AutoCalib: Automatic Calibration of Traffic Cameras at Scale

Romil Bhardwaj[†], **Gopi Krishna Tummala*,** Ganesan Ramalingam[†], Ramachandran Ramjee[†], Prasun Sinha*

[†]Microsoft Research, *The Ohio State University

Police cameras on US roads doubles in three years

The number of police CCTV cameras monitoring US roads has almost doubled in the last three years, and are projected to record 75 million images daily.

CCTV cameras on Britain's roads capture 26 million images every day

LSE News

Speed cameras reduce road accidents and traffic deaths, according to study

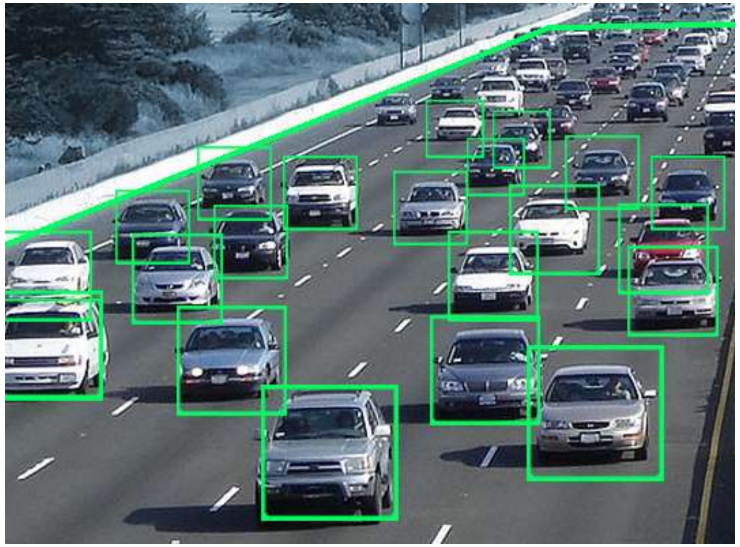
Conventional Traffic Camera Uses

Manual Surveillance

Post-facto Incident Review

Emerging Traffic Camera Use Cases

Vehicle Speed Measurement
(without dedicated sensors)

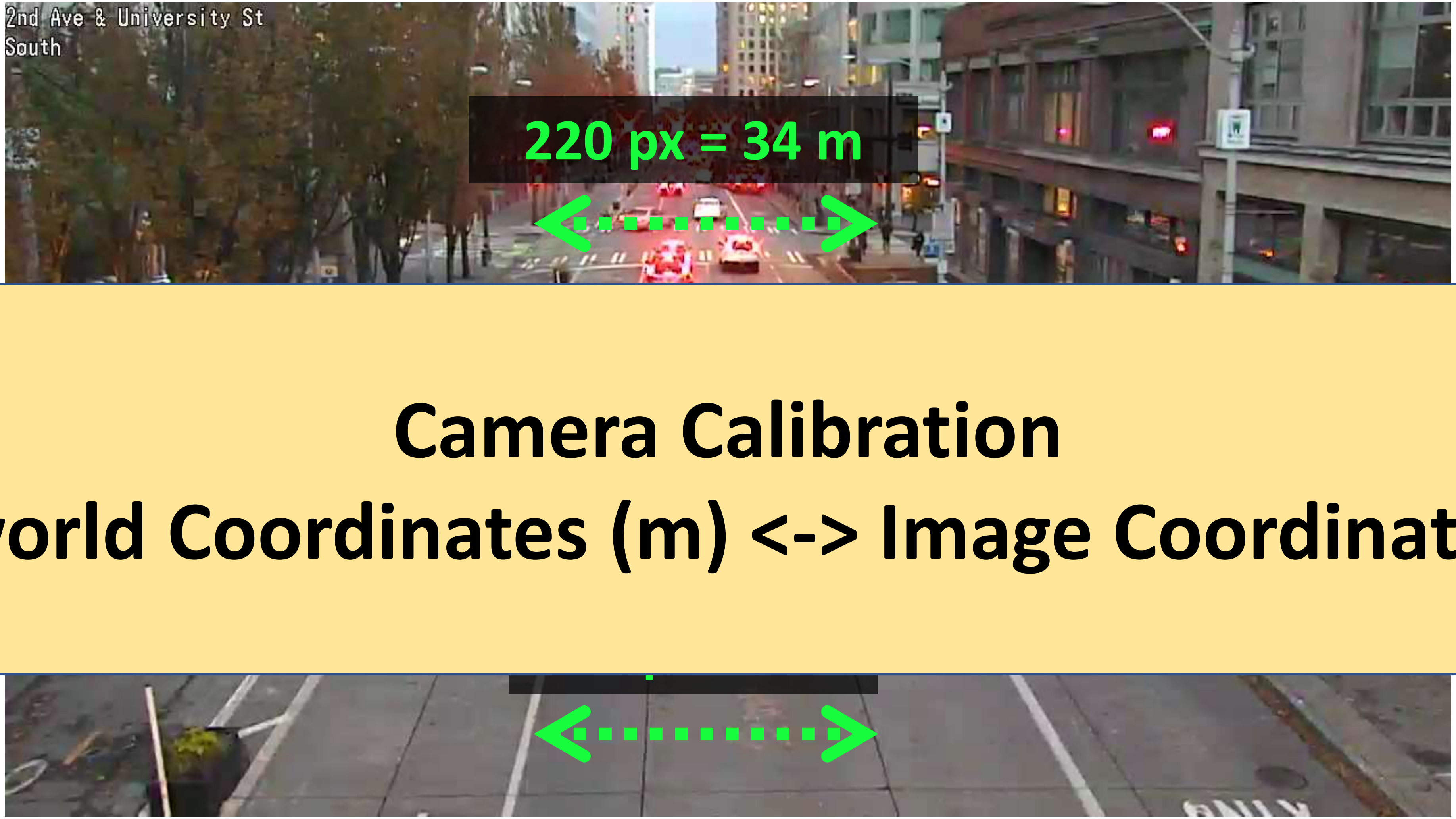


Traffic Analytics

Near Miss Stats

All require distance measurements in the scene

Measuring Distances in an Image



Camera Calibration

Real-world Coordinates (m) <-> Image Coordinates (px)

Camera Calibration

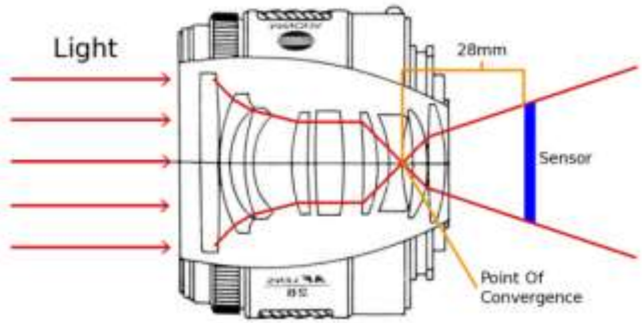
$$y = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} x$$

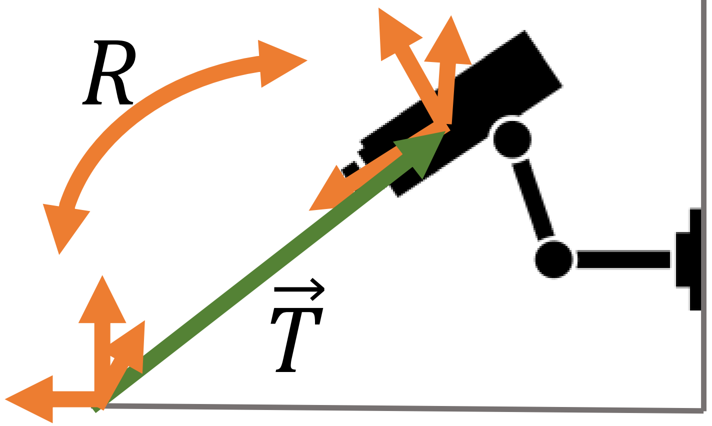
Image Coordinates

Intrinsic Matrix (Focal length, camera center)

Extrinsic Matrix (Rotation, Translation)

Real World Coordinates





"Hard" Calibration

$$y = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} x$$

Image Coordinates

Intrinsic Matrix (Focal length, camera center)

Extrinsic Matrix (Rotation, Translation)

Real World Coordinates

Not Scalable!

