



# Measuring Main Streets

## GOOGLE STREET VIEW & ARTIFICIAL INTELLIGENCE POWERED PUBLIC REALM & ACCESSIBILITY ASSESSMENT

August 2025



Canadian  
Urban  
Institute

Institut  
Urbain du  
Canada

# A Complex Problem




How to assess the *quality* of urban environments from a pedestrian or accessibility point of view?

How to accomplish this at scale, using a quantitative methodology?

Traditionally, measures have relied on proximity and density, which can have misleading results...



Dundas Street, Mississauga



**Very Walkable**  
Most errands can be accomplished on foot.

?

# A Solution: Computer Vision



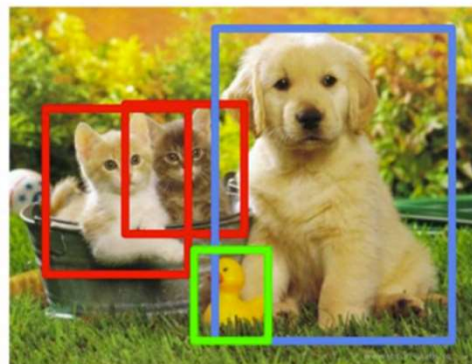
Teaching computers to gain high-level understanding from images or videos

**Classification**



CAT

**Object Detection**



2xCAT, DOG, DUCK

**Semantic Segmentation**



CAT, DOG, DUCK

**Instance Segmentation**



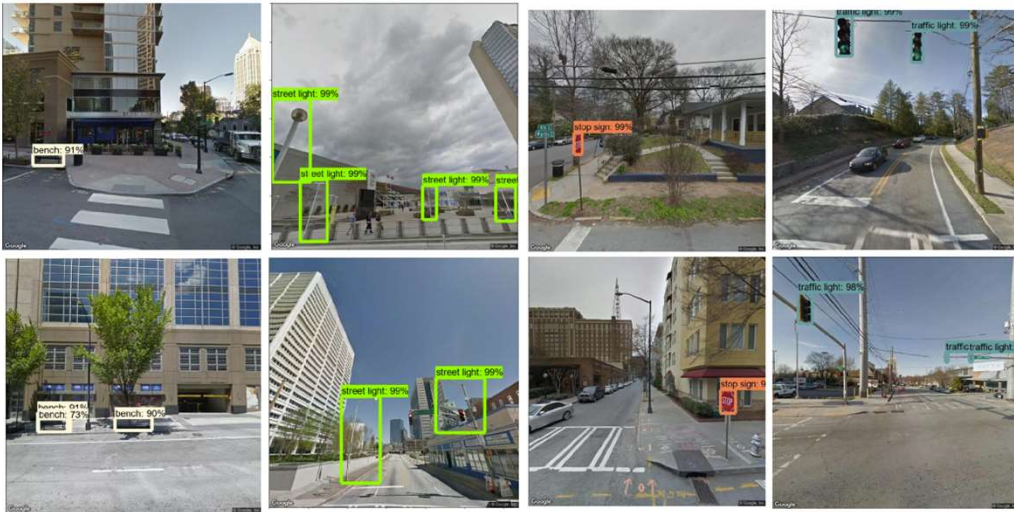
2xCAT, DOG, DUCK

Traditional Approaches

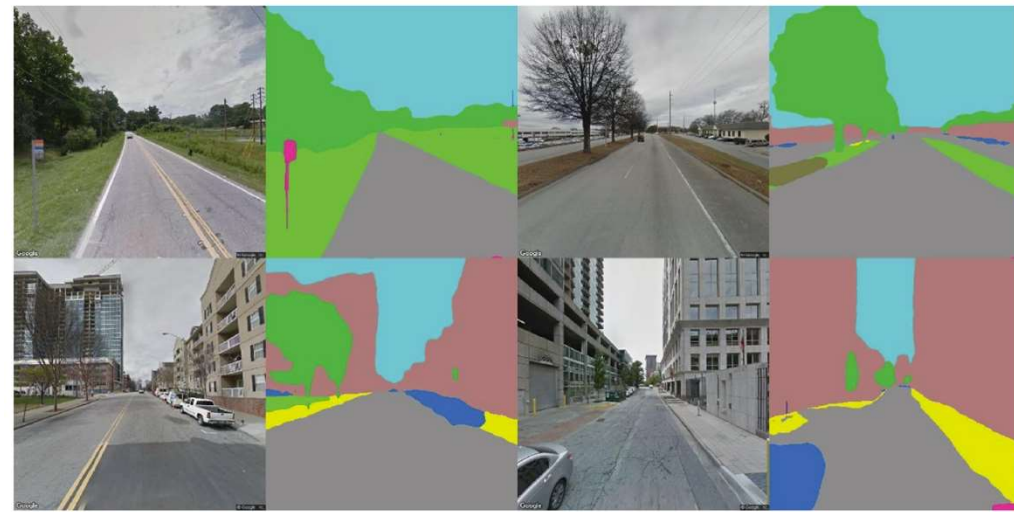
# What is Computer Vision?



