An Efficient YOLOv7x Based Automated Street Parking Space Detection for Smart Cities

Tala Bazzaza, Hamid Reza Tohidypour, Yixiao Wang, and Panos Nasiopoulos

The University of British Columbia, Canada
TELUS Communications Inc., Canada
Contact email: tbazzaza@ece.ubc.ca

Research

- Deep Learning
- Computer Vision
- Motion Detection

Education

- M.A.S.c 2020-present; UBC
- B.Sc. 2016-2020; UOS

Work

- Research Assistant; DML
- Teaching Assistant; UBC

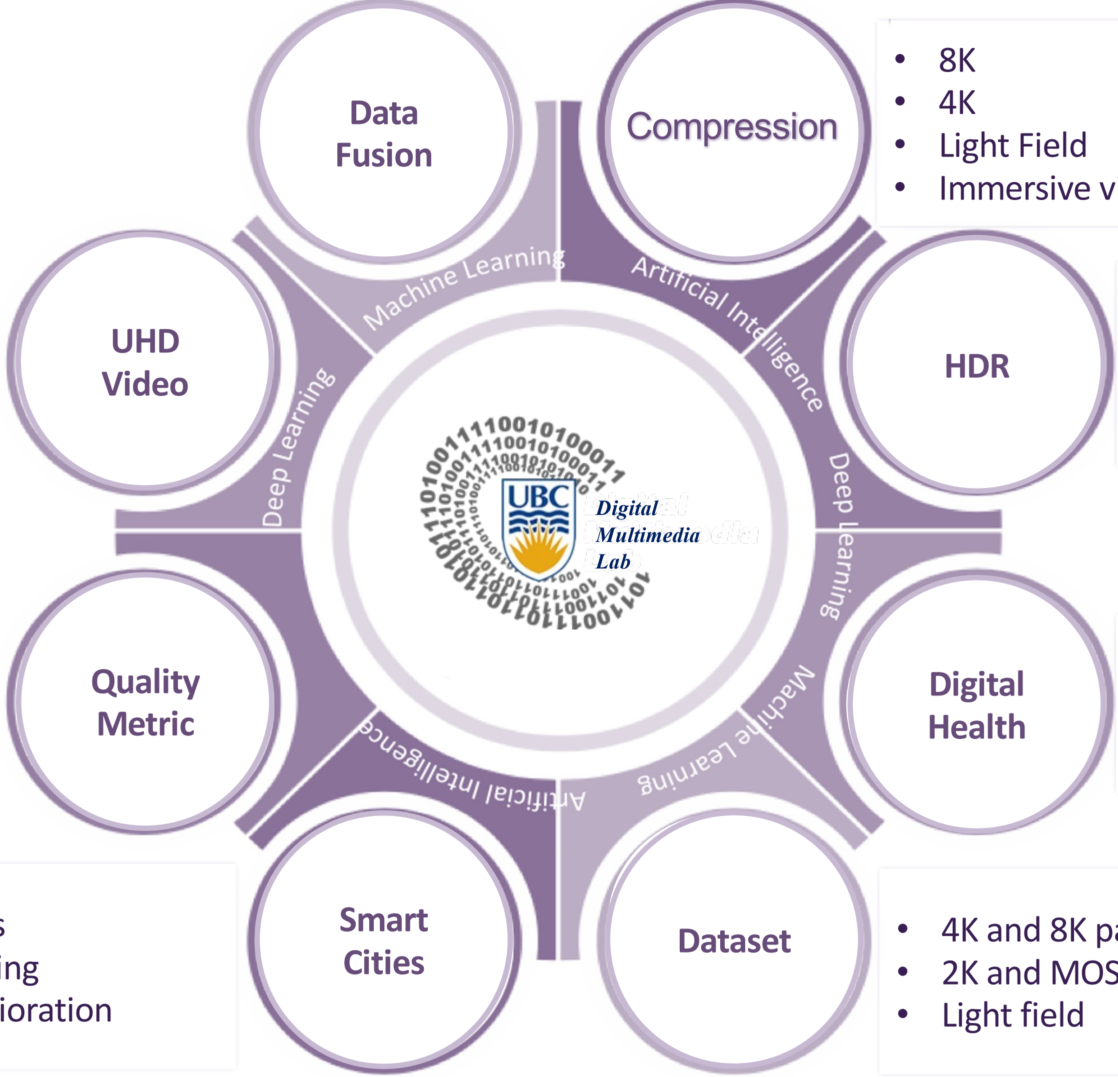
Skills

- Deep Learning
- UI design and front-end development
- Software development

Interests

- Movies
- Travelling
- Ice skating
- Skiing

ece Electrical and Computer Engineering



- 8K
- 4K
- Light Field
- Immersive video

- Tone Mapping
- Inverse Tone Mapping

- Fall Detection
- Medical Image processing

- 4K and 8K pairs
- 2K and MOS pairs
- Light field

- Street Signs
- Street Parking
- Road Deterioration