"Soft" Calibration

EPnP Solver

≈

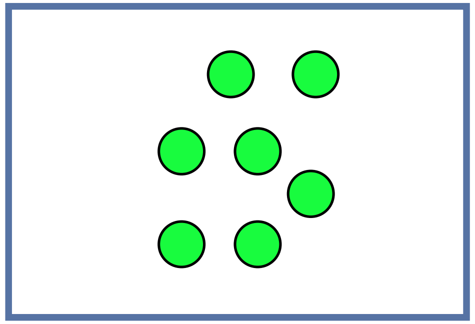
$$y = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} x$$

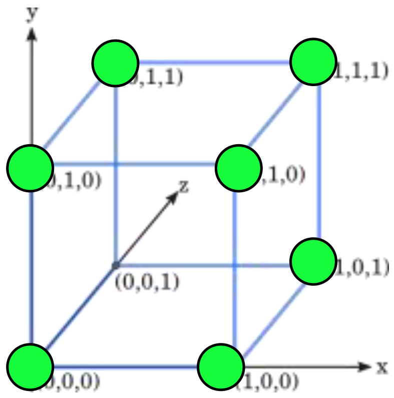
Image Coordinates

Intrinsic Matrix (Focal length, camera center)

Extrinsic Matrix (Rotation, Translation)

