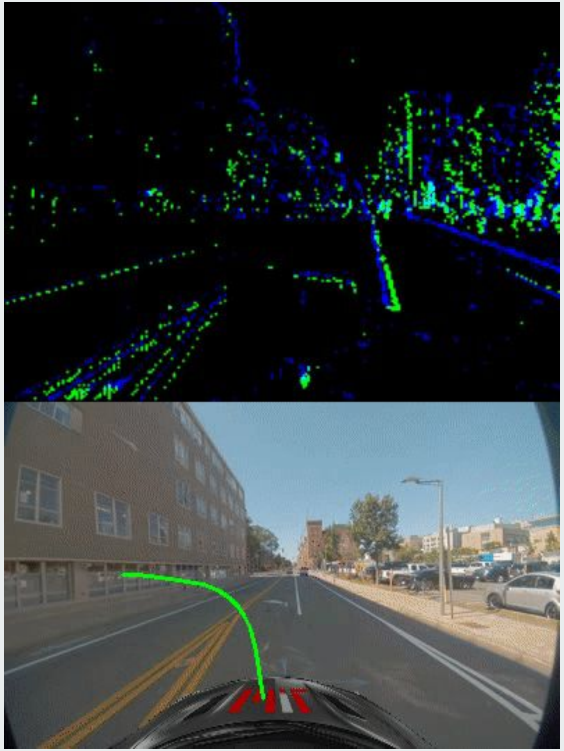
Real world application of event-based end to end autonomous driving

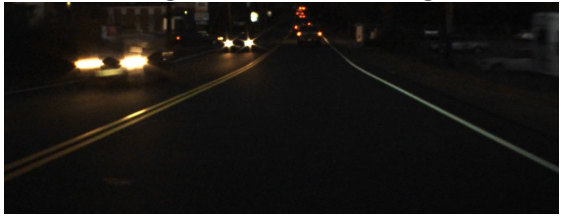


Yuxuan Chen
Mentors: Dr. Igor Gilitschenski, Alexander Amini

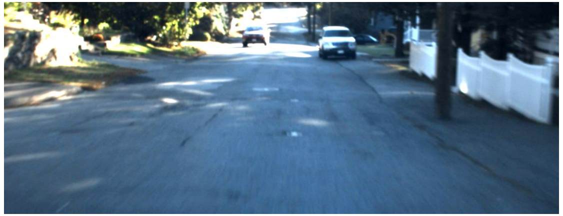
Motivation

Autonomous driving cars need to handle a wide range of scenarios

Night-time Driving



No Lane Markings



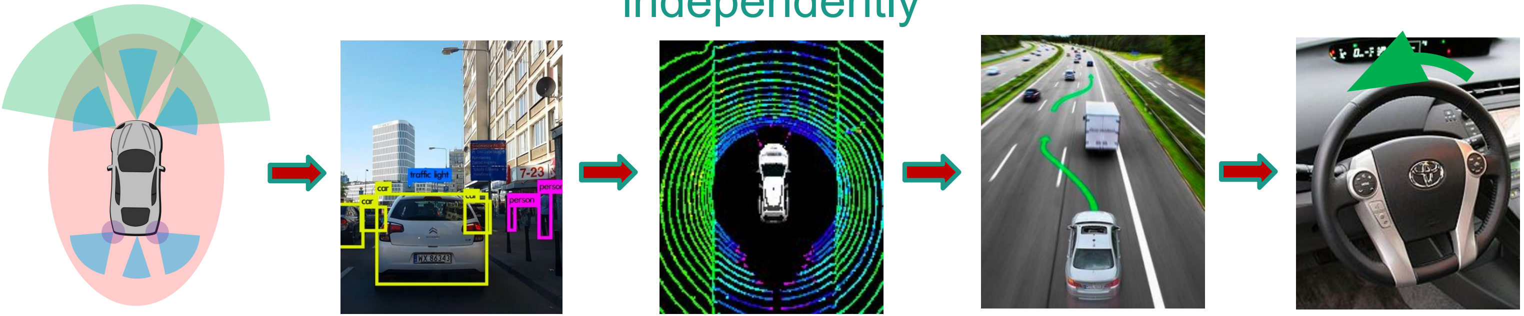
Rainy Weather

[Learning steering bounds for parallel autonomous systems, Amini et al.]

How do they do it?

