UBC Digital Multimedia Lab

Smart Urban Solutions: Alleviating Traffic Congestion with Innovations in the Detection and Localization of Free Street Parking Spaces through Artificial Intelligence and Motion Estimation

Tala Bazzaza
tbazzaza@ece.ubc.ca
Electrical & Computer Engineering
University of British Columbia

IARIA

About me

- Completed my M.A.Sc degree in Electrical and Computer Engineering at the University of British Columbia (UBC) in Vancouver, Canada.

- Currently, I hold the position of Scientific Engineer at UBC, where I oversee research initiatives at the Digital Multimedia Lab and provides supervision to a significant cohort of graduate students.

- My research interests lie in the dynamic field of Artificial Intelligence, with a specific focus on advancing Video Driven Information Applications. My contributions extend across various domains, including Smart Cities, Intelligent Transportation, Digital Health, and Entertainment.

Traffic Congestion

In metropolitan cities, a major contribution to traffic congestion is the lack of street parking availability

It is estimated that 30-50% of traffic congestion is caused by drivers searching for parking spots during peak hours in metropolitan cities

Our Objective

- Automatic detection and recognition of street parking
- Notify drivers of available parking spaces ahead, thus significantly alleviating traffic congestion

ece Electrical and Computer Engineering

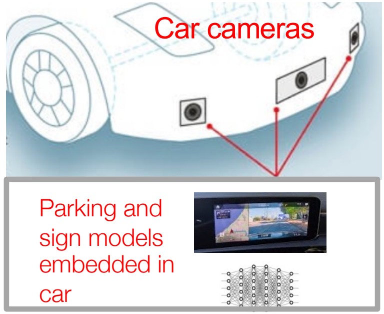
IARIA

Previous Work – Detection of Street Parking Spaces

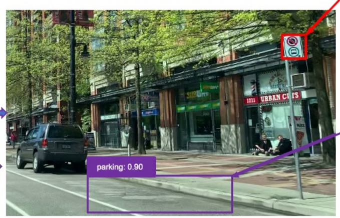
- Several methods rely on aerial views of street parking spots captured by surveillance cameras
- Limited detection area due to obstructions caused by trees
- Cannot clearly identify spaces between parked cars
- Impractical solutions

Holistic Approach

GPS of available parking location

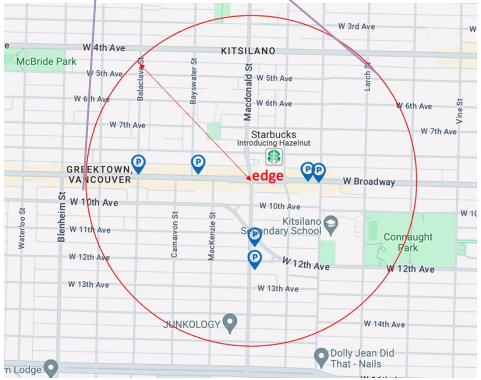


sign



Parking

Visual representation of the parking and sign detection

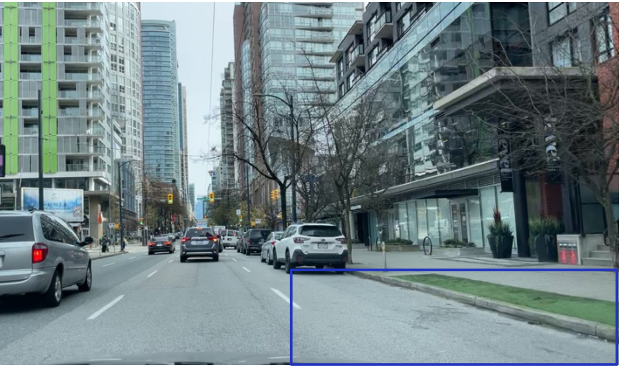


Edge sends information of free parking to cars within a radius of 3-4km

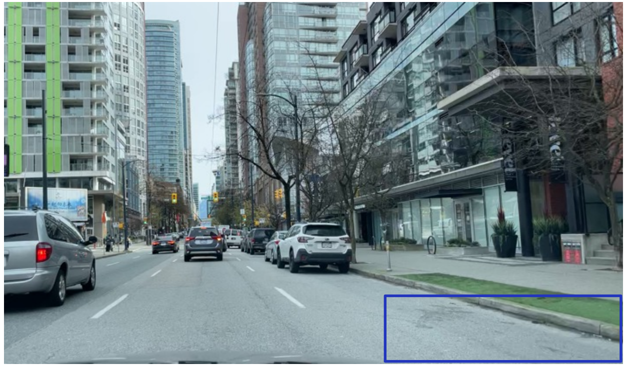
Available parking shown in navigation screen

Labeling is Everything!

A simple but naïve approach



A parking spot behind a parked car



One parking spot

Another parking spot

Parking Spot: empty space as one spot!

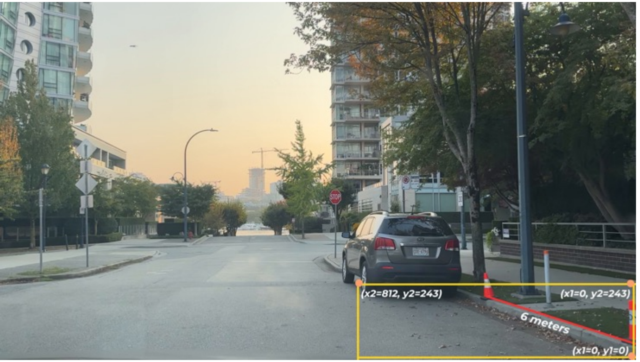
Labeling is Everything!

We want to identify all free parking spaces

Free parking

Free parking

Labeling is Everything!



Scenario 1:

Single parking spot behind a parked car

Labeling is Everything!

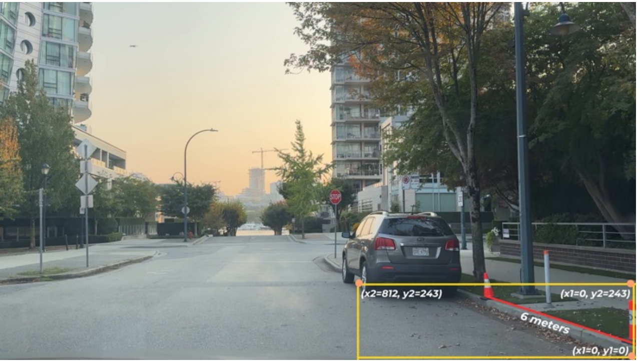
x,y coordinates: size of the parking spot.

