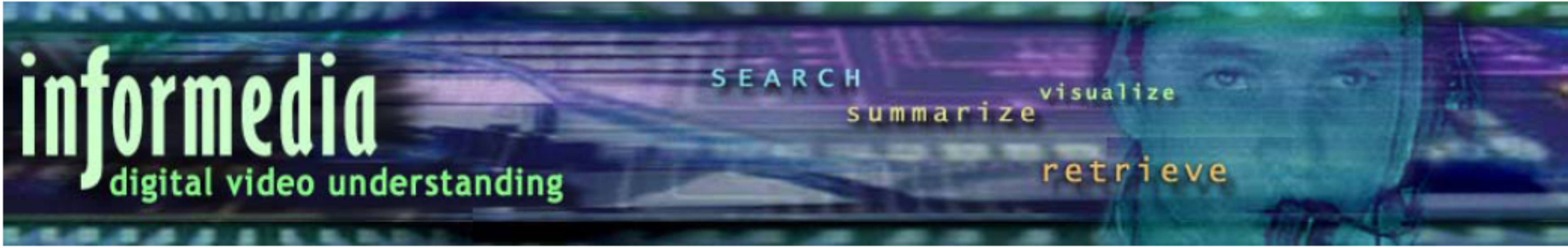


Zero-VIRUS: Zero-shot VehIcle Route Understanding System for Intelligent Transportation

Lijun Yu[1] Qianyu Feng[1,2] Yijun Qian[1] Wenhe Liu[1] Alexander G. Hauptmann[1]

