Hawkai

Deep Learning for Collision Detection

Saifullah Rais
Usha Challa
Daniel Elkin
Shahbakht Hamdani

W210 Capstone Project

Outline

1. Project Goal
2. Inspiration
3. Value Proposition
4. The Architecture
   - Data
   - Analytics
   - Serving
5. Performance Analysis
6. Demo
7. Q&A

Objective

- Use CCTV cameras and deep learning to automate the detection of motor vehicle collisions
- Reduce emergency service response times to save lives

Inspiration

- Survival rate dependent on Emergency Response
- More traffic cameras means less attention span
- Automated notifications can assist monitoring
- Improve emergency response time

1.3 mn
Yearly fatalities caused by car accidents globally

7-10
Median response time in U.S. in minutes (urban vs rural)

50 mn
Surveillance cameras in the US

37k
Fatalities in U.S.

13%
Fatalities prevented by reducing response time to the median

The Traffic Department's 'search' problem

24
Seconds
Per Fatality

13%
Reduced Fatalities
by improving
response time

---

50 mn
surveillance
cameras in US

3-35
Monitors per Operator

30%
Drop in Operator
accuracy (*)

Value Proposition

- Existing products focus on Automated Incident Detection, not specifically targeting vehicular collision, such as:
    - Congestion and stopped vehicles
    - Inclement weather
    - Lighting conditions

- We will focus on Motor Vehicle Collisions
    - Difficult problem due to rarity of events
    - Automated detection can decrease EMS dispatch times
    - Video can be used to determine severity of collision

The Architecture

Architecture: Data + Analytics + Serving

DATA

CCTV VIDEOS

- Caltrans
- YouTube

DATA REPOSITORY

- UNLABELLED FOOTAGE
- 1.6 MILLION VIDEOS

ANALYTICS

PRE PROCESSING

Background Subtraction	Motion Interaction Field
Grayscaling	Optical Flow

- SUPERVISED
- UNSUPERVISED

SERVING

- API
- BIGQUERY TABLE
- TABLEAU SERVER

The Data Layer:

- Youtube
- Live Traffic Cameras (Caltran)

The Data (lots of it!)

25 K
hours of live traffic footage recorded

150K
videos processed into optical flow versions

298 mn
Parameters trained in the Autoencoder

374
Caltran cameras recording live traffic feed

~1.5K
Car collisions within 1km radius of these cameras in July (estd.)

1600
US$ Outstanding on IBM Cloud Resources

Analytics Layer

Preprocessing Techniques - Examples

RAW IMAGE

GRAYSCALE

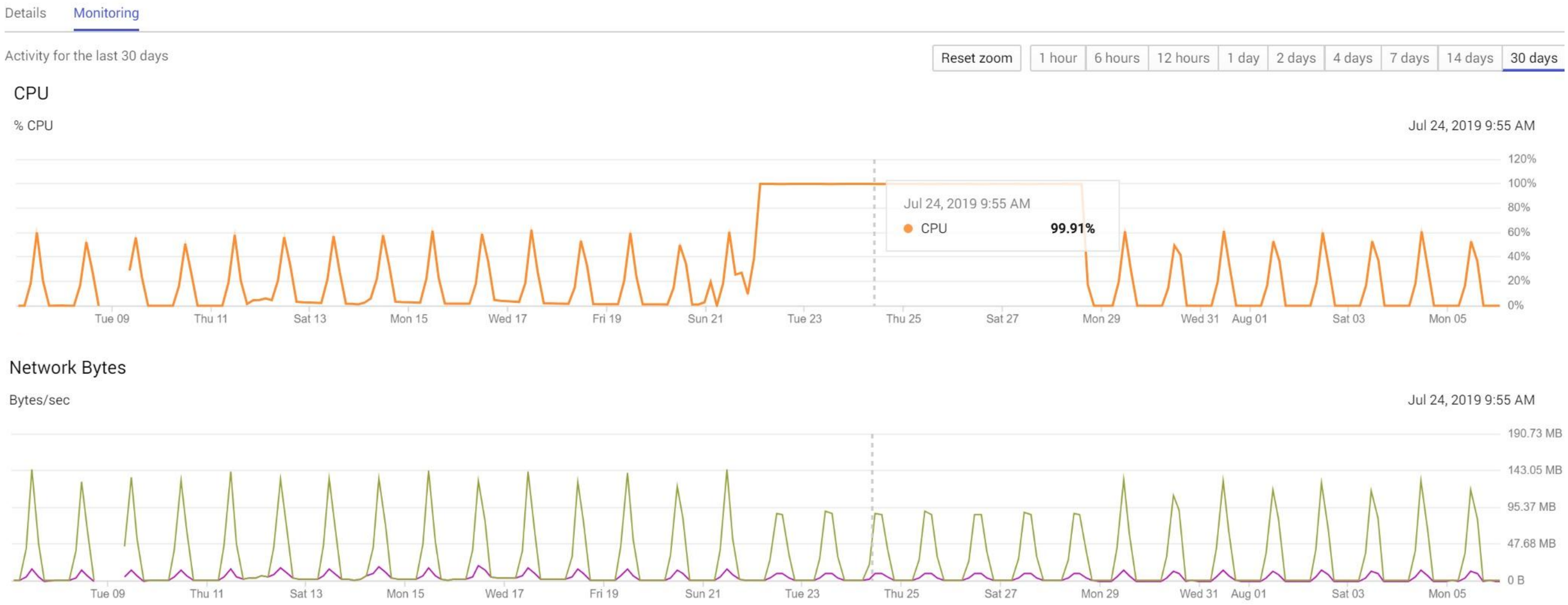
OPTICAL FLOW

BACKGROUND SUBTRACTED

Employed several preprocessing techniques to help improve our results, both for supervised and unsupervised learning.

