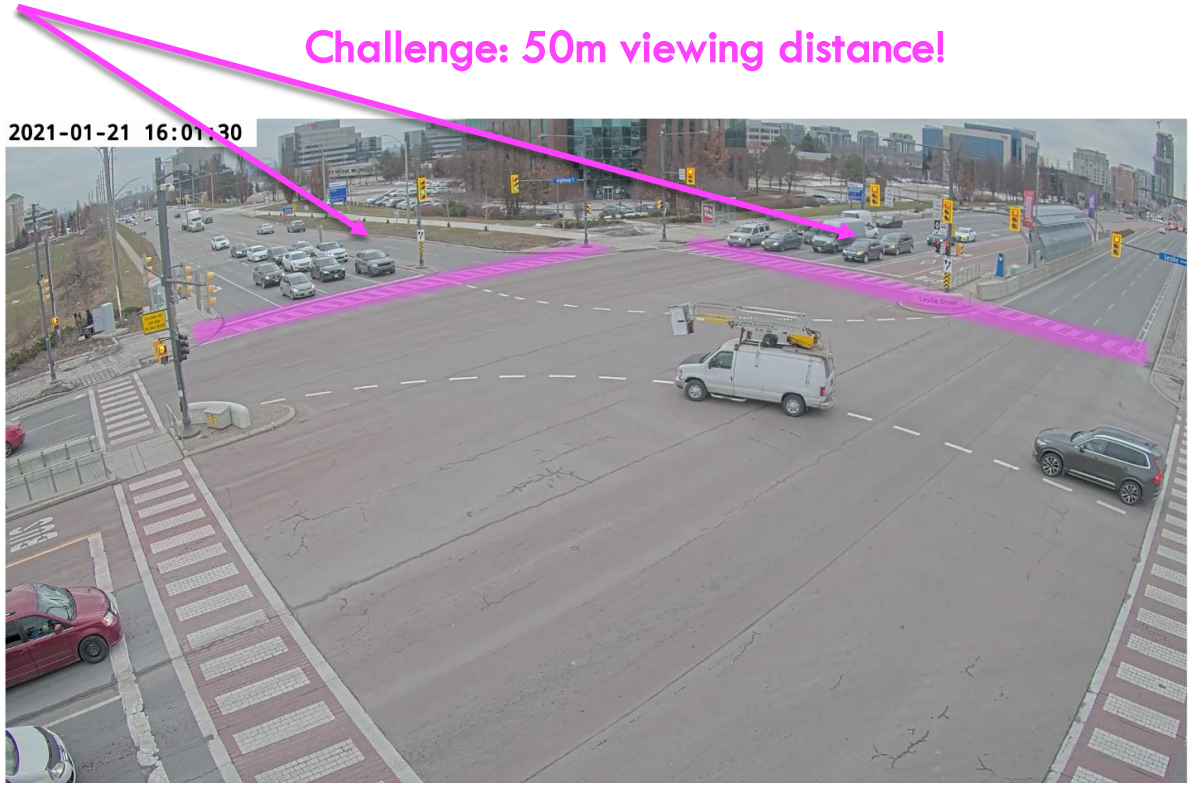
- ✓ 6. Traffic lights
- 7. Trees
- 8. Signs
- 9. Bus stops
- 10. Lamps
- 11. Electricity poles