[1]Carnegie Mellon University [2]University of Technology Sydney

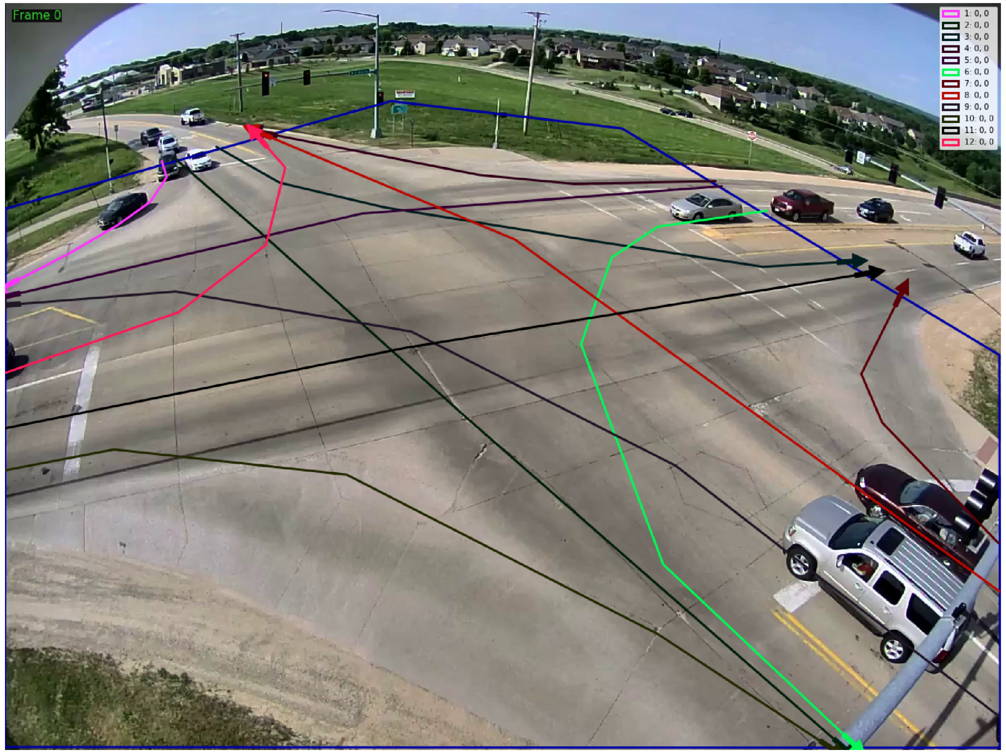
lijun@cmu.edu

Carnegie Mellon University
Language Technologies Institute

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

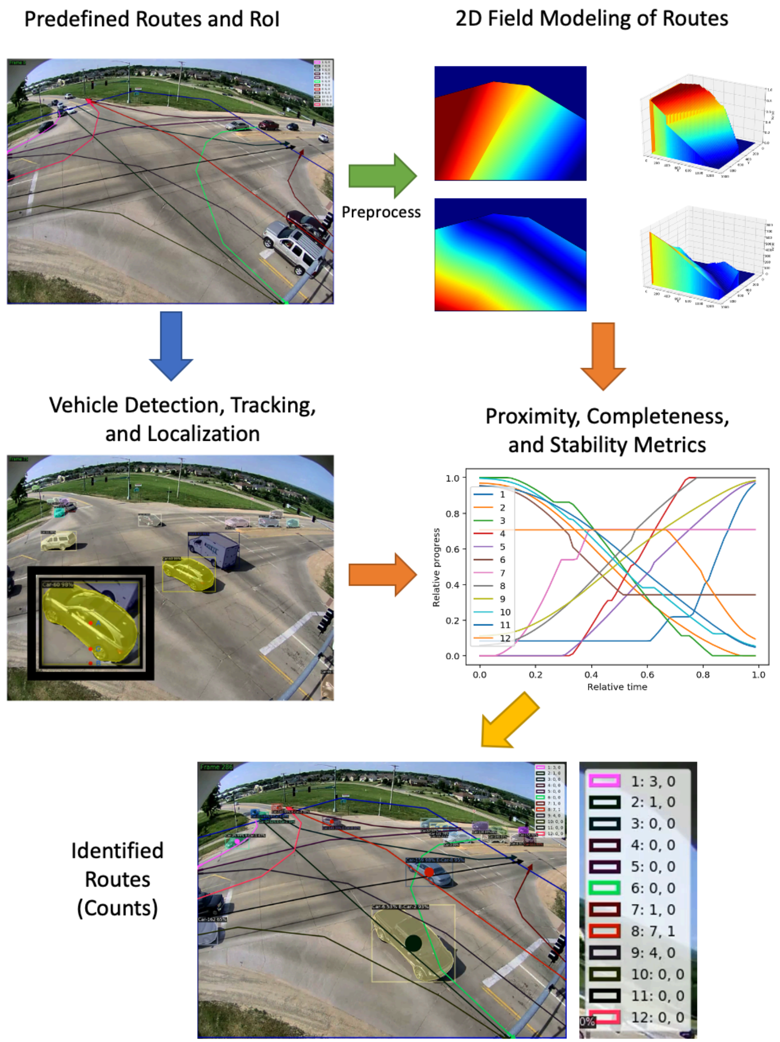
Introduction

- AI City Challenge 2020 Track 1:
  Multi-Class Multi-Movement Vehicle Counting
- Input:
  - Video from a stable surveillance camera view
  - Pre-defined region and movements
- Goal:
  - Report the movement ID of each vehicle at the time of exiting the region
- No ground truth provided