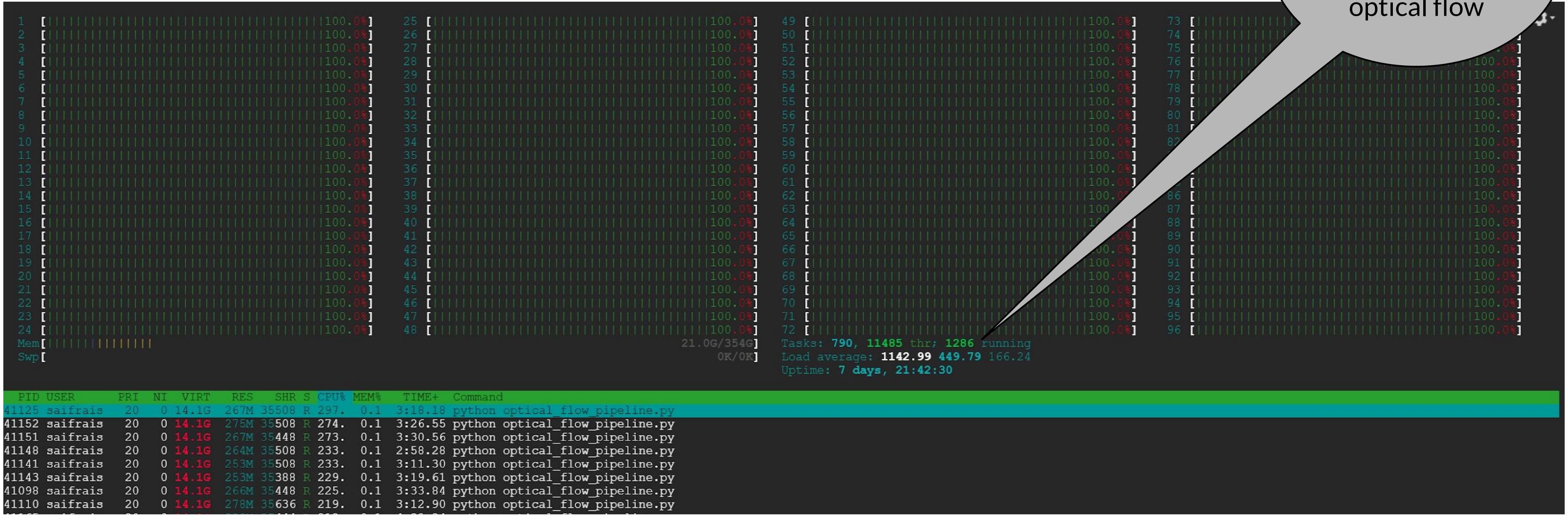
Computational Resources

- A 96-core server captured 1.6 mn minutes of traffic footage



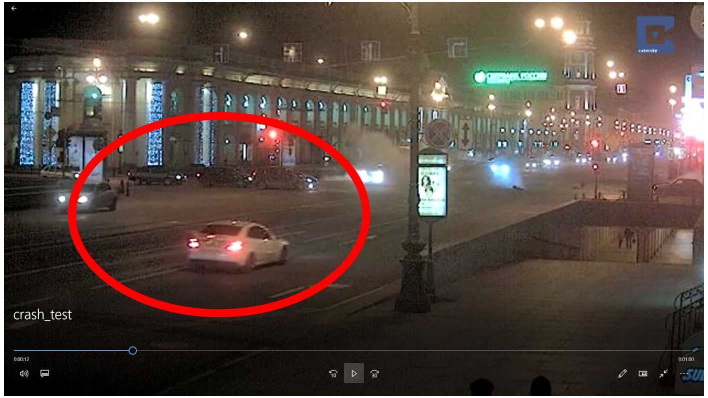
Computational Resources: Optical Flows

10% of our sample videos were converted into optical flow



Focusing on what matters!!

Optical flow captures erratic changes in motion and gives us triggers to detect anomalies



RAW IMAGE

OPTICAL FLOW

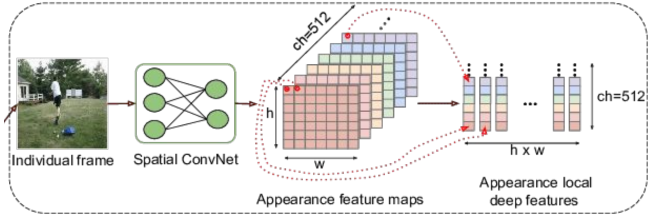
Analytics: Self-supervised Learning

Video Feature Extraction: YT8M Challenge

2,048
KB per
1-min video

YouTube | 8M

FEATURE EXTRACTOR

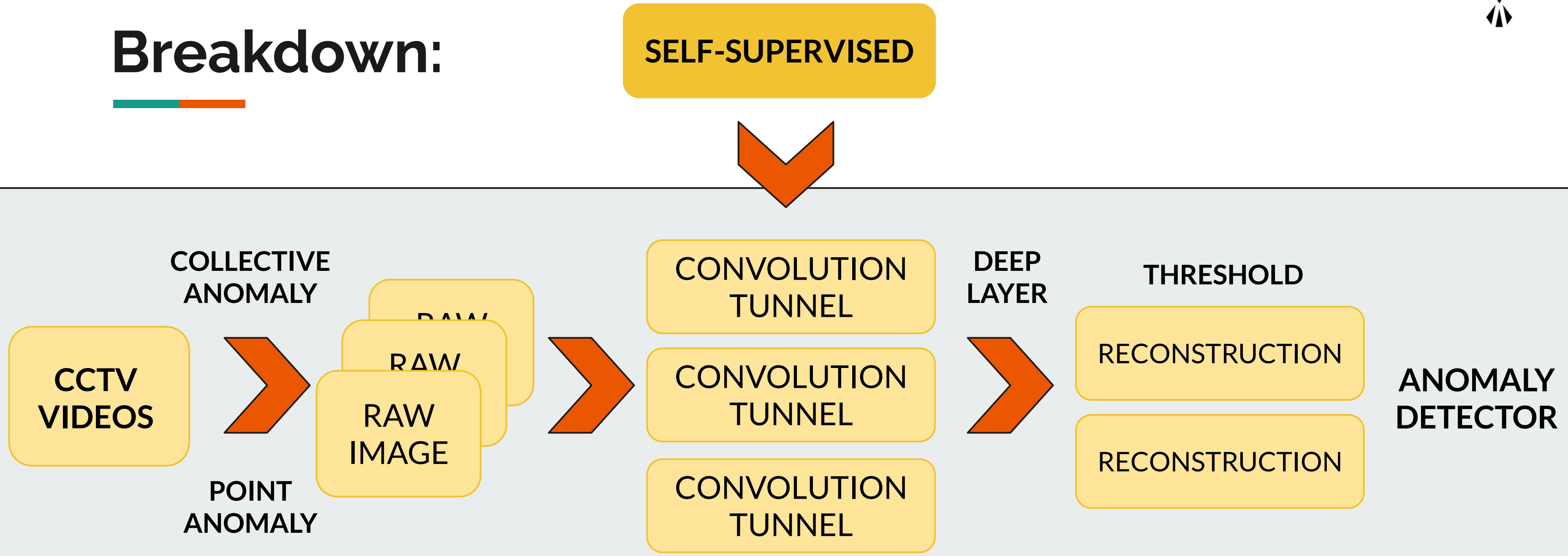


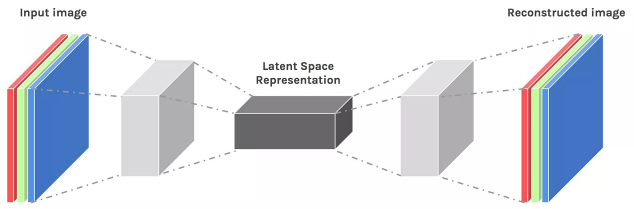
tfRecord
[1 x 1024]