- DL based video quality
- LF spatial and angular quality

- 4K, 8K, LF capture
- Calibration
- Rendering

Finding street parking spaces is becoming increasingly difficult, time and resource consuming

Electrical and Computer Engineering

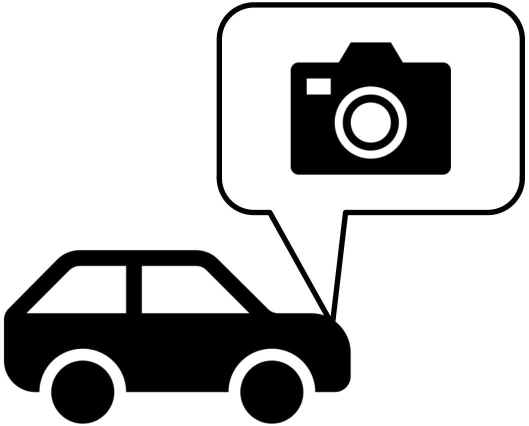
Introduction

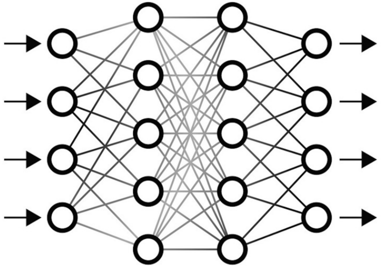
- Smart city deployment includes addressing the congestion in cities
- Automatic detection and recognition of street parking in metropolitan cities is a key component of a smart city infrastructure
- This is essential for autonomous driving
- It helps to notify drivers of available parking spaces ahead, thus significantly alleviating traffic congestion

© Drew Angerer/Getty Images

Our Approach

- Deep learning based real-time occupancy detection system, using a video feed from vehicle-mounted cameras.
    - to accurately detect the available parking spots in real-time.