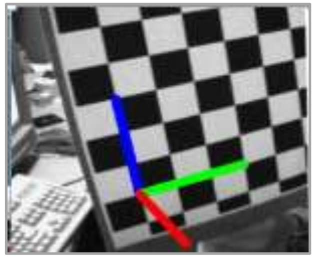
Real World Coordinates

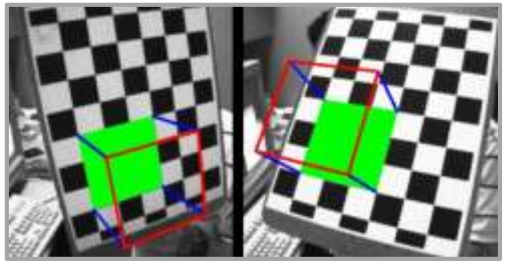




“Soft” Calibration - Prior Art

Chessboard Calibration



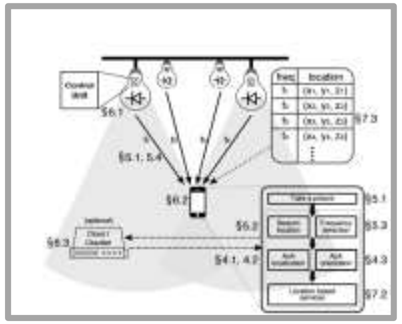


No Chessboard Patterns in Traffic Views

Vanishing Points

Assumption of Straight Line Motion

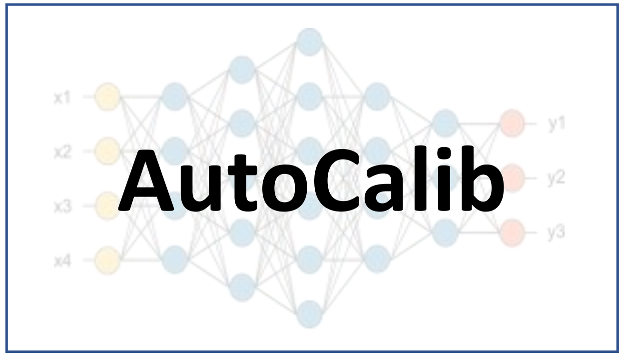
Geometric Landmarks

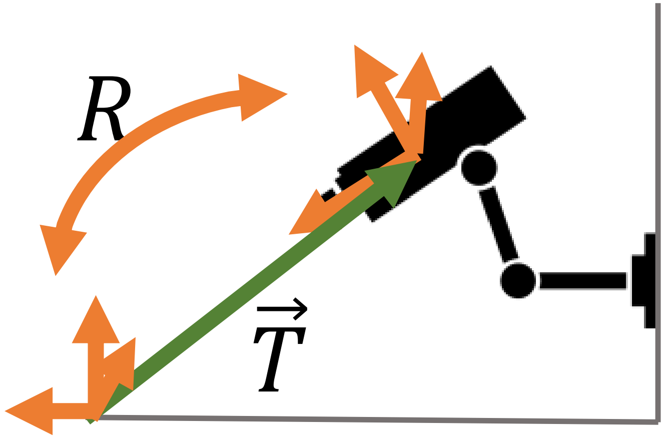


Assumption of Landmarks

AutoCalib Overview

Traffic Video

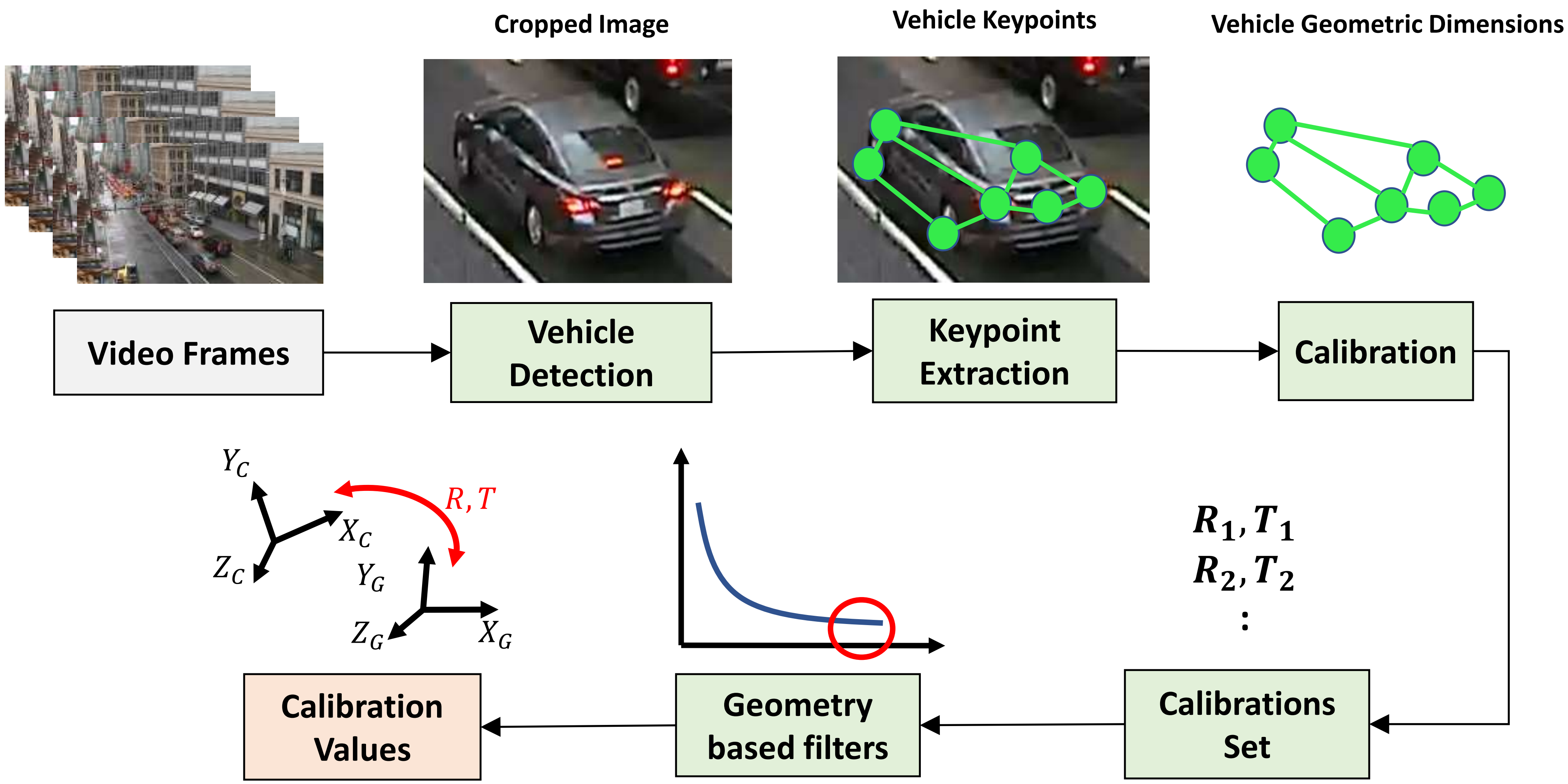




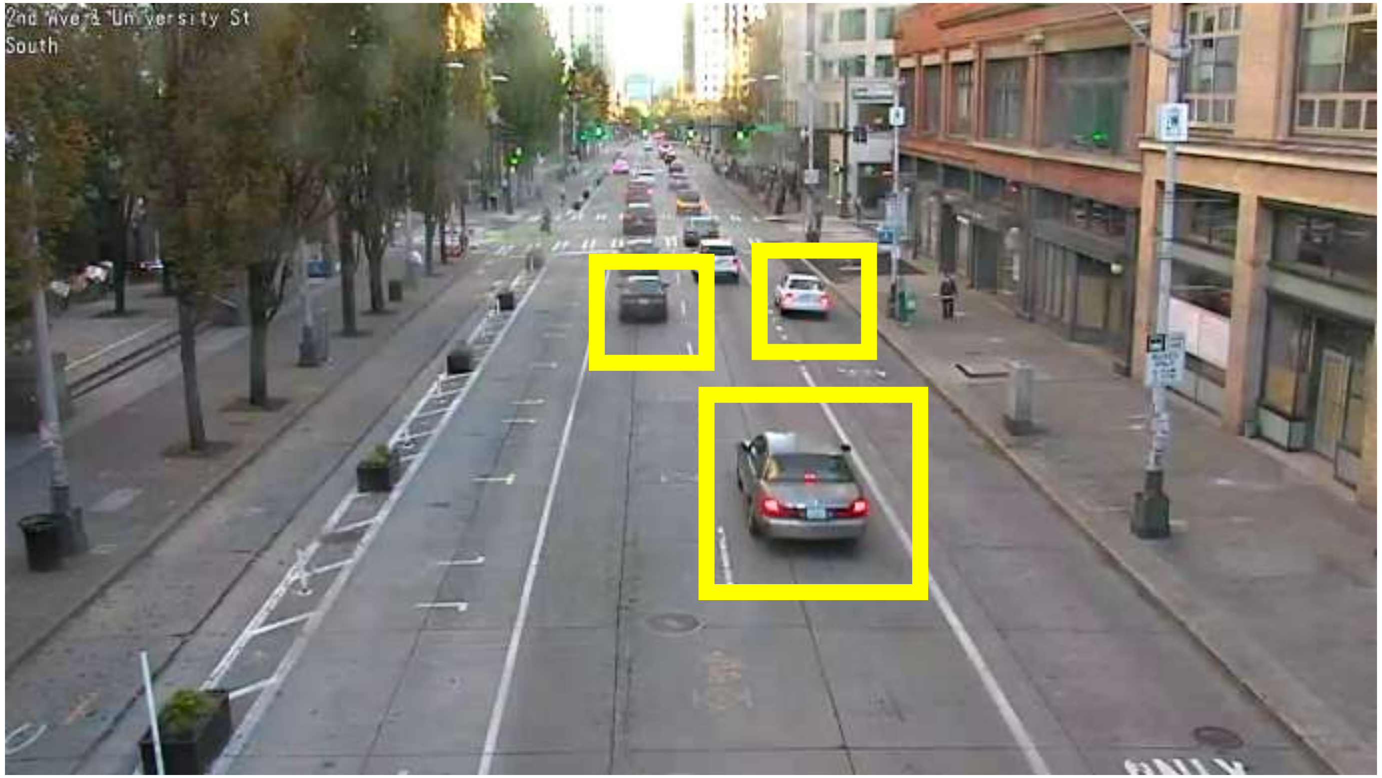
Calibration Estimate

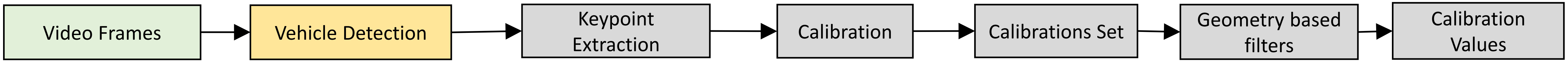
AutoCalib: no humans-in-the-loop, robust calibration

AutoCalib - Pipeline



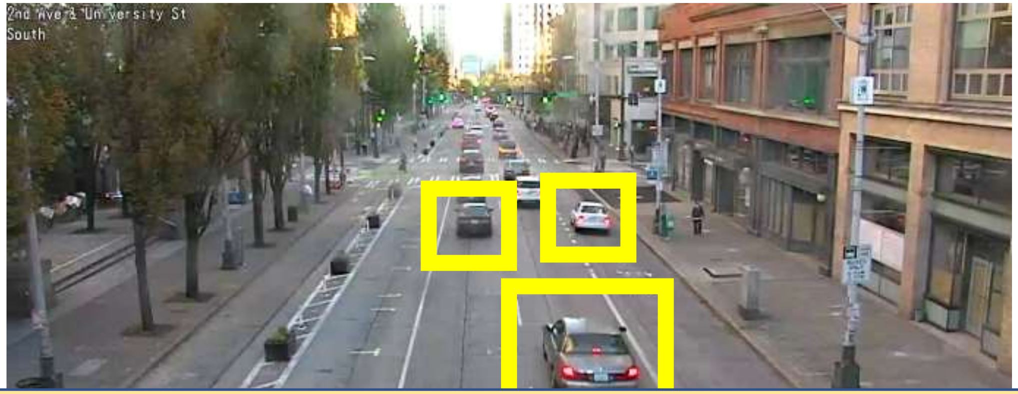
Vehicle Detection



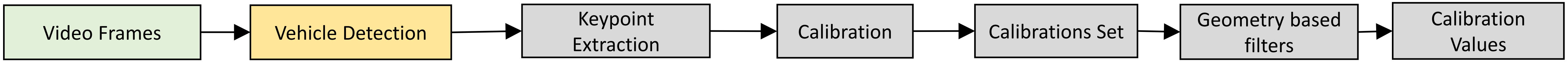


Vehicle Detection

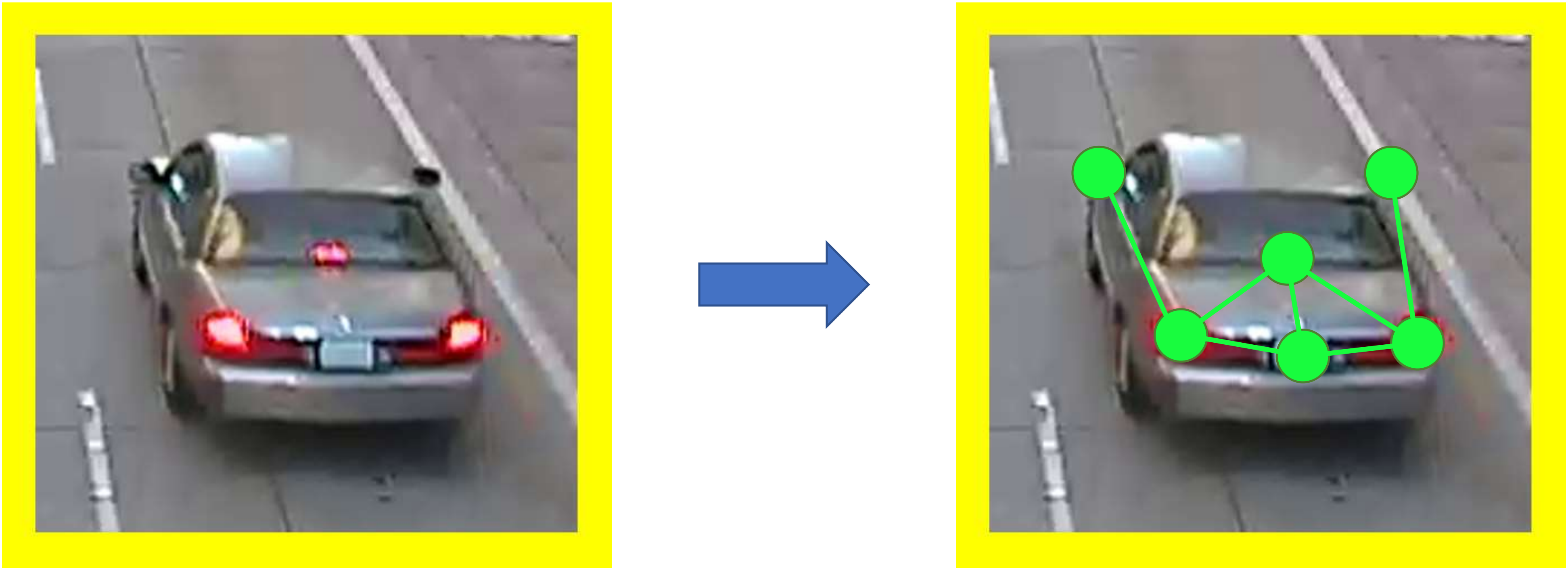
- Off-the-shelf DNNs (Fast-RCNN, YOLO) promise state of the art accuracy
  - Expensive, scene often empty
- Background Subtraction is fast
  - Inaccurate