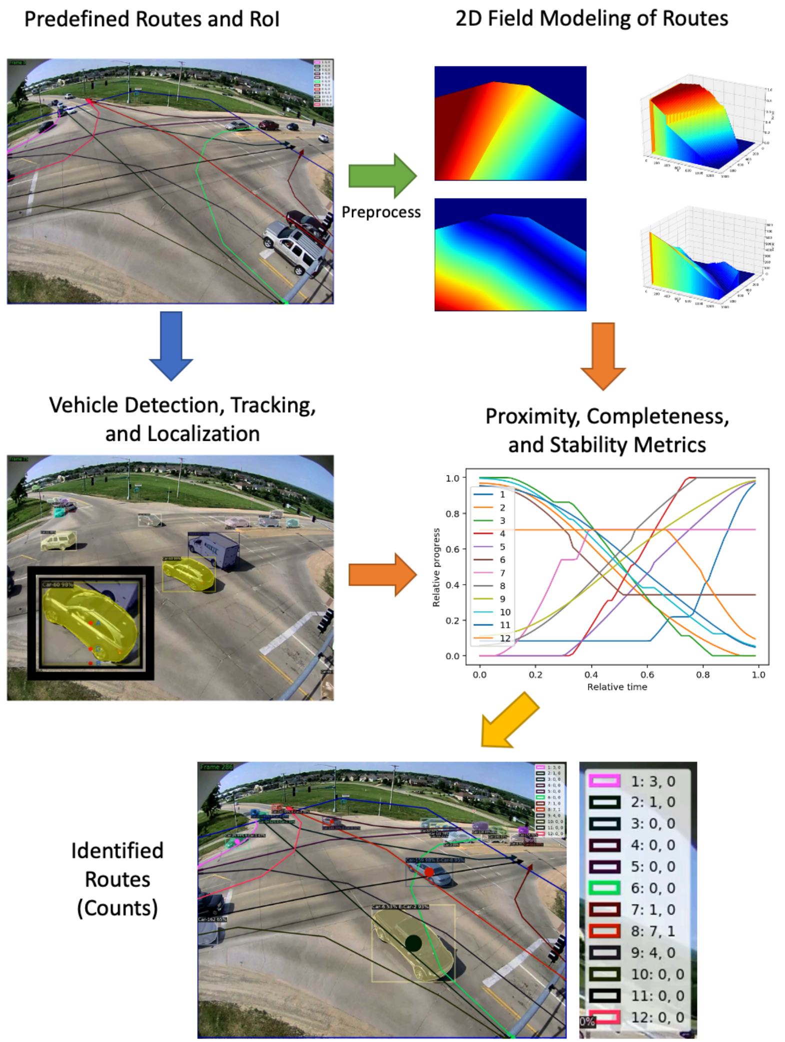
Approach Overview

- Vehicle trajectory
  - Detection, tracking, localization
- Pre-defined route
  - 2D field modeling as feature representation
- Trajectory vs. route
  - similarity metrics: proximity, completeness, stability
- Route classification and vehicle counting



Design Advantage

- No data required
  - No need for ground truth labels
  - No need for statistics
- Online processing
  - Frame by frame for deployment
  - No post-processing, e.g. clustering
- Easy adaption
  - New camera: just define the routes

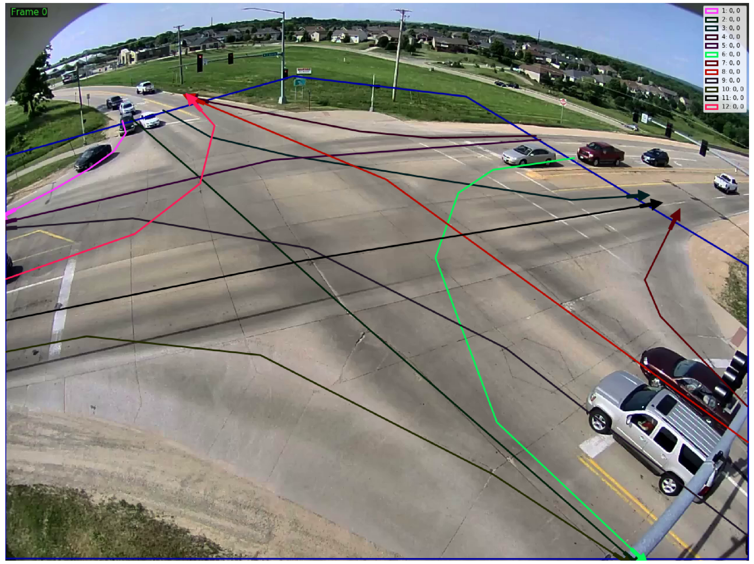


Vehicle Detection and Tracking

- Detection: Mask R-CNN
  - ResNeXt-101-FPN backbone, trained on COCO
- Tracking: Towards-Realtime-MOT
  - RoI feature from detection model
- Object classes: Car, Truck
  - inconsistent definition with COCO, e.g. pickups
  - Weighted inter-class non-maximum suppression
  - Tracklet label refinement

Trajectory Enhancement

- Localization with segmentation
  - Point C: bottom center of segmentation mask
  - Scale factor: diagonal of bounding box
- Interpolation and smoothing
  - Linear interpolation to fill in gaps
  - 1D-gaussian smoothing
- Movement and region filter
  - Filter out stopped period based on local average speed, e.g. traffic lights or jams
  - Within a pre-defined region of interest



Route Modeling

- Route: A polyline $\mathbf{R}_i = [P_1^i \quad P_2^i \quad \ldots \quad P_n^i]_i^T$
- Proximity field: distance to a route

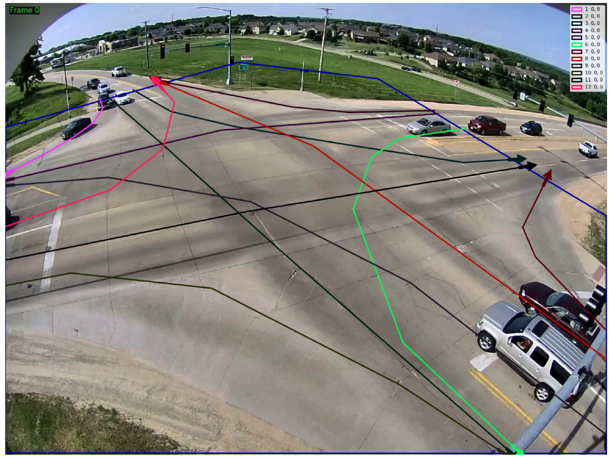
$$\mathbf{F}_{p,\mathbf{R}_i}(X) = \min_j d(X, \overline{P_j^i P_{j+1}^i})$$

