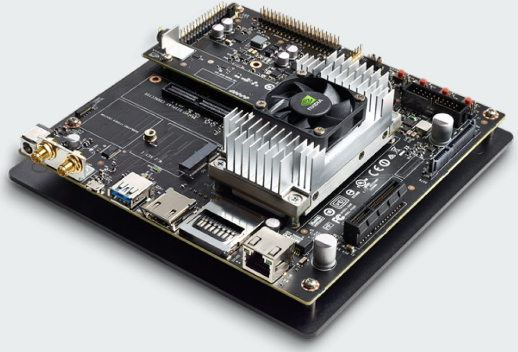
Accident Detection On the Edge



Matthew Potts, Saifullah Rais, Kyle MacIlvain

Outline

1. Problem Statement/Inspiration
2. Data Overview
3. Training
4. Testing
5. Demo
6. Error Analysis
7. Challenges
8. Final Thoughts
9. Q&A

Problem Statement

- Build an accident-detection algorithm for surveillance cameras and deploy a prototype on an edge device

Inspiration

- Survival rate dependent on Emergency Response
- Hit-and-run incidents rising with higher fatalities
- 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

30%
Higher fatality rates for hit-and-run accidents

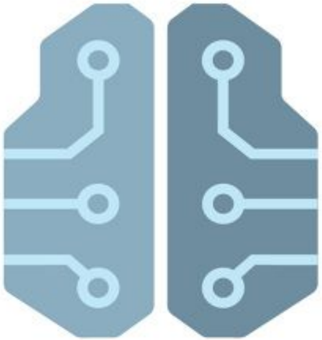
50 mn
Surveillance cameras in the US

Operationalization

- Supervised Learning
- Binary Image Classification
- Transfer Learning

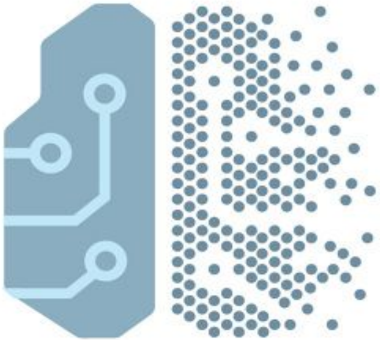
Tools

- **Google Cloud Platform** - Compute and Cloud Storage
- **IBM SoftLayer** - P100 GPUs
- **NVIDIA:** DIGITS & Transfer Learning Toolkit
- **Edge:** NVIDIA Jetson TX2 and Logitech USB C270 Camera
- **Others:** Google Vision API, ffmpeg