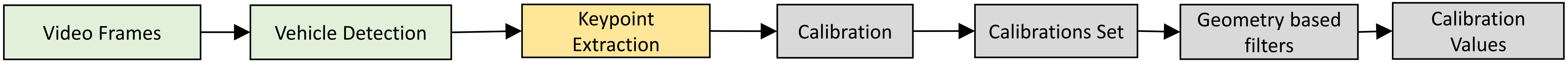


Solution - Trigger the DNN with Background Subtraction



Key-point Extraction

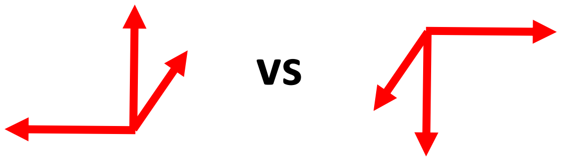


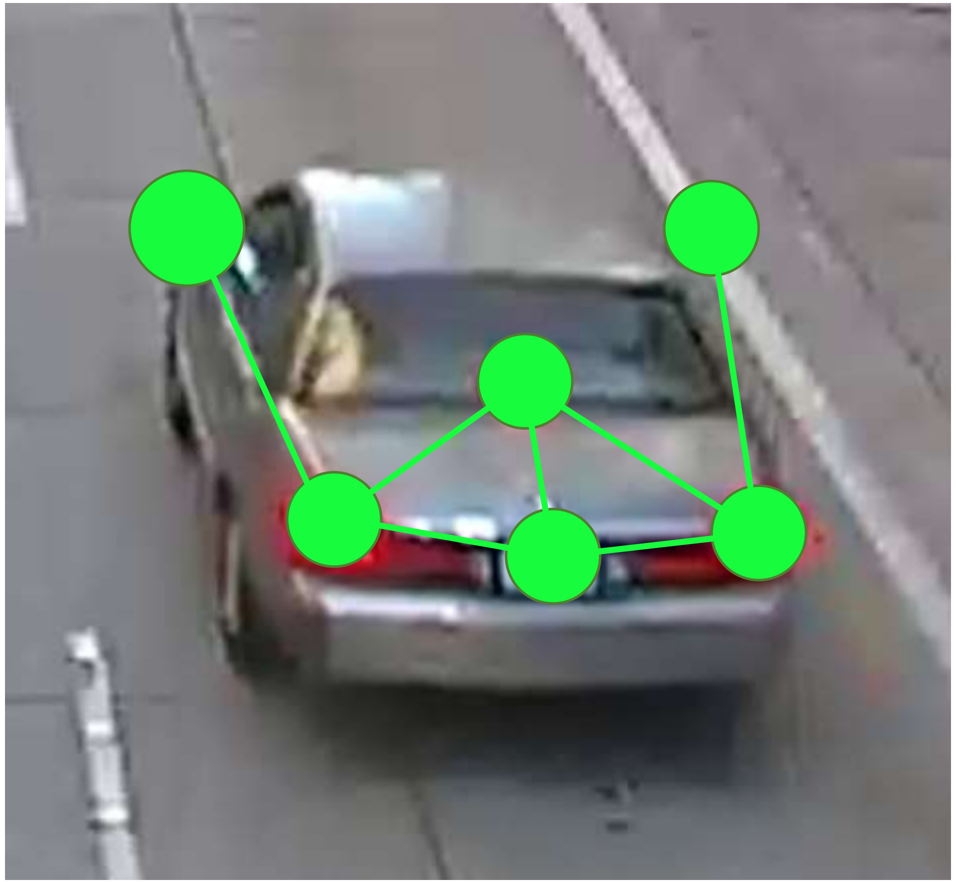


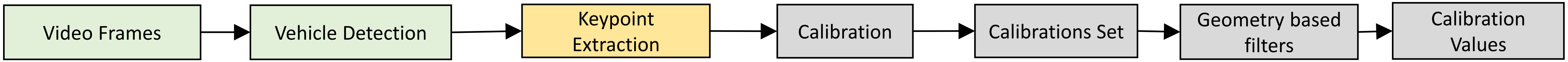
Key-point Selection

Desired Properties

1. Visually Distinct
   - Ease of detection
2. Non-planar
   - Robust Calibrations

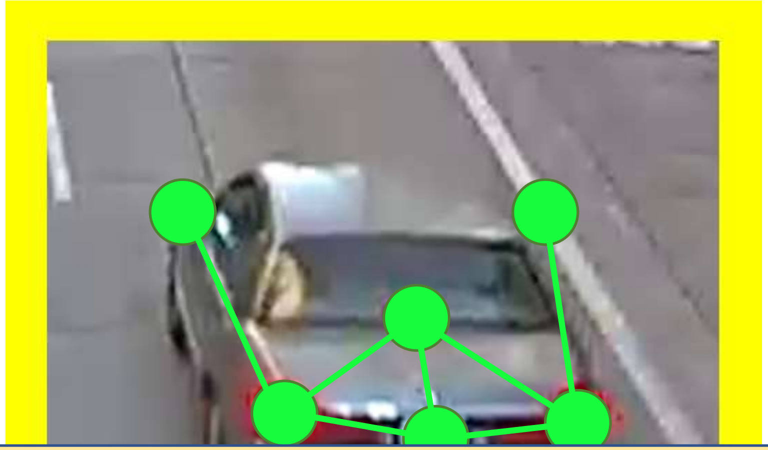






Key-point Extraction

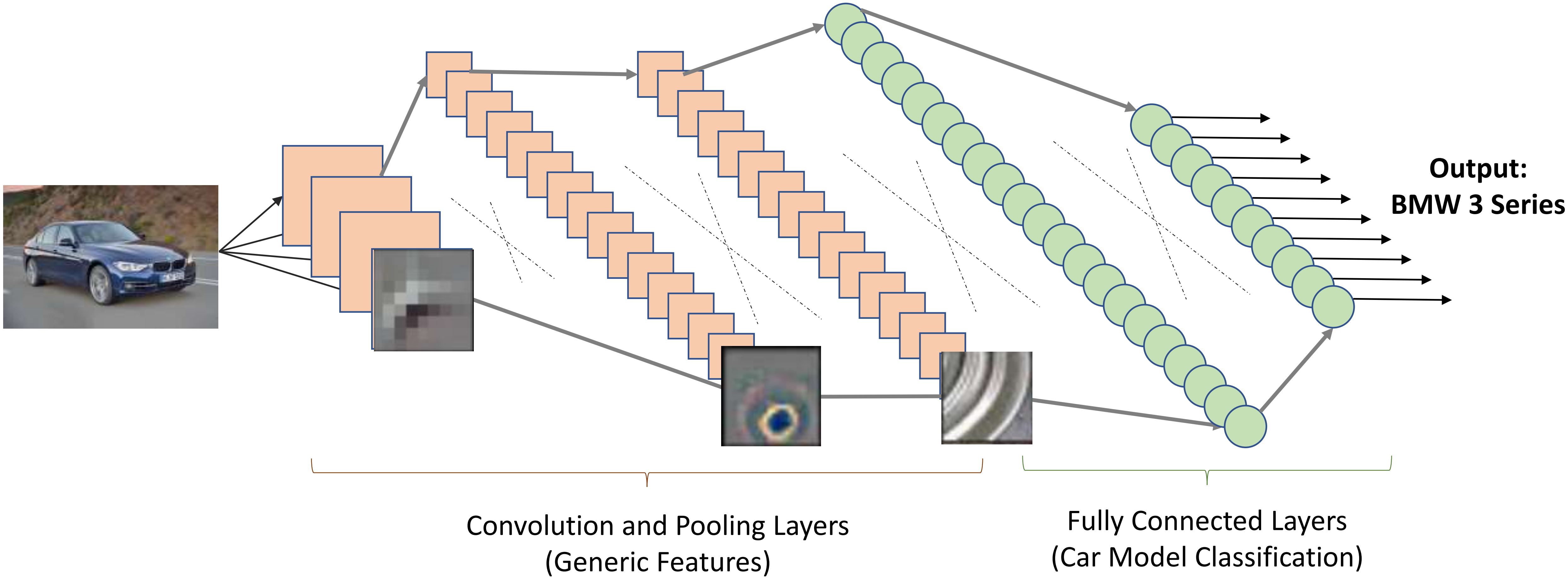
- Statistical vision based techniques aren't robust to lighting variations
- DNNs require a lot of labelled data
  - No datasets available



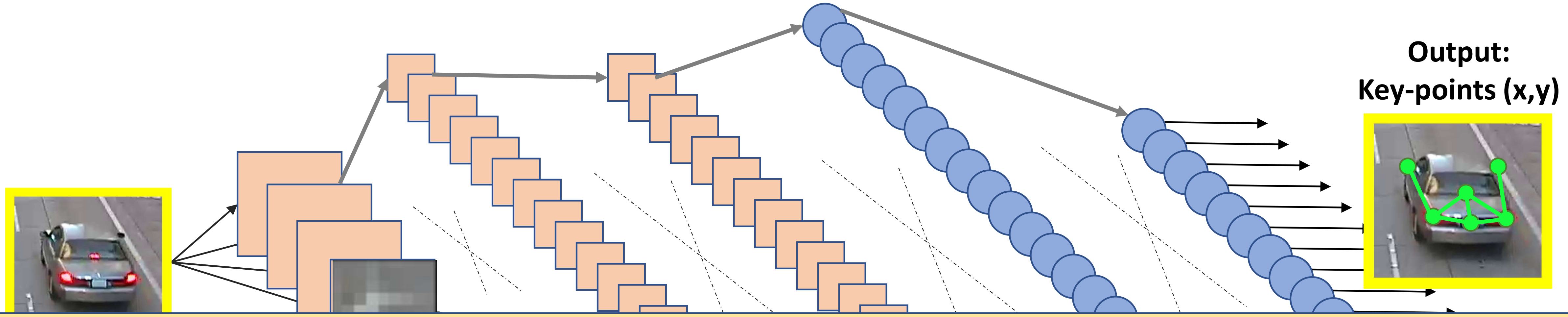
Transfer learn a DNN on a smaller dataset