## URBAN CONTEXT



Object Detection

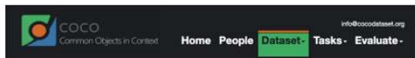


Semantic Segmentation



# Pre-trained models are limited...

- Require manual feature engineering and large labeled datasets.
- Identifiable categories are limited, training custom models is time-consuming and expensive
- Outputs raw detections (e.g., “crosswalk” “bench”).



COCO Explorer

COCO 2017 trainval browser (123,287 images, 856,284 instances). Crowd labels not shown.

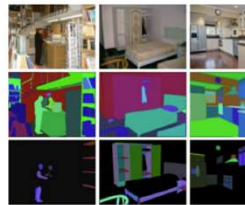


[COCO - Common Objects in Context](#)

80 object categories



The annotated images cover the scene categories from the SUN v3 segmentation, and parts segmentations.



The next visualization provides the list of objects and parts and their annotated instances and parts with more than 10 annotated instances

[ADE20K](#)

150 categories



## Class Definitions

Please click on the individual classes for details on their definitions.

Group	Classes
flat	road · sidewalk · parking* · rail track*
human	person* · rider*
vehicle	car* · truck* · bus* · on rails* · motorcycle* · bicycle* · caravan** · trailer**
construction	building · wall · fence · guard rail* · bridge* · tunnel*
object	pole · pole group* · traffic sign · traffic light
nature	vegetation · terrain
sky	sky
void	ground* · dynamic* · static*

[CITYSCAPES](#)

30 categories

# A new approach: Large Language Models



## How It Works

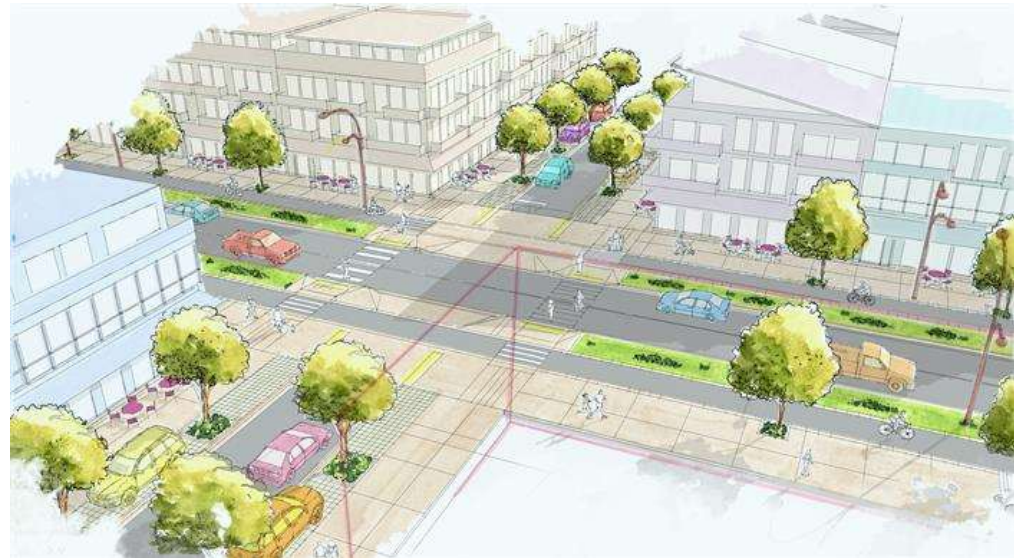
- Image Encoding: The model processes street-level images using **multimodal AI** (vision + language) models via an API.
- Contextual Evaluation: Using a **structured prompt**, it assesses urban elements
- Human-Like Insights: Instead of binary detections, it provides **scored evaluations**, qualitative analysis and **reasoning** (e.g., “The sidewalk is too narrow for two wheelchairs”).