Red line demonstrates the size of the parking spot being 6 meters

Case 1:

Single parking spot behind a parked car

Labeling is Everything!



Scenario 1:

Single parking spot behind a parked car

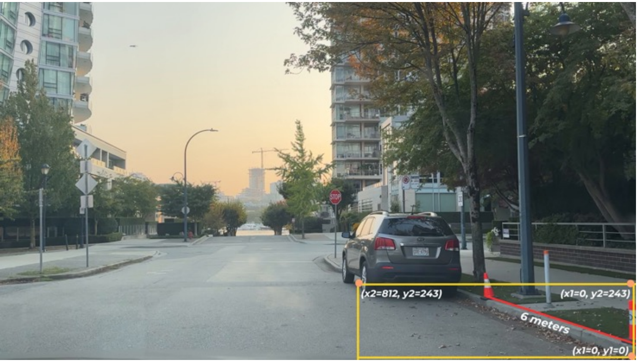
Scenario 2:

More than one empty parking spot in a frame

Labeling is Everything!

x,y coordinates: size of the parking spot.

Red line demonstrates the size of the parking spot being 6 meters



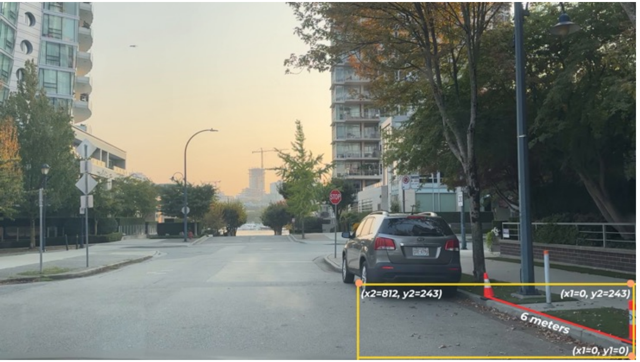
Scenario 1:

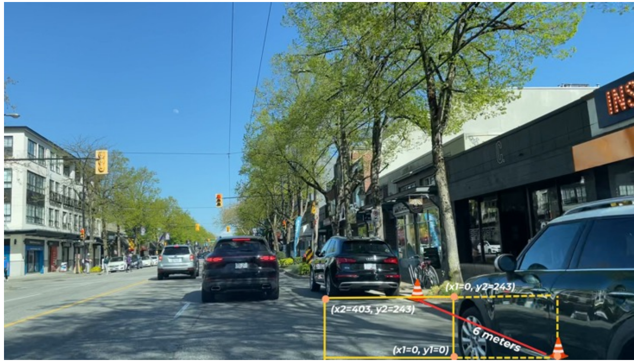
Single parking spot behind a parked car

Scenario 2:

More than one empty parking spot in a frame

Labeling is Everything!





Scenario 1:

Single parking spot behind a parked car

Scenario 2:

More than one empty parking spot in a frame

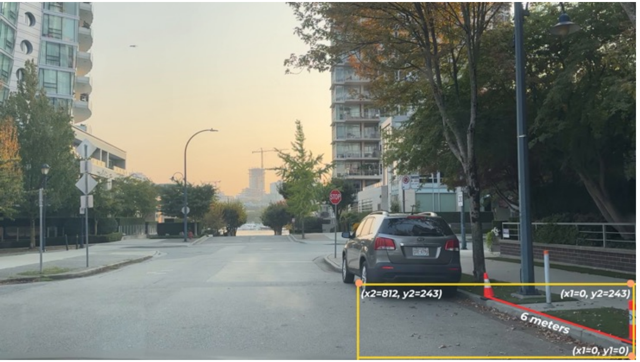
Scenario 3:

Parking spot in a frame is between two parked vehicles

Labeling is Everything!

x,y coordinates

Red line 6 meters



Scenario 1:

Single parking spot behind a parked car

Scenario 2:

More than one empty parking spot in a frame

Scenario 3:

Parking spot in a frame is between two parked vehicles

Choosing a Network

We tested 4 different networks: YOLOv4, Swin, YOLOv7 and YOLOv7-x

Choosing a Network

We tested 4 different networks: YOLOv4, Swin, YOLOv7 and YOLOv7-x

TESTING RESULTS OF THE NETWORKS

Model	mAP @ 0.5	Precision	Recall	F1 Score
YOLOv4	83.3%	0.84	0.81	0.82
SWIN	81.6%	0.81	0.74	0.77
YOLOv7	85.5%	0.89	0.77	0.83
YOLOv7-x	90.9%	0.92	0.82	0.87

Choosing a Network

We tested 4 different networks: YOLOv4, Swin, YOLOv7 and YOLOv7-x

TESTING RESULTS OF THE NETWORKS

Model	mAP @ 0.5	Precision	Recall	F1 Score
YOLOv4	83.3%	0.84	0.81	0.82
SWIN	81.6%	0.81	0.74	0.77
YOLOv7	85.5%	0.89	0.77	0.83
YOLOv7-x	90.9%	0.92	0.82	0.87

mAP:
Mean
Average
Precision

Accuracy of predicting the bounding box

Choosing a Network

We tested 4 different networks: YOLOv4, Swin, YOLOv7 and YOLOv7-x

TESTING RESULTS OF THE NETWORKS

Model	mAP @ 0.5	Precision	Recall	F1 Score
YOLOv4	83.3%	0.84	0.81	0.82
SWIN	81.6%	0.81	0.74	0.77
YOLOv7	85.5%	0.89	0.77	0.83
YOLOv7-x	90.9%	0.92	0.82	0.87

mAP:
Mean
Average
Precision

Accuracy of predicting the bounding box
&
Accuracy of detecting the correct class in the box

Choosing a Network

We tested 4 different networks: YOLOv4, Swin, YOLOv7 and YOLOv7-x

TESTING RESULTS OF THE NETWORKS

Model	mAP @ 0.5	Precision	Recall	F1 Score
YOLOv4	83.3%	0.84	0.81	0.82
SWIN	81.6%	0.81	0.74	0.77
YOLOv7	85.5%	0.89	0.77	0.83
YOLOv7-x	90.9%	0.92	0.82	0.87