Autonomous Driving Pipeline

Separate problem into smaller sub-modules, tackle each independently



Sensor Fusion
- What's happening around me?

Detection
- Where are obstacles?

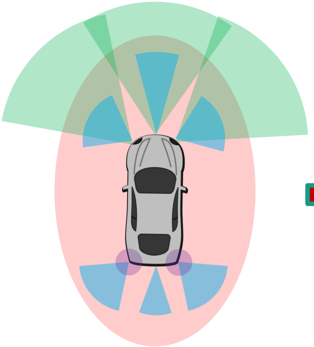
Localization
- Where am I relative to the obstacles?

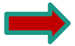
Planning
- Where do I go?

[Learning steering bounds for parallel autonomous systems, Amini et al.]

End-to-end Learning

Learn the control directly from raw sensor data





Deep Neural Network

Sensor Fusion
- What's happening around me?

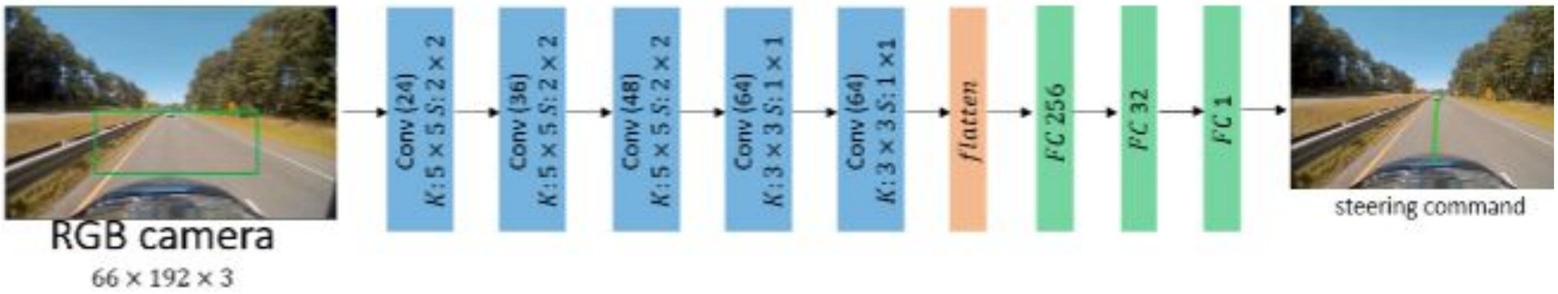
Learned Model
Underlying representation of how humans drive

Actuation
- What control signals to take?

[Learning steering bounds for parallel autonomous systems, Amini et al.]

PilotNet

Learn the steering directly from pixel values



[End to end learning for self-driving cars, Bojarskiet al.]

Problem with RGB cameras

Dynamic Range

Motion blur

Latency

What are event-based cameras

Novel bio-inspired sensors that capture motion in the scene

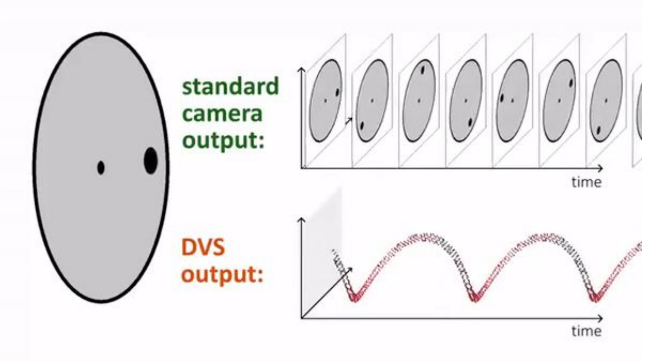
DAVIS240 from Inivation.com

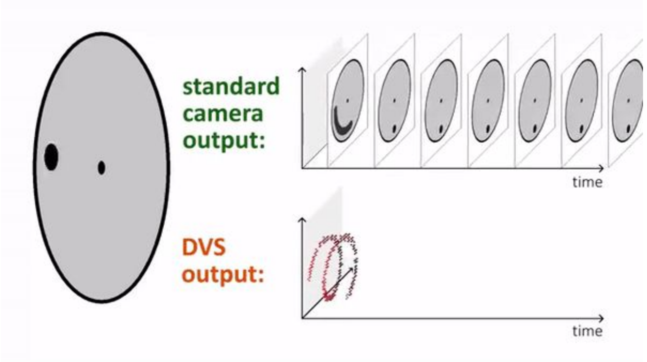
[Event-based Cameras: Challenges and Opportunities, Scaramuzza et al.]