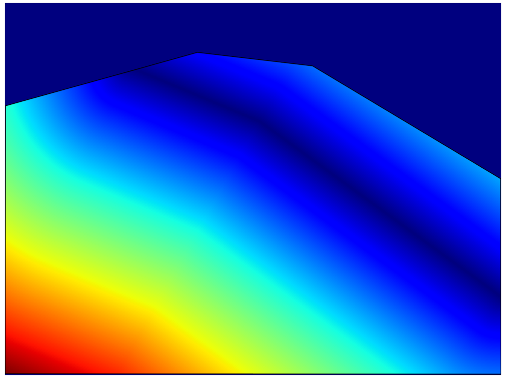
  - point-segment distance

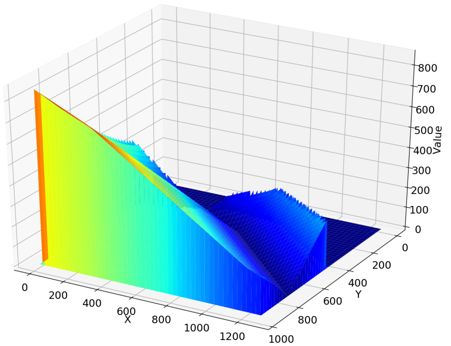
$$d(X, \overline{P_j^i P_{j+1}^i}) = \begin{cases} \|\overrightarrow{XP_j^i}\| & \alpha_j \leq 0 \\ \|\overrightarrow{XX_j^i}\| & 0 < \alpha_j < 1 \\ \|\overrightarrow{XP_{j+1}^i}\| & \alpha_j \geq 1 \end{cases}$$

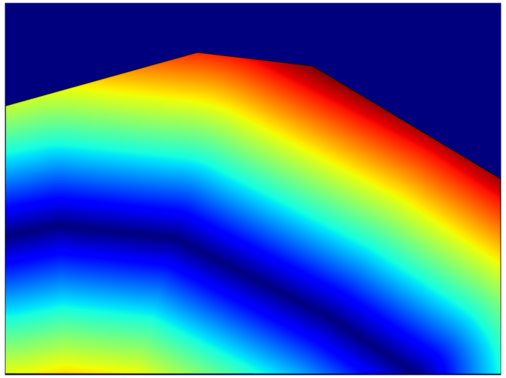
$$X_j^i = P_j^i + \alpha_j^i(X)\overrightarrow{P_j^i P_{j+1}^i}$$

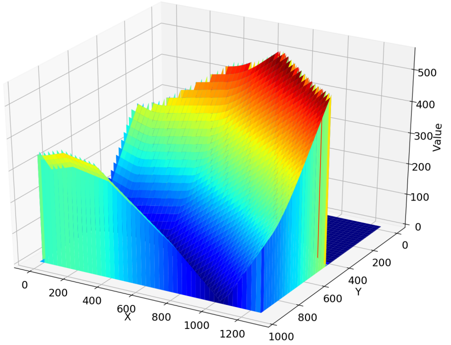
$$\alpha_j^i(X) = \frac{\overrightarrow{P_j^i X} \cdot \overrightarrow{P_j^i P_{j+1}^i}}{\|\overrightarrow{P_j^i P_{j+1}^i}\|^2}$$











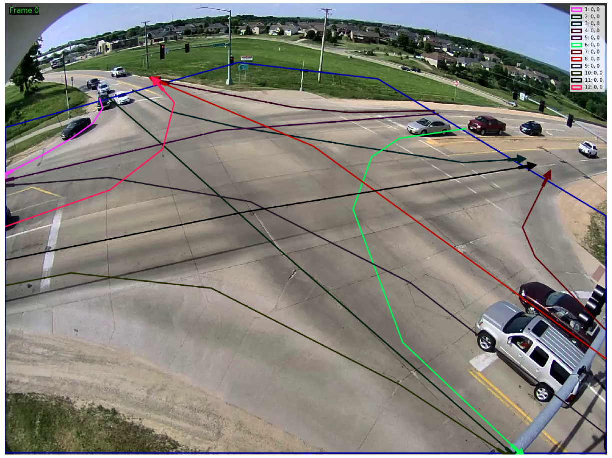
Route Modeling

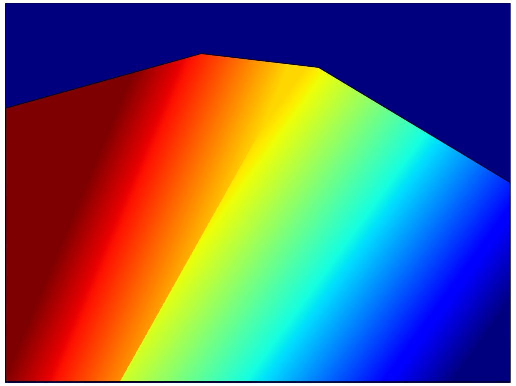
- Completeness field:
  relative location within a route

