KrishnaCam: Using a Longitudinal, Single-Person, Egocentric Dataset for Scene Understanding Tasks

Krishna Kumar Singh, Prof. Kayvon Fatahalian, Prof. Alexei A. Efros

Carnegie Mellon University
University of California, Berkeley
University of California, Davis

Opportunity: ubiquitous visual sensing will soon provide the opportunity to record a large fraction of life's events

KrishnaCam Dataset

Time Span: 9 months

Duration: 70 hours

Total Clips: 460
(10-20 minutes)

Locations: Various neighborhoods in Pittsburgh

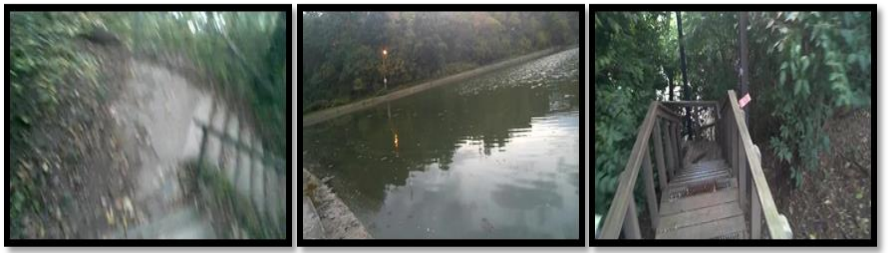
Device: Google Glass

Data: 720 p, 30 fps
Accelerometer,
Gyroscope,
Orientation, GPS

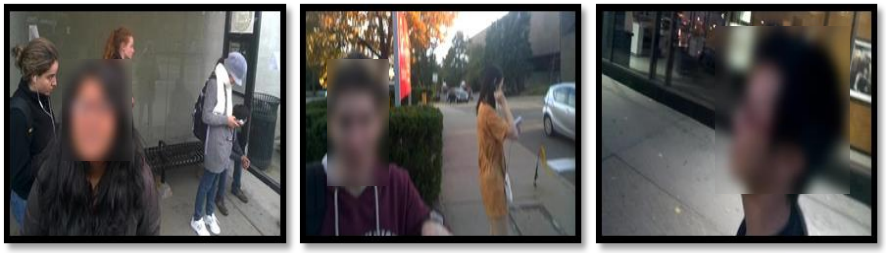
Dataset contains diverse life experiences

Walking in different neighborhoods

Visiting parks



Talking to people



Shopping and eating outdoor

Different time of day

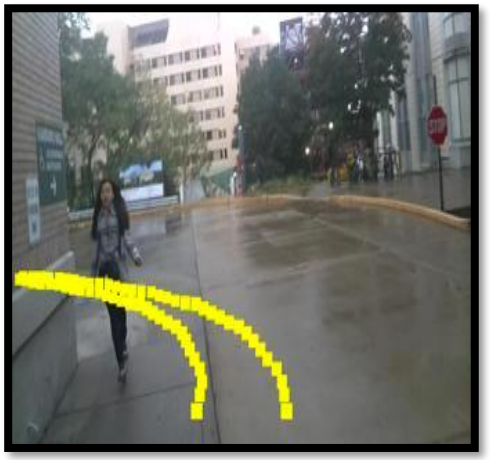
Different seasons

Prediction Task:
Where Will Krishna Move Next ?

Motion trajectories

- For a given frame, trajectory represents motion in the next 7 seconds.
- Trajectories are generated through sensor data and used as ground-truth.
- Yellow lines indicate movement.
- Red dots indicate stopping.
- Our goal is to predict these trajectories.