Yolo-v7 object detection network

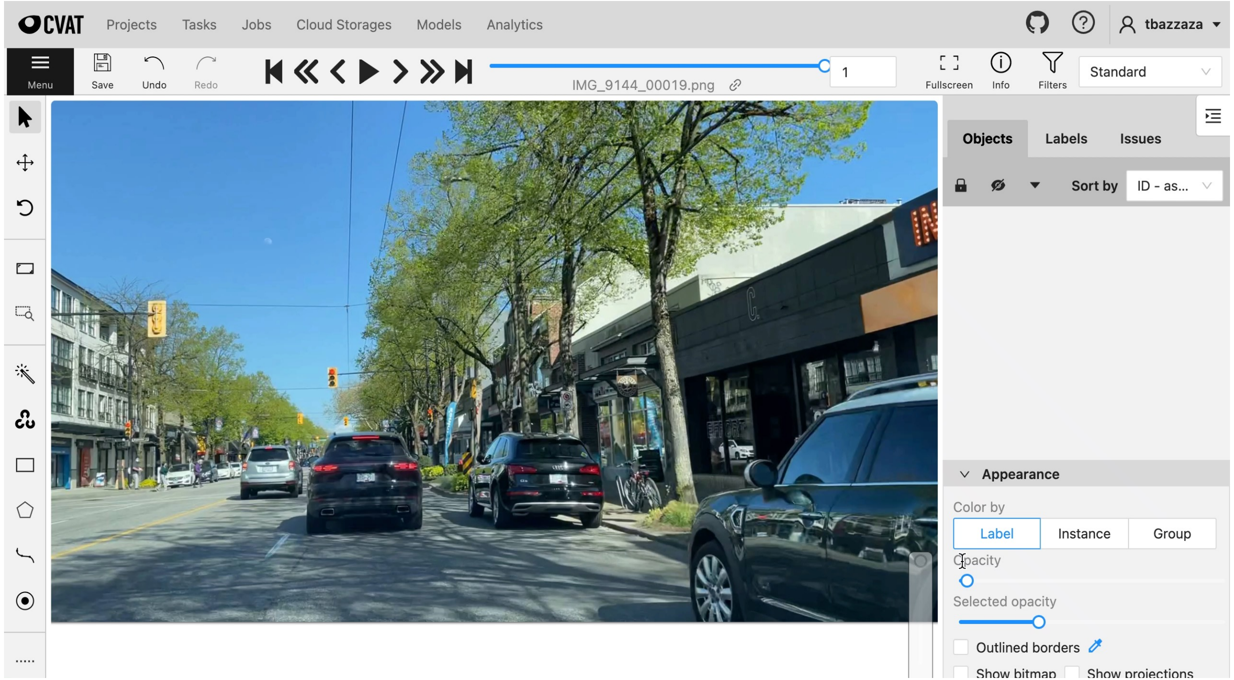
Dataset: Data Collection

- 53 videos were captured by driving around the streets of the city of Vancouver, Canada

- The dataset was carefully curated to include a diverse range of weather conditions (sunny, cloudy, rainy, snowy)

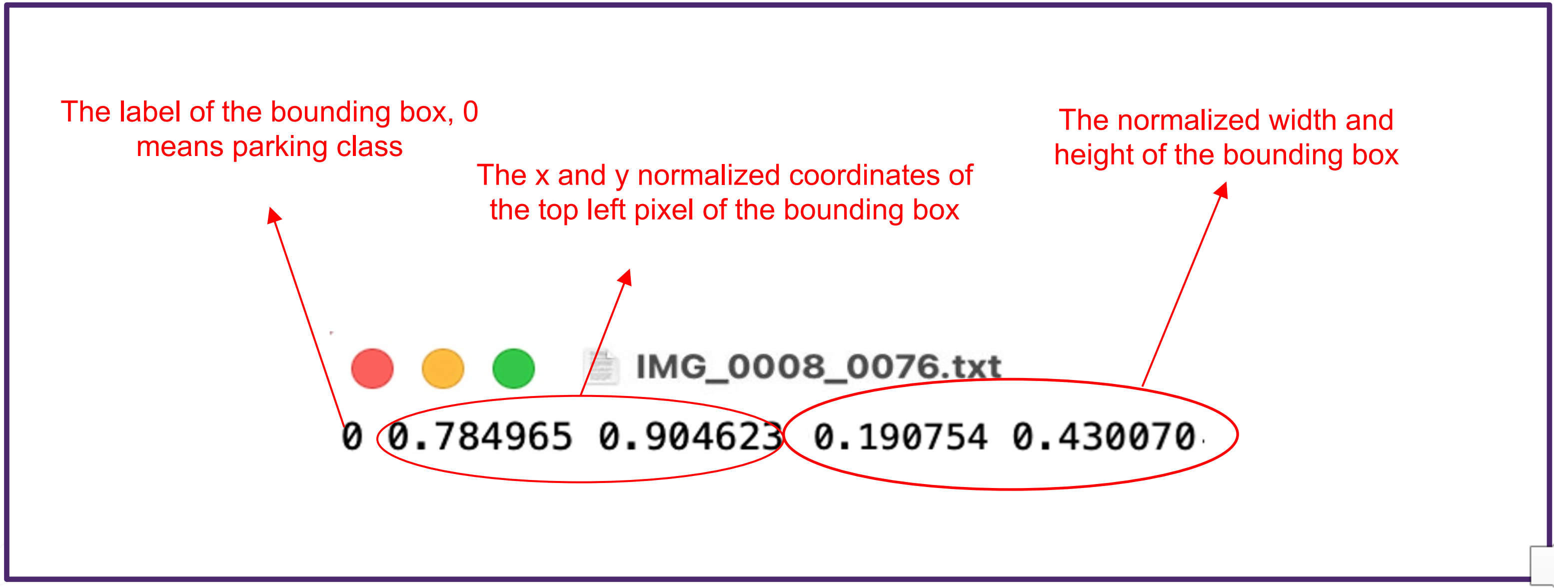
Dataset: Labeling Approach

- The frames were extracted using FFMPEG, and the Computer Vision Annotation Tool (CVAT) was used for the labeling.
- We drew bounding boxes around parking spots in the extracted frames and labeled them as parking. We used one class for available parking spots.