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Choosing a Network

We tested 4 different networks: YOLOv4, Swin, YOLOv7 and YOLOv7-x

TESTING RESULTS OF THE NETWORKS

Model	mAP @ 0.5	Precision	Recall	F1 Score
YOLOv4	83.3%	0.84	0.81	0.82
SWIN	81.6%	0.81	0.74	0.77
YOLOv7	85.5%	0.89	0.77	0.83
YOLOv7-x	90.9%	0.92	0.82	0.87

F1 score:

Measures precisely the network's ability to avoid misclassifying non-vacant spots and ensure that no free parking spaces are missed

Choosing a Network

We tested 4 different networks: YOLOv4, Swin, YOLOv7 and YOLOv7-x

TESTING RESULTS OF THE NETWORKS

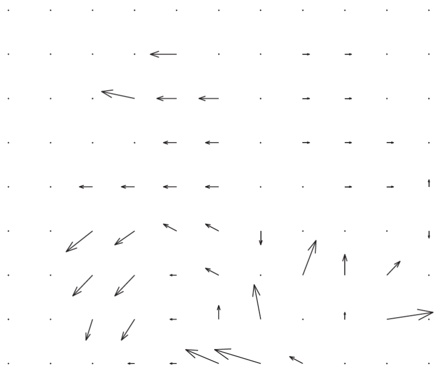
Model	mAP @ 0.5	Precision	Recall	F1 Score
YOLOv4	83.3%	0.84	0.81	0.82
SWIN	81.6%	0.81	0.74	0.77
YOLOv7	85.5%	0.89	0.77	0.83
YOLOv7-x	90.9%	0.92	0.82	0.87

F1 score:
Measures precisely the network's ability to avoid misclassifying non-vacant spots and ensure that no free parking spaces are missed

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$\text{GMV}=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$

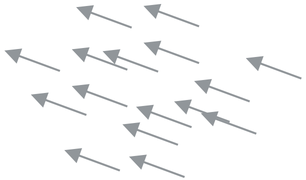


Motion Vectors

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$GMV=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$



Motion Vectors showing car movement

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$GMV=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$

Global vector

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$GMV=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$

- Not all blocks in a frame are important and useful in calculating global motion

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$\mathrm{GMV}=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$

- Not all blocks in a frame are important and useful in calculating global motion

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$\text{GMV}=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$

- Not all blocks in a frame are important and useful in calculating global motion

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$\text{GMV}=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$

- Not all blocks in a frame are important and useful in calculating global motion

Accurate Measurement of Parking Space

- We base our approach on the global motion vector for each frame:

$$\mathrm{GMV}=\frac{\sum_{i=1}^{N} s_i mv_i}{\sum_{i=1}^{N} s_i}$$

- Not all blocks in a frame are important and useful in calculating global motion

Accurate Measurement of Parking Space

Motion vectors of blocks used for global estimation in blue

Accurate Measurement of Parking Space

Motion vectors of blocks used for global estimation in blue

This will give us distance in pixels!

Translating MV (pixels) into distance (meters)

Cones placed at distances of 0.5 meters (11.5 m total)

Translating MV (pixels) into distance (meters)

Cones placed at distances of 0.5 meters

Recorded different video sequences while driving across these cones at various speeds ranging from 5 to 50 km/h.

Translating MV (pixels) into distance (meters)

Cones placed at distances of 0.5 meters

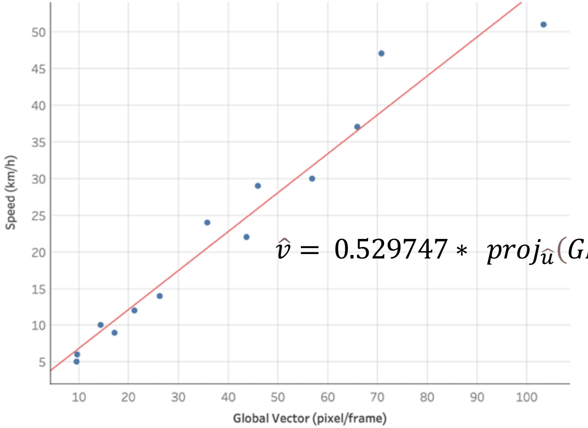
Recorded different video sequences while driving across these cones at various speeds ranging from 5 to 50 km/h.

THE CORRELATION BETWEEN SPEED AND GLOBAL MOTION

Speed	Global Vector
5	9.5
6	9.9
9	17.1
10	14.4
12	21.1
14	26.2
22	35.8
24	43.7
29	46
30	56.9
37	65.9
47	70.8
51	103.4

Translating MV (pixels) into distance (meters)

Cones placed at distances of 0.5 meters

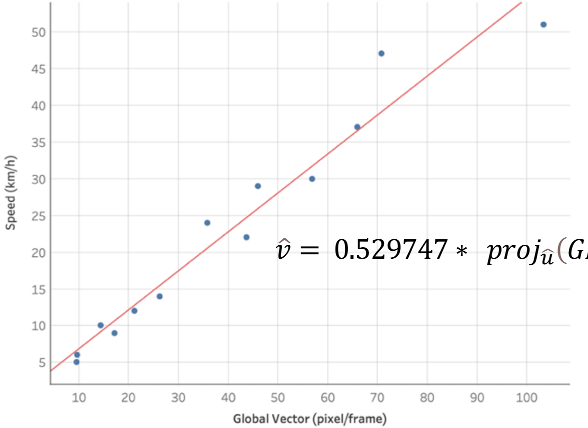


$$\hat{v} = 0.529747 * proj_{\hat{u}}(GMV) + 1.57119$$

Linear relationship between speed and global vector

Translating MV (pixels) into distance (meters)

Cones placed at distances of 0.5 meters



$$\hat{v} = 0.529747 * \ proj_{\hat{u}}(GMV) + 1.57119$$

Linear relationship between speed and global vector

Now we can compute the number of meters the car moved per frame: $$d_m = \frac{1}{\frac{30\ frames}{second} * \frac{second}{meter}}$$