Forward



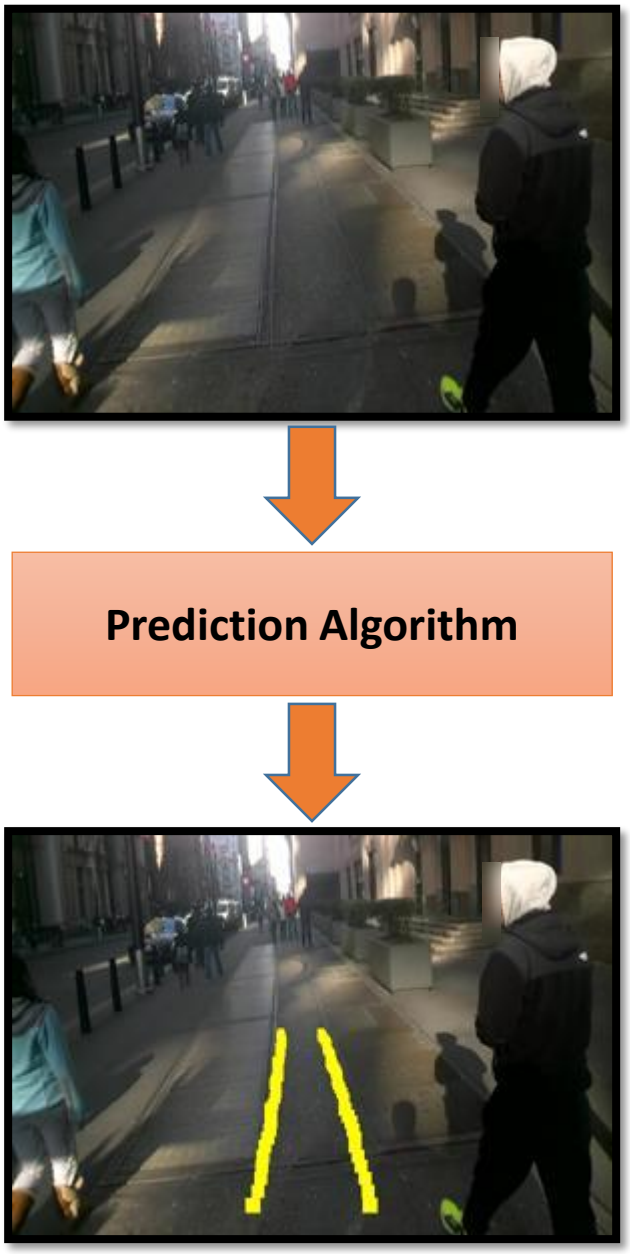
Turn

Stationary

About to stop

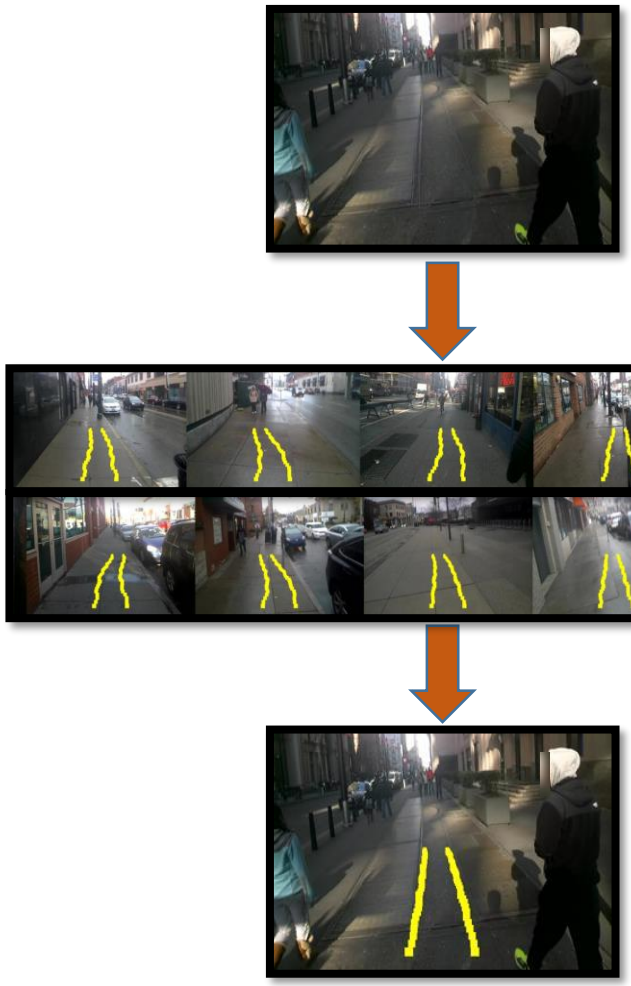
Ground-truth Trajectory examples

Prediction Problem



Trajectory prediction using nearest neighbors

- Goal: Predict future trajectory for a given frame.