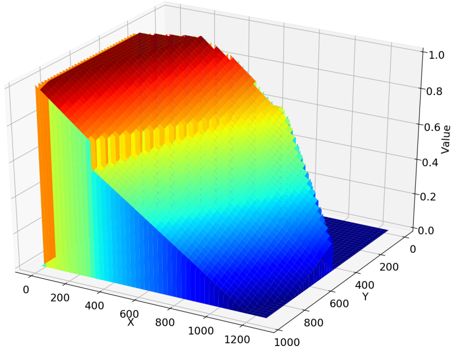
$$\mathbf{F}_{c,\mathbf{R}_i}(X) = \frac{\alpha^i_{j^*}(X)\|\overrightarrow{P^i_{j^*}P^i_{j^*+1}}\| + \sum_{j=1}^{j^*-1}\|\overrightarrow{P^i_jP^i_{j+1}}\|}{\sum_{j=1}^{n-1}\|\overrightarrow{P^i_jP^i_{j+1}}\|}$$

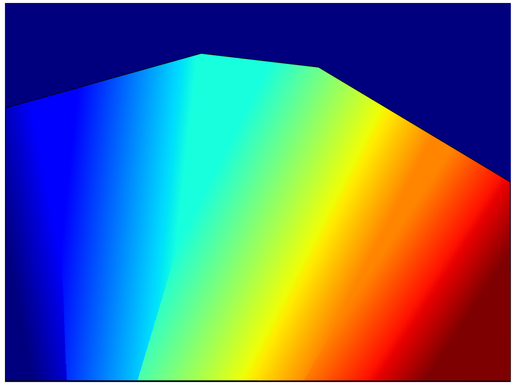
$$j^* = \underset{j=1,2,\ldots,n-1}{\arg\min}\; d(X, \overline{P^i_jP^i_{j+1}})$$

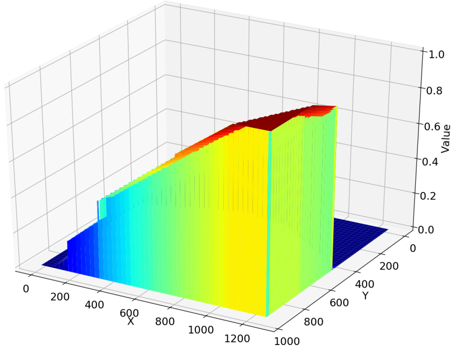
$$\alpha^i_j(X) = \frac{\overrightarrow{P^i_jX}\cdot\overrightarrow{P^i_jP^i_{j+1}}}{\|\overrightarrow{P^i_jP^i_{j+1}}\|^2}$$











Route Identification

- Proximity metric: scale-normalized average distance

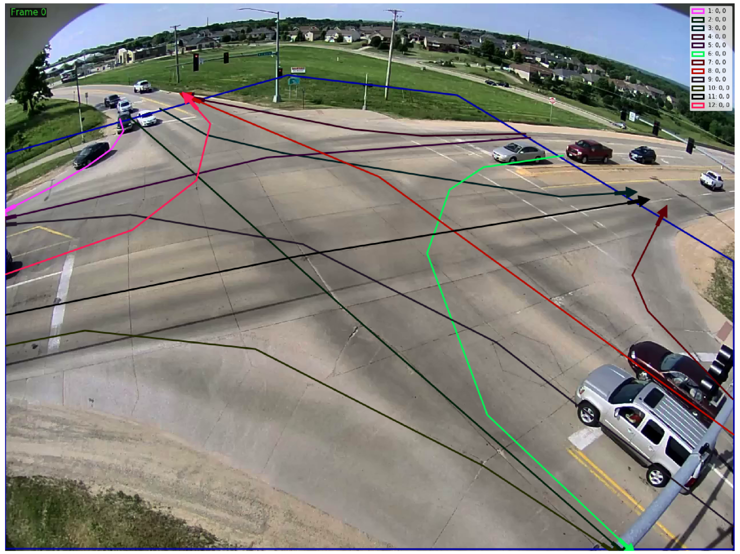
$$d(\mathbf{T}_{x,j}, \mathbf{R}_i) = \frac{\mathbf{F}_{p,\mathbf{R}_i}(\mathbf{P}_{x,j})}{\mathbf{S}_{x,j}}$$