0	0	1	1

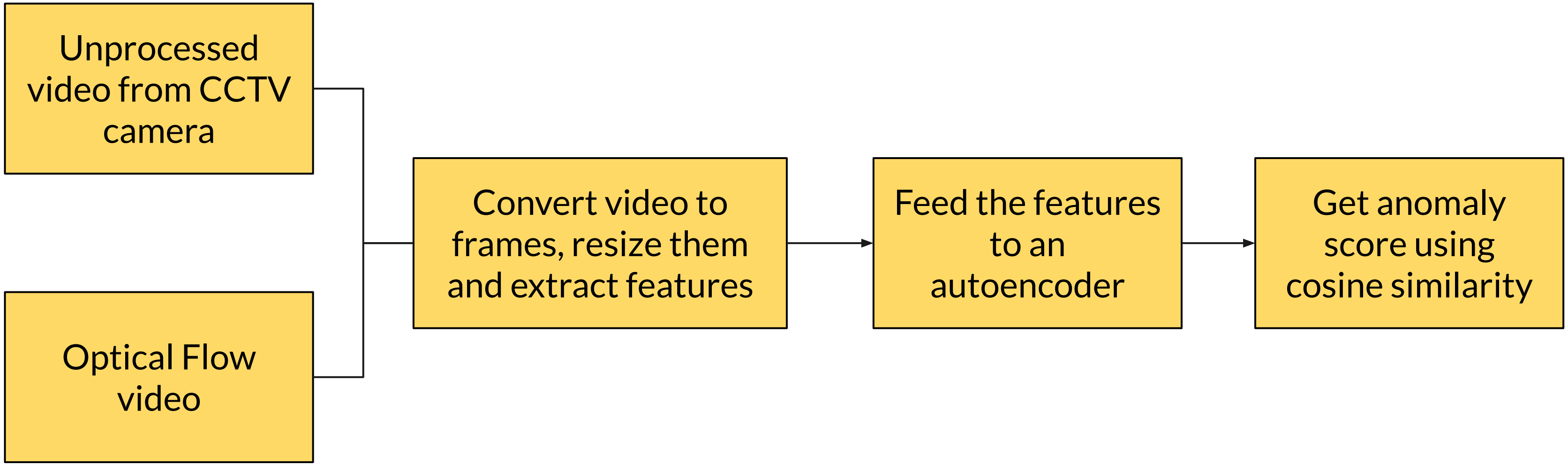
60
KB per
1-min video

Breakdown:


SELF-SUPERVISED
CCTV VIDEOS
COLLECTIVE ANOMALY
POINT ANOMALY
RAW
RAW IMAGE
CONVOLUTION TUNNEL
CONVOLUTION TUNNEL
CONVOLUTION TUNNEL
DEEP LAYER
THRESHOLD
RECONSTRUCTION
RECONSTRUCTION
ANOMALY DETECTOR


Input image
Latent Space Representation
Reconstructed image

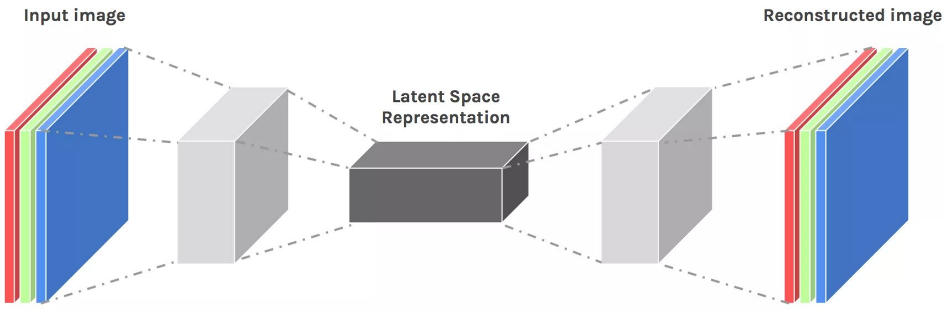
Self-supervised Learning - Overview



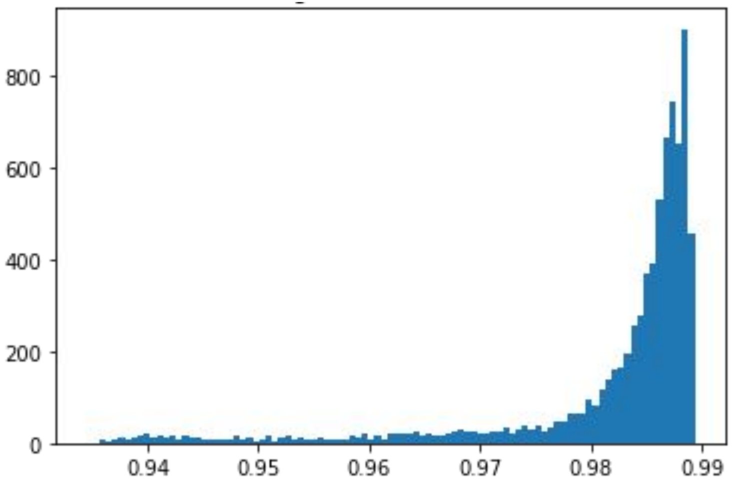
Anomaly Detection: CNN AutoEncoders

- Model is trained to reconstruct a sequence of 3 frames
- Cosine Similarity Score calculated between actual and reconstructed frame sequences
- Higher reconstruction error implies anomalous behavior

RECONSTRUCTION ERRORS



RECONSTRUCTION ERRORS

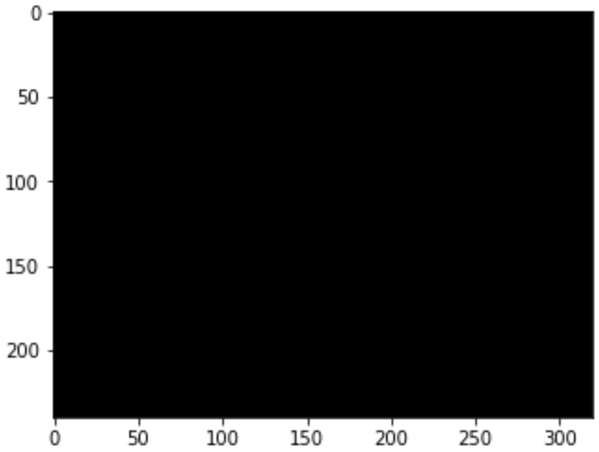


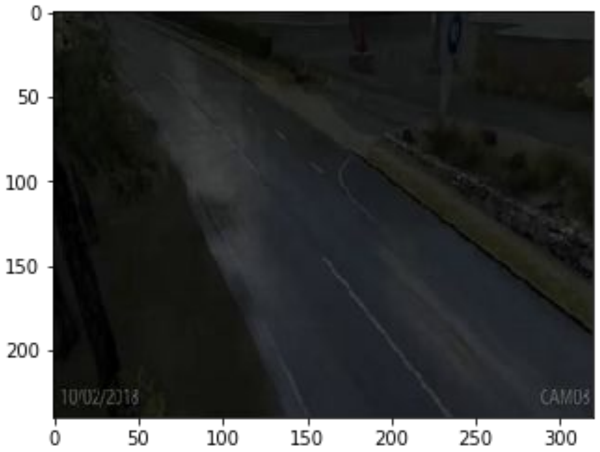
The Thin Red Line: Anomaly vs. Normality

- Reconstruction Error is lower for normal traffic footage..but just about!