Ground Truth Trajectory

Top 10 nearest neighbors using fifth layer deep features and cosine similarity as distance kernel

Predicted trajectory is average of trajectories of top 10 nearest neighbors

Predicting common human behaviors

People walk on sidewalk

Predicting common human behaviors

Ground Truth Predicted Top 10 Nearest Neighbors

People remain stationary while eating

Ground Truth Predicted Top 10 Nearest Neighbors

People stop soon after approaching a traffic button

Prediction of Krishna specific behaviors

Ground Truth **Predicted** **Top 10 Nearest Neighbors**

Turning left at particular intersection

Ground Truth **Predicted** **Top 10 Nearest Neighbors**

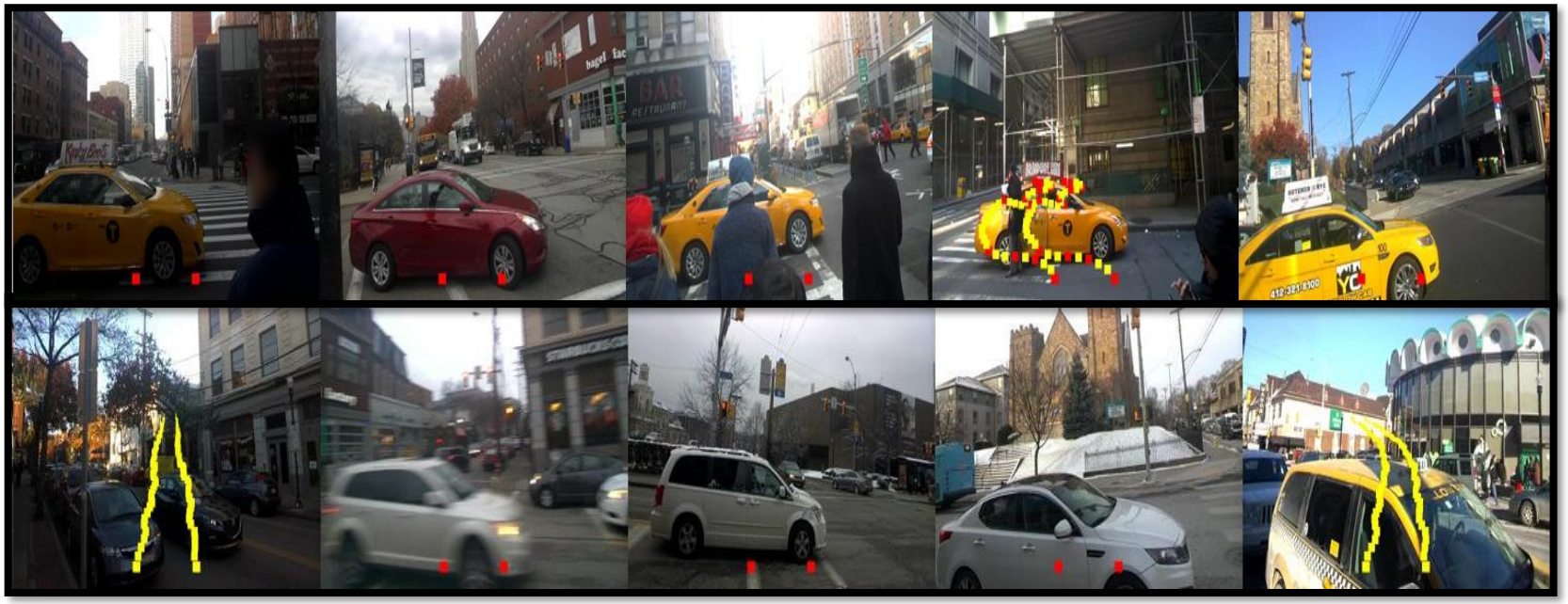
Turning right outside my house

Predicting behavior due to transient objects

Ground Truth

Predicted

Top 10 Nearest Neighbors



Predicting stop, if car is in front

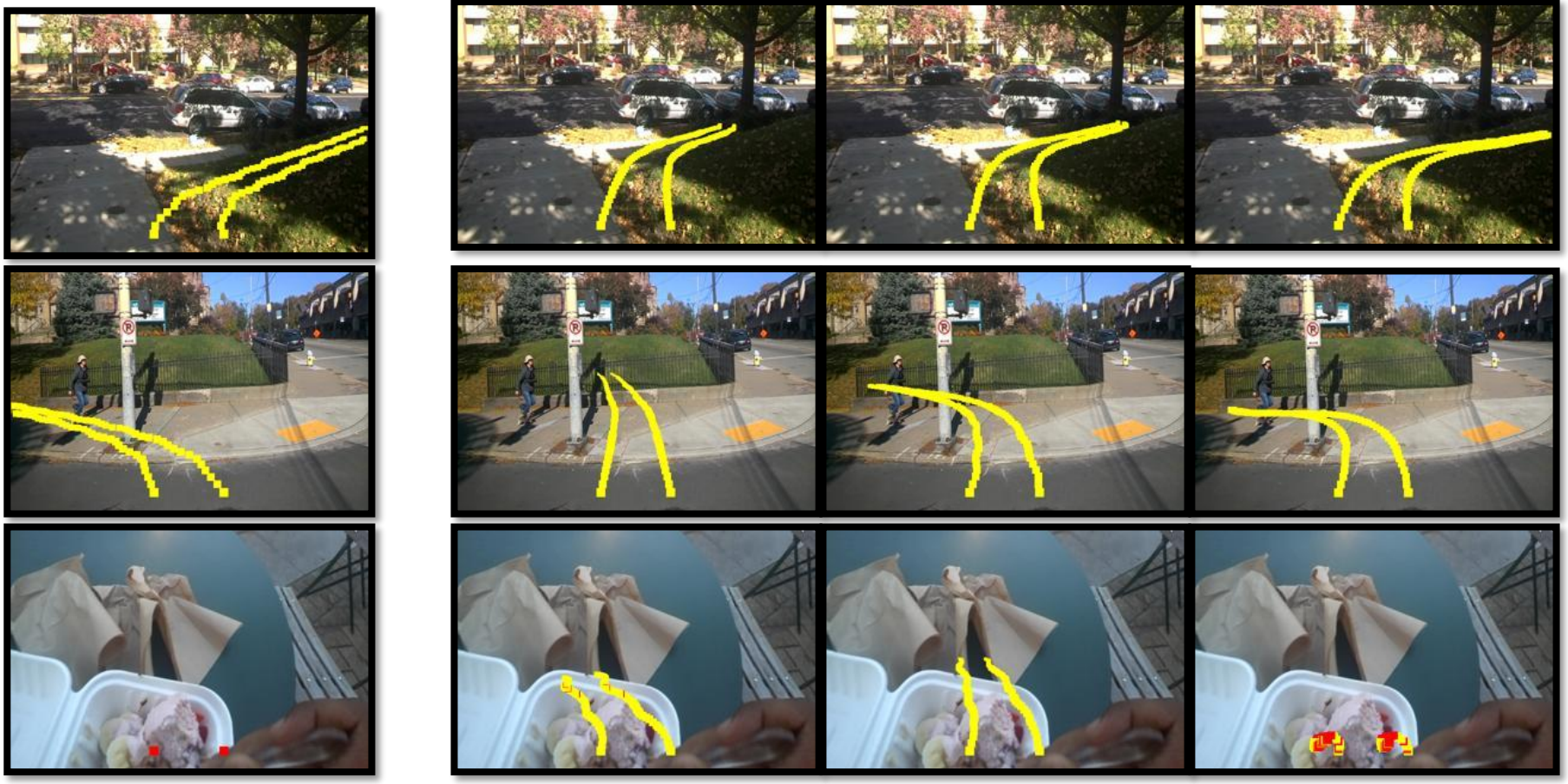
Is big data actually necessary for this task? (how much data do you really need?)