Experimental Results

- Accuracy of measuring parking size: **91.4%**

Speed (km/h)	Ground Truth Distance (meters)	Accumulated Distance using Motion (meters)	Accuracy (%)
5	11.5	13.2	85.22
6	11.5	13.1	86.09
9	11.5	12.9	87.83
10	11.5	12.4	92.17
12	11.5	10.9	94.78
14	11.5	11.9	96.52
22	11.5	12.1	94.78
24	11.5	10.2	88.70
29	11.5	10.9	94.78
30	11.5	12.4	92.17
37	11.5	11.7	98.26
47	11.5	10.2	88.70
51	11.5	12.9	87.83
			Average: 91.4%

Experimental Results

- Accuracy of measuring parking size: **91.4%**

Speed (km/h)	Ground Truth Distance (meters)	Accumulated Distance using Motion (meters)	Accuracy (%)
5	11.5	13.2	85.22
6	11.5	13.1	86.09
9	11.5	12.9	87.83
10	11.5	12.4	92.17
12	11.5	10.9	94.78
14	11.5	11.9	96.52
22	11.5	12.1	94.78
24	11.5	10.2	88.70
29	11.5	10.9	94.78
30	11.5	12.4	92.17
37	11.5	11.7	98.26
47	11.5	10.2	88.70
51	11.5	12.9	87.83
			Average: 91.4%

- Overall accuracy of detecting available street parking space: **87.5%**

YOLOv7-X model → Global MV → Distance measure

Our end-to-end system

Digital Multimedia Lab

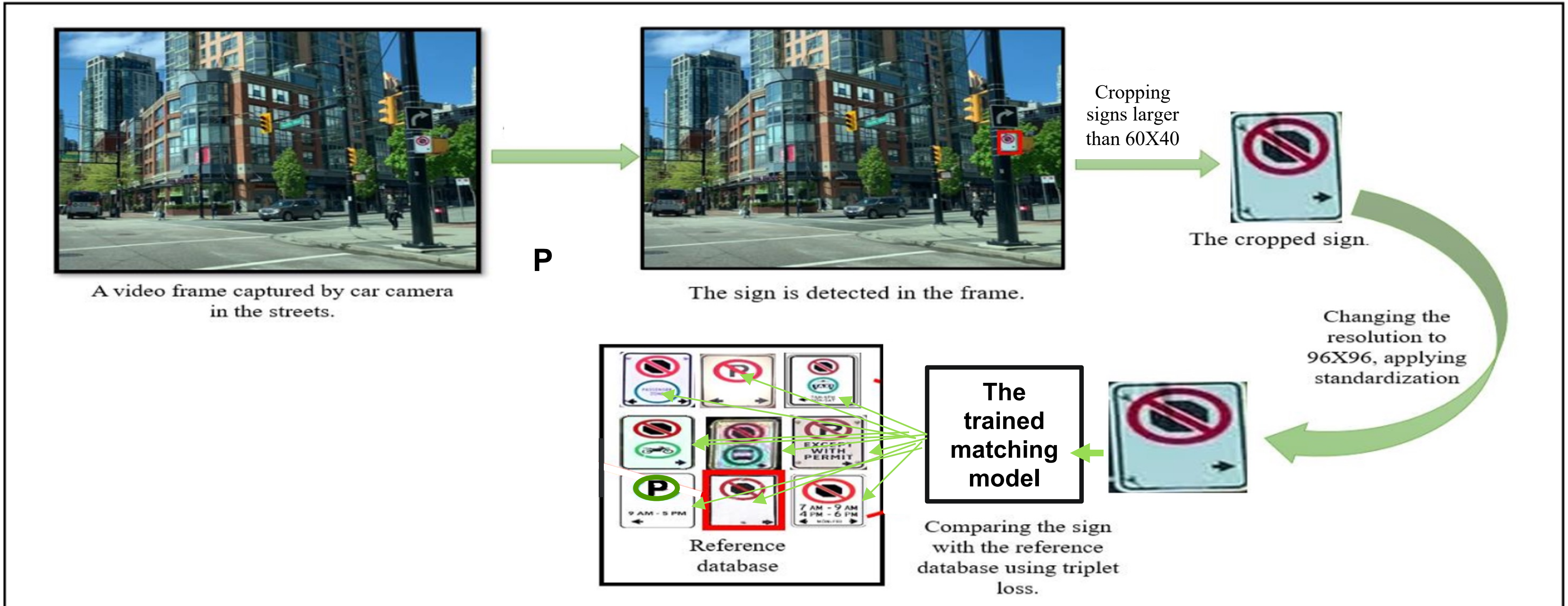
Combining Parking Detection with Parking Sign Recognition

Electrical and Computer Engineering

IARIA

Detecting Parking Signs on a Test Video

Parking Sign Detection & Identification Pipeline

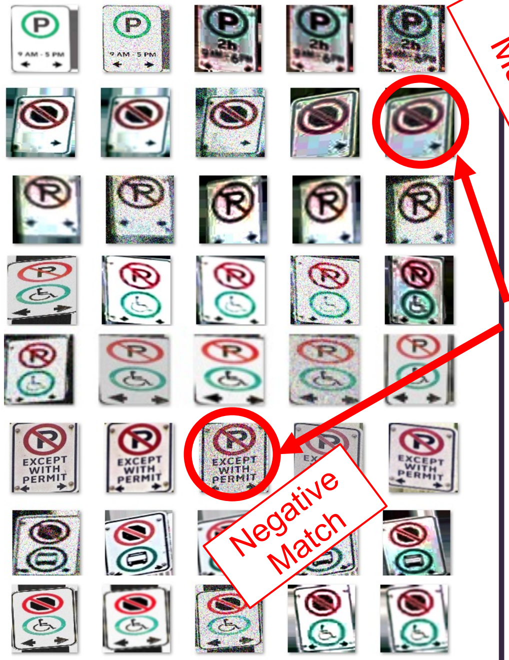


Parking sign detection and identification pipline.

Parking Sign Detection & Identification Pipeline

Training Phase

Cropped Signs Dataset
(3651 cropped signs - Labeled)



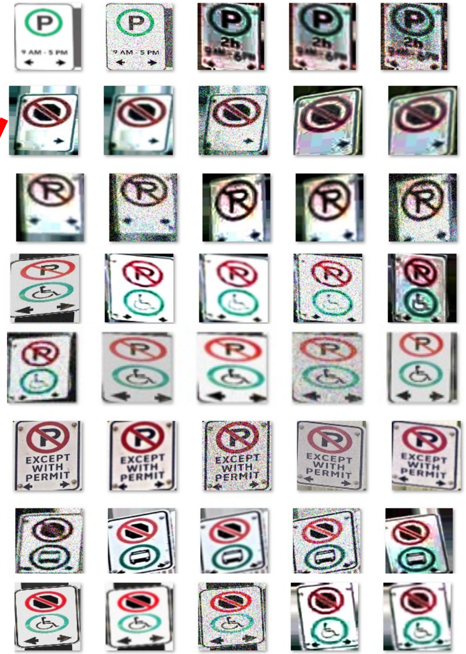
.....
.....

