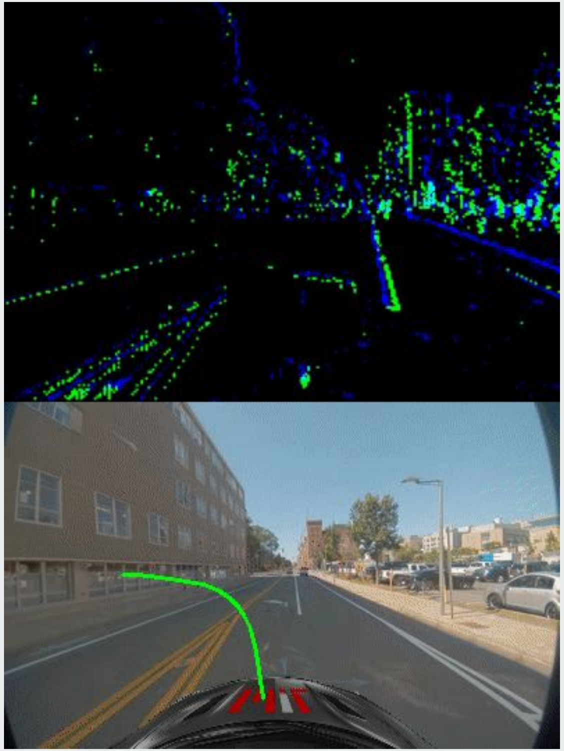
A Deep Learning Approach to End-to-end Autonomous Driving Using Event-based Vision



Yuxuan Chen
Mentors: Dr. Igor Gilitschenski, Alexander Amini

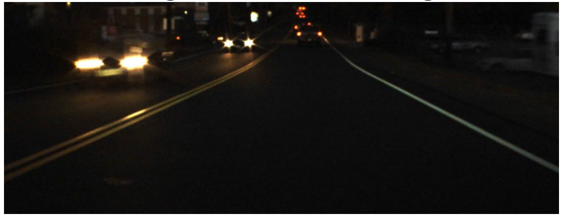
Overview

- Motivation
- Brief introduction to event-based vision
- Our goal
- Related works
- Our works
- Experiments

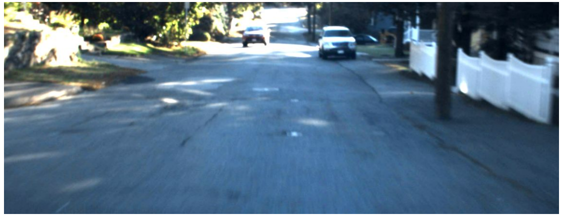
Motivation

Autonomous driving cars need to handle a wide range of scenarios

Night-time Driving



No Lane Markings



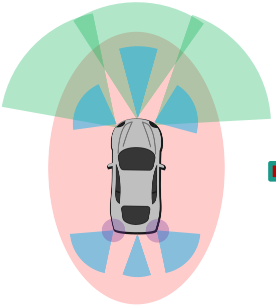
Rainy Weather

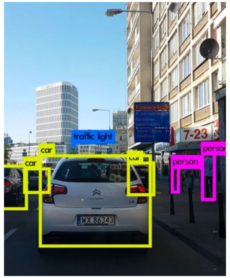
[Learning steering bounds for parallel autonomous systems, Amini et al.]

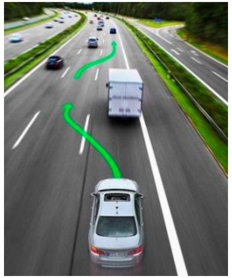
How do they do it?

Autonomous Driving Pipeline

Separate problem into smaller sub-modules, tackle each independently







Sensor Fusion
- What's happening around me?

Detection
- Where are obstacles?

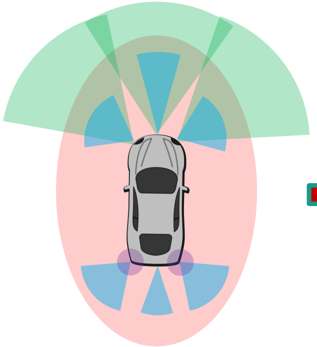
Localization
- Where am I relative to the obstacles?

Planning
- Where do I go?

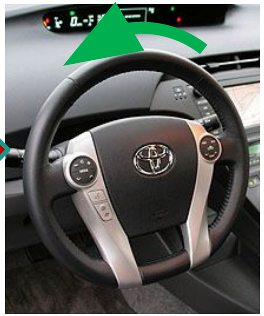
[Learning steering bounds for parallel autonomous systems, Amini et al.]