What are event-based cameras

Novel bio-inspired sensors that capture motion in the scene

DAVIS240 from Inivation.com





[Event-based Cameras: Challenges and Opportunities, Scaramuzza et al.]

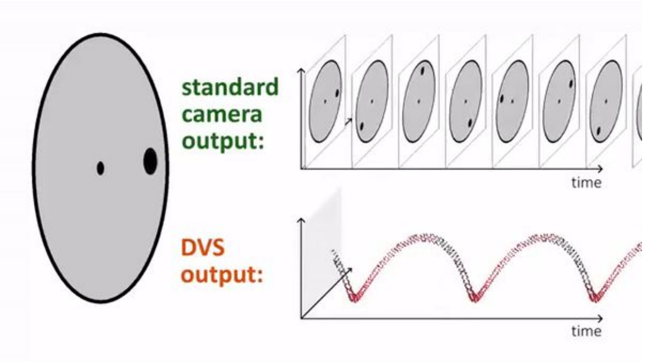
What are event-based cameras

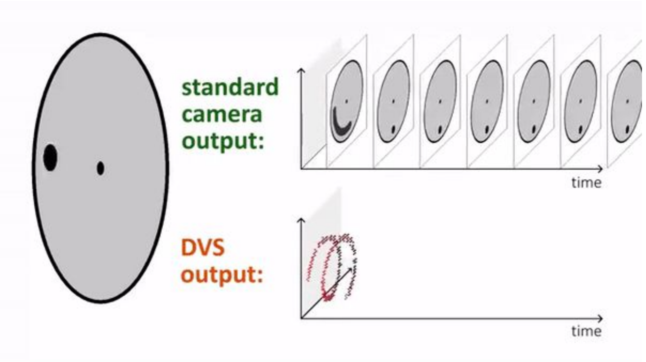
Novel bio-inspired sensors that capture motion in the scene

DAVIS240 from Inivation.com

Benefits:

- Low latency (~ 1 microsecond)
- No motion blur
- High dynamic range (140 dB instead of 60dB)





[Event-based Cameras: Challenges and Opportunities, Scaramuzza et al.]

What are event-based cameras

DAVIS240 from Inivation.com

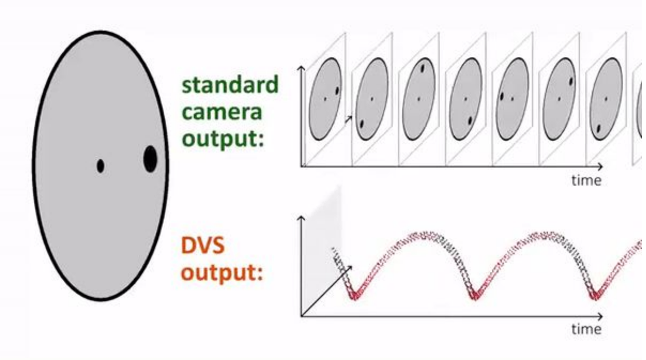
Novel bio-inspired sensors that capture motion in the scene

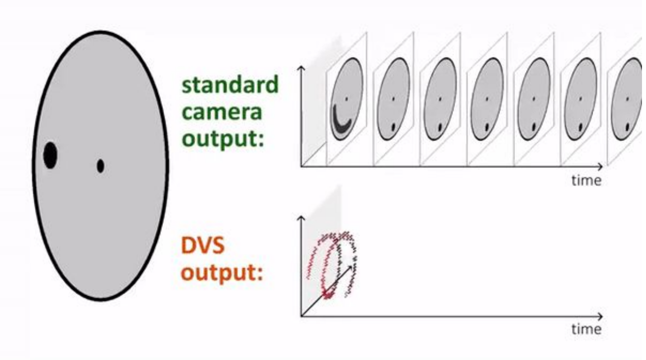
Benefits:
- Low latency (~ 1 microsecond)
- No motion blur
- High dynamic range (140 dB instead of 60dB)

Challenges:
- Data format of events

$$e_k = (x_k, y_k, t_k, p_k)$$

- Monochromatic
- Low resolution





[Event-based Cameras: Challenges and Opportunities, Scaramuzza et al.]