Prediction of rare events require more training data

Ground Truth

Prediction on different training data size

10 days **30 days** **50 days**

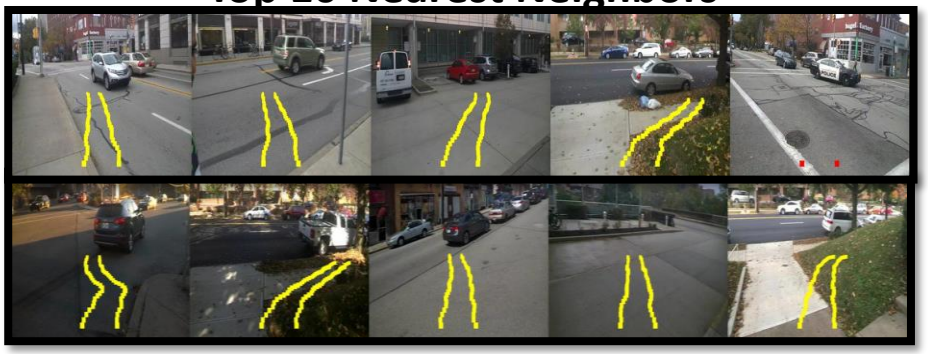


Prediction of rare events require more training data

Ground Truth

Prediction using 2 months of training data

Top 10 Nearest Neighbors



Prediction using 4 months of training data

Top 10 Nearest Neighbors

Prediction failure cases: bifurcations

Ground Truth Predicted Top 10 Nearest Neighbors

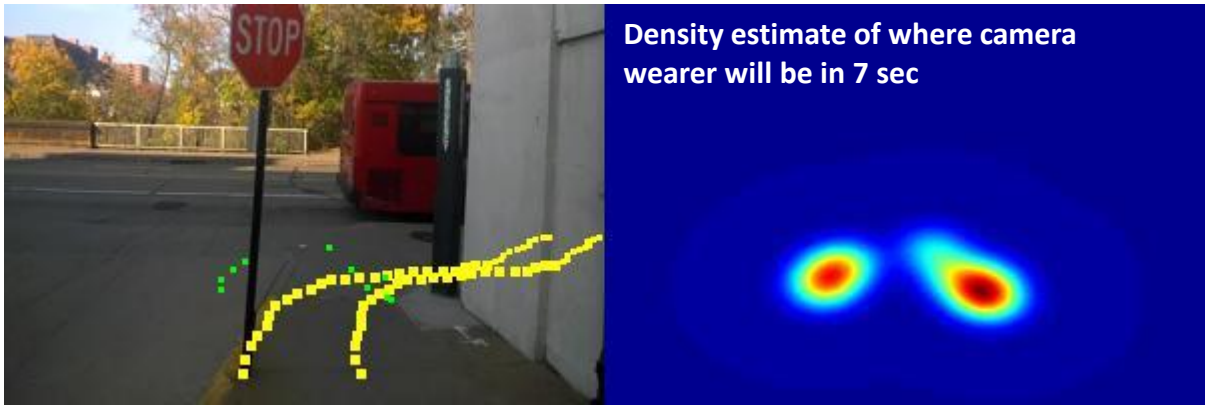
Junction where both left and right turn possible

