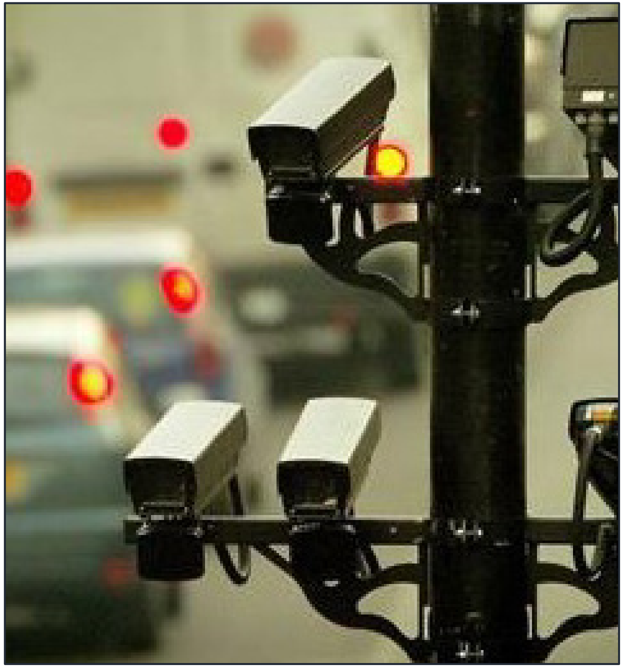
The AI Revolution of Traffic Analytics

David C. Anastasiu
danastasiu@scu.edu

May 18, 2023

Why Traffic Analytics?



1+ billion cameras worldwide
10's of exabytes of data per day
30 billion frames per second

AI brings real-time insights, sensor fusion & control.

Applications in Traffic Management, Public Safely, Transit, Retail Analytics, etc.

Security and Ethical Implications

Qualifications

David C. Anastasiu

Sounds like "so nice to see you…"

- PhD in CS from University of Minnesota
- Assistant Professor at Santa Clara University
  - Teach Data Mining, Machine Learning, Deep Learning, Parallel Computing
- Research in Machine Learning/AI, Computational Genomics, HPC
- Co-Organizer and Evaluation Chair of the yearly AI City Challenge since 2017

Current Sensors

Induction Loop Detectors

https://upload.wikimedia.org/wikipedia/commons/thumb/8/8c
/Inductance_detectors.jpg/330px-Inductance_detectors.jpg

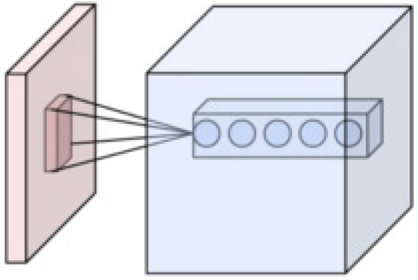
- Installed during road construction
- Can only detect presence of a vehicle
- Maintenance requires breaking and re-building the road
- Limited use of the data

https://i.ytimg.com/vi/KvzJn09DqaM/hqdefault.jpg

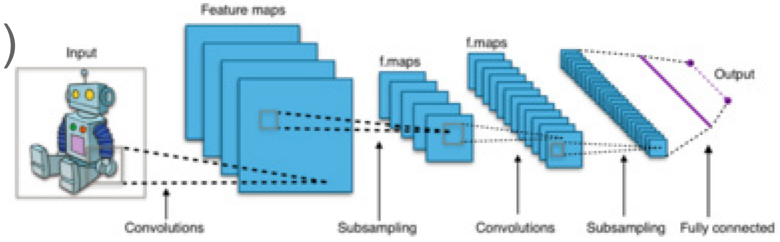
What Made It Possible?

Convolutional Neural Networks

- Inspired by biological research on the cat visual cortex
  - Neurons focus on receptive fields
- Multilayer Perceptron/Stochastic Gradient Descent (1960s/1970s)
- Neocognitron by Kunihiko Fukushima (1980)
  - Rectified Linear Unit activation (1969)
- Back-propagation for CNNs (1988-1989)
- GPU training – AlexNet/ImageNet (2012)

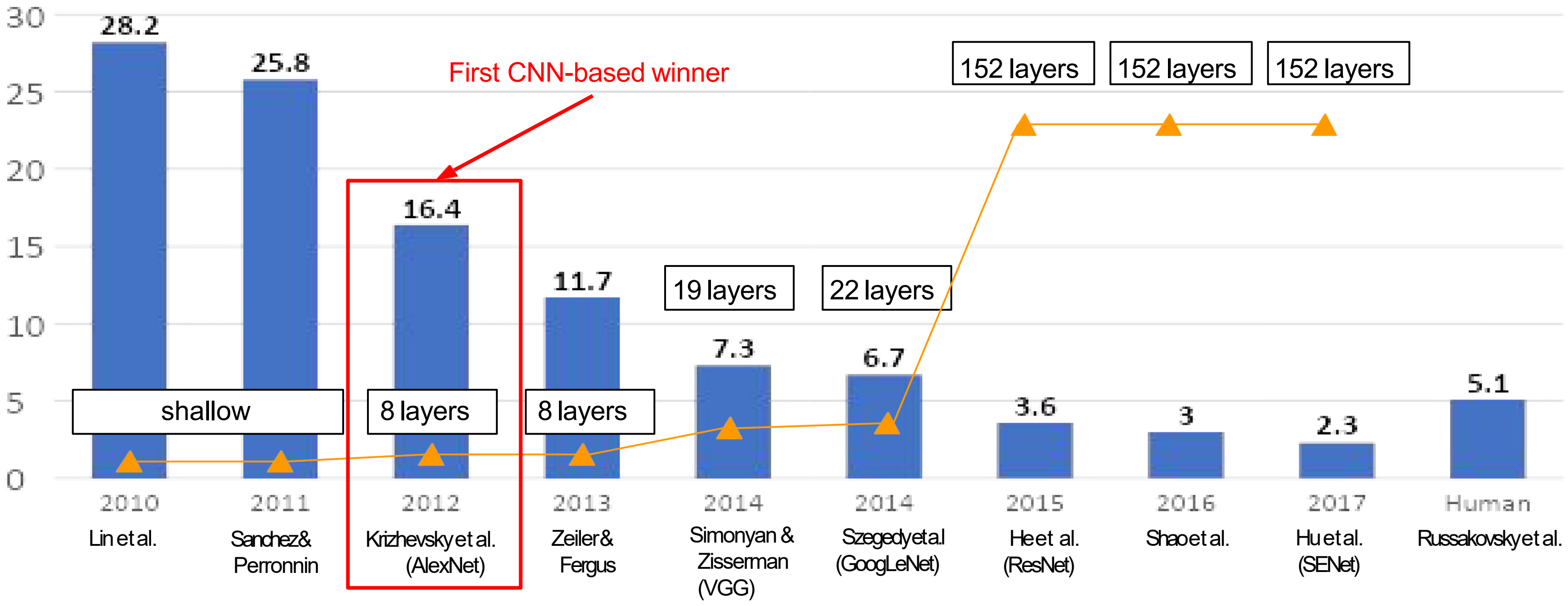


https://en.wikipedia.org/wiki/File:Conv_layer.png



https://upload.wikimedia.org/wikipedia/commons/thumb/6/63/Typical_cnn.png/593px-Typical_cnn.png