Artificial Intelligence

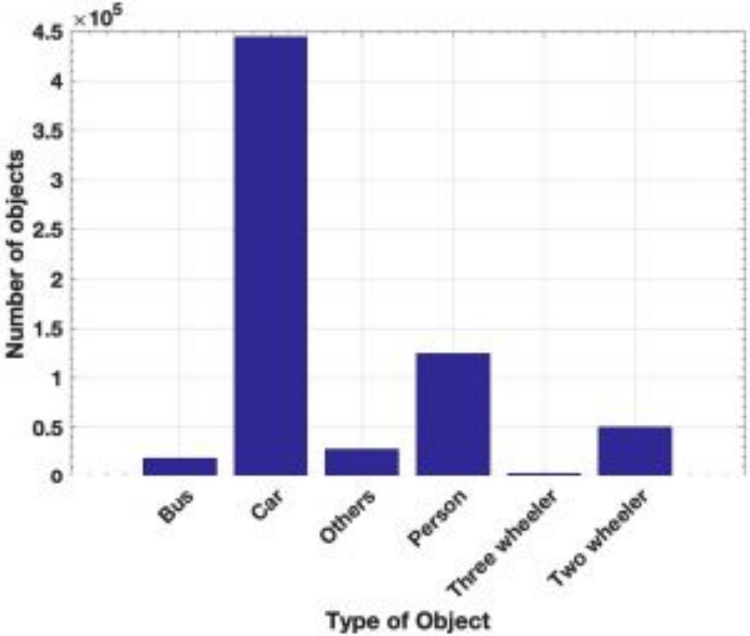
Machine Learning



Deep Learning

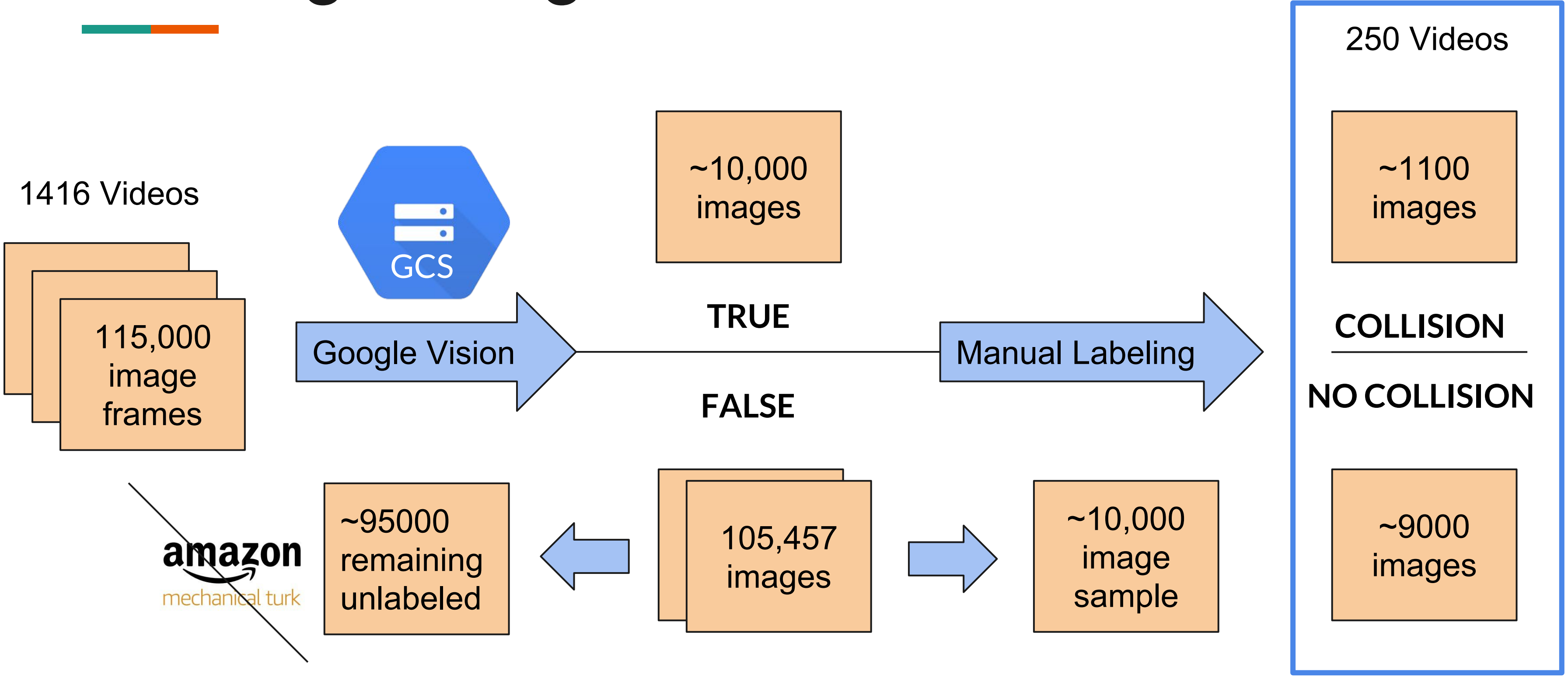
Dataset

- **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

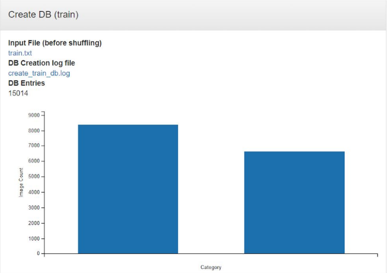


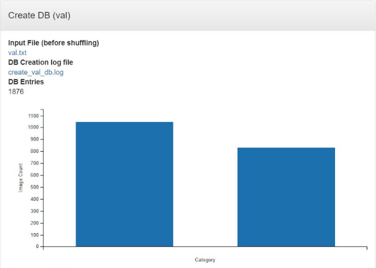
(a) Number of objects by categories

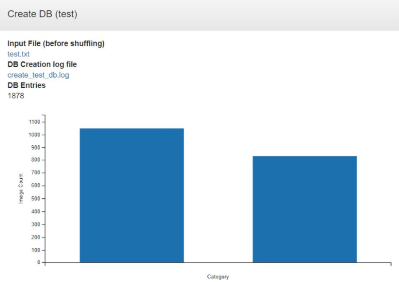
Data Engineering and Annotation

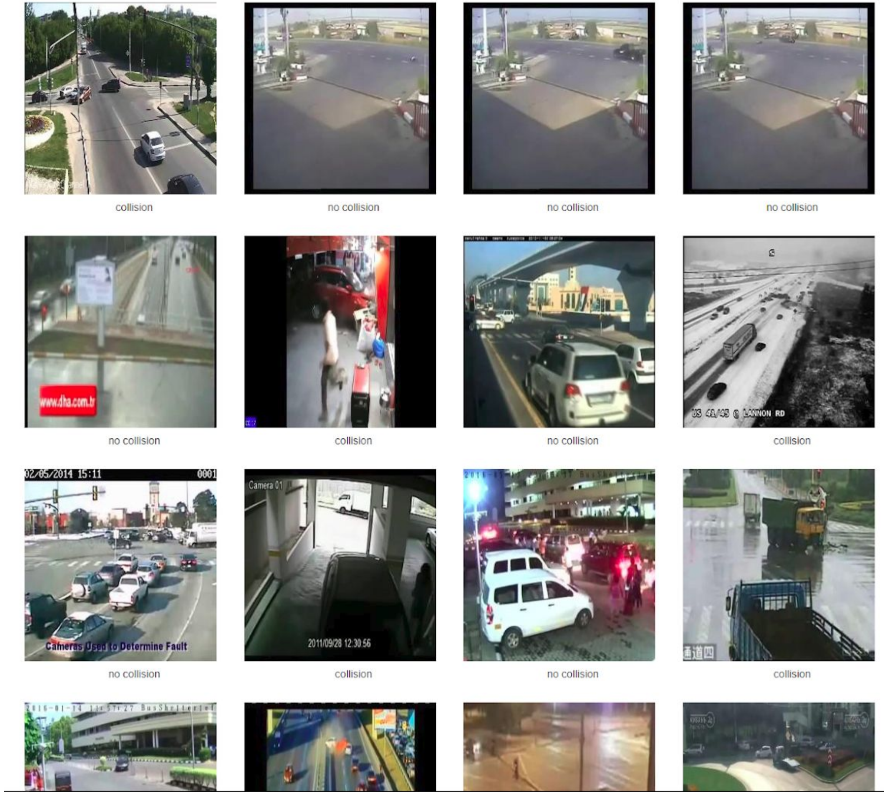


Exploring The Data