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

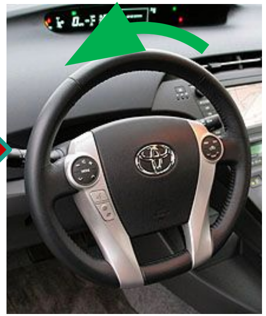
Actuation
- What control signals to take?

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

End-to-end Learning

Learn the steering directly from pixel values

Deep Neural Network



Raw images: a front facing RGB camera

Learned Model
Underlying representation of how humans drive

Actuation
- What control signals to take?

[Learning steering bounds for parallel autonomous systems, Amini et 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

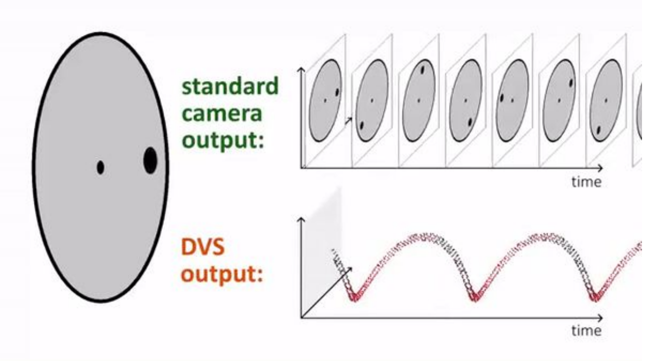
DAVIS240 from Inivation.com

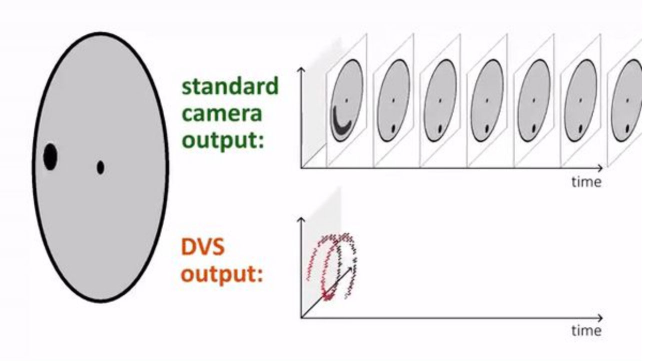
[Event-based Cameras: Challenges and Opportunities, Scaramuzza et al.]

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.]

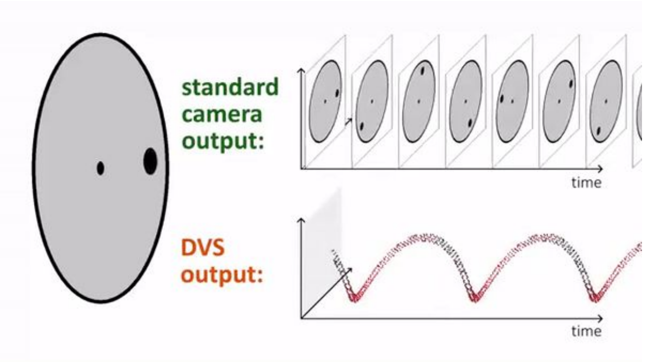
What are event-based cameras

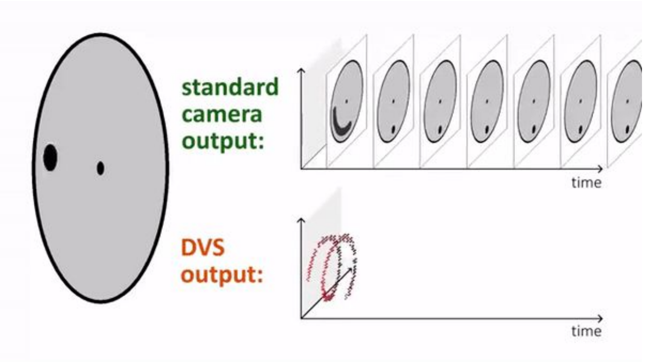
Novel bio-inspired sensors that capture motion in the scene

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

Novel bio-inspired sensors that capture motion in the scene

DAVIS240 from Inivation.com

Benefits:

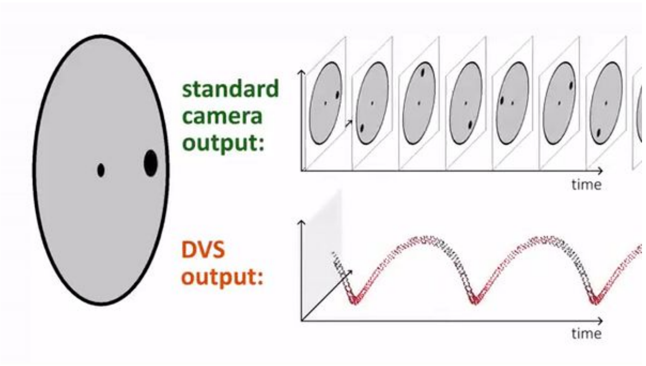
- Low latency (~ 1 microsecond)
- No motion blur
- High dynamic range (140 dB instead of 60dB)