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



The AI City Challenge

Object Detection

People Tracking

MTMC Tracking

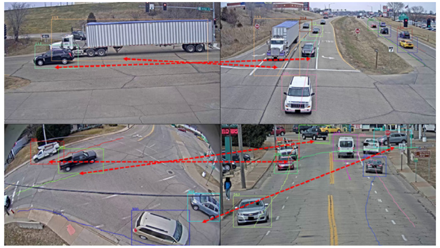
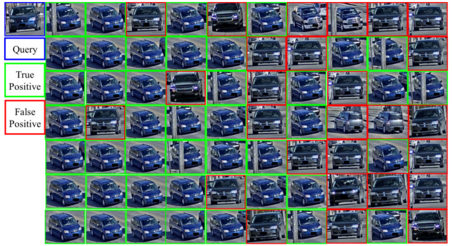


Image-based ReID



Anomaly Detection

Activity Recognition

AI City Challenges

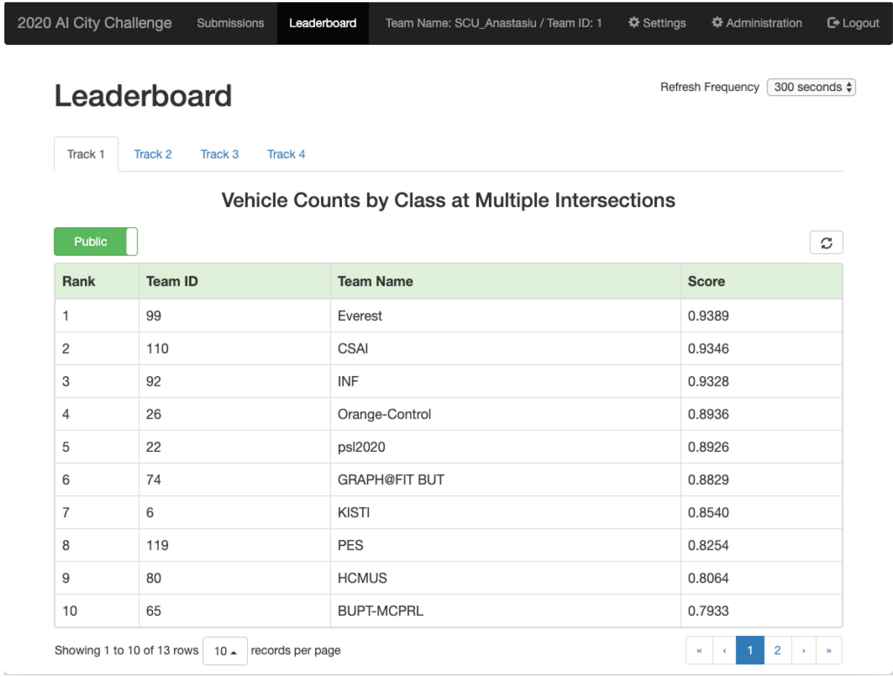
Annual Editions

	2017	2018	2019	2020	2021	2022
Challenge Tasks	Vehicle Detection, Classification, Tracking	Speed Estimation, Anomaly Detection, Vehicle ReID	Anomaly Detection, MTMC Tracking, Vehicle ReID	Turn Counting, Vehicle ReID, MTMC Tracking, Anomaly Detection	Turn Counting, Vehicle ReID, MTMC Tracking, Anomaly Detection, NL Based Retrieval	MTMC Tracking, NL Based Retrieval, Driving Action Recognition, Product Counting
Dataset	130 hours of labeled intersection data from 3 states	35+ hours of labeled intersection & highway data from 2 states	50+ hours of labeled multi-camera intersection & highway data	60+ hours of labeled multi-camera intersection & highway data along with synthetic data	60+ hours of labeled multi-camera intersection & highway data along with synthetic data	60+ hours of labeled multi-camera intersection & highway data; 14+ hours of naturalistic driving videos; 116,500 synthetic images with classification and segmentation labels on 100+ merchandise items
Associated Conference	SmartWorld 2017	CVPR 2018	CVPR 2019	CVPR 2020	CVPR 2021	CVPR 2022
Awards	2 NVIDIA Jetson TX2 2 NVIDIA Titan Xp	4 NVIDIA Jetson TX2s 1 NVIDIA Titan Xp 1 NVIDIA Quadro GV100	1 NVIDIA Titan GV 100 3 NVIDIA Titan RTX 2 NVIDIA Jetson Xavier AGX	1 NVIDIA Titan GV 100 4 NVIDIA Titan RTX 3 NVIDIA Jetson Xavier AGX	2 NVIDIA RTX 3090 5 NVIDIA RTX 3080 3 NVIDIA Jetson Xavier NX	4 NVIDIA RTX 3070TI 4 NVIDIA Jetson Xavier NX

My Role in the AI City Challenge

Automated Evaluation System

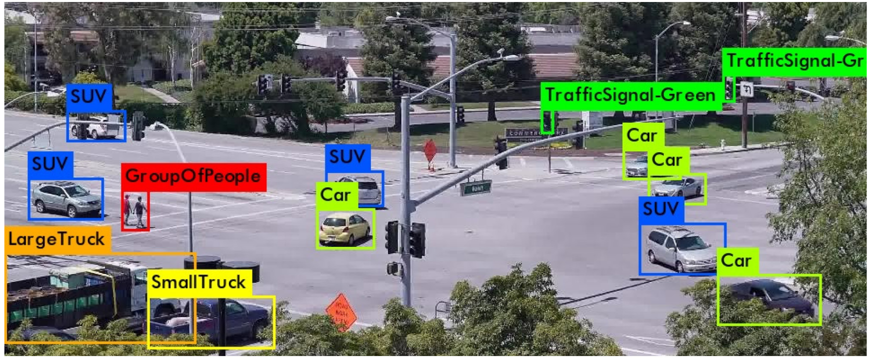
- 5x submissions/day/track/team
- Max total submissions (10 or 20 submissions, depending on track)
- Validated submission format & sent submission for processing
- Individual team scores available right after processing (5-30s)
- Leaderboard showing the team's own rank + anonymized top-3 results for 50% test subset
- Full leaderboard available after submissions close
- Choices between *Public* & *General* boards