CNN AutoEncoders: Largest Anomalies Detected

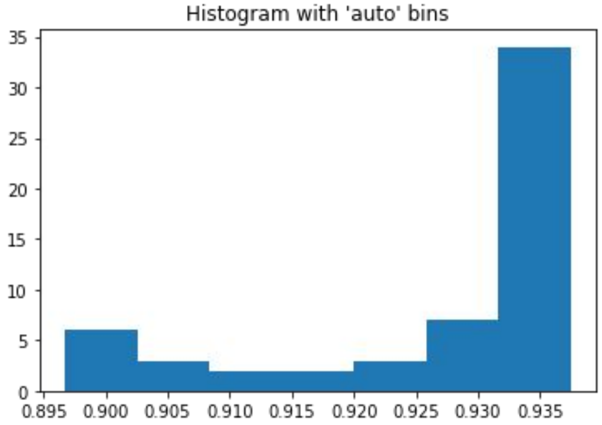
- In the absence of crashes caught in footage, we had to look for ‘out of context’ crash videos
- These included graphics and aberrations from the standard output anticipated from CCTV

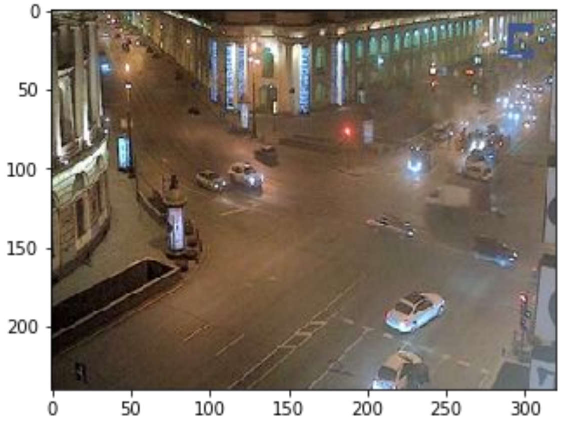


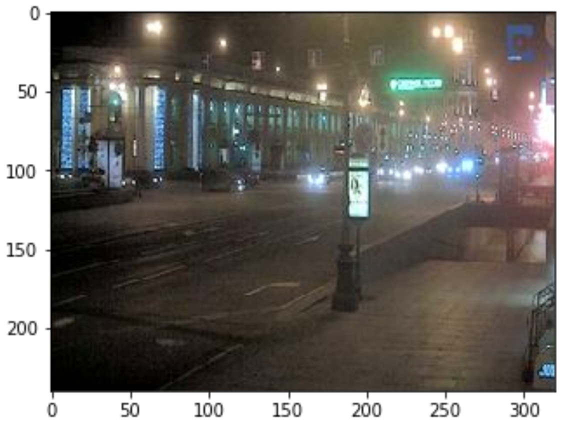


How much error is too much error?

- Anomaly, by construct, is reconstruction error.
- **Scene change** caused anomalies due to the training context during sequence
- ***Camera pose*** can cause reconstruction error (as in the case of incident below)