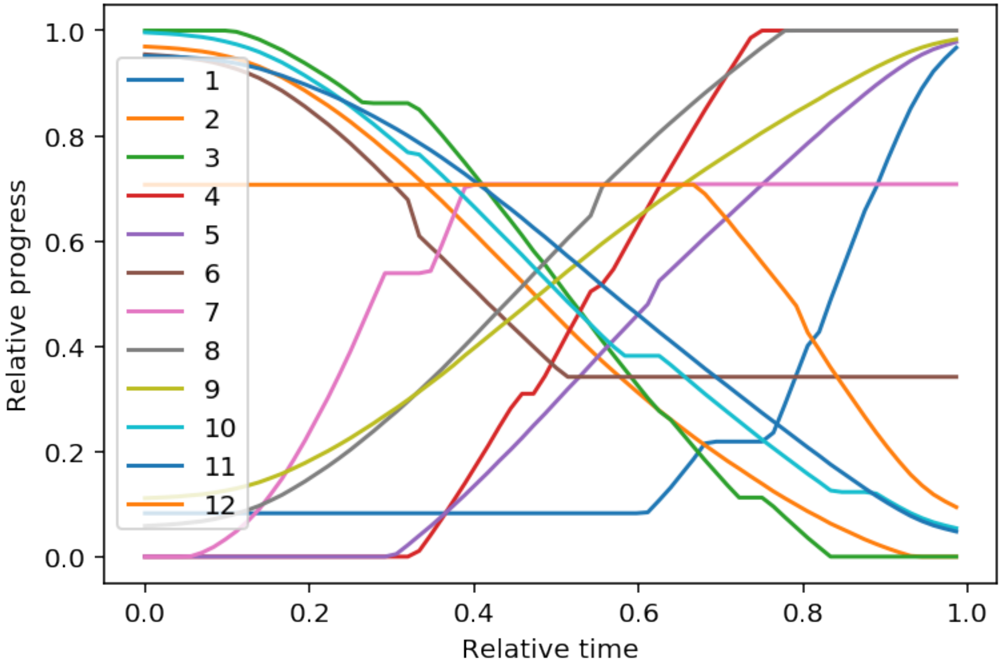
$$M_p(\mathbf{T}_x, \mathbf{R}_i) = \sigma(a - b\frac{1}{n}\sum_{j=1}^{n} d(\mathbf{T}_{x,j}, \mathbf{R}_i))$$

- Completeness metric: how a vehicle goes along a route
  - Linear model to approximate change of completeness. A perfect slope should be 1

$$\mathbf{F}_{c,\mathbf{R}_i}(\mathbf{P}_{x,j}) = c_{x,i}\frac{j}{n} + d_{x,i}$$

$$M_c(\mathbf{T}_x, \mathbf{R}_i) = \min(c_{x,i}, \frac{1}{c_{x,i}})$$





Route Identification

- Stability Metric: does vehicle go along the route at constant distance
  - Linear model to approximate the change of scale-normalized distance. A perfect slope should be 0.