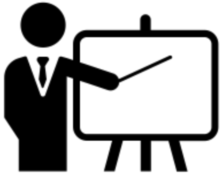








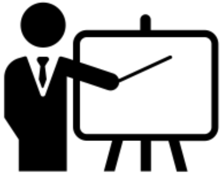
Training



- **APPROACH:** Transfer Learning via Pretrained Models
- **FRAMEWORK:** DIGITS, Tensorflow, and Nvidia Transfer Learning Toolkit
- **MODEL ARCHITECTURE:** ResNet18, ResNet50, VGG-16, Inception v3 and Googlenet

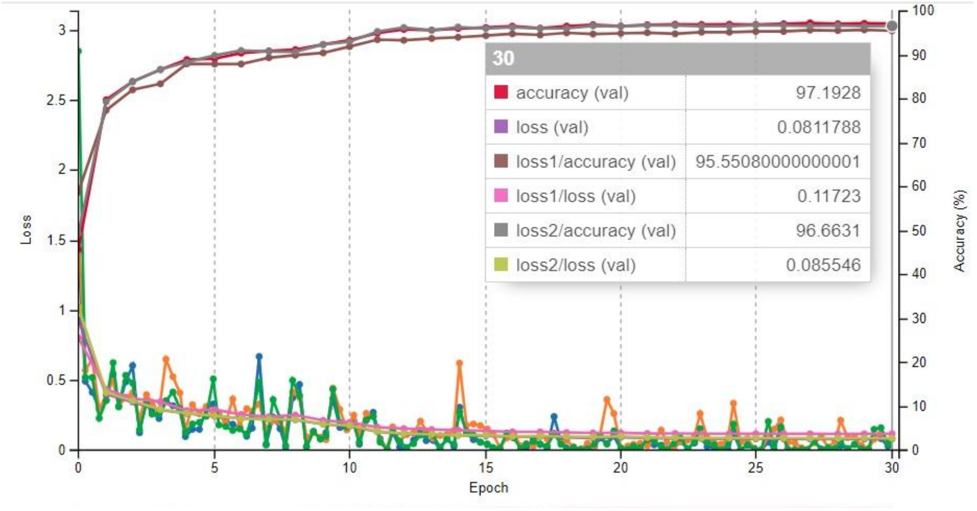
FRAMEWORK	TOOLKIT	MODEL ARCHITECTURE	DEPLOYMENT
Caffe2	NVIDIA	GOOGLENET	
	NGC TLT DIGITS	RESNET-18 RESNET-50 VGG16	
	TENSORFLOW	INCEPTION-V3	JETSON DL SDK