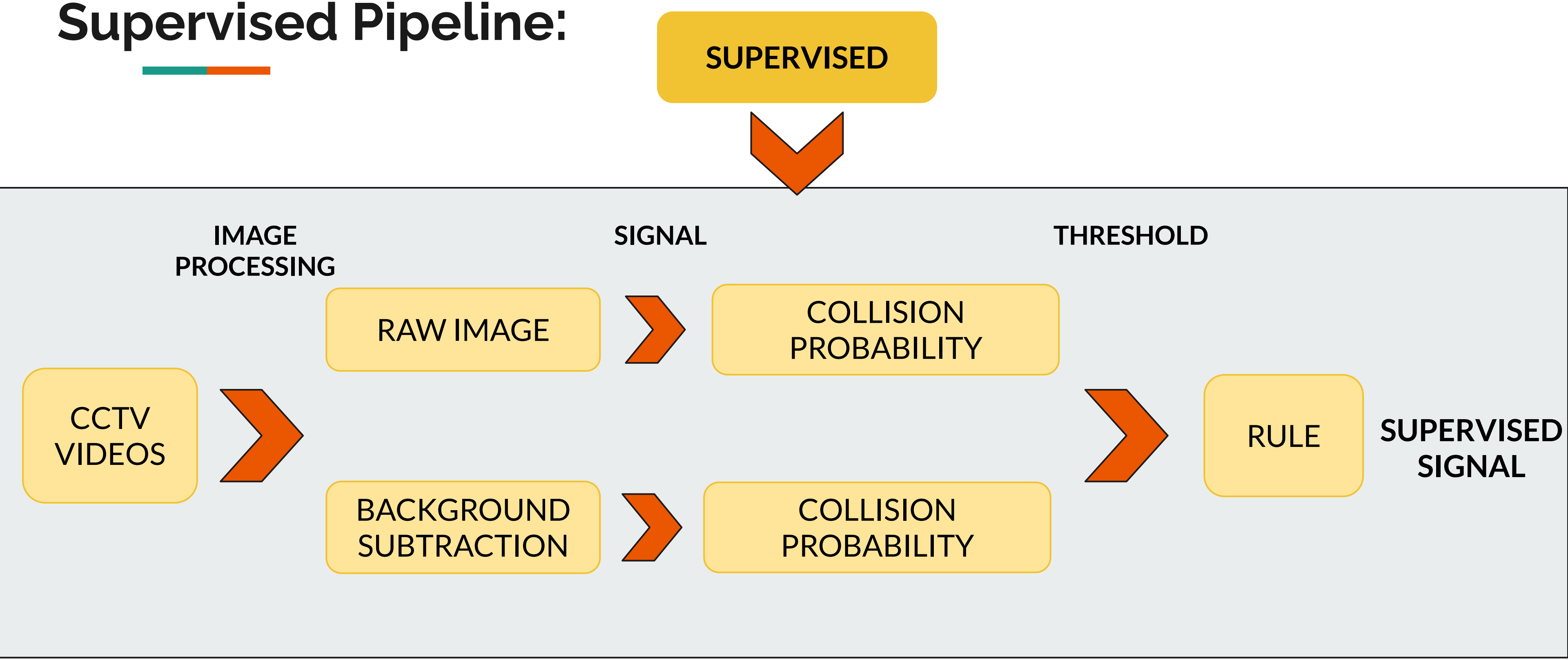




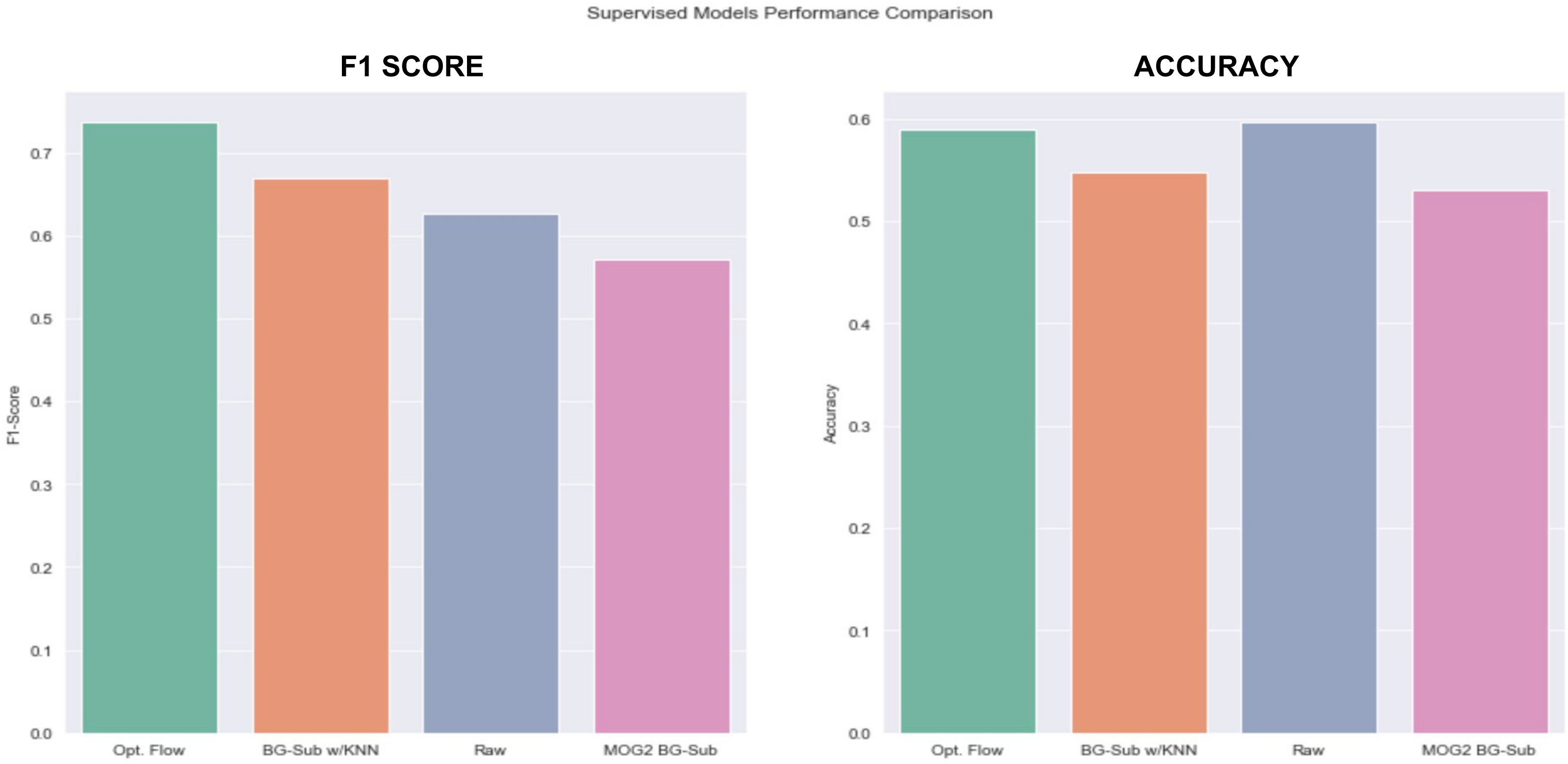


Analytics: Supervised Learning

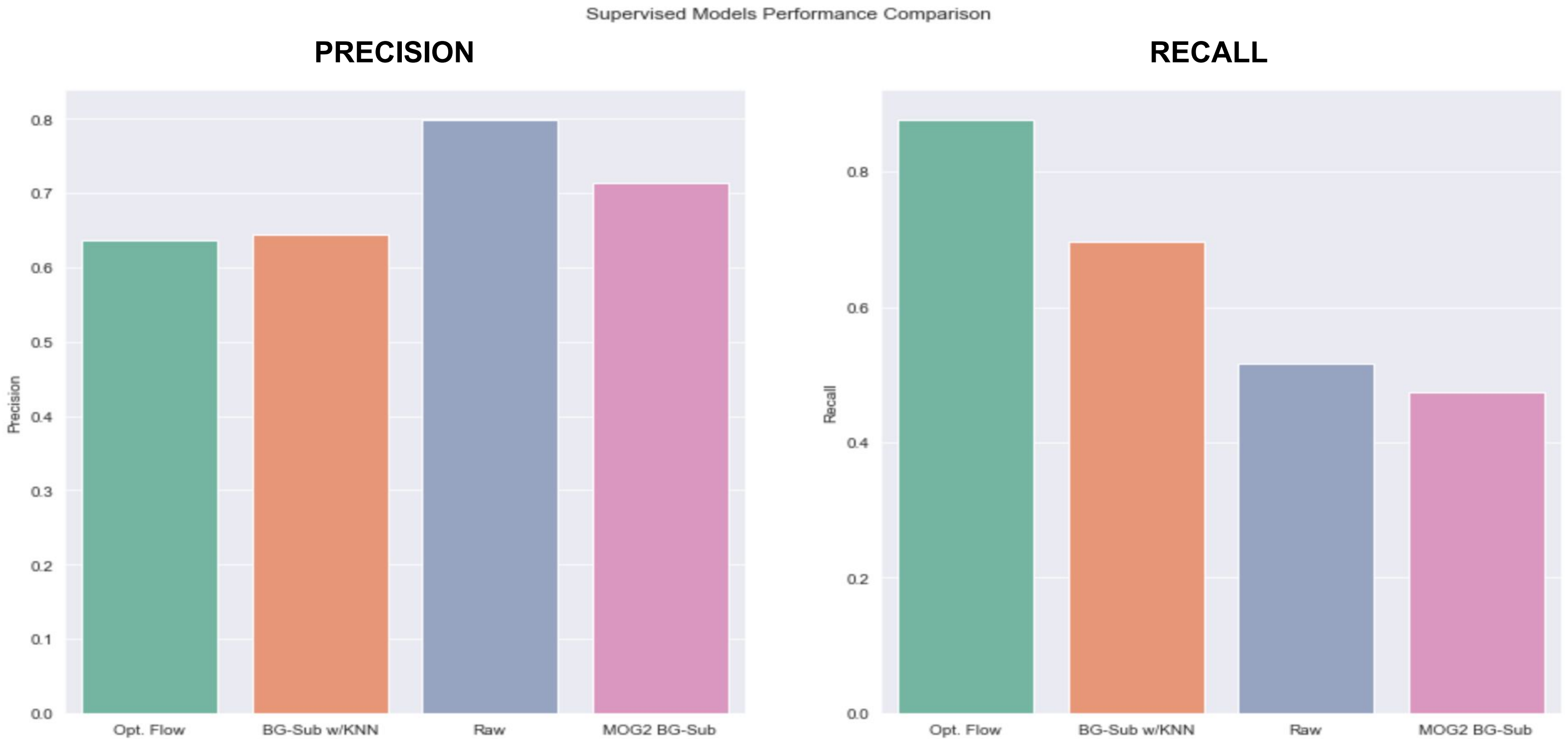
Supervised Pipeline:



Supervised Learning: Performance Comparison

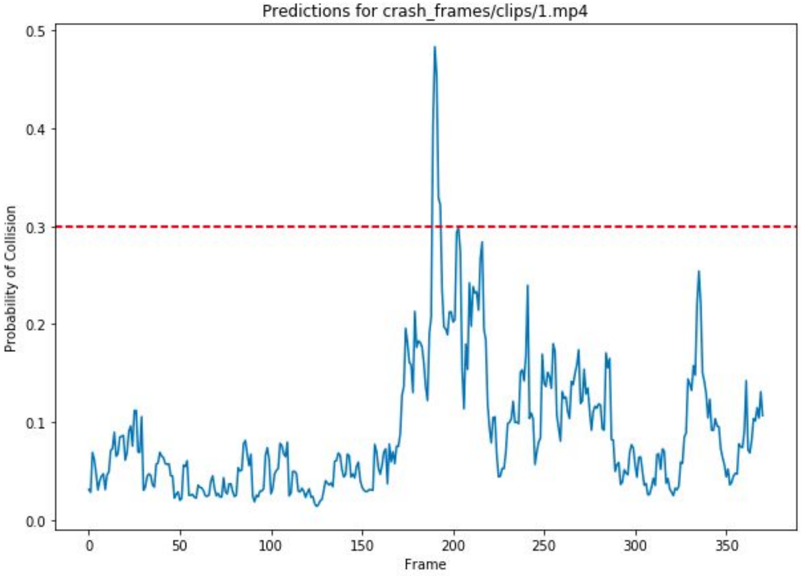


Supervised Learning: Performance Comparison
Precision & Recall

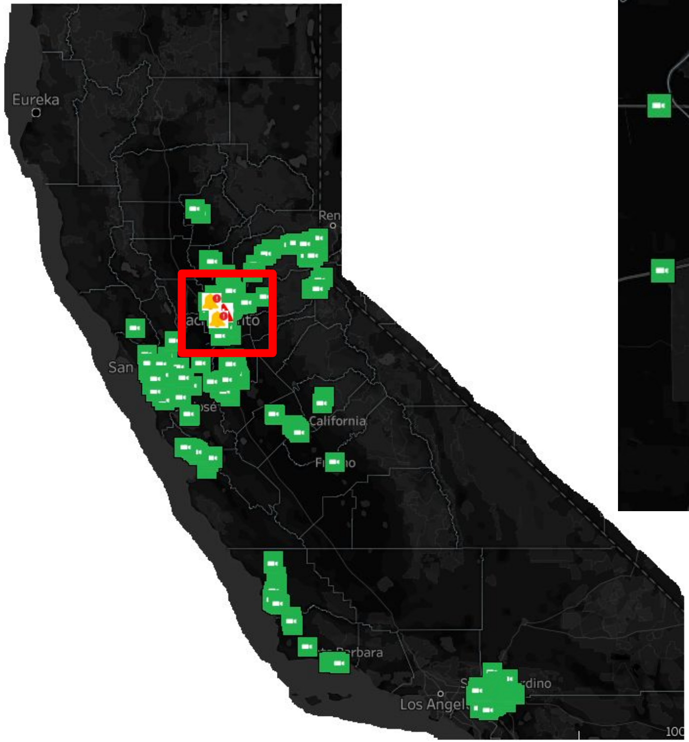


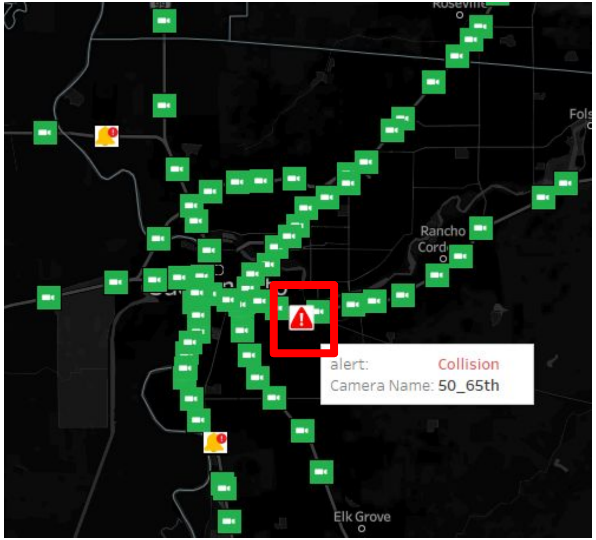
Serving: Web Resource and Edge Device

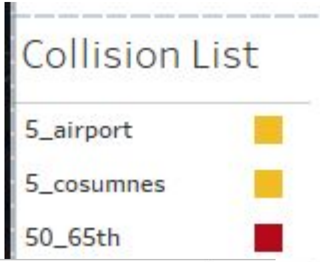
Supervised Model Predictions



Serving Layer #1: Real-time Incident Alerts



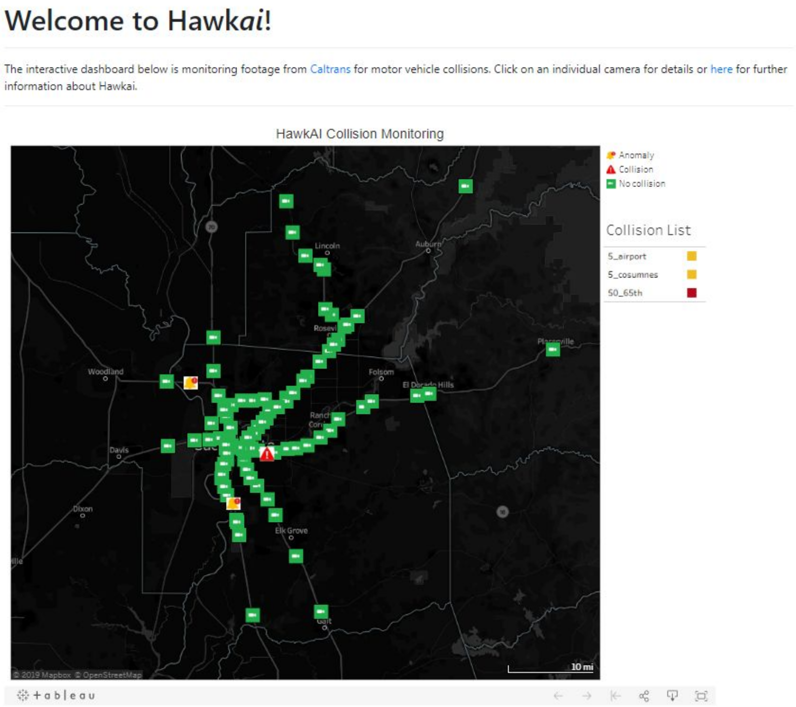




Demo Link

HawkAI - Questions?

PRESENT

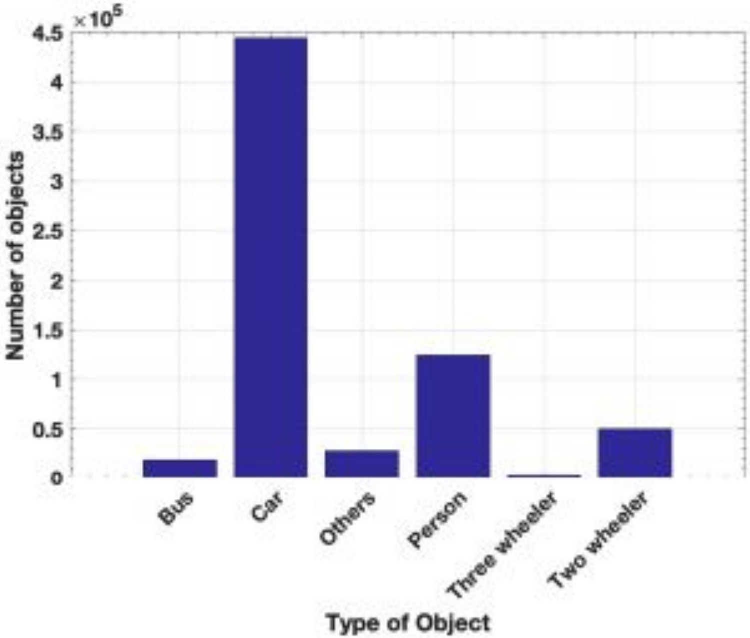