LOD 3



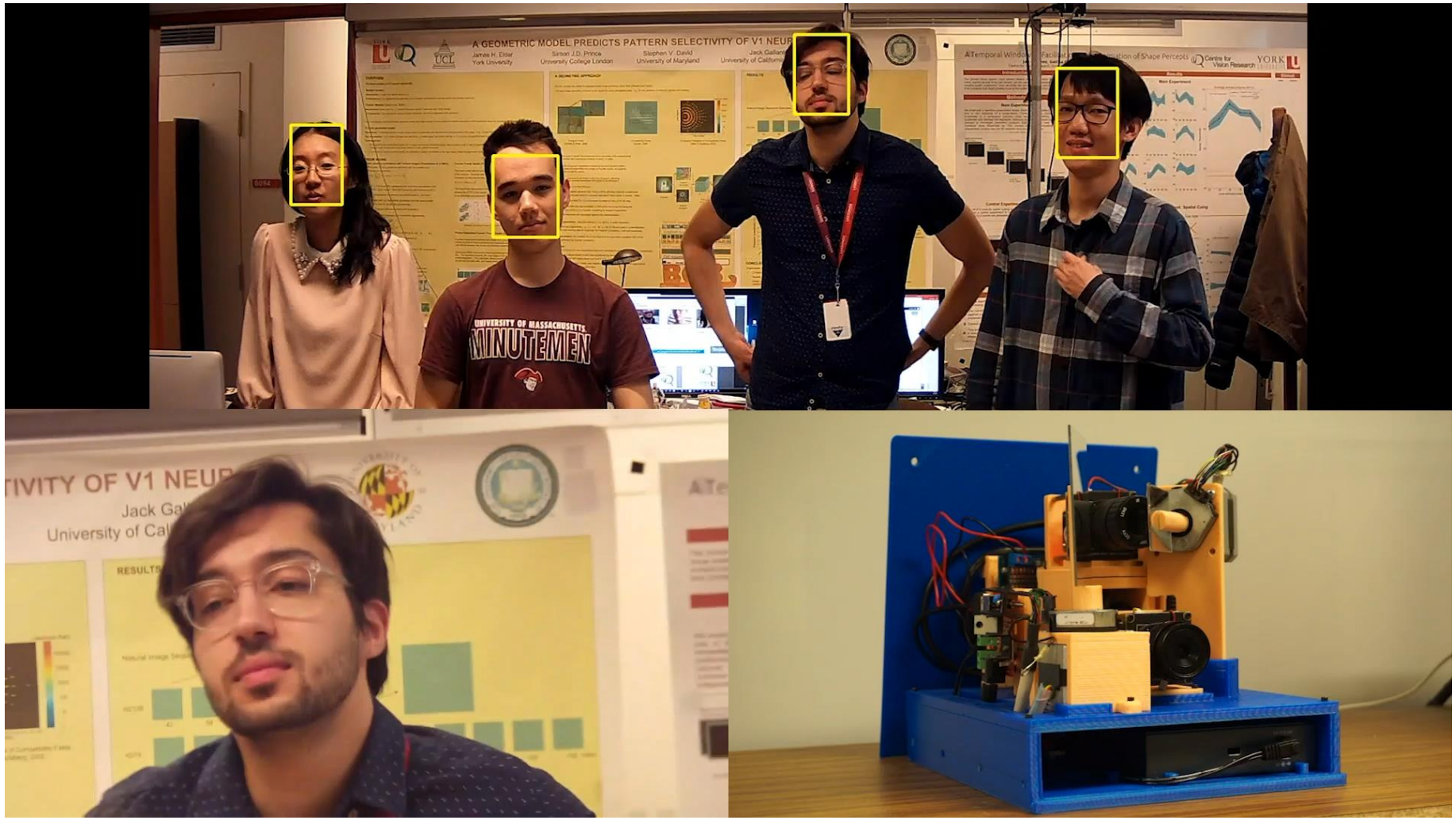
Attending to Vulnerable Road Users

- Goal: Detailed detection and fine-grained assessment of vulnerable road users over two crosswalks with a single camera system
- Example conditions of interest:
  - Wheelchair
  - Cane
  - Other mobility challenge
  - Elderly
  - Children
  - Stroller
  - Distracted (e.g., texting)

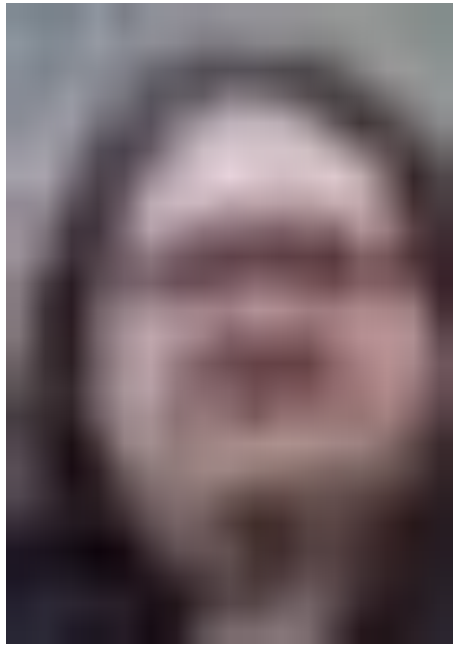


Attentive Sensor: Indoor Prototype

- Pre-Attentive:
  - 1920 x 1080p
  - 75 deg FOV
  - Using MTCNN face detector (Zhang et al 2016)
- Attentive:
  - 1920 x 1200p
  - 8 deg FOV
  - Operates at 6 fixations/sec
  - Human saccadic rate: 2-3 fixations/sec