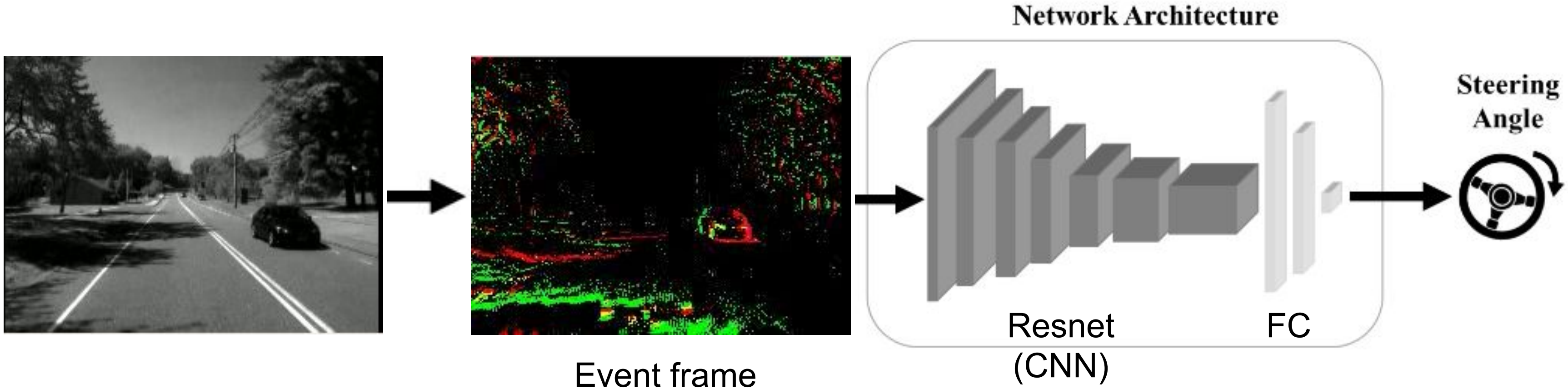
Our Goal

Use an event camera to drive a car in real time

DAVIS240 from Inivation.com

Deep Neural Network

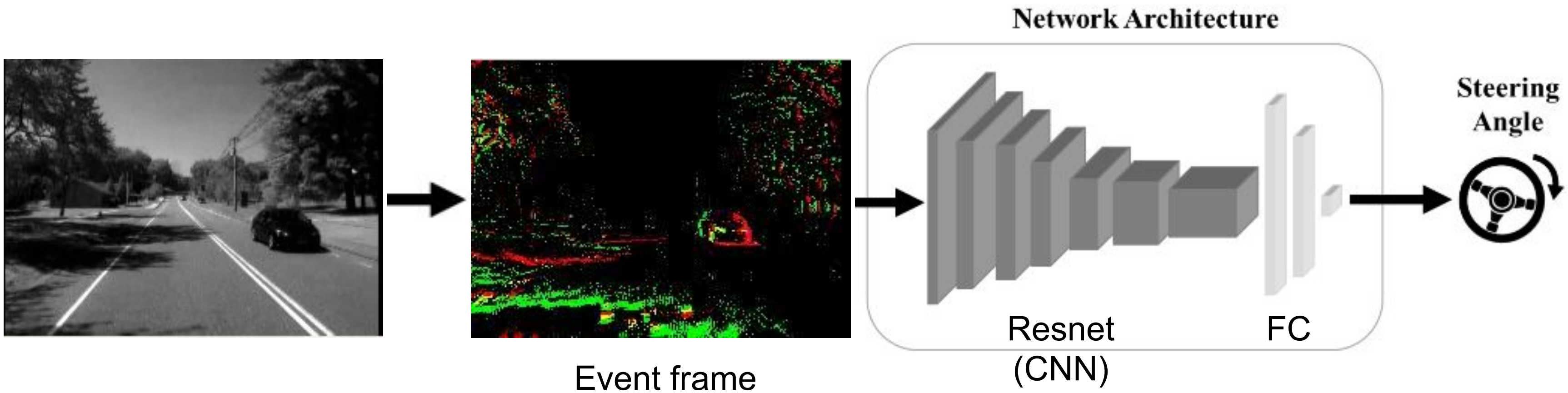
Related Work: Frame-based models



Event frame
R: positive
G: negative

[Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars, Manqueda et al.]