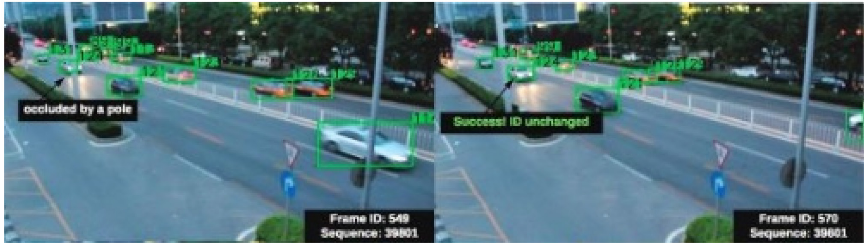


Rank	Team ID	Team Name	Score
1	99	Everest	0.9389
2	110	CSAI	0.9346
3	92	INF	0.9328
4	26	Orange-Control	0.8936
5	22	psl2020	0.8926
6	74	GRAPH@FIT BUT	0.8829
7	6	KISTI	0.8540
8	119	PES	0.8254
9	80	HCMUS	0.8064
10	65	BUPT-MCPRL	0.7933

Our Humble Beginnings

- Object Detection
  - 15 traffic-related classes
  - 125 hours of traffic videos
  - 29 teams from 4 continents
  - Online annotation system
  - Collaboratively annotated >153K key-frames
  - >1.4M annotations





Our Humble Beginnings

NVIDIA AI City Challenge | Home | Summary | Team 21 - User 1 | Logout

- Trained detection models at 3 resolutions
- Automatic evaluation system
- Inference on NVIDIA Jetson
- Methods:
  - YOLO
  - SSD
  - ImageNet Transfer Learning
- Ethical Considerations
  - Identity of pedestrians/drivers

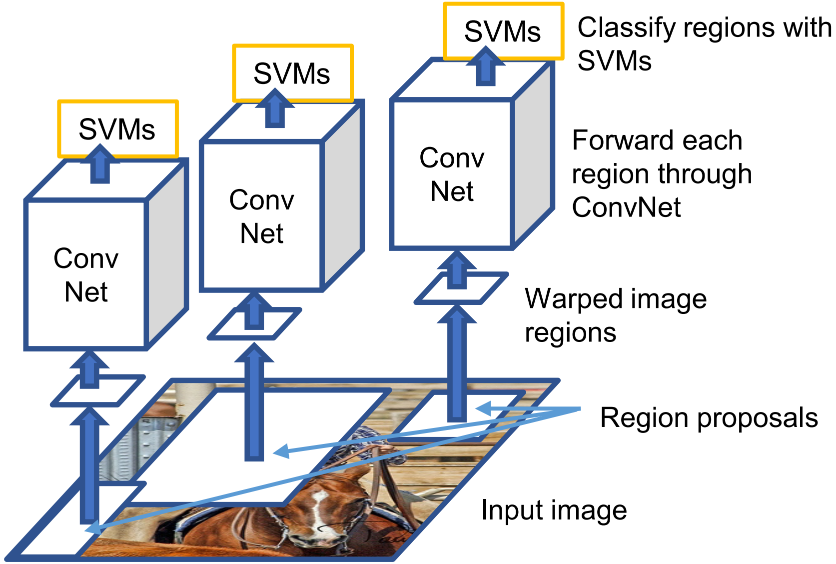
AIC480: Average Precision

Team	mAP	Car	SUV	SmTruck	MdTruck	LgTruck	Bus	Van	Bicycle	Ped	Loc
Team24_24_480	0.52	0.86	0.75	0.72	0.41	0.45	0.63	0.4	0.25	0	0.91
Team5_SUNY_CVML	0.45	0.81	0.66	0.54	0.37	0.39	0.62	0.32	0	0	0.88
Team21_480_filter	0.44	0.82	0.66	0.62	0.32	0.31	0.52	0.26	0.08	0	0.89
Team6_RDFCN_ISU	0.41	0.76	0.61	0.56	0.37	0.35	0.5	0.29	0	0	0.82
Team10_tyutcie4	0.37	0.7	0.54	0.53	0.28	0.22	0.45	0.23	0.17	0	0.82
Team19_aic480_3	0.35	0.7	0.51	0.48	0.21	0.24	0.39	0.19	0	0	0.81
Team4_480comb	0.34	0.75	0.52	0.45	0.19						
Team23_T23_480	0.33	0.66	0.54	0.45	0.21						
Team2_Ensembl480	0.15	0.5	0.17	0.08	0.07						

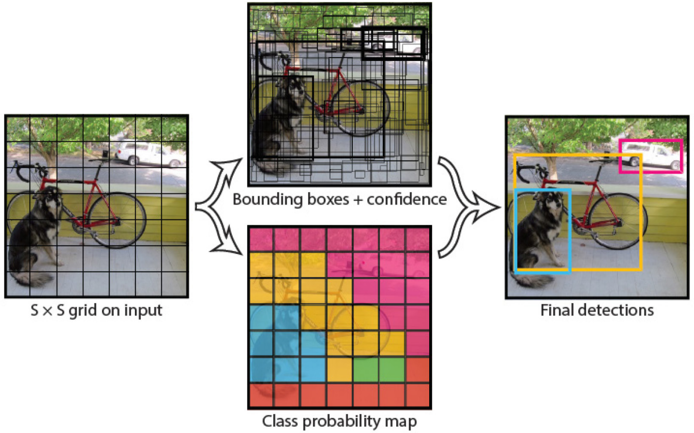
Note: Click on the class or model name to display the associated Precision-Recall g

Detection strategies: Summary

- Before ~2010: dominated by sliding windows
- 2010-2013: proposal-driven
- Recent methods: dense predictions

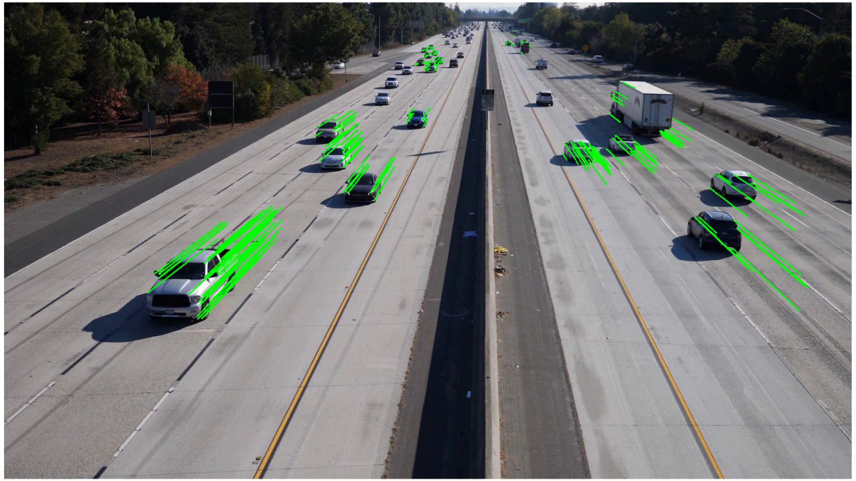


R. Girshick, J. Donahue, T. Darrell, and J. Malik, [Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation](), CVPR 2014



J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, [You Only Look Once: Unified, Real-Time Object Detection](), CVPR 2016