Triplet Loss

Our matching network (SqueezeNet)

Anchor

Cropped Signs Dataset
(3651 cropped signs - Labeled)

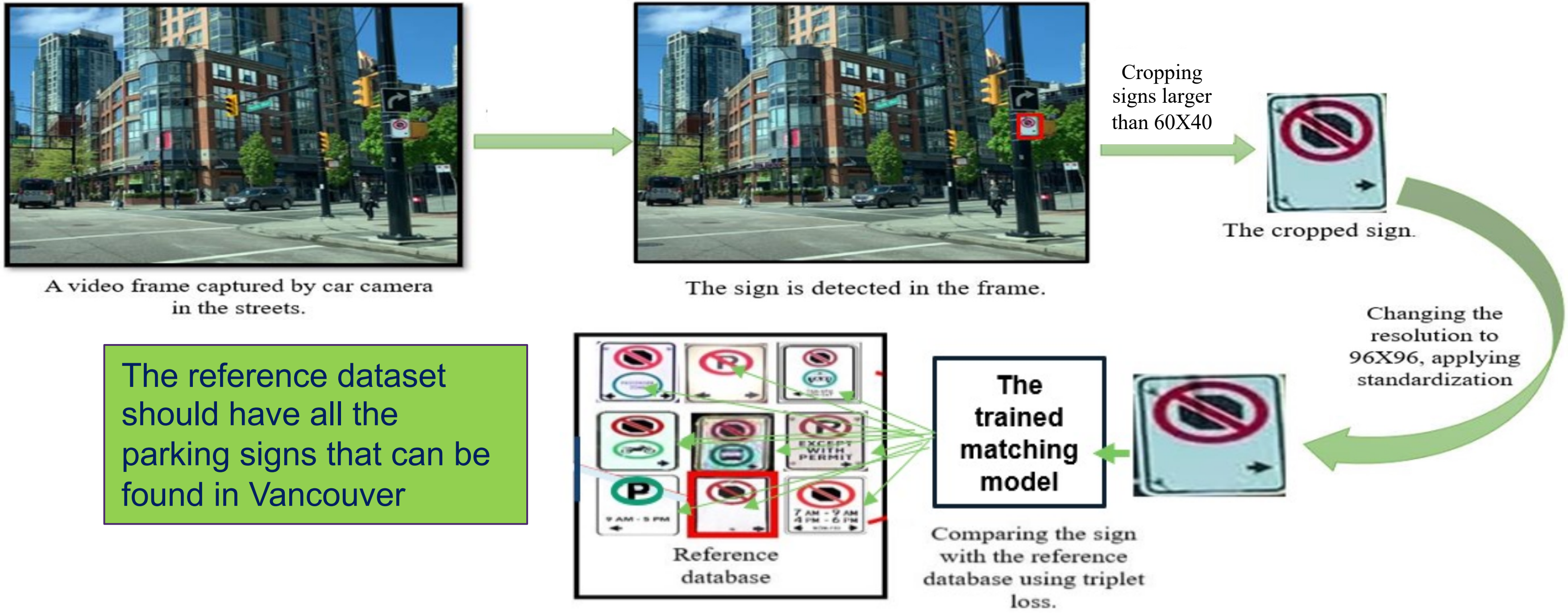


.....
.....

Triplet Loss

1. Force similarity between Anchor and Positive images
2. Force distance between Anchor and Negative images

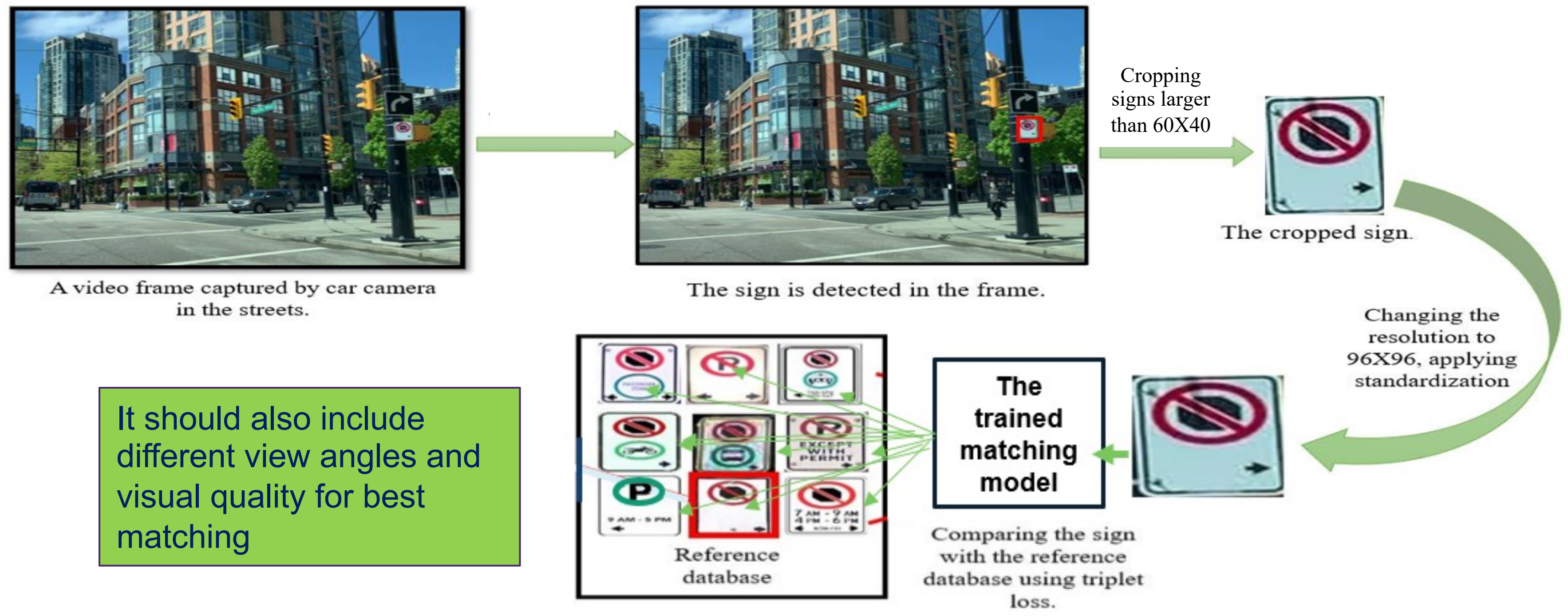
Parking Sign Detection & Identification Pipeline



Parking sign detection and identification pipline.

