Traffic Violation Detection

Joseph Rosenblum

DSI-US-13 EC Capstone

Contents

Executive Summary

Context	• Many municipalities make material investments into traffic monitoring cameras to reduce the man-hours of traffic police, increase revenues though accurate detection of violations, and save pedestrian lives. Yet, without significant updates to software, it is challenging to unlock the full potential of traffic cameras.
Objective	• Utilize object detection software to interpret moving vehicle traffic violations from available camera data.
Methodology	• A data format resembling one currently available to The City of New York (of even lower resolution and quality) was used to train an object-detection model. The model was trained to specifically identify vehicles and signals, and the output was used to assess whether vehicles were stopped in a pedestrian crosswalk during a red light.
Findings	• This approach is viable. While not robust enough on its own, there is a clear path forward to improving the model's performance, which would enable it to both become more reliable and to detect additional violations.

Problem

Today's traffic monitoring cameras are limited by their software

- Municipalities today are cash-strapped, with $360B in losses expected between 2020 and 2022
- There is a need to do more with less, and automation presents a material opportunity
- Traffic cameras can save municipalities policing labor costs, increase revenues from ticketed violations, and save lives
  - Some municipalities report red light cameras generating revenues in line with an average business
  - In addition, traffic cameras reduced red-light running fatalities by close to 25% in 14 of the most populated cities in the U.S.
- Despite the potential value-add of this infrastructure, the cameras can have high installation and operating costs, with an average price tag of ~$100K
- Municipalities must therefore weigh the cost-benefit analysis of purchasing such hardware
- The strongest case to be made for red light cameras is an ongoing improvement to the efficacy of the software they run on
- Today's camera software operates on sensors and can only detect a limited number of offenses
- Neural networks, however, can "upgrade" the functionality of cameras by enabling them to detect an expanded range of moving violations

Methodology

Create a video surveillance environment like one possible for a municipality to achieve and use it to identify at least one type of traffic violation.	
1	Selected dataset closest to real-life scenario based on image resolution and content
2	Trained the model to identify objects necessary to identify if a violation exists within a still image
3	Identified objects in a test set of images
4	Identified frames in which all components of a potential violation are present

Data

WebCamTDataset

- Sequences of images taken 1s apart from webcam videos
- Low resolution of 352 x 240 with grainy compression
- Many images of intersections from traffic-control-device height

Footage

Data Selection

- Images to train were selected based primarily on whether they include a pedestrian signal
- Image were also restricted to those which show cars approaching and preferably if they are driving through a crosswalk
- Out of 60K frames contained within the dataset, this project was trained on just over 1K

Annotations

Annotations are the labels put on objects to identify them for the model.

Vehicle annotations were converted, signal annotations were created

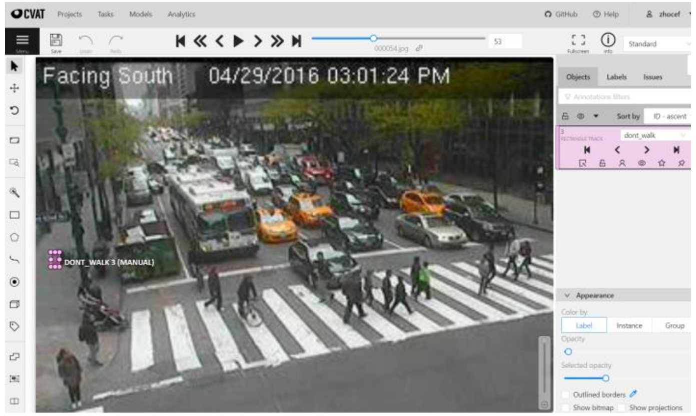
XML
Absolute coordinates
Structured with tags

YOLO
Relative coordinates
Structured as a table

XML is a ‘tree’ of tagged data and absolute coordinates

Because the original XML annotations are not in the correct form for the model, they were converted to YOLO

YOLO is a list of relative coordinates



YOLOv5x Model

Input

We are using the largest version of the model available, as it is the most accurate and still provides fast enough inference for this project