Dataset: Labeling Approach

The labeled data was exported in YOLO format. Therefore, for each image in the dataset, there is a text file containing the coordinates of the bounding box and a number indicating the label(class) of the bounding box



An example of the exported CVAT annotation text file

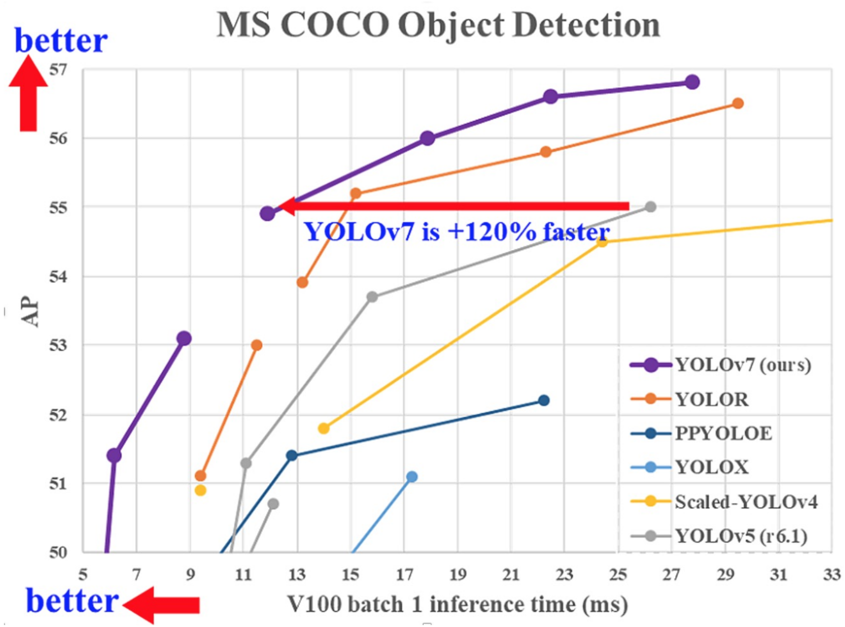
Our Solution

- We chose YOLOv7x as the object detection and classification neural network
  - Tested against other Deep Learning architectures
  - Showed the best performance for this task

Object Detection Network: You Only Look Once

- YOLO is a neural network architecture that is designed to perform object detection in real-time with high accuracy.
- It is a single, unified deep neural network that takes an entire image as input and predicts the locations and class probabilities of objects in the image
- In this paper, we compared Yolov4 against different Yolov7 versions

- YOLOv7 has a more efficient architecture that reduces the number of parameters by 75%, thus requiring 36% less computation and achieving 1.5% higher Average Precision (AP)



C.-Y. Wang and A. AB implementation of YOLOv7

Our YOLOv7 and Yolov7x model:

- We train two versions of the YOLOv7 family, the original YOLOv7 and the latest YOLOv7x.
- The main differences between the two is that YOLOv7 uses the method of stack scaling on the neck, which is a technique to increase the capacity of a model by adding more layers to it. In this case, this technique is applied to the "neck" of the model
- YOLOv7 uses an auto-anchor algorithm borrowed from YOLOv5, which adapts to the scale of the objects in an image by using a single anchor box that can adjust to different scales
- We also train the older Yolov4 model to perform a comparison, and further discuss its limitations on our results.

Training

- We extracted a total of 3381 frames from the 53 videos that we collected
- We randomly assigned 80% and 20% of our 3381 unique frames into training frames and validation frames, respectively.
- While Yolov7 uses an auto anchor algorithm, Yolov4 uses k-means clustering for generating anchors which requires manually calculation of anchors. We generated anchor boxes before starting the training process