# Structured Prompt



## Flexible and Transparent

- Tree cover & shade
- Traffic Calming Measures
- Sidewalks
- Bicycle Infrastructure
- Street Furniture & Lighting
- Public Amenities
- Accessibility
- Building Characteristics



Presence, Condition, Quality → overall score and sub-scores for each category

# Case Study: Arbutus Station (Kitsilano), Vancouver



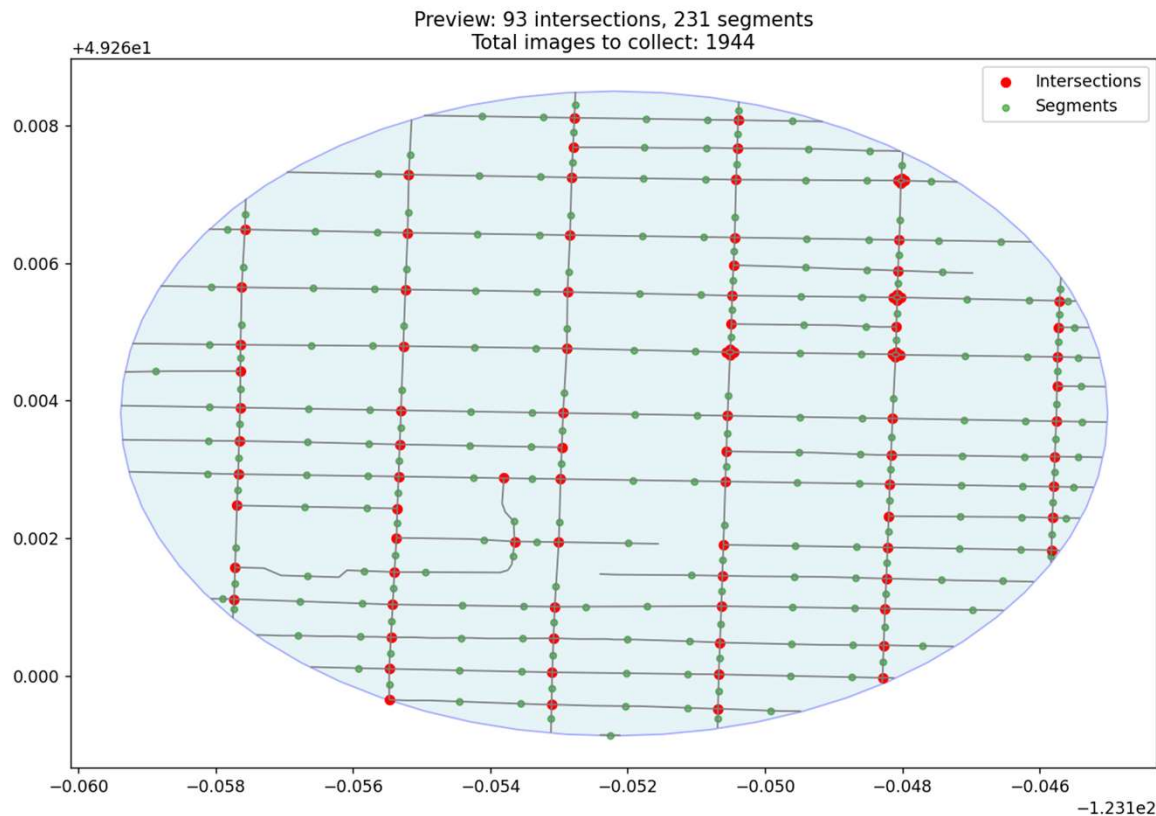
"Sidewalks are wide enough for mobility devices, but the lack of TWSI may hinder some users."

"The presence of trees and greenery along the median enhances the aesthetics and provides shade."

"The buildings have a human scale that fits well with the street width."