The reference dataset should have all the parking signs that can be found in Vancouver

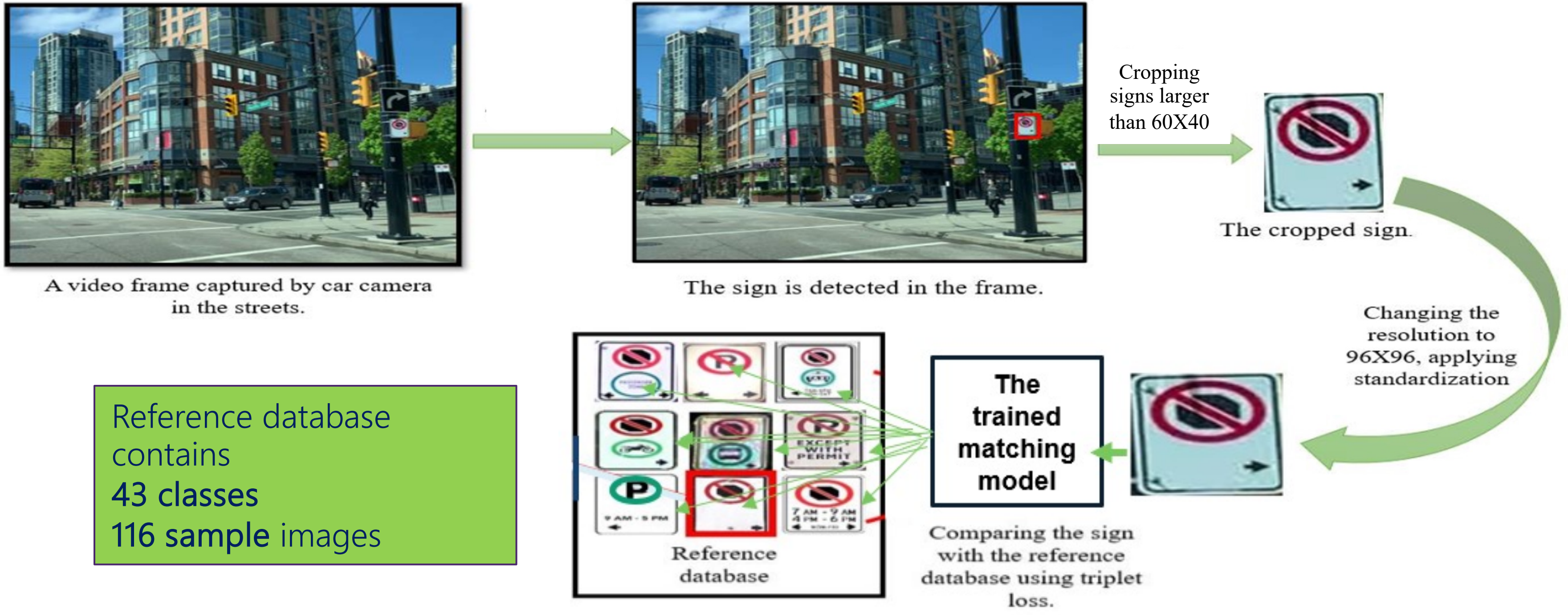
Parking Sign Detection & Identification Pipeline



It should also include different view angles and visual quality for best matching

Parking sign detection and identification pipline.

Parking Sign Detection & Identification Pipeline

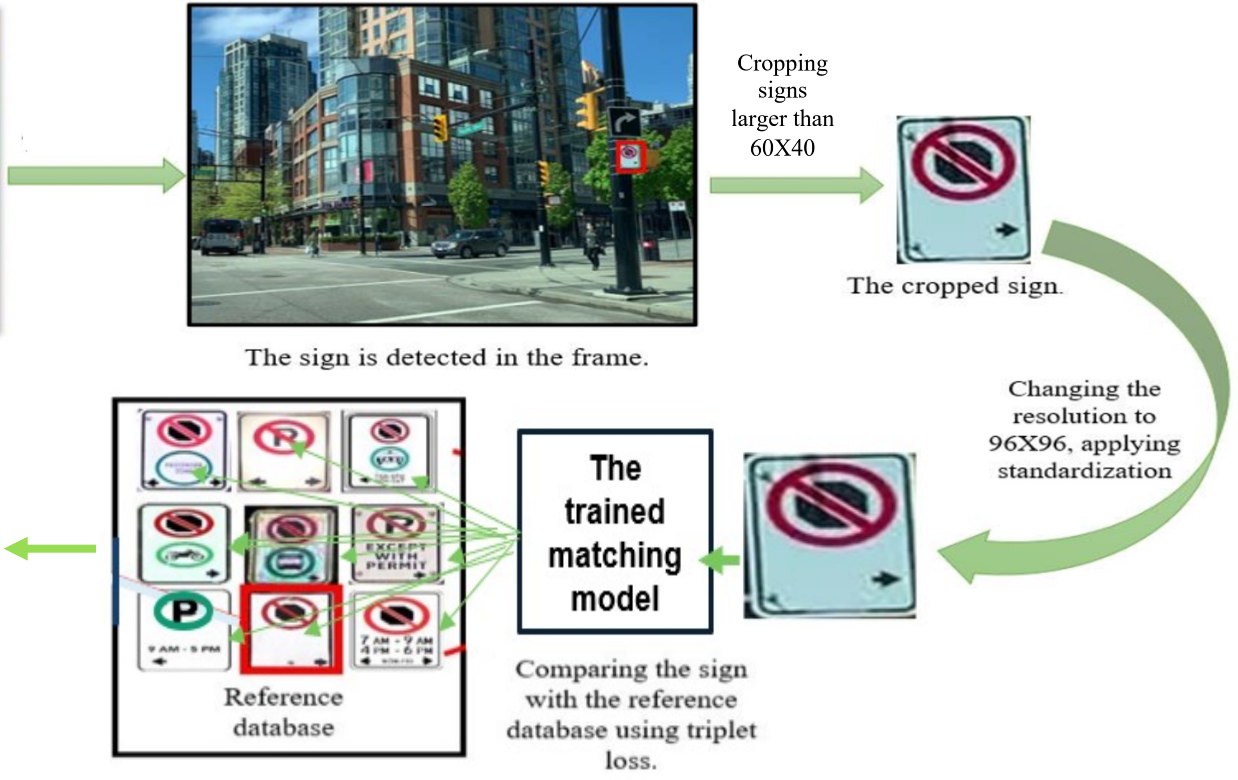


Parking sign detection and identification pipline.