Traffic Flow Analysis



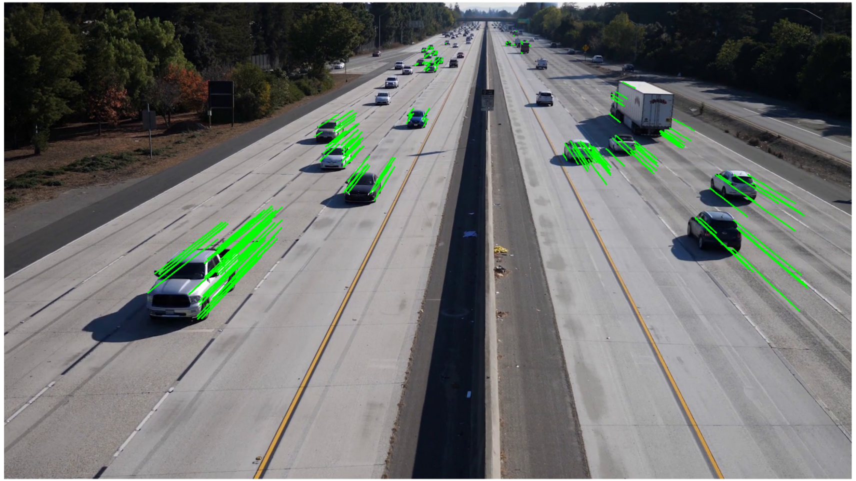
AIC 2018

- Vehicle Speed Estimation
  - Created dataset w/ 27 1-min videos
  - Ground truth set by control vehicles
  - Evaluate based on detection rate and speed prediction
  - Detect-then-Track using a variety of detectors (YOLOv2, DenseNet, Mask R-CNN, Faster R-CNN) + statistical association methods
  - Speed estimated from pixel distance increments
- Ethical Considerations
  - Speeding in public dataset

Traffic Flow Analysis

Our Solution

- Detection
  - Transfer Learning from AIC 2017
- Tracking
  - Tracklets from associated bounding boxes with min ϵ overlap
  - Extract Shi-Thomasi corners from all boxes
  - Compute optical flow based on corner features
  - Merge tracklets with continuing flow
- Speed estimation
  - Assume constant speed and direction of travel towards the camera
  - Series of estimators based on corner point movements within tracks



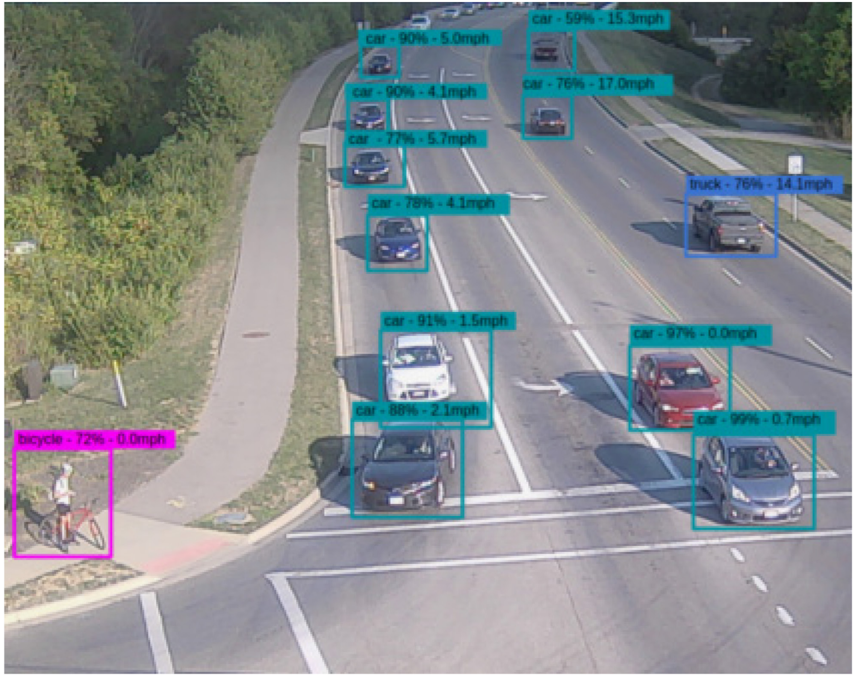
Shuai Hua, Manika Kapoor & David C. Anastasiu. Vehicle Tracking and Speed Estimation from Traffic Videos. In 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'18), 1:153-1537, 2018.

Traffic Flow Analysis - IRL

NoTraffic Traffic Light Optimization

https://blogs.nvidia.com/blog/2019/10/31/notraffic-ai-startup-intersections/

- Video + radar sensor instead of induction loops
- NVIDIA Jetson IoT to fuse data and derive AI-based decisions, 15 fps
- Can integrate dedicated short-range communications (DSRC) and cellular vehicle-to-everything (CV2X) communication
- Sends results to cloud + some data for further processing
- Security considerations
  - Wireless communication, cloud transmission



Traffic Flow Analysis - IRL

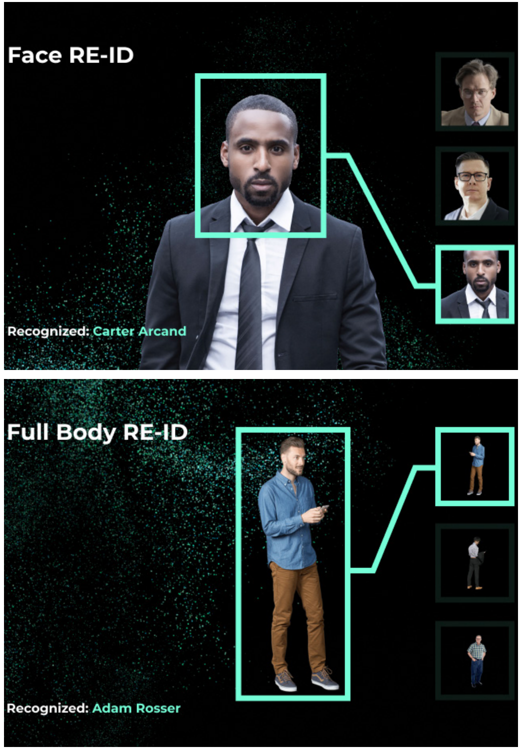
Tapway Toll Monitoring

https://blogs.nvidia.com/blog/2022/05/09/ai-malaysia-highways/

- Video-based analysis on the edge
- Detect vehicle class, make, color, and license plate number
- 50 millisecond detection time @ up to 40 km/h
- One server > 20 video streams
- Ethical considerations
    - False positive identifications