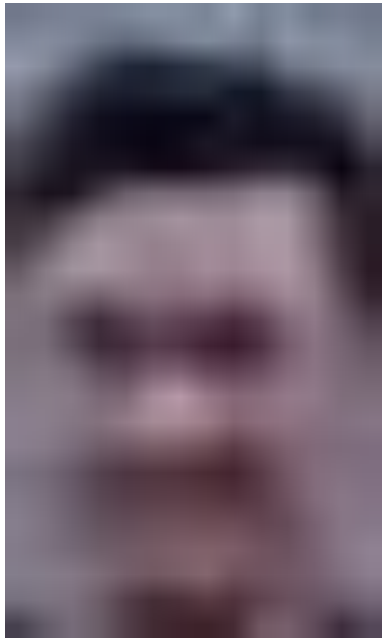


Attentive Sensing Extends Visual Range



Pre-Attentive

Attentive

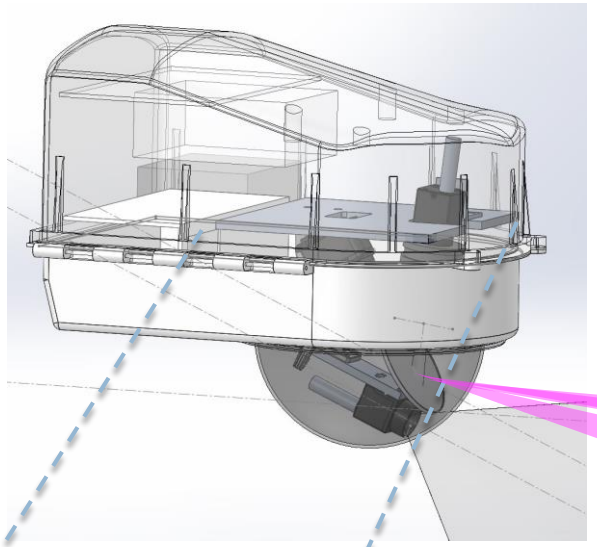


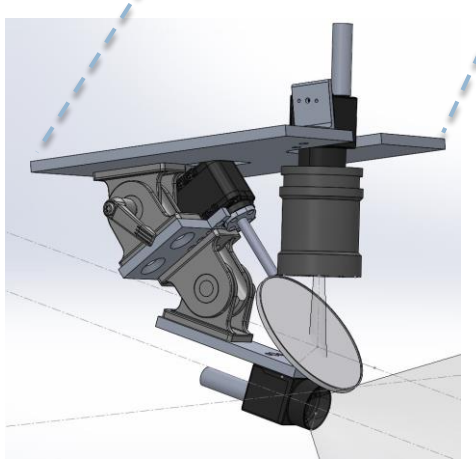
Pre-Attentive

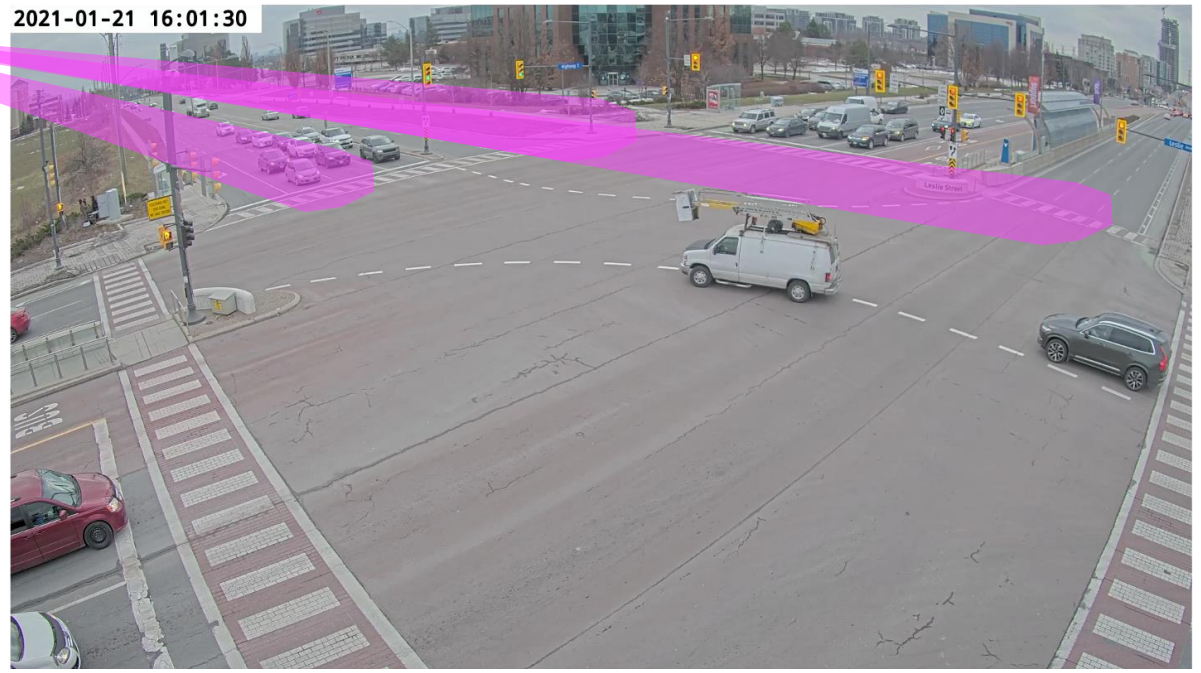
Attentive

Range: 27 metres

Attentive Sensor for Pedestrian Analytics: Design







Kartikeya Bhargava

Next Steps

- Generalization to busy downtown intersection with dense pedestrian traffic.