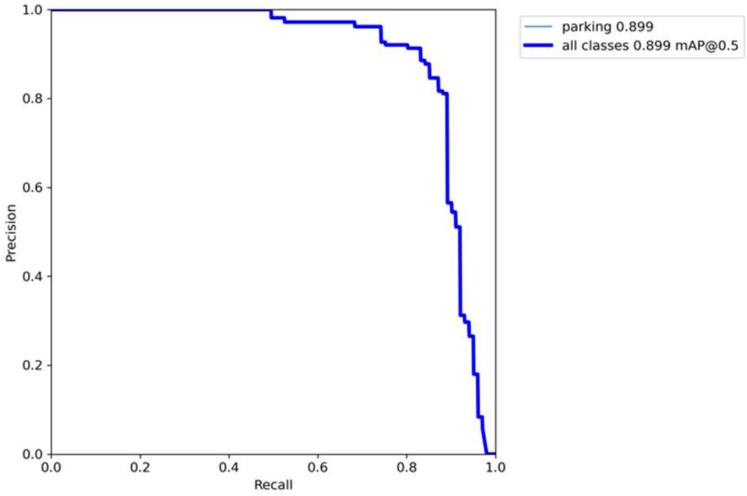
Our validation results

- Performance evaluation and accuracy of the three models was done using the mean average precision (mAP) metric.
- Average precision is calculated by measuring the precision and recall of the model at different intersection-over-union (IoU) thresholds, which is the ratio of the area of overlap between the predicted bounding box and the ground-truth bounding box to the area of the union of the two boxes.

Model	mAP@0.5
Yolov4	82%
Yolov7	84%
Yolov7x	90%

Test Performance

- In order to evaluate the performance of our models against that of YOLOv4 for unseen test data, we tested all of them on 278 previously unseen frames captured by our lab in the city of Vancouver.
- We observe that YOLOv7x achieves the best performance, reaching a mAP of 89% at a threshold of 50% overlap between the labeling bounding box and the predicted one.



Yolov7x Precision-Recall curve

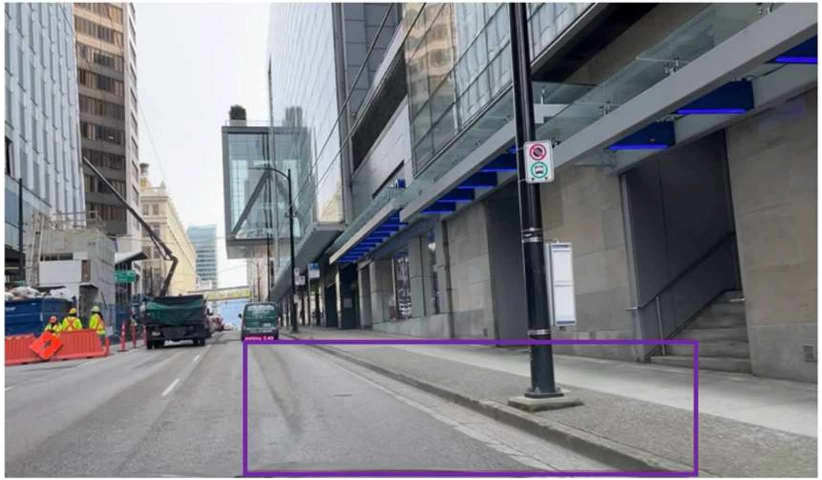
Test Performance Comparison

- The table below shows results of detecting vacant parking spots on the three models. We observe that YOLOv7x achieves the best performance.

Model	Precision	Recall	mAP @ 0.5
Yolov4	0.84	0.81	83.3%
Yolov7	0.90	0.79	86.5%
Yolov7x	0.91	0.81	89.9%

An Example of Detecting Parking Spots on a Test Video

Example 1: Yolov4 vs Yolov7



An example where YOLOv4(left) testing failed to identify some parking spots whereas YOLOv7x(right) successfully detected parking spaces.

Example 2: Yolov4 vs Yolov7

An example where YOLOv4(left) detected a parking area more than once whereas YOLOv7x(right) successfully detected a single parking area.

Conclusion

- We introduced a deep learning-based detection method for **detecting street parking spots**.
- We developed a **unique labeling scheme**, which allowed us to **detect** vacant spaces in the parking lot.
- We compared different versions of Yolov7 and Yolov4 object detection networks
- The Yolov7x achieved a detection accuracy of **89.9%** for the parking class on the test set.
- As our training and validation dataset was only 3381 frames, we are confident that this network can achieve even higher accuracy with a larger dataset and the same labeling scheme.

Thank you!

Q/A

tbazzaza@ece.ubc.ca