Reference database contains
43 classes
116 sample images

Parking Sign Detection & Identification Pipeline

A video frame captured by car camera



Name of the sign presented in the reference dataset	Similarity Scores
No stopping before the sign	*6.63*
No parking before the sign	*6.43*
No stopping on both sides	6.12
…	…

81%

Parking sign detection and identification pipline.

"n" consecutive frames showing the same sign captured by a car camera

Detected signs in the frames.

Time axis

“n” consecutive frames showing the same sign captured by a car camera

YOLOv7X Object detection model

Time axis

Detected signs in the frames.

Crop signs larger than 60X40

Cropped signs.

Change resolution to 96X96

Compare signs with the reference dataset using triplet loss.

Reference Dataset
116 reference pictures

81%

Which sign matches best the input?

"n" consecutive frames showing the same sign captured by a car camera

Time axis

YOLOv7X Object detection model

Detected signs in the frames.

Time axis

Crop signs larger than 60X40

Cropped signs.

Change resolution to 96X96

Compare signs with the reference dataset using triplet loss.

Which sign matches best the input?

Reference Dataset
116 reference pictures

81%

Summing similarities over the cropped pictures

Name of the sign presented in the reference dataset	Accumulated Similarity Scores (116)
Time-limited no stopping On the right side	5.33 *(highest score)*
Time limited no parking on the tight side	*5.1*
No stopping on the right side	4.96
…	…
….	…

85%

“n” consecutive frames showing the same sign captured by a car camera

Time axis

YOLOv7X Object detection model

Time axis

Detected signs in the frames.

Crop signs larger than 60X40

Cropped signs.

Change resolution to 96X96

Compare signs with the reference dataset using triplet loss.

Which sign matches best the input?

Reference Dataset
116 reference pictures

81%

Summing similarities over the cropped pictures

Name of the sign presented in the reference dataset	Accumulated Similarity Scores (116)
Time-limited no stopping On the right side	5.33 *(highest score)*
Time limited no parking on the tight side	*5.1*
No stopping on the right side	4.96
…	…
….	…

85%

Applying Text Detection

Time limited no stopping

88%

ece Electrical and Computer Engineering

IARIA

Sign Identification: Text Detection

- We used **MMOCR** framework for text detection

"n" consecutive frames showing the same sign captured by a car camera

Time axis

YOLOv7X Object detection model

Detected signs in the frames.

Time axis

Crop signs larger than 60X40

Cropped signs.

Change resolution to 96X96

Enhance the quality of images

Compare signs with the reference dataset using triplet loss.

Which sign matches best the input?

The Reference Dataset (118 reference pictures)

81%

Summing similarities over the cropped pictures

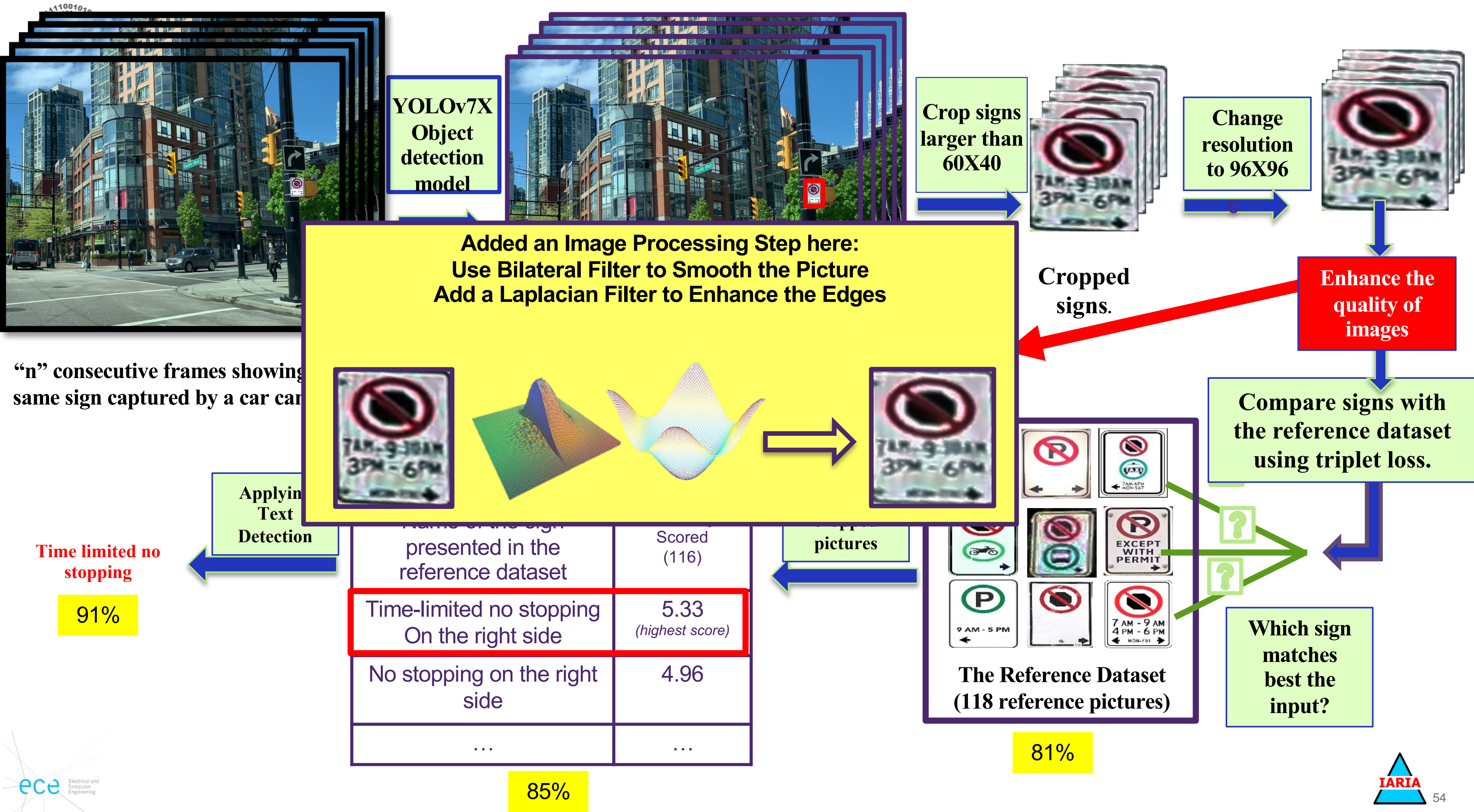
Name of the sign presented in the reference dataset	Accumulated Similarity Scores (116)
Time-limited no stopping On the right side	5.33 *(highest score)*
Time limited no parking on the tight side	*5.1*
No stopping on the right side	4.96
…	…
….	…