Ground Truth Predicted Top 10 Nearest Neighbors

Stopping at an intersection, waiting at a traffic light, or continuing to walk

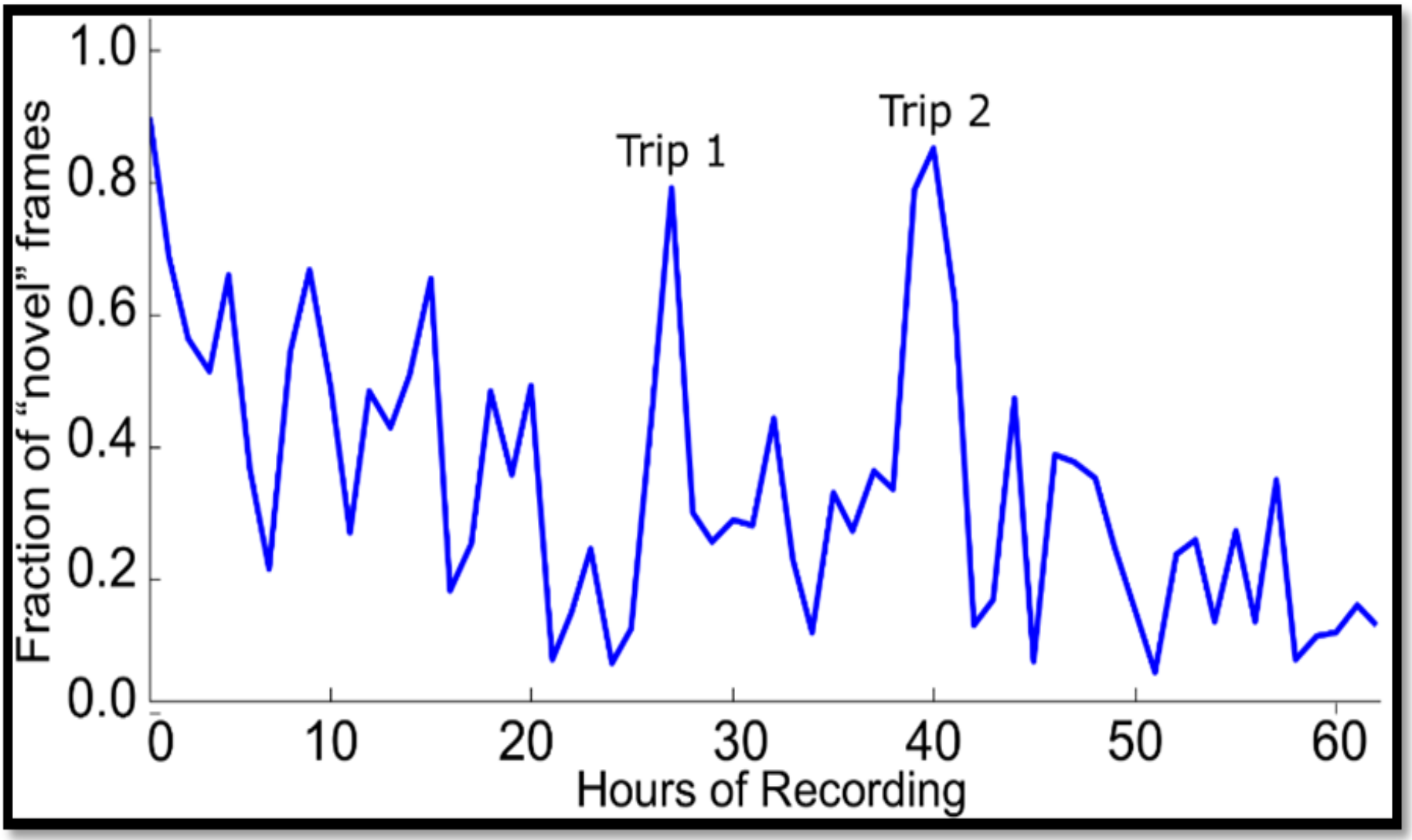
Density Estimate

Left-Right Bifurcation



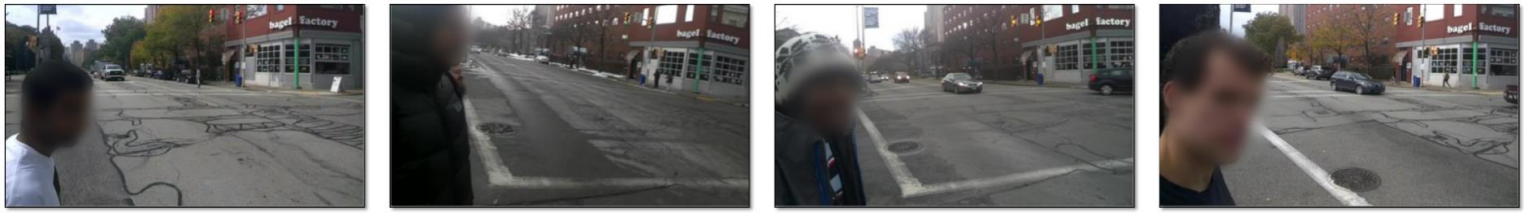
Stop-Go Bifurcation

Amount of novel data decreases with time



Virtual Webcam (capture changes)