FUTURE

(Archived Slides)

Original Dataset

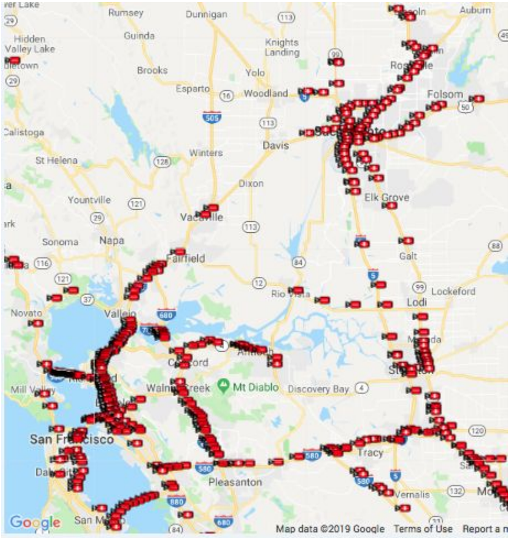
- **Road Collision Videos:** sourced from Youtube
- Collated by Ankit et al, Carnegie Mellon University
- 5.2 hours of footage
- 45GB
- 1,416 videos
- 518,256 extracted video frames
- Variety of weather and lighting amongst frames



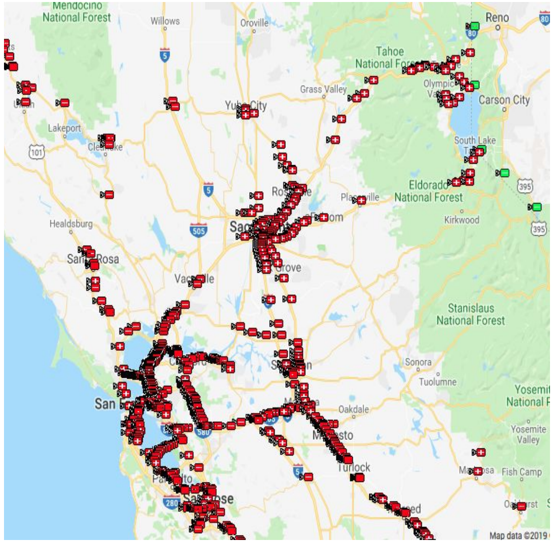
(a) Number of objects by categories

Caltrans

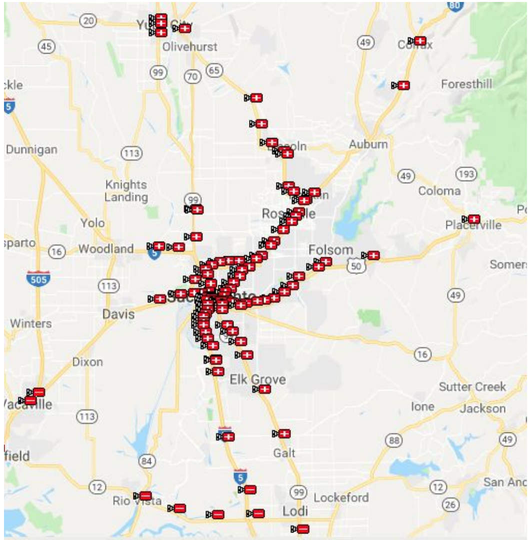
There are 374 CCTV cameras are available for streaming as open data