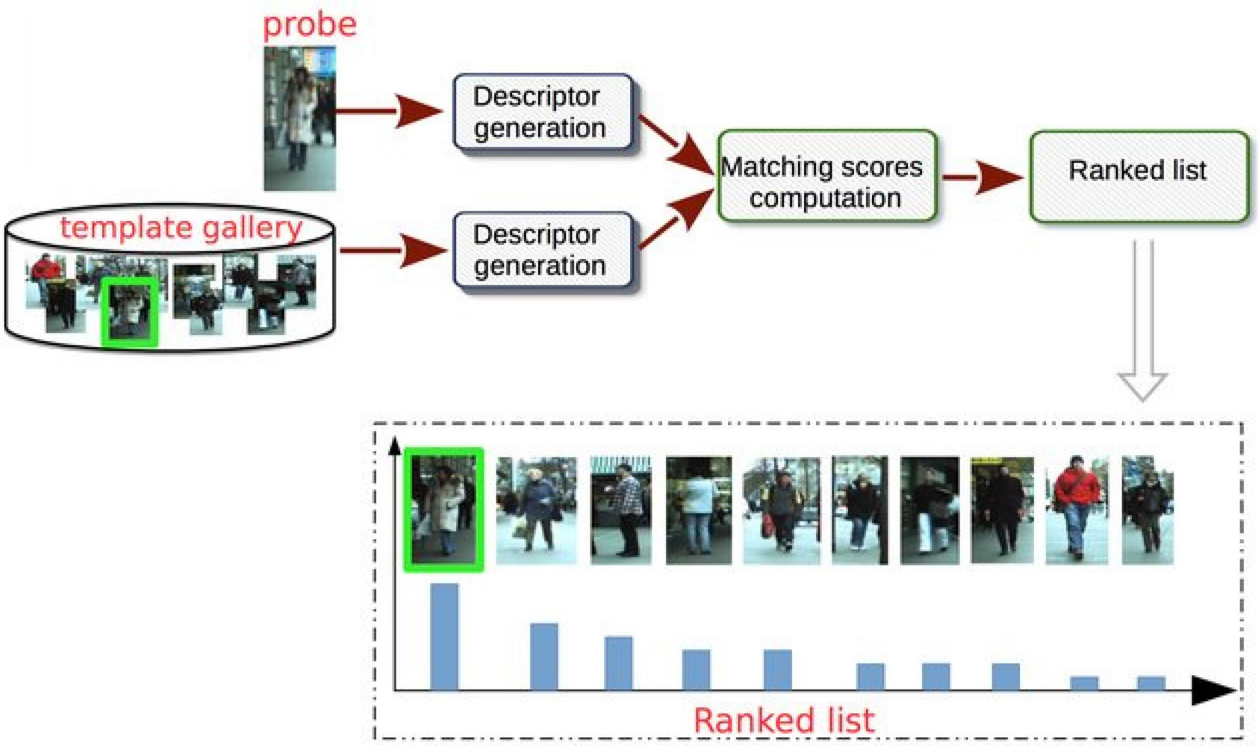
Object Re-Identification: Why Useful?

- **Authorization**
  - Is person an employee authorized in this building?
- **Safety**
  - Find lost child
- **Personalization**
  - How can we personalize the service for this customer?
- **Statistics**
  - Avoid double-counting objects
- **Algorithmic kernel**
  - Tracking requires re-ID of object from previous frames
  - Clustering
  - De-duplication



https://opencv.org/multiple-object-tracking-in-realtime/

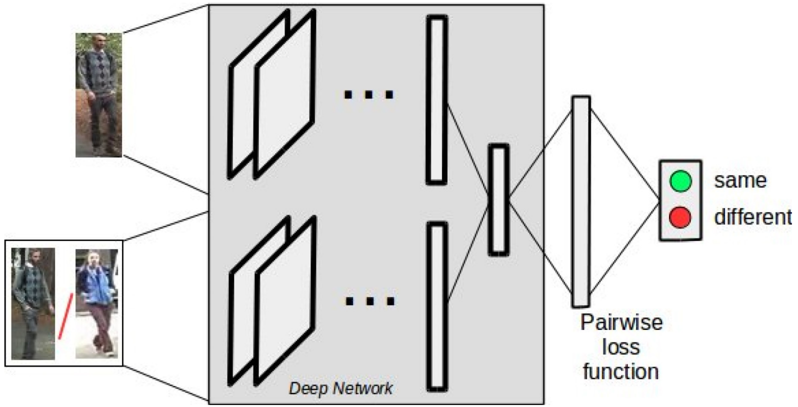
Object Re-ID General Approach



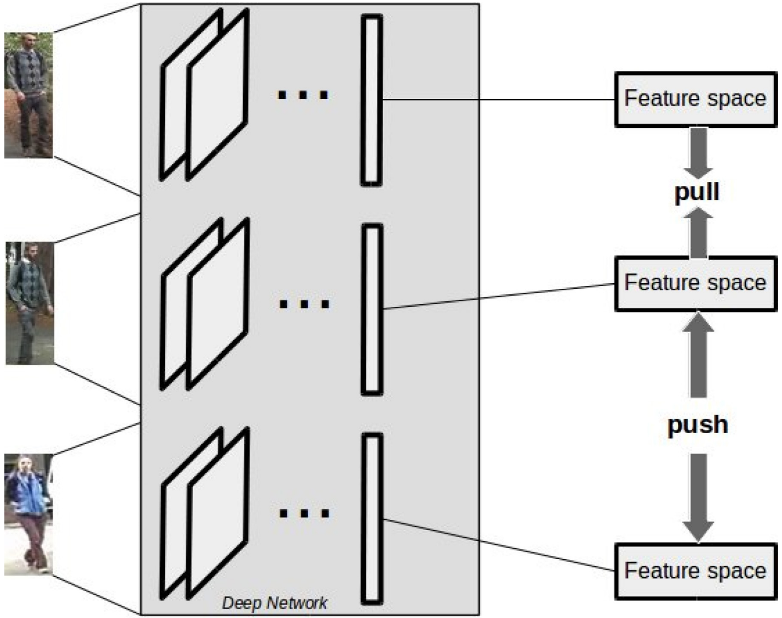
Lavi et al. 2018. Survey on Deep Learning Techniques for Person Re-Identification Task

Object Re-ID General Approach

Pairwise-loss feature learning

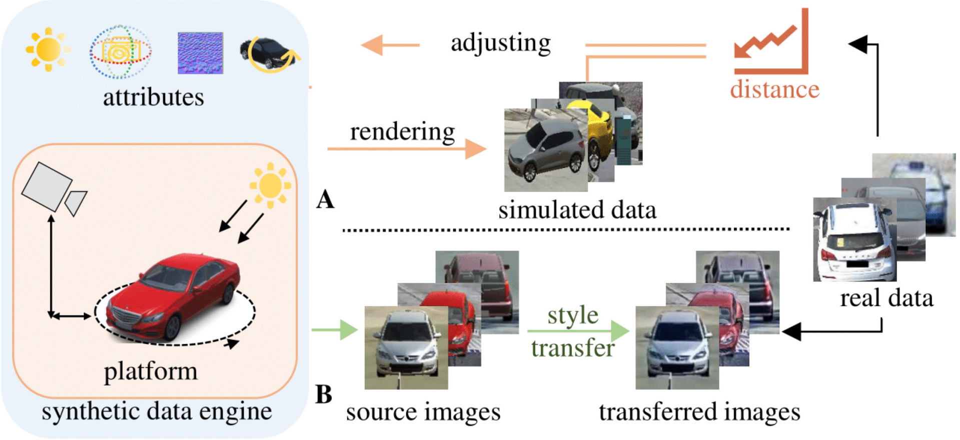


Triplet-loss feature learning



Lavi et al. 2020. Survey on Reliable Deep Learning-Based Person Re-Identification Models: Are We There Yet?