Person



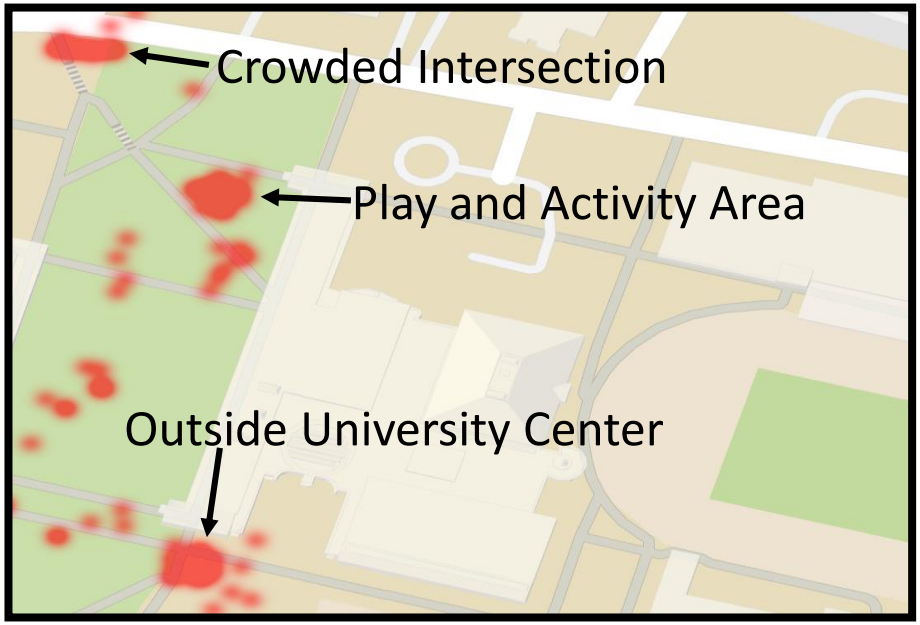
Bicycle stand

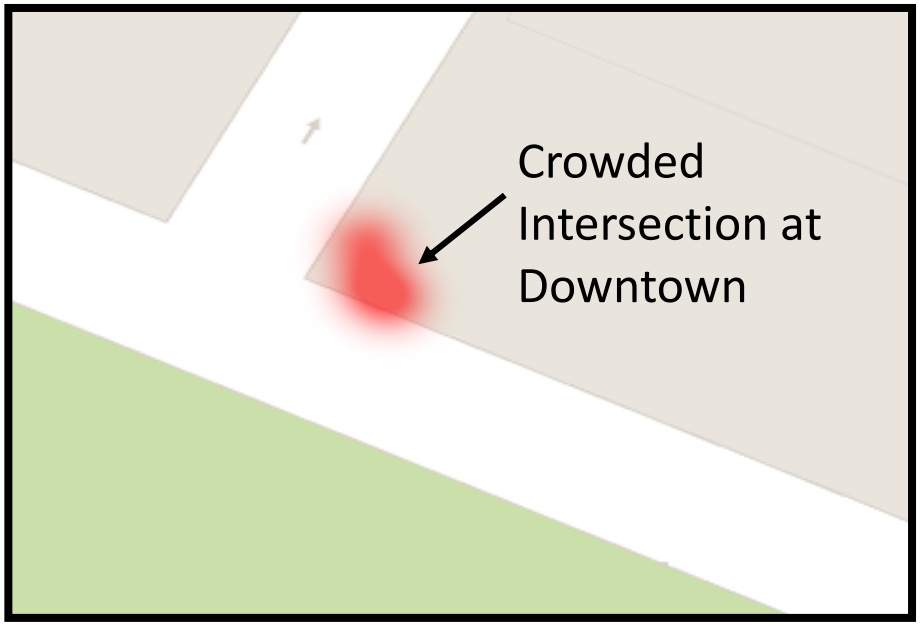
Car

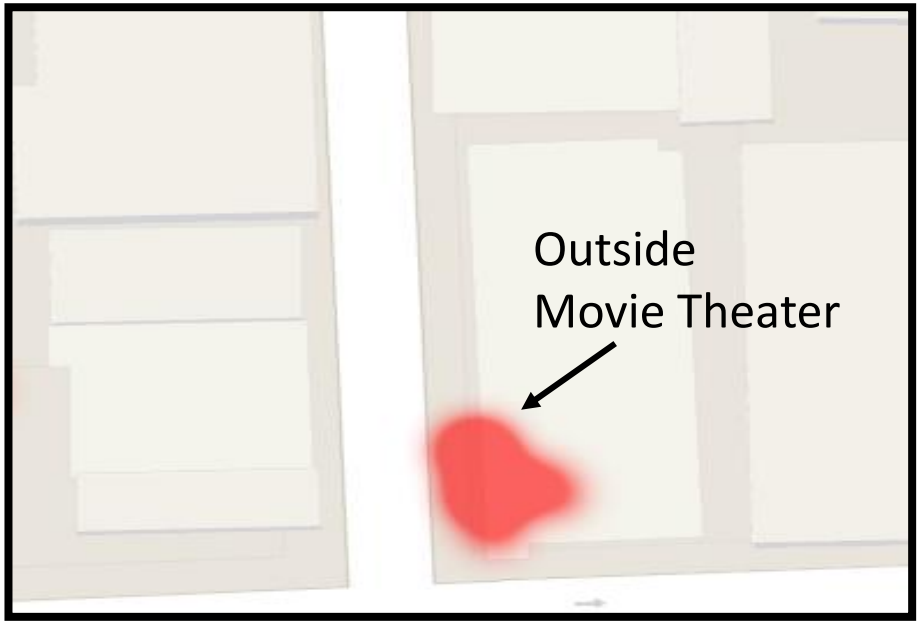
Season

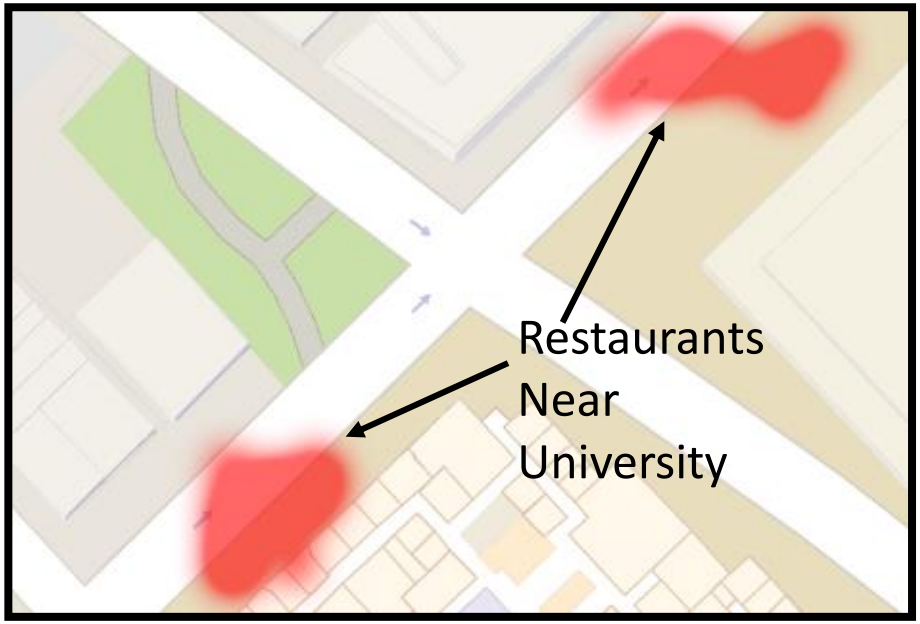
Time of day

Crowded locations in dataset









Summary

- Simple nearest neighbor predictions are effective due to heavy redundancy in dataset.
- Simple nearest neighbor is able to generalize for the novel places for which we have seen diverse set of examples.
- Prediction of rare events require long term training data.

Thank You