Related Work: Frame-based models



Event frame
R: positive
G: negative

Problems:
- passive training not tested on a real vehicle
- unable to capture the whole scene at low speed

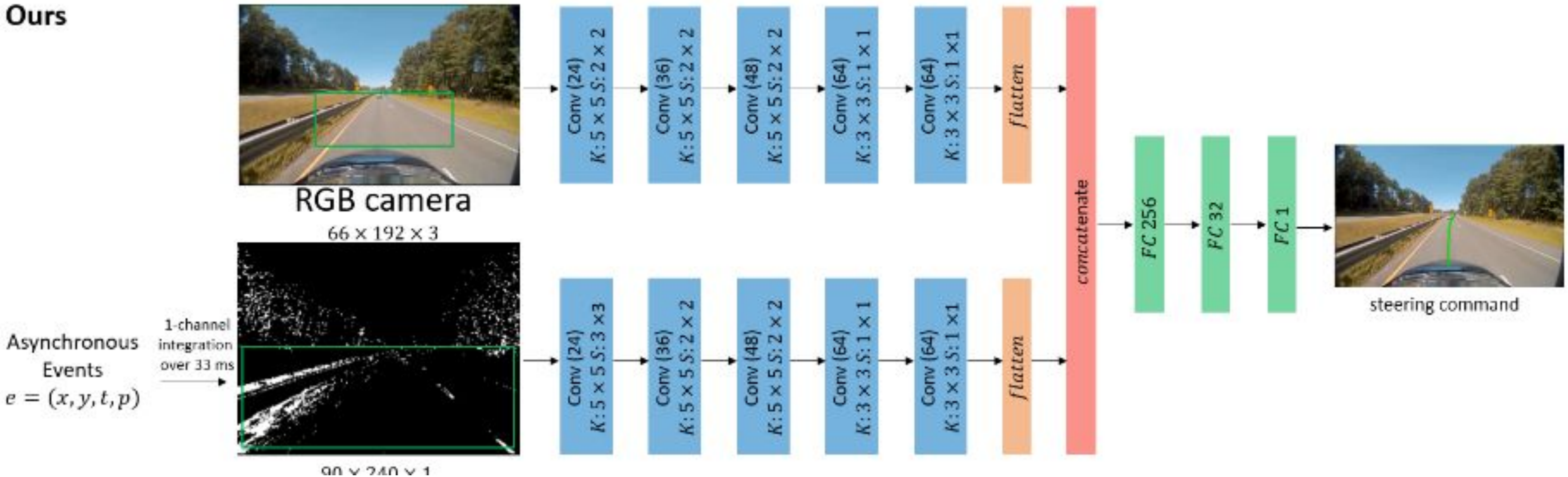
[Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars, Manqueda et al.]