Video Frames → Vehicle Detection → Keypoint Extraction → Calibration → Calibrations Set → Geometry based filters → Calibration Values

Transfer Learning - Primer



Transfer Learning - Primer



Transfer Learning - Less Data, Faster Training

Key-point DNN Dataset

- Manually labelled key-points on 486 car images
- Image Augmentation

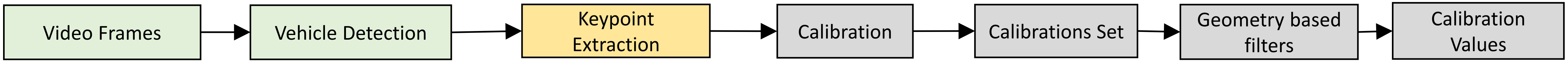
Original Crop | Original Rotate | Original Img | Horz Mirror | Horz Mirror Rotate | Horz Mirror Crop

Total of 10,344 images post augmentation

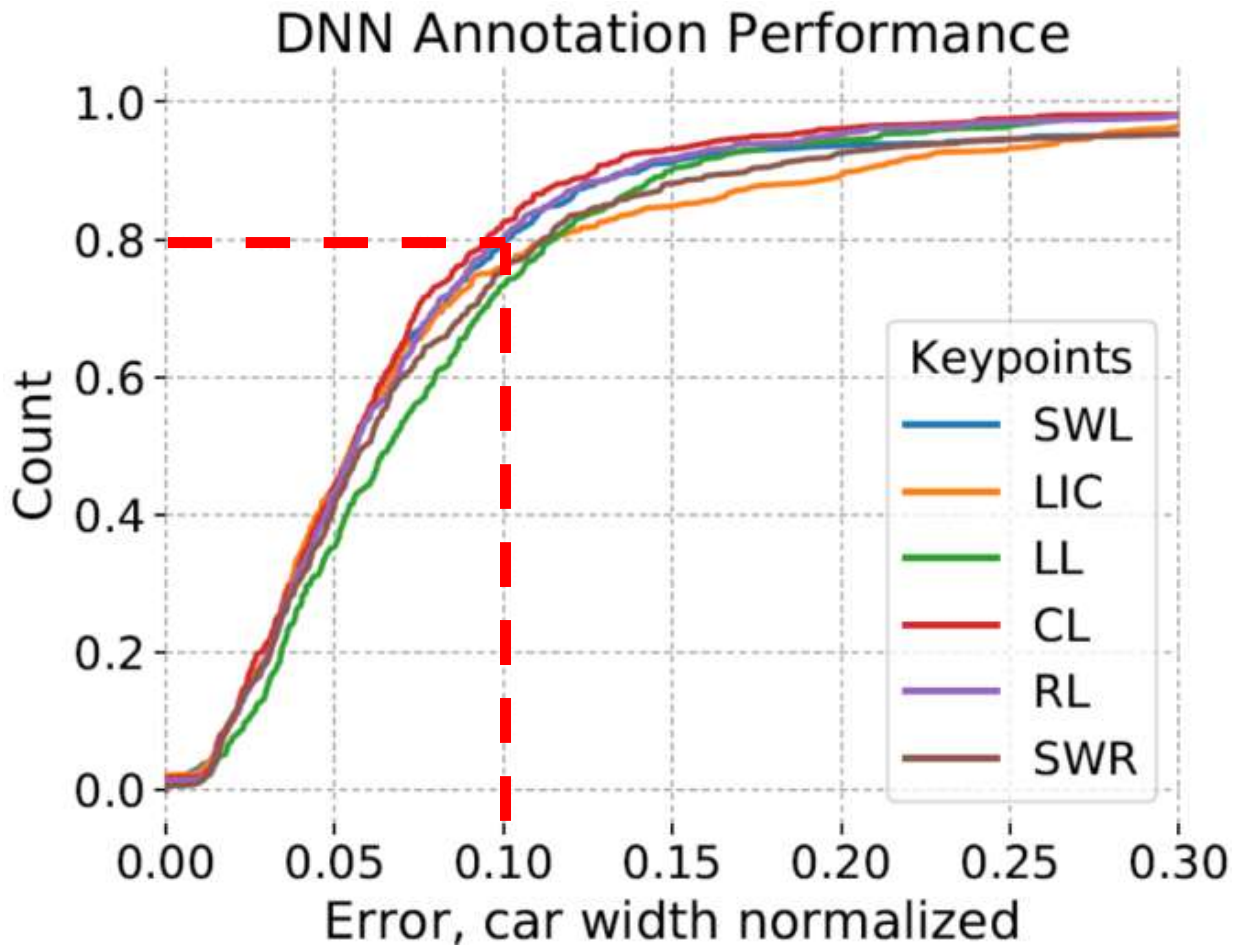
Video Frames → Vehicle Detection → Keypoint Extraction → Calibration → Calibrations Set → Geometry based filters → Calibration Values

Key-point DNN Training

- GoogLeNet architecture trained on CUHK CompCars dataset (CVPR ‘15) for Car make/model classification
- Replaced last two fully connected layers with keypoint regression outputs



Key-point DNN Performance



~80% of Key-points < 10% error

Calibration Estimation

$$y = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} x$$

Image Coordinates

Intrinsic Matrix (Focal length, camera center)

Extrinsic Matrix (Rotation, Translation)

Real World Coordinates

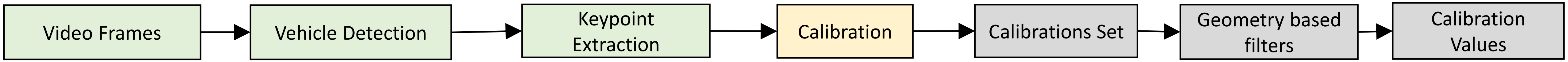
Video Frames → Vehicle Detection → Keypoint Extraction → Calibration → Calibrations Set → Geometry based filters → Calibration Values

Vehicle Identification at low resolution…



… is hard!

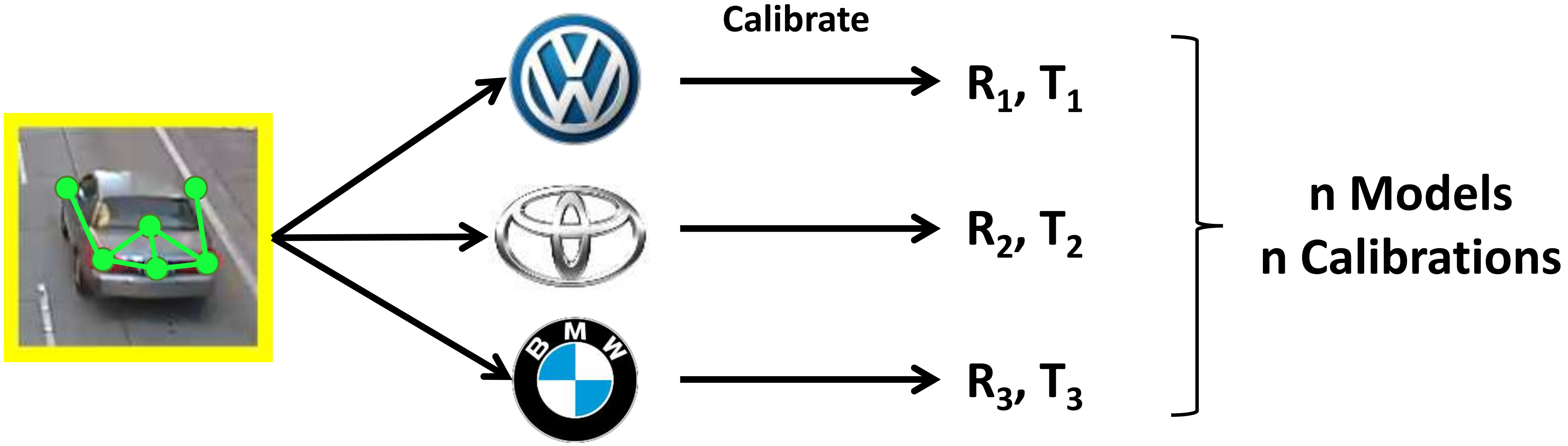
(for both, humans and machines)

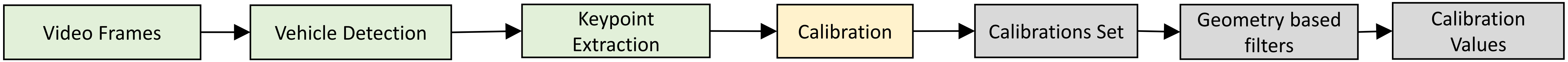


Can’t identify… so, approximate!