Training augmentation includes horizontal flips and mosaic

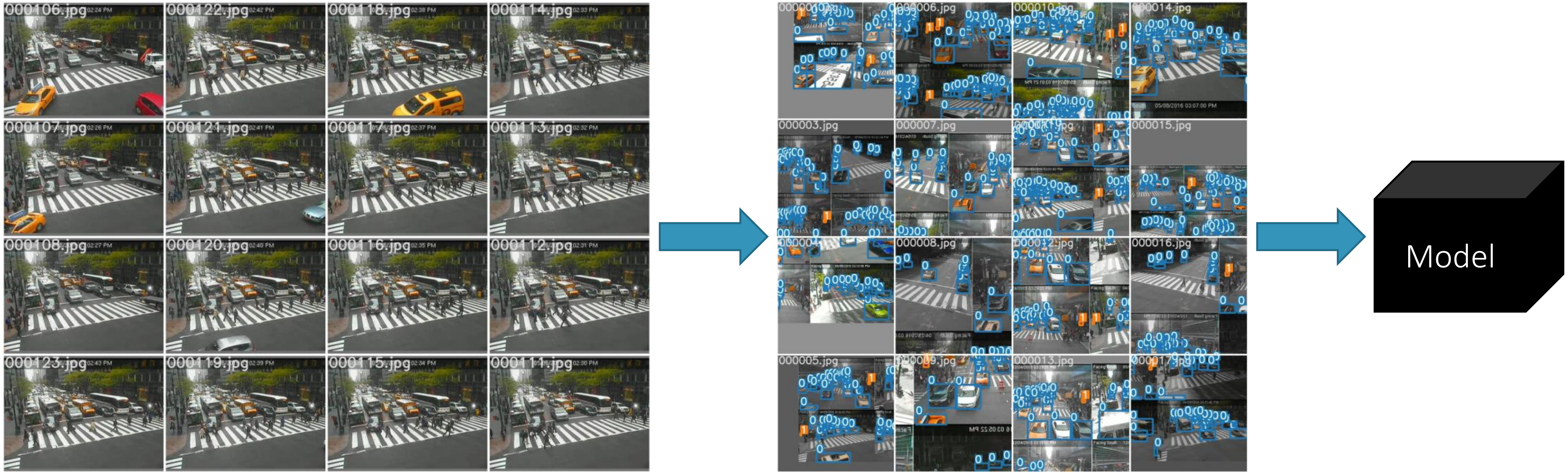
Output

Model will generate bounding boxes for objects identified

Evaluation

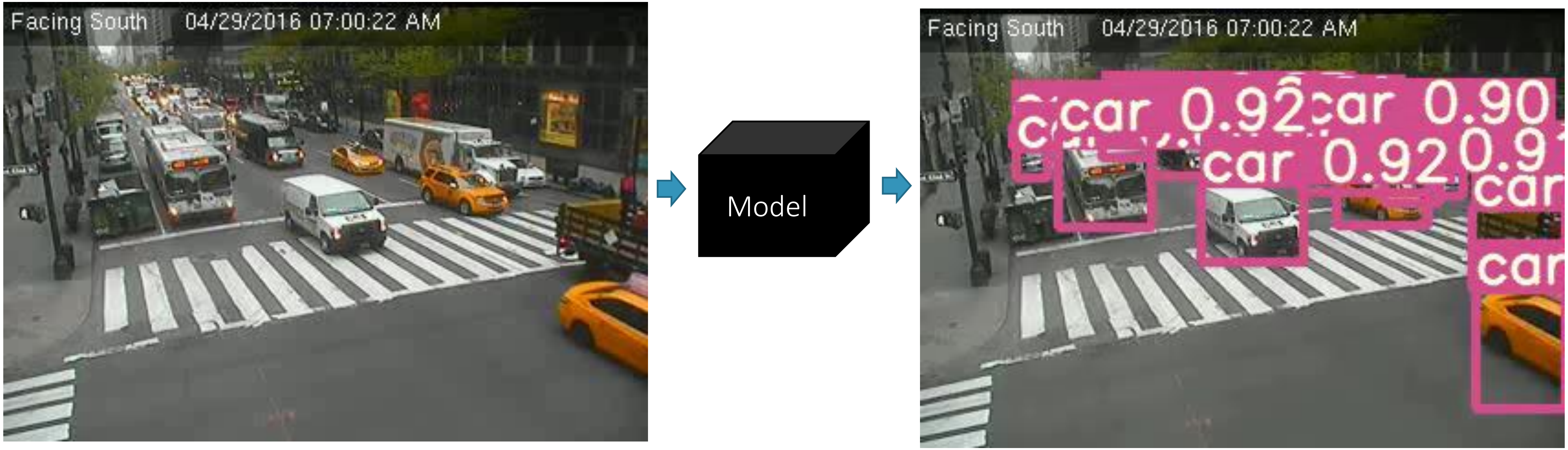
Model is evaluated by mean average precision (mAP)