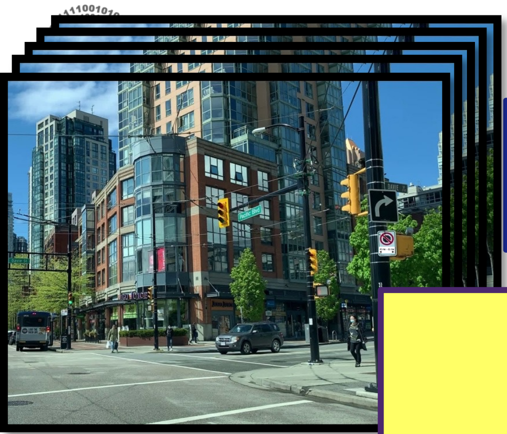
85%

Applying Text Detection

No Text!

88%


YOLOv7X Object detection model
Crop signs larger than 60X40
Change resolution to 96X96
Added an Image Processing Step here:
Use Bilateral Filter to Smooth the Picture
Add a Laplacian Filter to Enhance the Edges
Cropped signs.
Enhance the quality of images
"n" consecutive frames showing
same sign captured by a car car
Compare signs with the reference dataset using triplet loss.
Applyin
Text
Detection
Time limited no stopping
91%
presented in the reference dataset
Scored (116)
Time-limited no stopping On the right side
5.33 (highest score)
No stopping on the right side
4.96
…
…
85%
pictures
EXCEPT WITH PERMIT
9 AM - 5 PM
7 AM - 9 AM 4 PM - 6 PM MON-FRI
The Reference Dataset (118 reference pictures)
81%
Which sign matches best the input?



“n” consecutive frames showi
same sign captured by a car ca

YOLOv7X Object detection model

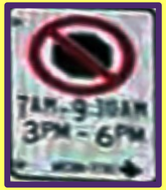
Crop signs larger than 60X40

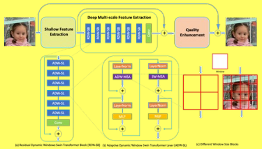
Cropped signs.

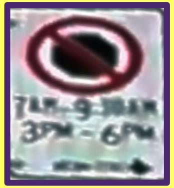
Change resolution to 96X96

Enhance the quality of images

Added Deep Learning Method:
Used SwinIR to clean and enhance the images
(same result as the image processing step)







Compare signs with the reference dataset using triplet loss.

The Reference Dataset (118 reference pictures)

81%

Which sign matches best the input?

pictures

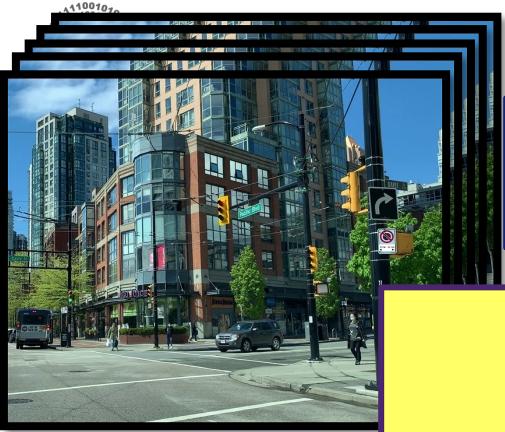
…	
time-limited no stopping	5.33 (highest score)
No parking	4.8
Time-limited no stopping sign	5.23
…	

85%

Apply Text Detection

Time limited no stopping

91%



"n" consecutive frames showin[g the]
same sign captured by a car ca[mera]

YOLOv7X Object detection model

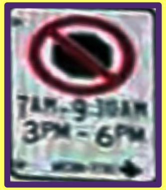
Crop signs larger than 60X40

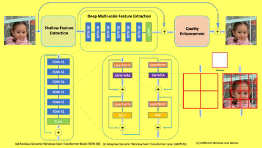
Cropped signs.

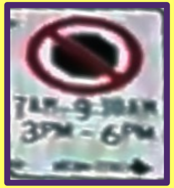
Change resolution to 96X96

Enhance the quality of images

Added Deep Learning Method:
Used SwinIR to clean and enhance the images
(same result as the image processing step)







Compare signs with the reference dataset using triplet loss.

The Reference Dataset (118 reference pictures)

81%

Which sign matches best the input?

[Cropped] pictures

Applyi[ng] Text Detection

…	
time-limited no stopping	5.33 (highest score)
No parking	4.8
Time-limited no stopping sign	5.23
…	

85%

Time limited no stopping

91%


Home
Restaurants
Coffee
W 16th Ave
W 16th Ave
word Fern Trail
Douglas Fir Trail
mpton Pl
Google
Rufus Guitar Shop
Lessons
Repairs
2621 ALMA ST
604-22


Parking sign: 0.90
Parking: 0.97
NORTH SOUTH TRAV
Home
Restaurants
Coffee
23°
W 16th Ave
W 16th Ave
Sherry-Sakamoto-Trail-(30)
Google

Thank you!!

tbazzaza@ece.ubc.ca