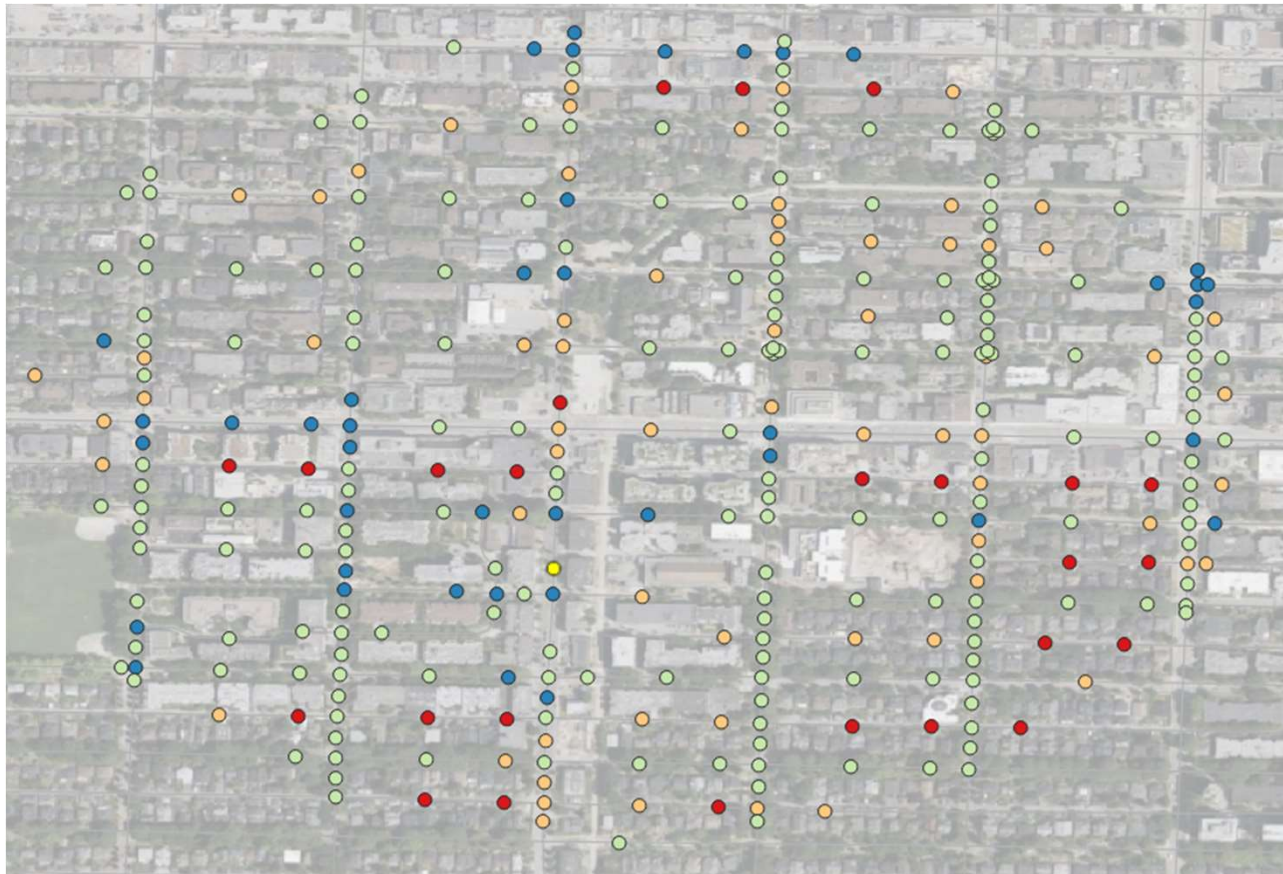
Overall Score: 4/5

# Case Study: Arbutus Station (Kitsilano), Vancouver



- 800m radius
- Sample points at every intersection and 100m in between intersections
- Sample 6 images at each point to capture a 360° view
- 1944 Google Streetview Images captured