$$d(\mathbf{T}_{x,j}, \mathbf{R}_i) = e_{x,i}\frac{j}{n} + f_{x,i}$$

$$M_s(\mathbf{T}_x, \mathbf{R}_i) = \exp(-\frac{1}{2}e_{x,i}^2)$$

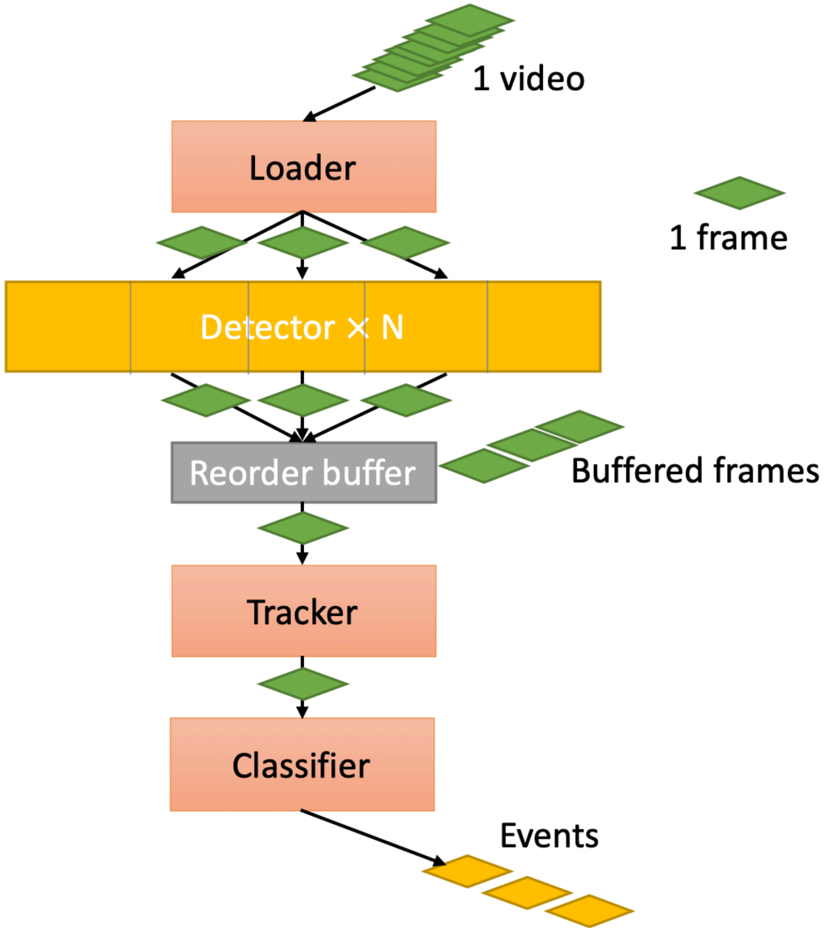
- Aggregation and classification

$$\begin{aligned} S(\mathbf{T}_x, \mathbf{R}_i) = & \min(1, \max(0, w_p M_p(\mathbf{T}_x, \mathbf{R}_i))) \\ & + \min(1, \max(0, w_c M_c(\mathbf{T}_x, \mathbf{R}_i))) \\ & + \min(1, \max(0, w_s M_s(\mathbf{T}_x, \mathbf{R}_i))) \end{aligned}$$

$$C(\mathbf{T}_x, \mathbf{R}) = \arg\max_i S(\mathbf{T}_x, \mathbf{R}_i)$$

System Implementation

- Mask R-CNN from Detectron2 in PyTorch
- Toward-Realtime-MOT from its author
- Pipelined system
    - Frame-level parallelism
    - Out-of-order execution for bottleneck
    - 9 fps on single 2080Ti GPU
    - Support up to 8 GPUs



Dataset and Metrics

- 2020 AI City Challenge Track 1 dataset split A
  - 5 hours video
  - 20 unique cameras in different light and weather conditions
  - Diagrams of routes with text descriptions of each camera
- Our annotation: route polylines of each camera
- Official metric:
  - Efficiency: running time and a base factor of hardware
  - Effectiveness:

$$wRMSE = \sqrt{\sum_{i=1}^{k} \frac{i}{\sum_{j=1}^{k} j} (\hat{x}_i - x_i)^2}$$

Results

- No ground truth provided
- Only leaderboard results
- Full test set

Table 1. Summary of the Track 1 leader board.

Rank	Team ID	Team name (and paper)	Score
1	99	Baidu [20]	0.9389
2	110	ENGIE [27]	0.9346
3	92	CMU [46]	0.9292
6	74	BUT [37]	0.8829
7	6	KISTI [5]	0.8540
9	80	HCMUS [43]	0.8064
13	75	UAlbany [6]	0.3116
N/A (*General*)	60	DiDi [2]	0.9260
N/A (*General*)	108	VT [1]	0.8138