DISTRICT 2



DISTRICT 6

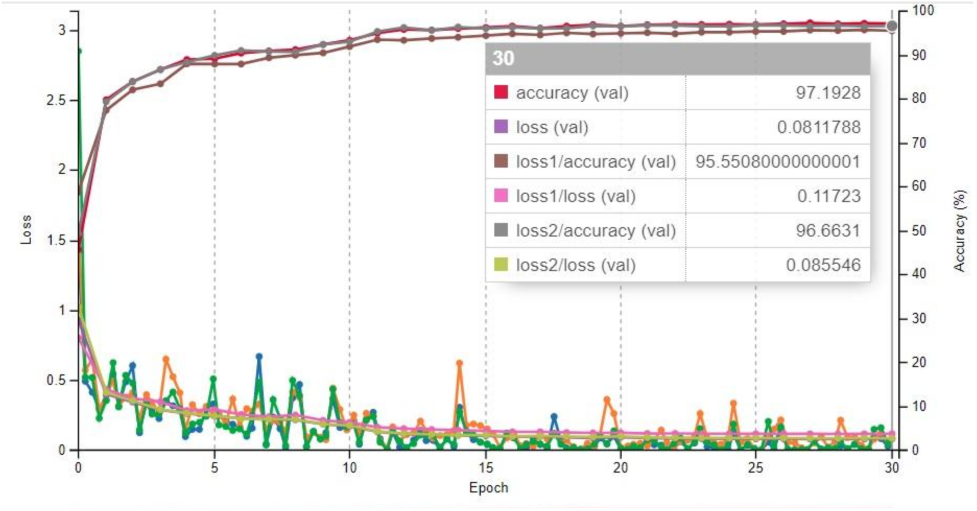


DISTRICT 10

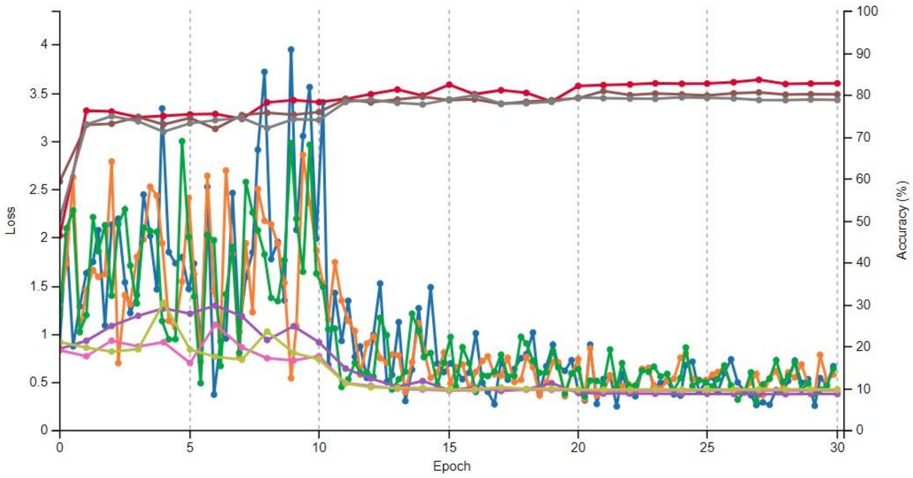
Supervised Learning: Training

- **GOOGLENET:** Retraining all the layers outperforms
- **EPOCHS:** 10-30; **OPTIMIZER:** SGD; **LEARNING RATE:** 0.005 - 0.01

LEARNING CURVE: GOOGLENET (UNFIXED LAYERS)



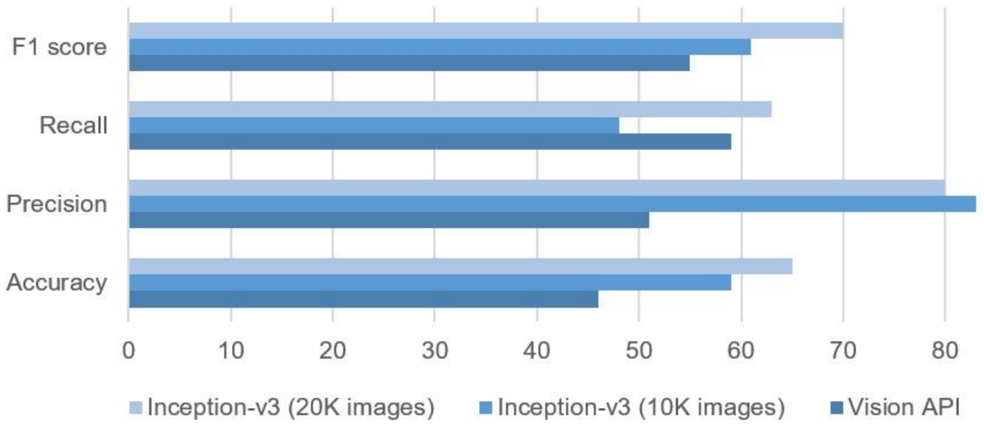
LEARNING CURVE: GOOGLENET (FIXED LAYERS)



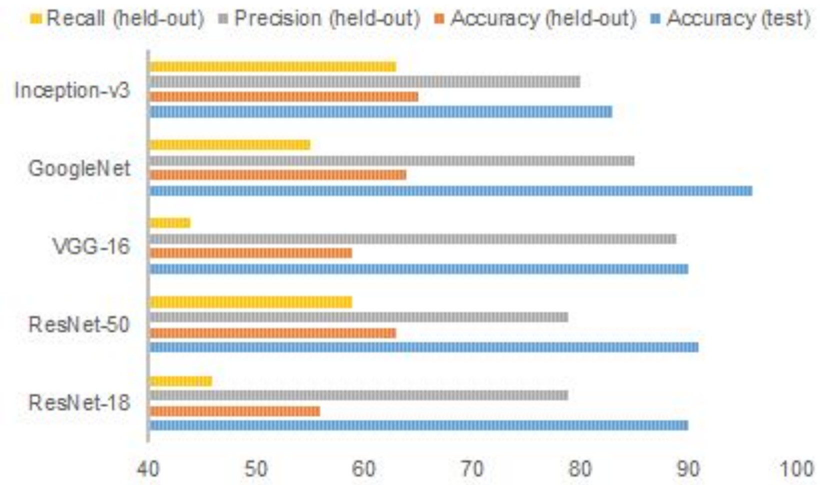
Supervised Learning: Validation and testing

- Baseline Vision API: 45% accuracy
- Recall rates improve by 15% as data doubles
- Inception V3 has the highest f1 score @ 0.7; Googlenet has the highest precision at 89%

EFFECT OF DATA SIZE ON MODEL PERFORMANCE



ARCHITECTURE-WISE MODEL PERFORMANCE



Supervised Learning: Caltrans Performance

- Model based on labeled frames from YouTube videos (non-Caltrans)
- Very low false positive rate for nighttime videos / videos with sparse traffic

33%

False Positive Rates for daytime videos / videos with dense traffic

48,000

Total -minute video footage captured from 59 videos

0

Accidents captured on Traffic Cameras

Serving Layer #2 Demo Video: Jetson TX2