Our proposed model

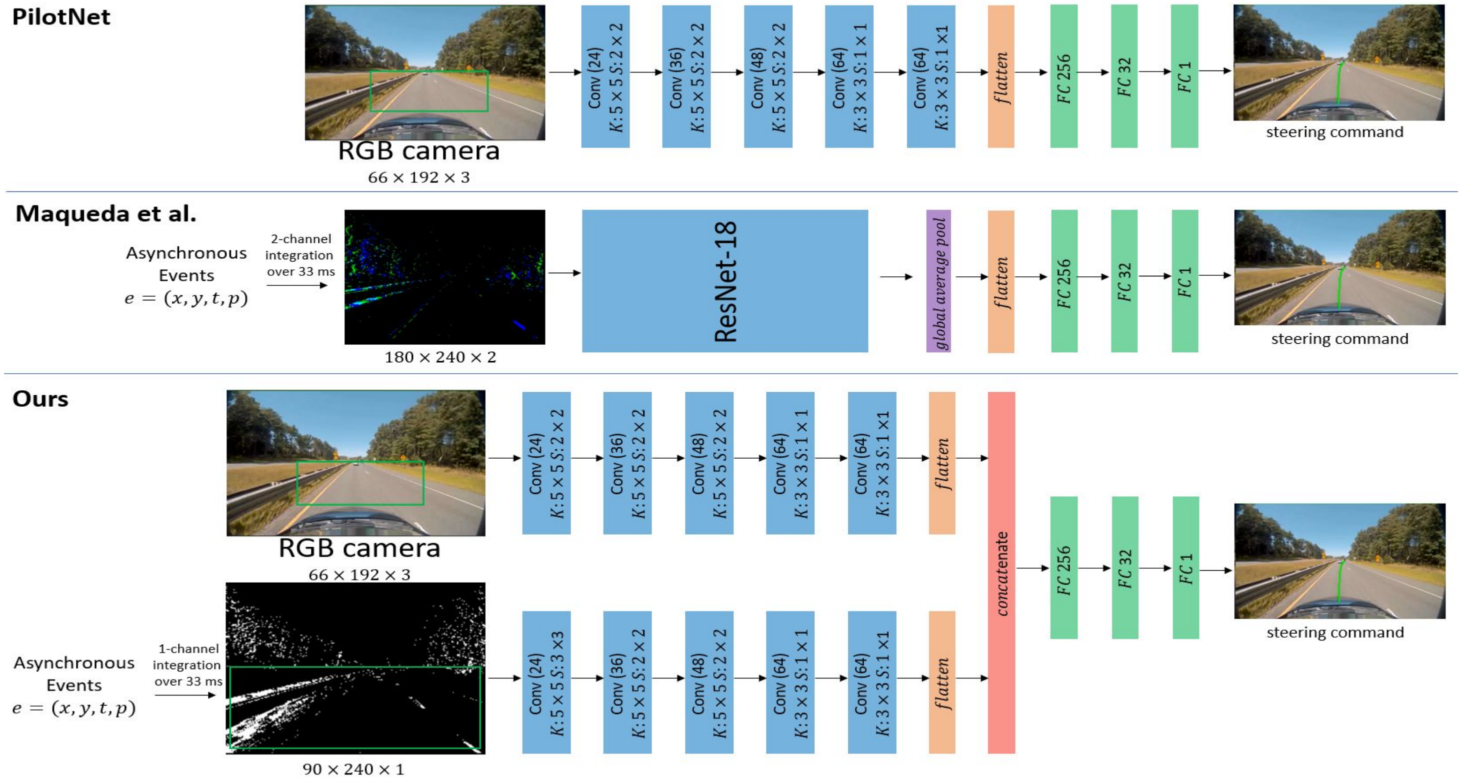
Thoughts: Augment the event-based model with inputs from a traditional RGB camera, so that the combined model perform at least as well as the best of the RGB-based and event-based models.

Our proposed model

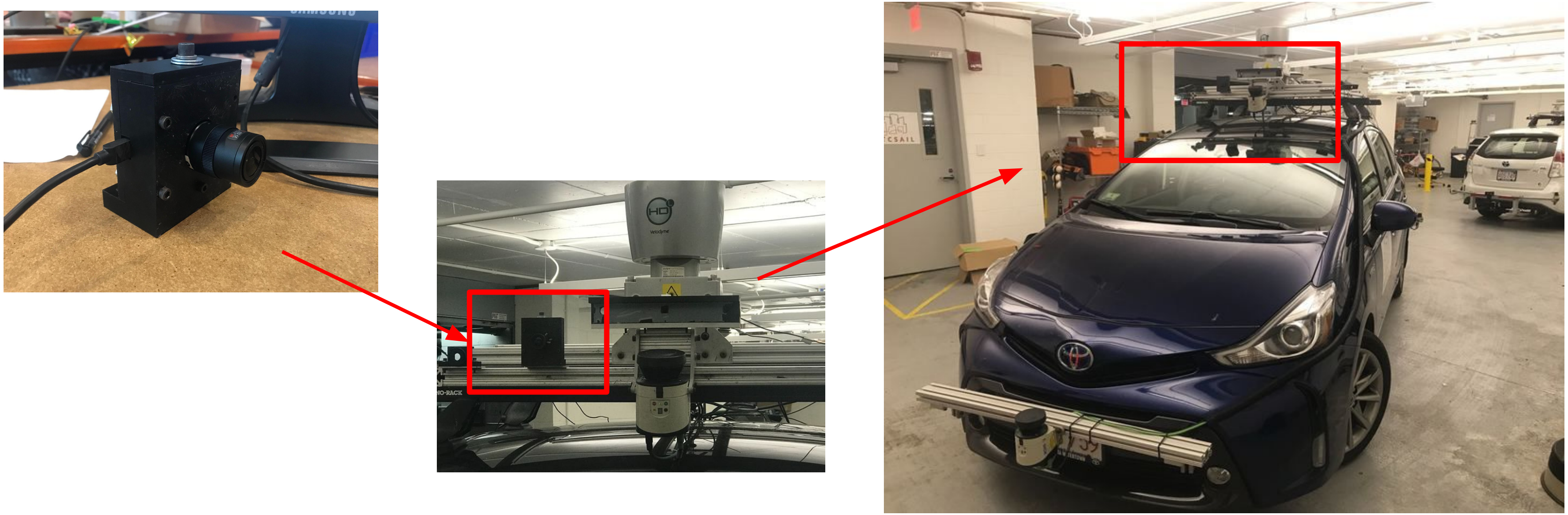
Thoughts: Augment the event-based model with inputs from a traditional RGB camera, so that the combined model perform at least as well as the best of the RGB-based and event-based models.