City-Scale Multi-Camera Vehicle ReID



- **Real dataset**
  - **880** real identities
  - **52,717** training crops
  - **31,238** testing crops
  - **1,103** image queries
- **Synthetic dataset**
  - **1,362** synthetic identities
  - **190,000+** synthetic crops
  - Synthetic **attributes** available: Type, color, lighting, orientation, etc.

City-Scale Multi-Camera Vehicle ReID

- **What most teams did**
  - Training supervision: Identity classification (cross-entropy loss) + metric learning (triplet loss)
  - Generation of synthetic data and application of domain adaptation
  - Use of synthetic data: (1) Direct training data augmentation; (2) Training attribute classifiers
- **What leading teams did**
  - Use of identity clustering to generate pseudo-labels on the test data to expand the training set
  - Application of spatial, temporal, and channel-wise attention mechanisms
- **What we learned year over year**
  - Stronger models by using circle loss, synthetic data pretraining, and effective post-processing
  - mAP improved by synthetic data is at least 3% (Baidu-UTS) and can be as large as 10% (RuiyanAI)

Object Tracking: Why Useful?

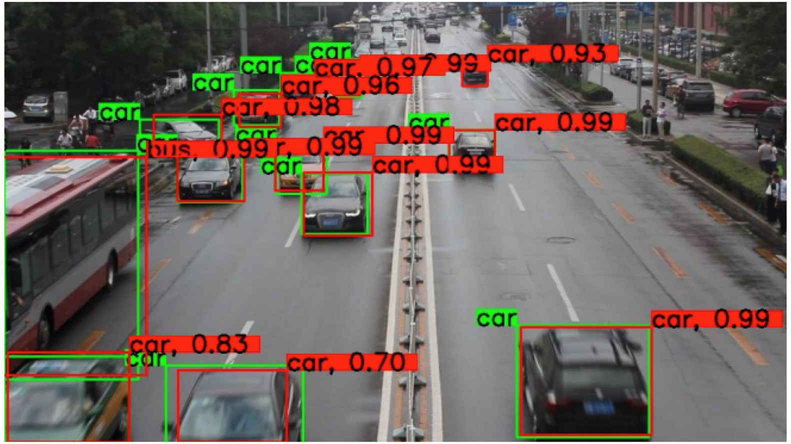
- **Understand behavior**
  - Are people social distancing?
  - What do customers gravitate towards in the store?
  - How many cars vs. trucks turn right at the next intersection?
- **Prevent bad behavior or accidents**
  - Warning: exceeding speed limit in zone
  - Breaking due to imminent collision
  - Modify traffic patterns to alleviate congestion
- **Render aid**
  - Dispatch aid car when accident detected
- **Provide service**
  - Delivery robot navigation
  - Self-driving car

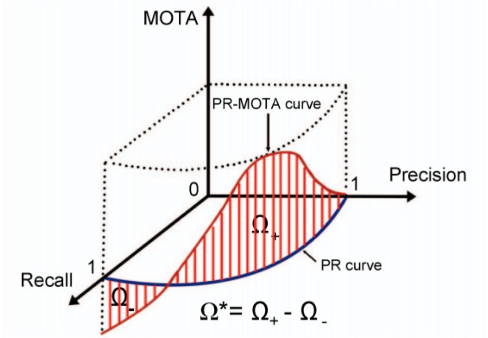
https://opencv.org/multiple-object-tracking-in-realtime/

Measuring Performance

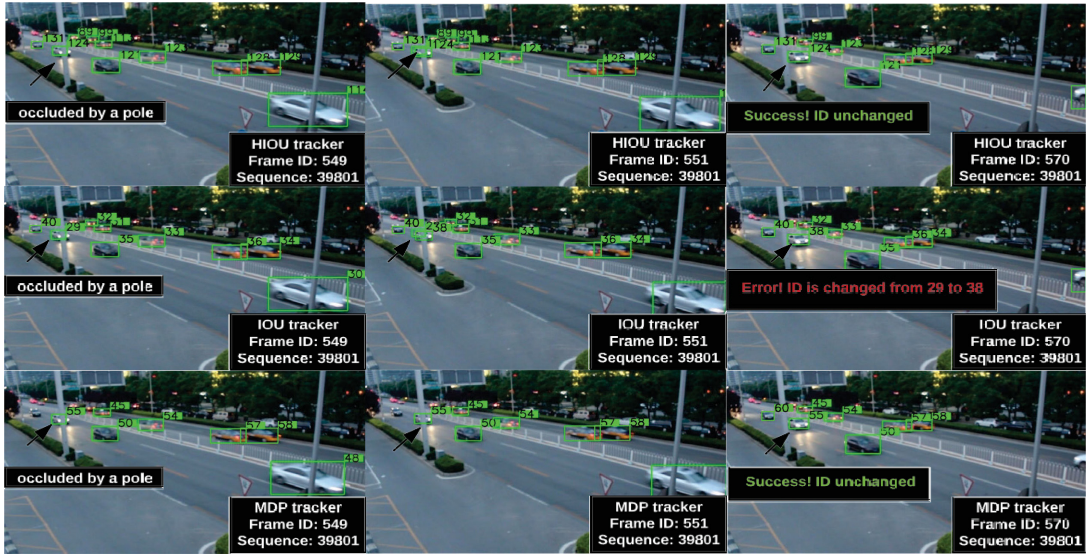
- **What is a correct prediction?**
  - Class must match
  - Bounding box must match in each frame
    - Intersection over union (IoU score)
  - Maintain tracking ID
    - Count number of ID switches
  - Track completion
    - Track at least N% of the GT track
    - Percent of tracks that were fragmented
  - Multi-object tracking accuracy (MOTA)
  - Multi-object tracking precision (MOTP)
  - Other metrics

$$MOTA = 100 \times \left(1 - \frac{\sum_t (FN_t + FP_t + IDS_t)}{\sum_t GT_t}\right)$$





Vehicle Tracking



Shuai Hua & David C. Anastasiu. Effective Vehicle Tracking Algorithm for Smart Traffic Networks. In Thirteenth IEEE International Conference on Service-Oriented System Engineering (SOSE) (SOSE 2019), IEEE, 2019

Our Solution

- History-Based IOU Tracker (HIOU)
  - Bounding-box association problem
  - Uses only historical bounding box location data
  - Overlap computed as Intersection-Over-Union
  - Expected overlap lowered linearly historically
- Performance
  - Outperformed IOU and MDP trackers overall
  - Significantly outperformed in rainy and foggy conditions
  - Only slight performance penalty compared to IOU