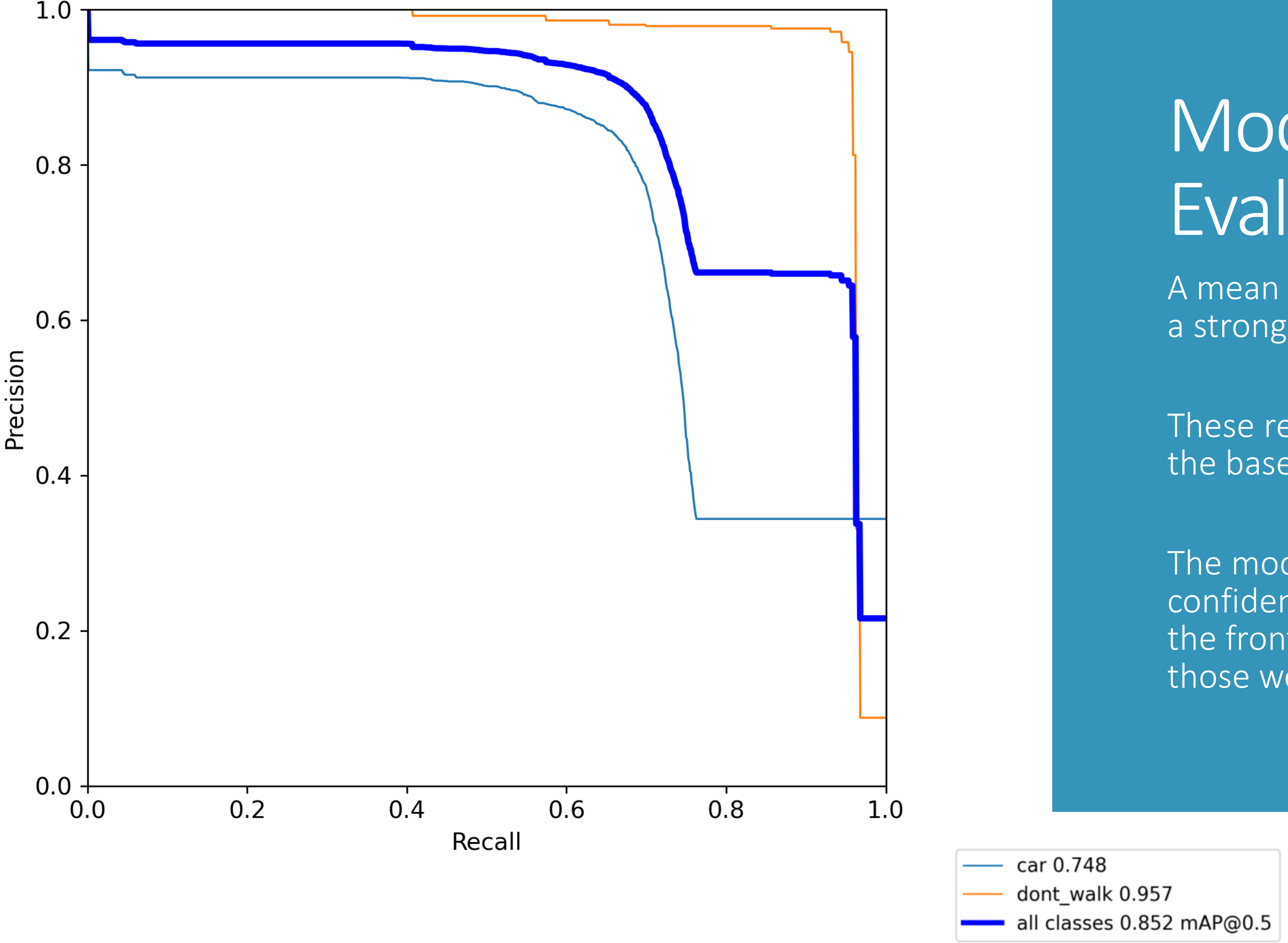
Training



000106.jpg
000122.jpg
000118.jpg
000114.jpg
000107.jpg
000121.jpg
000117.jpg
000113.jpg
000108.jpg
000120.jpg
000116.jpg
000112.jpg
000123.jpg
000119.jpg
000115.jpg
000111.jpg
000006.jpg
000010.jpg
000014.jpg
000003.jpg
000007.jpg
000015.jpg
000008.jpg
000012.jpg
000016.jpg
000005.jpg
000013.jpg
Model

Output


Facing South
04/29/2016 07:00:22 AM
Model
Facing South
04/29/2016 07:00:22 AM
car 0.92
car 0.90
car 0.92
car
car



Model Evaluation

A mean average precision 0f .852 is a strong result for this data.

These results should be considered the baseline.

The model has much higher confidence in classifying vehicles at the front of intersections, which are those we are most interested in.

Project Evaluation

Established criteria can be monitored and met:

- Vehicle within crosswalk
- Don’t walk signal is lit
- Conditions existed over x seconds

Conclusion

This goal of this project was to create a system that is capable of autonomously identifying traffic violations. This system, in its current state, does come very close, in that it narrows down frames of interest.

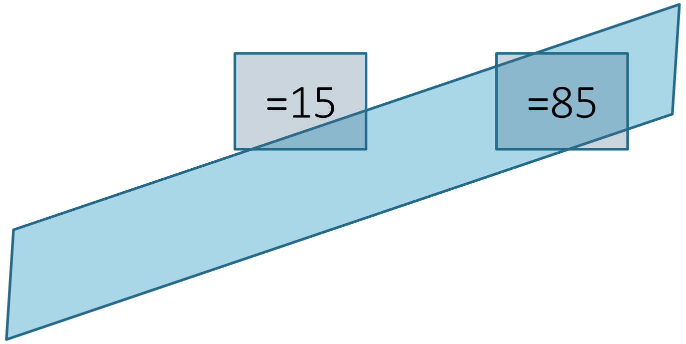
This establishes that there is much more possibility from this method, but it does not quite reach our goal of finding these occurrences completely without human validation.

I will continue to work on this project and already have some next steps I am excited to work on.

Next Steps

- Represent crosswalks as linear equation to account for crosswalks not being straight lines

- Use intersection score to judge how deep into the crosswalk a vehicle is rather than a binary determination
- Adding a method to tracking objects, rather than trying to determine action from a single image.
- Detect more objects, such as busses, bikes & pedestrians



Even Further Next Steps

- Detect the crosswalk or even street itself; use perspective to deepen understanding of what is happening
- There needs to be a plate reader incorporated to be able to issue a ticket. Much higher-resolution images would be needed to achieve this, which could be:
  - another layer of object detection that occurs after vehicle of interest is identified. Use a mask based on the bounding box of the vehicle to look for license plate / characters.
  - a vehicle sorted by its attributes can be attached to a license plate, once that has been established. Possibly from another camera. A pool of vehicles in an area can be tracked this way.
- Experiment with adding a classification model such as a gradient boost to evaluate violations