Comparison between the three models

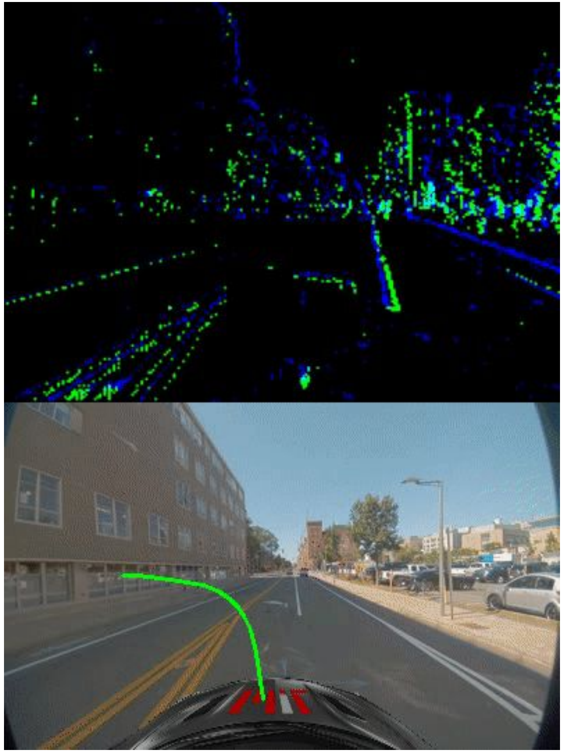


Experiment Vehicle Setup



Experiment Dataset

2 hours of human driving around Boston on urban roads
Supervise on curvature (1 / radius)



Experiment Metrics

Given ground truth value α and prediction value $\hat{\alpha}$

(Rooted Mean Sqare Error) $$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{a}_j - a_j)^2}$$

(Explained Variance) $$EVA = 1 - \frac{Var(\hat{a} - a)}{Var(a)}$$

Experiment Result on dataset

Model	urban RMSE	urban EVA
PilotNet [15]	0.00694	0.108
Ours	0.00665	0.182
Maqueda et al. [18]	0.00624	0.275
Maqueda et al. with ROI cropping	0.00707	0.109
Maqueda et al. with 1-channel integration	0.00666	0.175
Maqueda et al. with both	0.00707	0.0907

The original event-based model performs the best, but ROI-cropping and 1-channel integration decreased its performance.

Our model, which could be seen as a mixture PilotNet and Maqueda et al. with both ROI cropping and 1-channel integration, indeed perform better than either of them.

Experiment result on real cars

Model	autonomy
PilotNet [15]	66%
Maqueda et al. [18]	0%
Ours	45%

metric: $$autonomy = (1 - \frac{(number\ of\ interventions) \cdot 6\ seconds}{elapsed\ time\ [seconds]}) \cdot 100$$

Discussion

PilotNet also uses Imitation learning, so why does it work better?

Discussion:

Challenges

- Event-based cameras provide structure of the scene and the motion of the camera
- The model turns out to predict the existing motion of the car rather than learning how to drive

Discussion:

Challenges

- Event-based cameras provide structure of the scene and the motion of the camera
- The model turns out to predict the existing motion of the car rather than learning how to drive

Potential solutions for the future

- Use Deep Reinforcement Learning for the model to learn the correct causation
- Work on a event-based simulation platform

Thank you! Questions?

- My mentors: Dr. Igor Gilitschenski and Alexander Amini
- Prof Daniela Rus, Distributed Robotics Lab, MIT CSAIL
- MIT PRIMES
- My parents