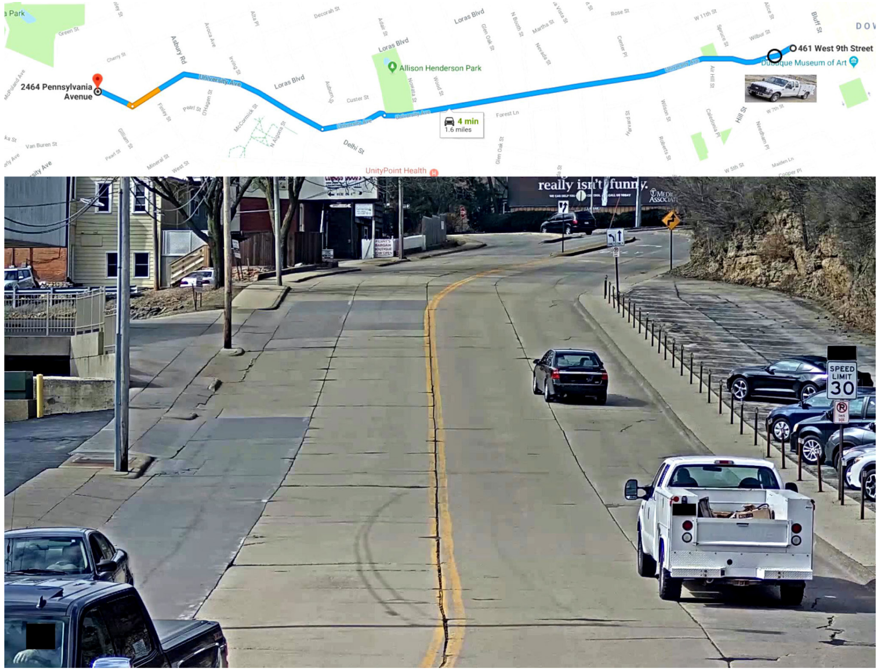
Vehicle Tracking - IRL

GovComm

- Wrong-way vehicle detection
- IoT-based system
- Alerts driver by lighting wrong-way sign
- Alerts traffic managers and law enforcement
- Captures confirmation images

https://govcomm.us/wrong-way-video-detection-systems/

City-Scale Multi-Camera Vehicle Tracking

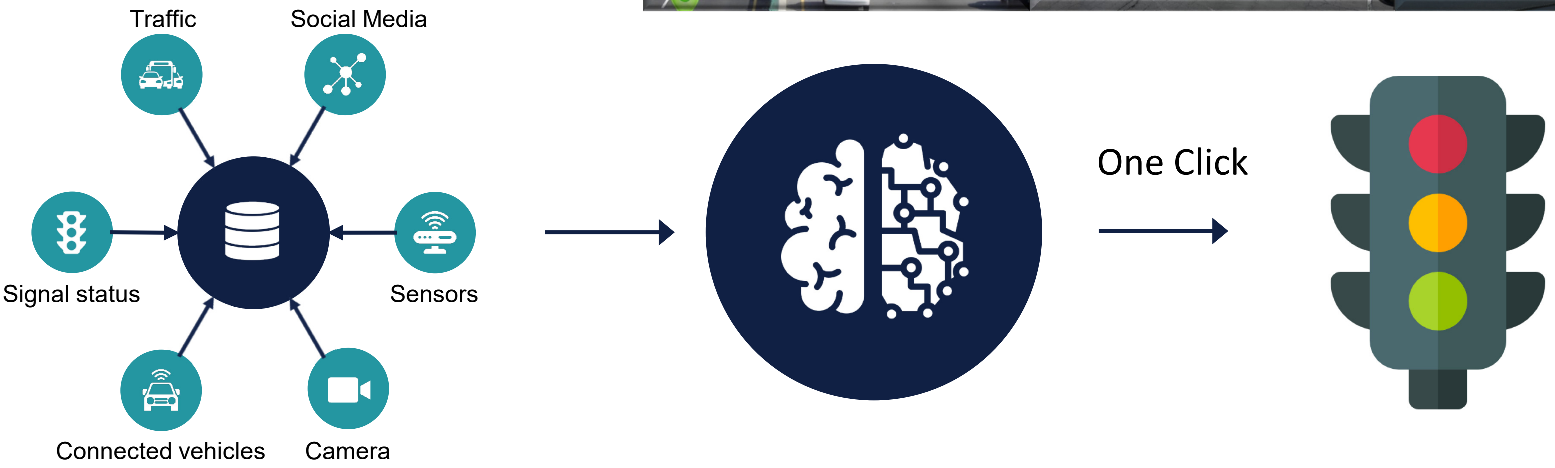


- **6** synchronized scenarios
- **215.03** mins of videos
- **46** camera views
- **16** intersections
- **4** km between furthest simultaneous cameras
- **300K+** bounding boxes
- **880** vehicle identities
- **Camera calibration & single-camera tracking results** available

City-Scale Multi-Camera Vehicle Tracking

- **What most teams did**
  - Detection -> ReID feature extraction -> Single-camera tracking -> Inter-camera association
  - Inter-camera association using similarity matrices with appearance and spatio-temporal information
- **What leading teams did**
  - Application of the state-of-the-art detectors (e.g., YOLOv5) and trackers (e.g., DeepSORT)
  - Bag of tricks such as offline re-link, forward/backward fusion, box-grained distance matrix, etc.
- **What we learned year over year**
  - The major improvement is from refining single-camera tracklets and hierarchical clustering.
  - Synthetic data and domain adaptation are useful for enhancing ReID features.

Vehicle Tracking - IRL



Continuous data monitoring

AI Engine automatically identify signals

Optimal signal retiming

Traffic Anomaly Detection

Dataset

- **250** video clips
- **15** mins duration each
- **800x410** resolution
- **30** fps
- **21** anomalies
- Various **weather & lighting** conditions

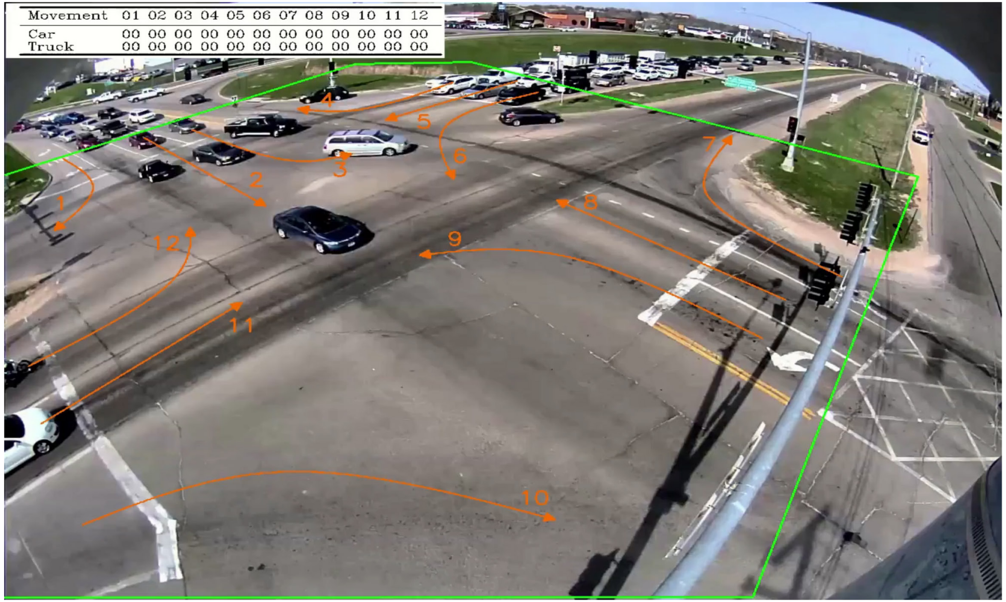
Anomalies

- Vehicles stopped on the side of the road
- Driving on the wrong side of the road
- Accidents
- Stalls