Calibrate with most popular cars

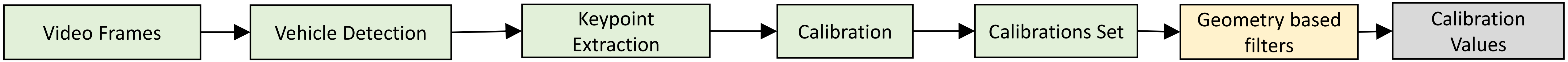
(Toyota Prius, Toyota Corolla, Honda Civic, Volkswagen Jetta, BMW 320i, Audi A4, etc.)



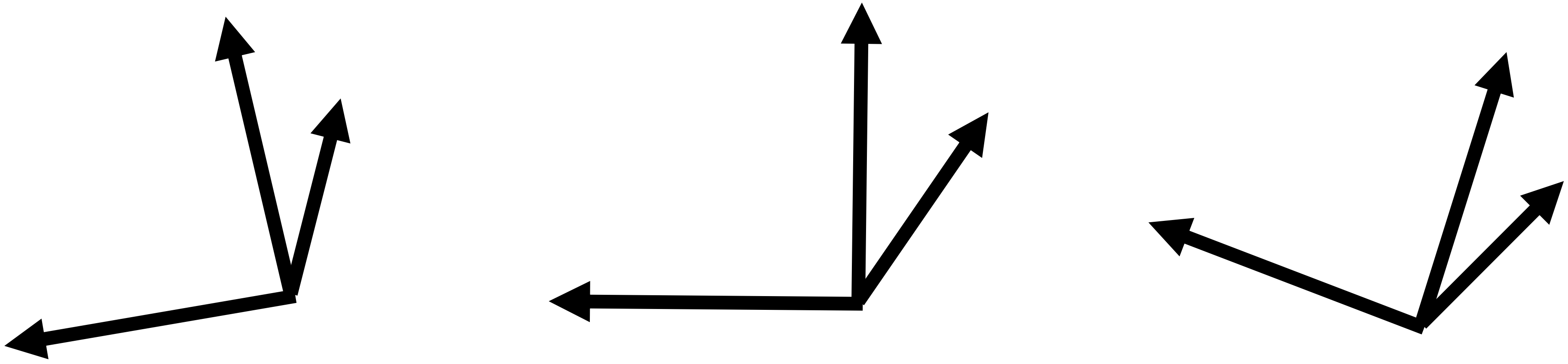


Errors in Calibration

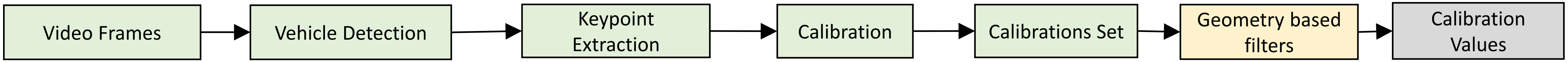
Statistical filters to remove outliers and average



Key Insight 1

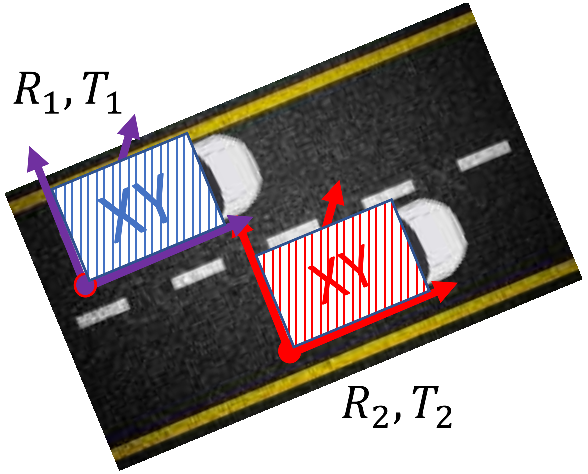


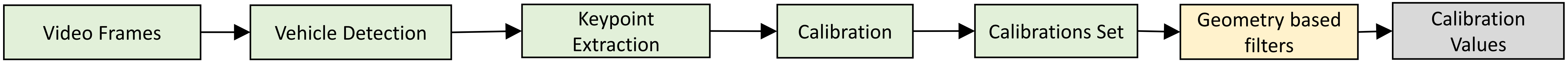
Ground plane should be consistent across all Calibrations



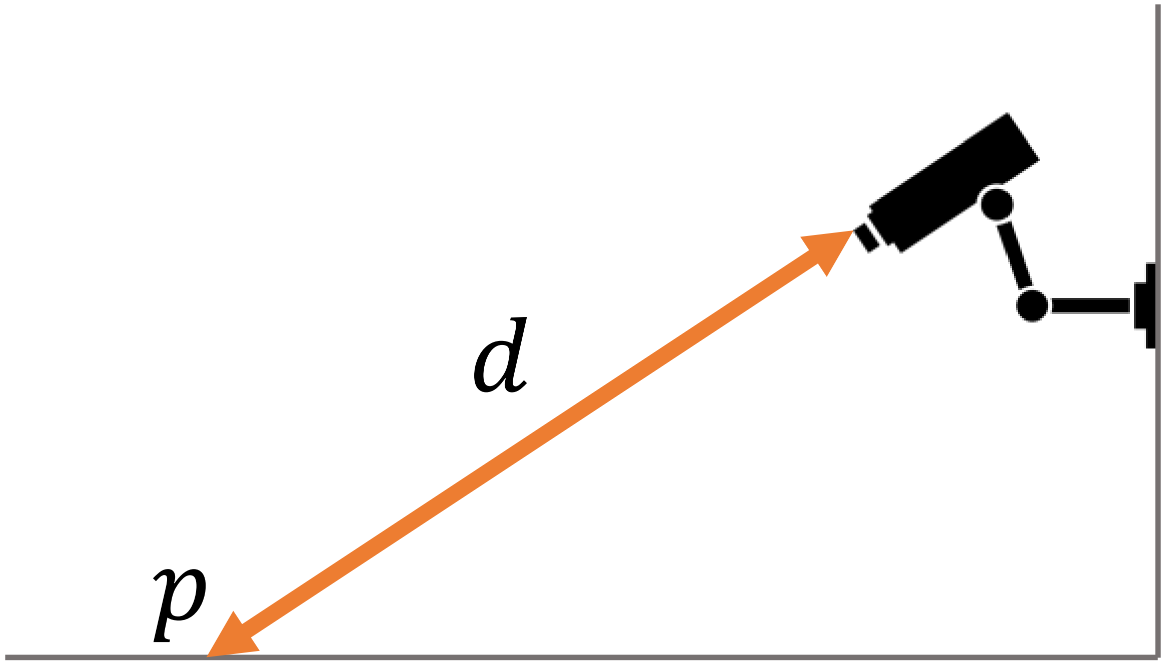
The Orientation Filter

1. For calibration (R^i, T^i), its Z-axis orientation $\vec{z}$ is defined by vector $R^i_{*,3}$
2. Let $\vec{z}_{avg} = Average(R^i_{*,3})$
3. Pick $n\%$ calibrations with the least deviation between $\vec{z}$ and $\vec{z}_{avg}$