# Case Study: Arbutus Station (Kitsilano), Vancouver



Overall Score

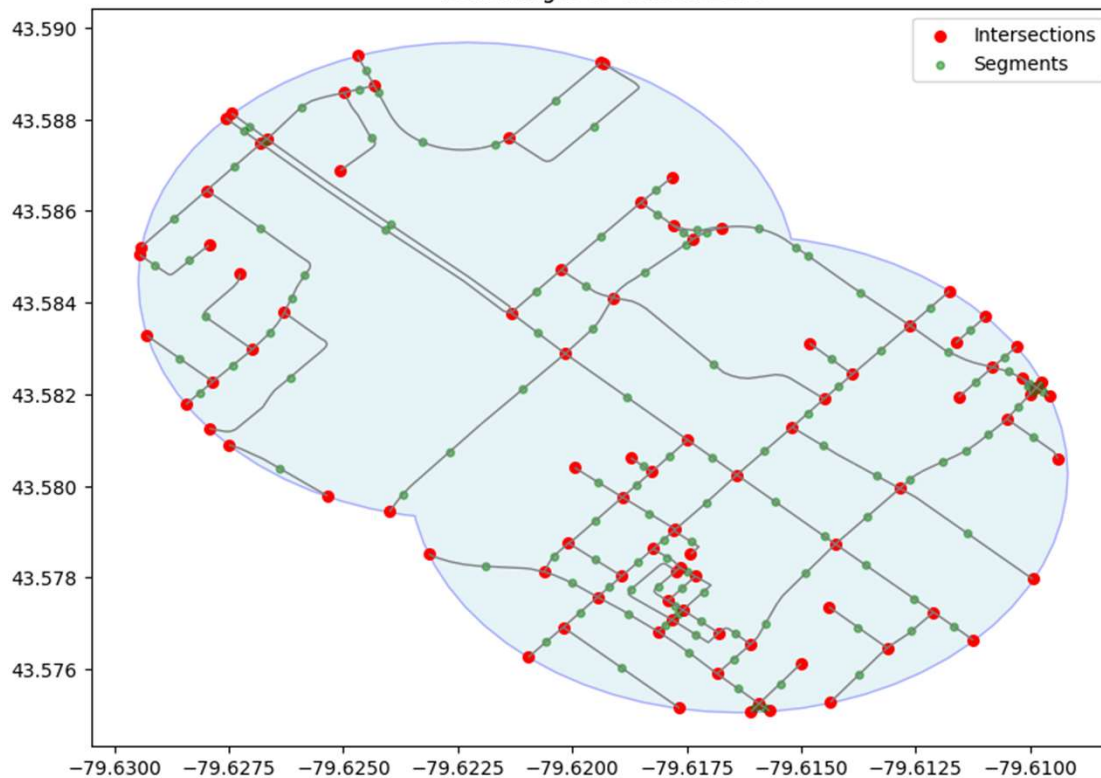
- 1 - 1.75
- 1.75 - 2.5
- 2.5 - 3.25
- 3.25 - 4

+ sub scores for each category

# Case Study: Cooksville, Mississauga



Preview: 92 intersections, 123 segments  
Total images to collect: 1290



- 800m radius
- Sample points at every intersection and 100m in between intersections
- Sample 6 images at each point to capture a 360° view
- 1290 Google Streetview Images captured

# Case Study: Cooksville, Mississauga



"No visible curb ramps at most crosswalk locations."

"Crosswalks are not clearly marked."

"Minimal landscaping visible; a few trees along the street."

Overall Score: 2/5

# Case Study: Cooksville, Mississauga





Overall Score

- 1 - 1.75
- 1.75 - 2.5
- 2.5 - 3.25
- 3.25 - 4

+ sub scores for each category

# Limitations



- Source of imagery (Google Street View) isn't consistently updated, image resolution is limited
- Models are currently limited in being able to accurately spatially place or measure objects within images (ex: “The sidewalk is wide”  vs. “The sidewalk is 5m wide” )
- Models may struggle to differentiate between specific or similar objects (sign indicating transit stop vs. sign indicating parking hours)

# Next Steps & Possibilities



- **Varying prompt and criteria based on type of street (residential vs. commercial)**
- Identifying alternative sources and methods of imagery / data collection.
- Combining traditional computer vision (object detection and segmentation) approaches with large language model-based approach.
- ‘Fine-tuning’ generic models to answer more specific questions and make more accurate, local judgements
- ... and many more!

Thank You!