Traffic Anomaly Detection

- **What most teams did**
  - Removal of stationary parked vehicles and abnormal vehicle tracking
  - Pre-processing and detection of vehicles using DCNNs
  - Backtracking optimization to refine anomaly prediction
- **What leading teams did**
  - Dynamic tracking that utilizes spatio-temporal status and motion patterns to determine timestamps
  - Multi-granularity strategy consisting of a box-level and a pixel-level tracking branches
  - Unsupervised system using *k*-means clustering to identify potential anomaly regions
- **What we learned year over year**
  - The expansion of test set makes a more realistic simulation of situations in real world.
  - Anomaly detection in real-world remains challenging.

Multi-Class Multi-Movement Vehicle Counting



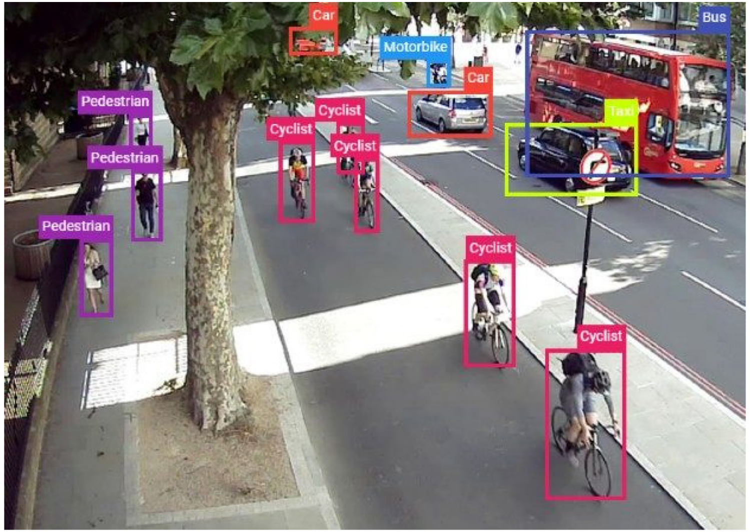
- **31** video clips
- **9** hours in total
- **20** unique camera views
- **960p+** resolution
- **10** fps
- **ROIs & MOIs** provided
- **2** vehicle types (*car* & *truck*)

	Dataset_A	Dataset_B
#cameras	20	20
#videos	31	31
total length	5 hours	4.5 hours
total # frames	185446	166964

Multi-Class Multi-Movement Vehicle Counting

- What most teams did
  - Deep-learning-based vehicle detection and single-camera tracking
  - Movement assignment: (1) ROI/tripwire-based rules; (2) trajectory template + similarity matching
- What leading teams did
  - Trajectory enhancement methods including smoothing, filtering, etc
  - Modeling of trajectories based on clustering for creating data-driven ROIs
- What we learned
  - Trajectory enhancement helps improve the effectiveness
  - Simpler tracker performs reasonably well in vehicle tracking
  - The best accuracy achieved by team is around 94%
  - The vehicle counting task can be run end-2-end at real-time speed. Some teams achieved ~35fps with reasonable accuracy.

Vehicle/Pedestrian Counting - IRL



https://resources.nvidia.com/en-us-metropolis-software-success-stories/embedded-vivacity-la

- VivaCity traffic sensors in UK, Australia, USA, 7+ others
- Real-time anonymous data on how road spaces are used
- >97% accuracy counting cyclists
- Monitoring cyclist counts can help identify areas in need or more bike lanes
- Detect 99.5% of all stationary and slow vehicles in tunnels
- Ethical considerations
  - Anonymity of pedestrians/cyclists/drivers

Naturalistic Driving Action Recognition

Camera	Location
Dash Cam 1	Dashboard
Dash Cam 2	Behind rear view mirror
Dash Cam 3	Top right -side window

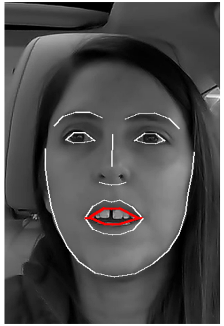
- **18** distracted driver activities
- Approx. **10** hours of videos from **10** diverse drivers
- **3** synchronized camera views
- Manually annotated labels of approx. **5** hours of videos(30 clips) for training
- Recorded at **30** frames/sec
- **1920 x 1080** resolution
- With and without appearance block
- Random order and duration of activities

1 Normal forward driving
2 Drinking
3 Phone call (right)
4 Phone call (left)
5 Eating
6 Texting (right)
7 Texting (left)
8 Hair / makeup
9 Reaching behind
10 Adjusting control panel
11 Picking up from floor (driver)
12 Picking up from floor (passenger)
13 Talking to passenger at the right
14 Talking to passenger at backseat
15 Yawning
16 Hand on head
17 Singing with music
18 Shaking or dancing with music

Naturalistic Driving Action Recognition

Our Solution

- Focus on movement of key points on the driver's body
- Heuristics to differentiate driver from passenger
- Per-frame activity classification
  - Pre-trained pose and facial expression estimation models to identify key points
  - Extract features from pose based on point locations and distances, key segment angles, position and angle shifts, and facial features
- Merge activity over time
  - Merge consecutive same-activity segments
  - Report maximal consecutive class segment of at least *MinLen* frames



Arpita Vats & David C. Anastasiu. Key Point-Based Driver Activity Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pages 3274-3281, 2022.

Naturalistic Driving Action Recognition

- **What most teams did**
  - Classification of different distracted activities such as eating, yawning etc.
  - Temporal Action Localization to determine the start and end time for each activity
- **What leading teams did**
  - Utilizing a 3D Action Recognition Model X3D, Multi action transformers or ConvNext model for classification
  - Localization and sliding window classification was used for determining the start and end points
- **Future directions**
  - Annotations for each activity can be more refined for better localization.
  - Cleaner ground-truth labels can improve the development and evaluation of the research progress
  - Increasing the data size with adequate representation can improve classification results (going from 5 drivers to 30-50 drivers)

Conclusions

- Video cameras are ubiquitous on our roads and in our cities
- AI + video sensors → improved traffic safety and reliability
- Technology is mature and currently being implemented in a city near you…

Questions?