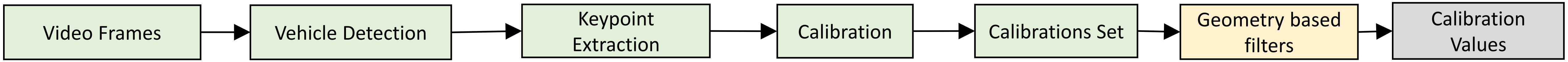




Key Insight 2

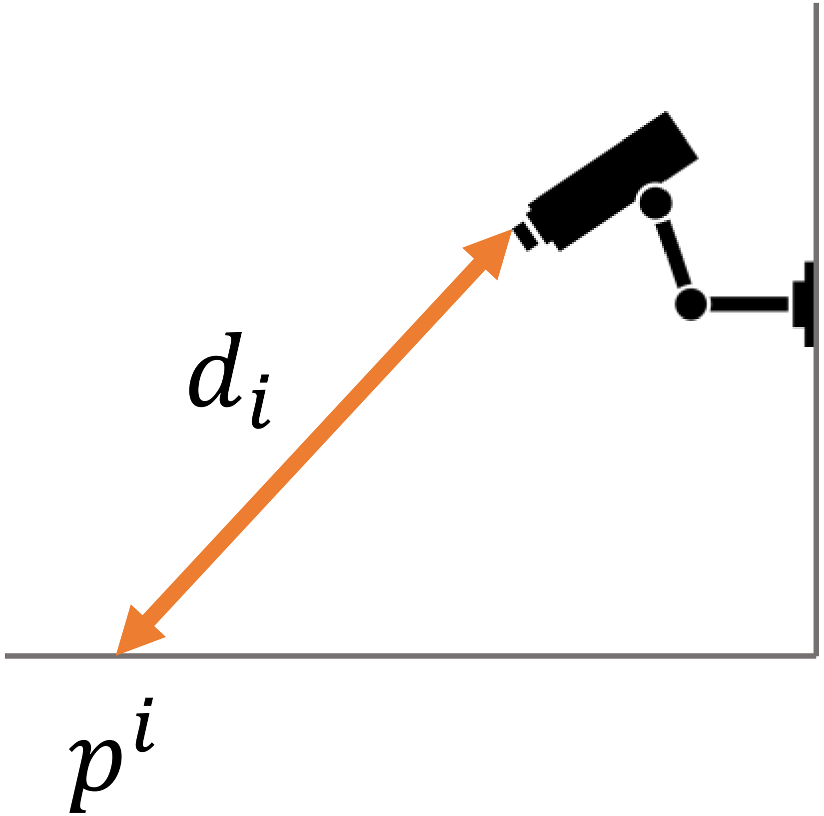


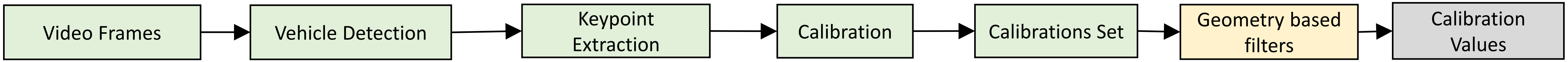
Distance to a fixed point must be consistent across Calibrations



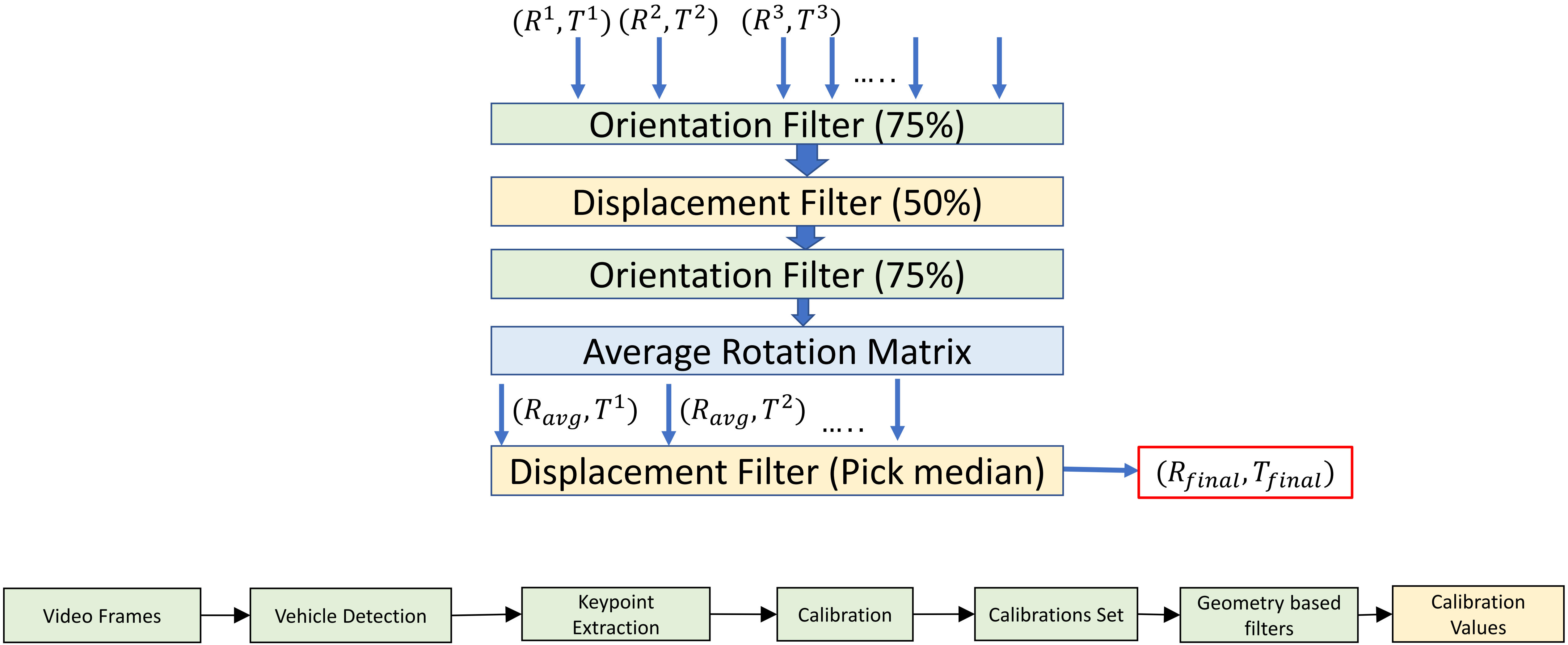
The Displacement Filter

- Focus region: Region where cars are detected
- For each Calibration:
  1. Point p^i = projection of center of focus region on the ground plane using (R^i, T^i)
  2. d_i = Distance of p^i to camera
- Pick middle n% and filter the rest





Filtering Overview



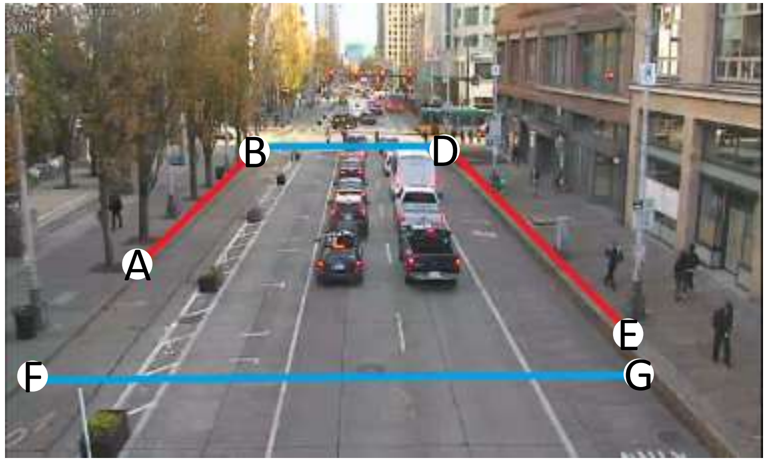
Implementation

Azure Service – 4 Tesla K80s, 224 GB RAM

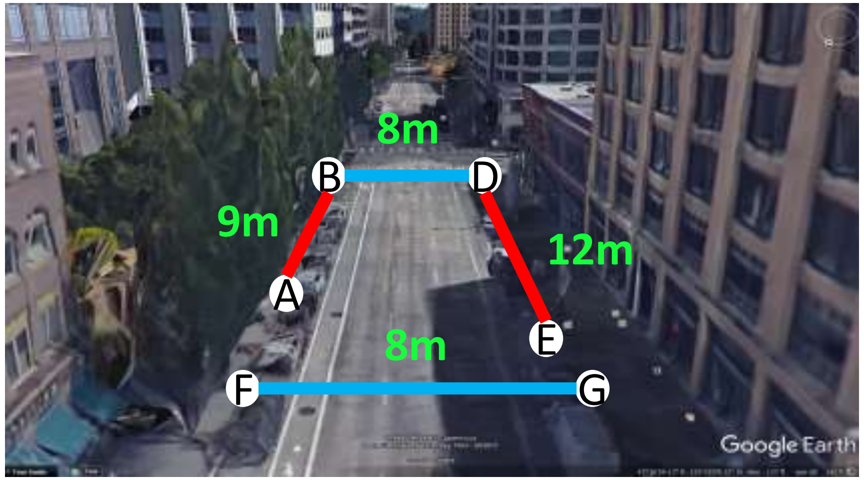
< 12% error with ~8 minutes of video

Evaluation - Dataset

- 350+ hours from 10 traffic cameras in Seattle
- Resolution - 640x360 to 1280x720
- Ground truth distances and calibration estimated using Google Earth