- 50% test set
  - Maybe overfitted, maybe imbalanced subset

Rank	TeamID	Score
1	**Ours**	**0.9444**
2	99	0.9415
3	110	0.9381

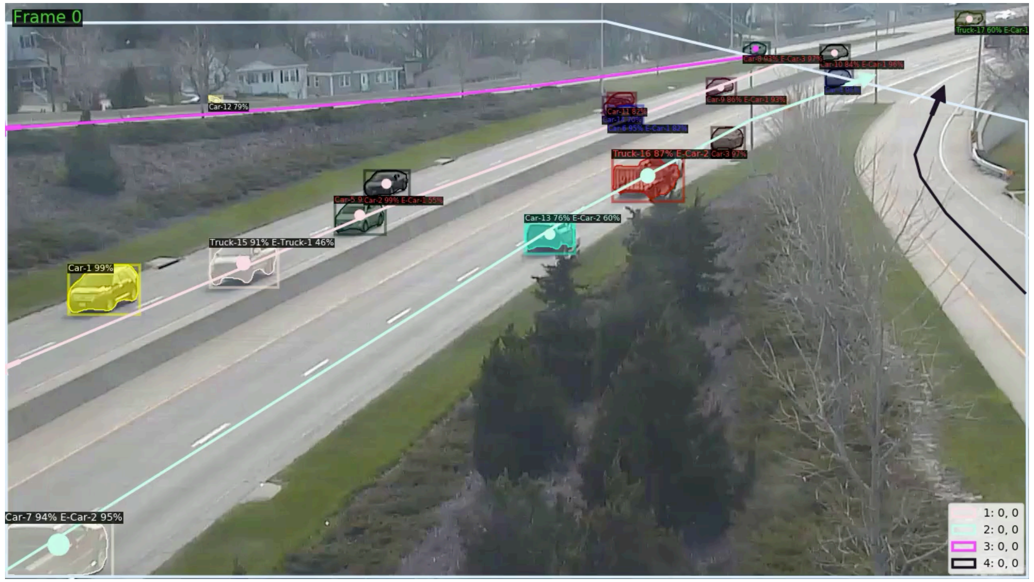
Ablations

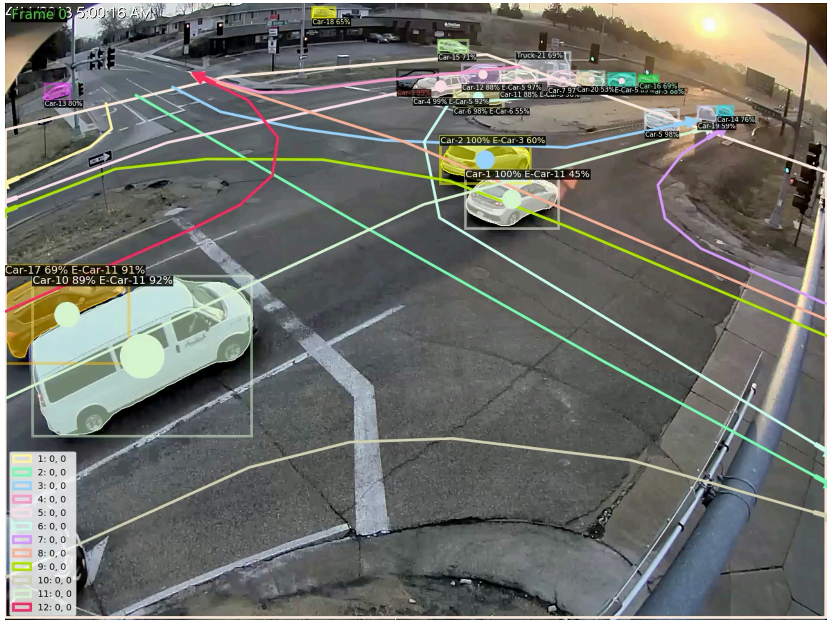
- Effectiveness of three metrics on a 60-second clip
  - Proximity metric, completeness metric, stability metric

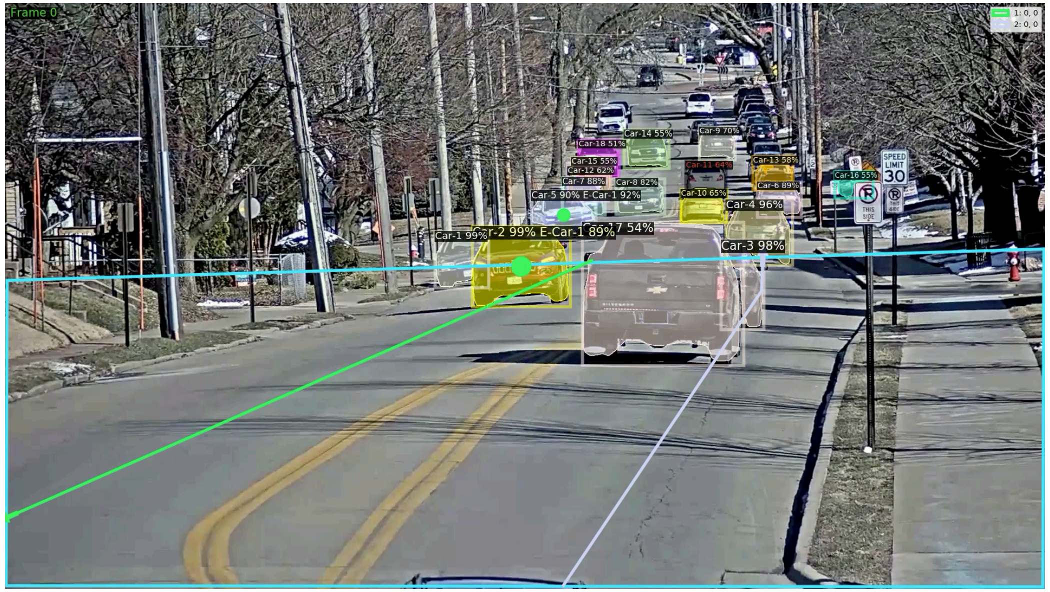
Metrics	Effectiveness Score
M_p	0.8903
M_p, M_c	0.9455
M_p, M_c, M_s	**0.9554**

Qualitative Results

https://drive.google.com/drive/folders/1s3TPykPa3JTaPOHUVOQF8S4iUi3SduAN?usp=sharing







Conclusion

- Zero-shot vehicle route identification
  - Minimal manual effort: define routes
  - Effective and efficient
- Future improvement
  - Detector: finetune between Car and Truck ----Biggest problem now
  - Trajectory: use ground trajectory if cameras are calibrated
- Zero-VIRUS, no coronavirus!