Training

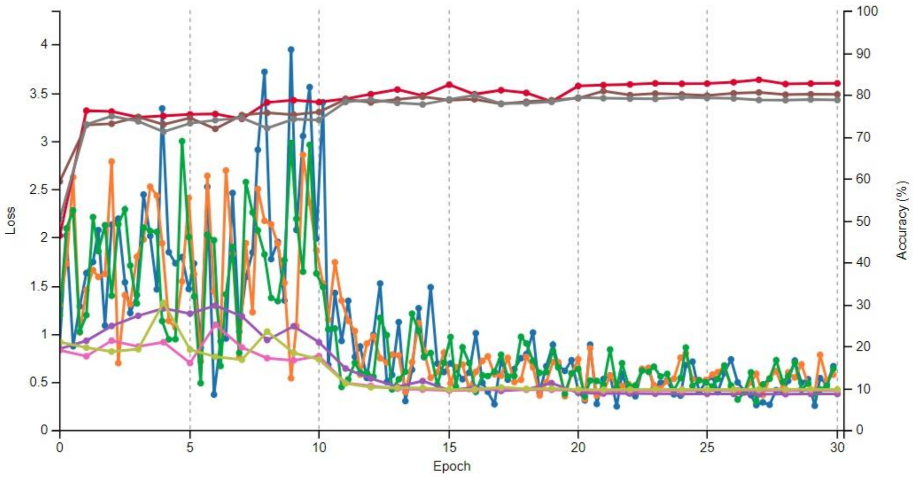


- **GOOGLENET:** Retraining all the layers outperforms
- **EPOCHS:** 10-30; **OPTIMIZER:** SGD; **LEARNING RATE:** 0.005 - 0.01

LEARNING CURVE: GOOGLENET (UNFIXED LAYERS)



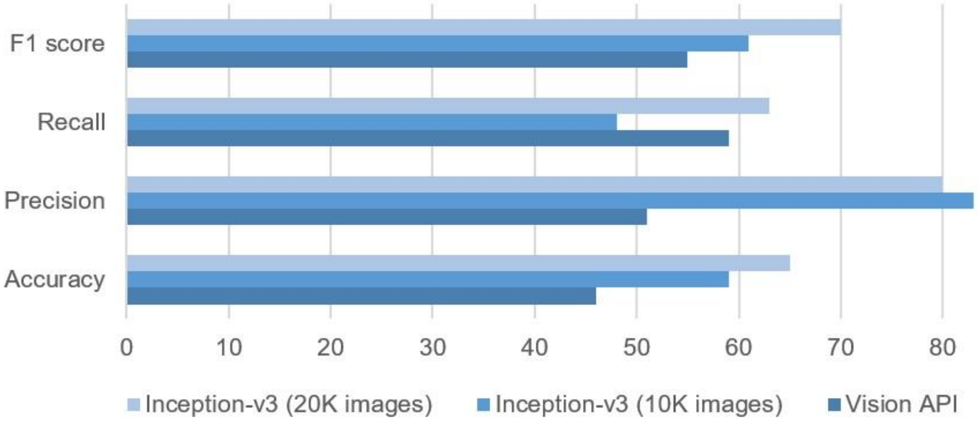
LEARNING CURVE: GOOGLENET (FIXED LAYERS)



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

Error Analysis

- Post-impact
- Motorcycles
- Edge of frame

Googlenet: Signs of Improvement

- We see an improvement in detecting these incidents
- Although, we believe there is room for improvement here

Predictions:

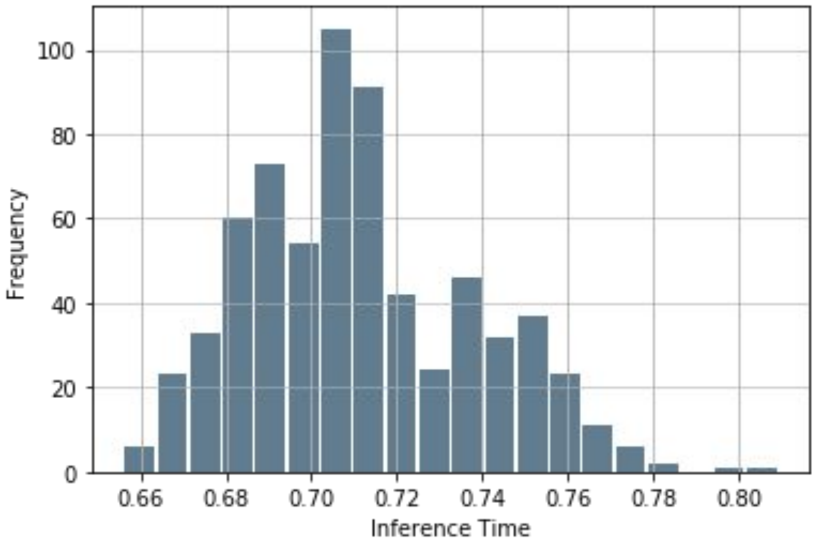
Collision	94.81%
No collision	5.19%

Predictions:

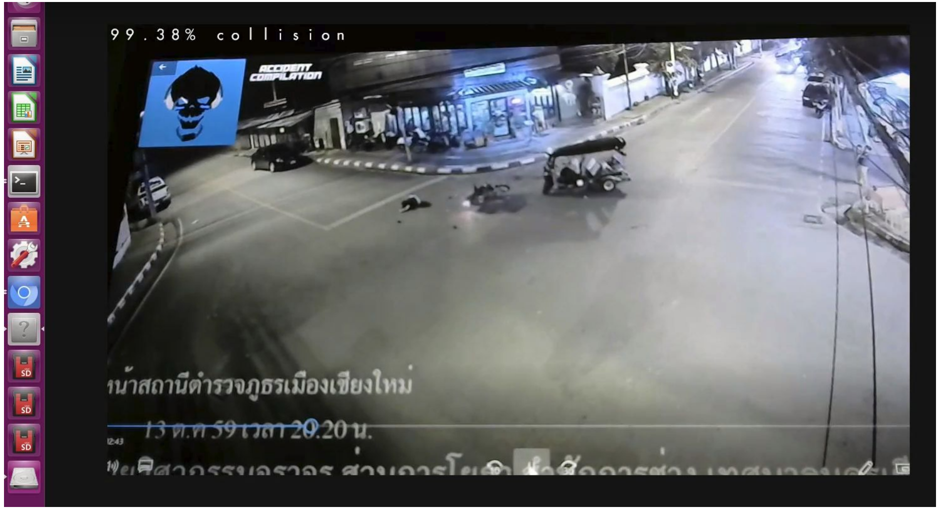
Collision	92.17%
No collision	7.83%

Inference Time

INFERENCE TIME ON INCEPTION V3



REAL-TIME INFERENCE USING JETSON TX2



Demo

Real-time

Challenges

- **Need More Data:** data augmentation could assist
- **HIT Was A Miss:** Human Intelligence Tasks (on MTurk) saw 32% errors in two-class labelling
- **Unintentional Bias:** caused by semi-supervised labelling (Vision API) towards motorcycles
- **Err On The Side Of Caution:** need higher recall rates
- **Inference Time:** For the best model (Inception V3), the fastest frame rate is 1fps

Discussions & Further Work

- **Decision Support vs. Automated Solution**
- **Multi-Class Approach:** to evaluate solution’s impact and post-impact performance
- **Background Subtraction:** Focus on what matters
- **Video Understanding:** Anomaly Detection Using Video Embeddings

Concluding Remarks

- **Transfer Learning outperforms** baseline across all model architectures
- ***More data*** means ***more accuracy*** (especially recall rates)
- **Model Performance:** High precision; Low recall rates
- **Inception-v3:** highest f1 score; **Googlenet:** highest precision
- **Video Understanding** (through video embeddings) vs. Frame-level Image Classification
- **Inference Time:** best model has inference time of 0.7 sec/frame

Q&A