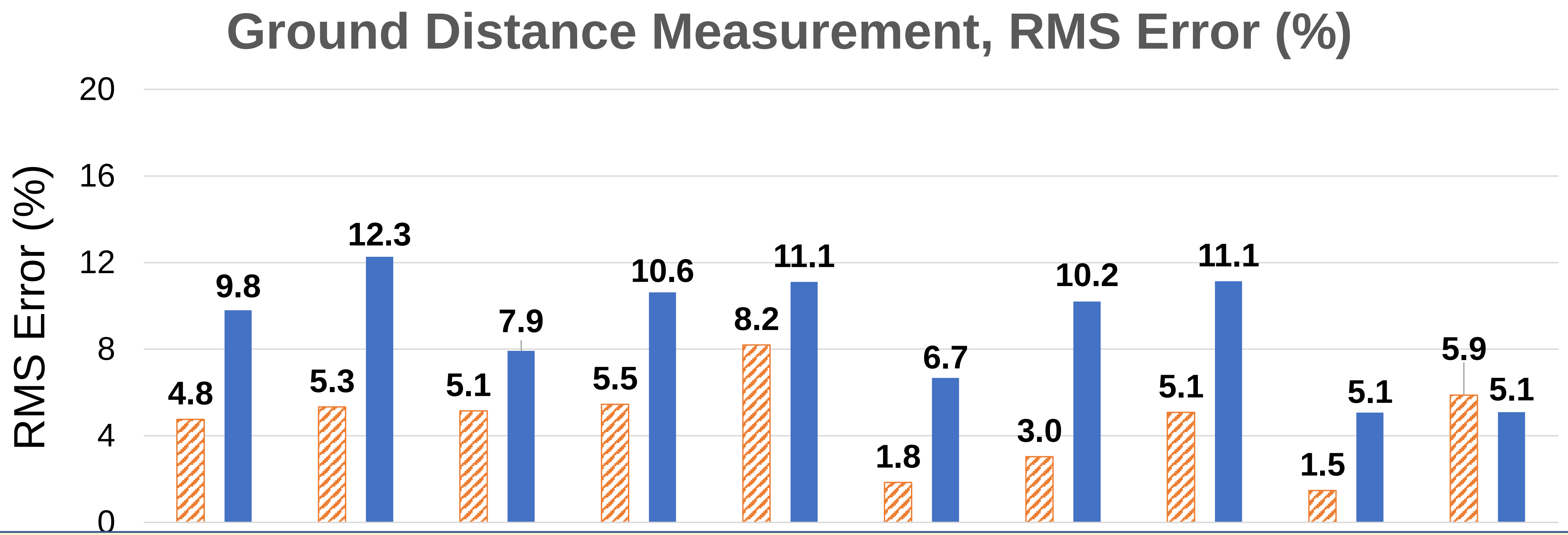
Camera Image



Google Earth View

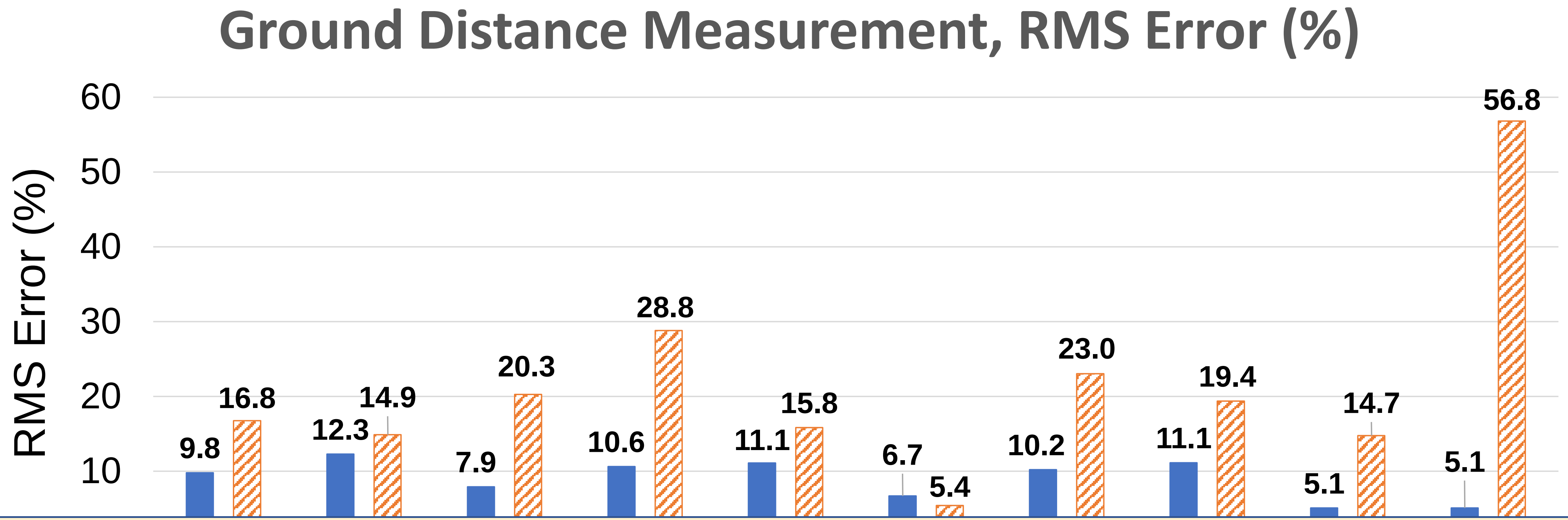
Evaluation

AutoCalib vs Manual Calibration



AutoCalib achieves <12% RMS error in measuring distances

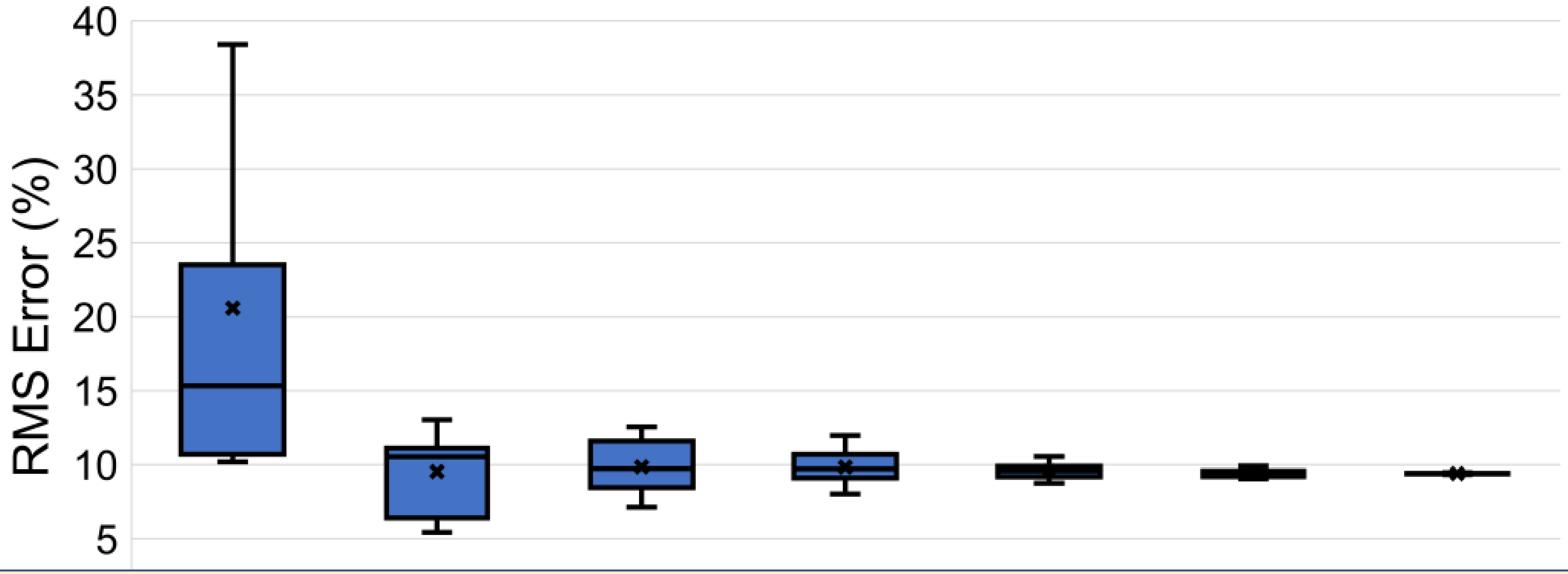
AutoCalib vs Prior Art



AutoCalib outperforms prior state of the art approaches

[1] Dubská et al., Fully automatic Roadside Camera Calibration for Traffic Surveillance. IEEE ITS 2015

Does more video data help?



AutoCalib converges with increasing vehicle detections

Application – Speed Measurement

AutoCalib Summary

- **Camera Calibration**
  - Enables distance measurements
  - Highly manual today
- **AutoCalib**
  - Scalable automatic calibration
  - Uses DNNs to analyze vehicle geometry
- **Experiments**
  - **< 12%** error in measuring distances
  - Calibrates with few hundred detections