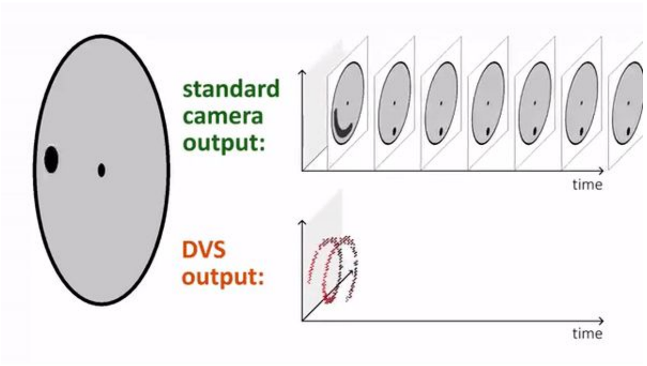
Challenges:

- 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.]

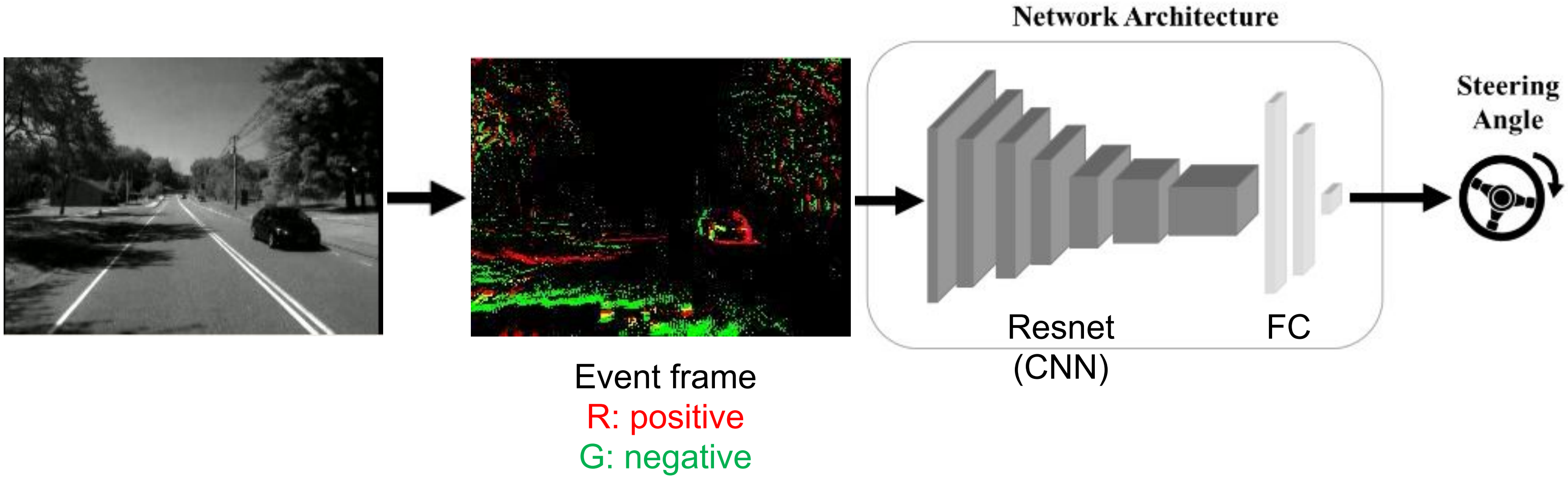
Our Goal

Use an event camera to drive a car in real time

DAVIS240 from Inivation.com

Deep Neural Network

Related Work: Frame-based models

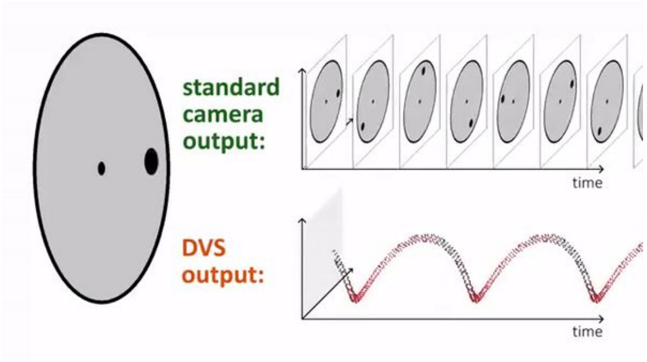


[Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars, Manqueda et al.]

If we use frame-based model,
why don’t we use RGB cameras instead?

PointNet-based models

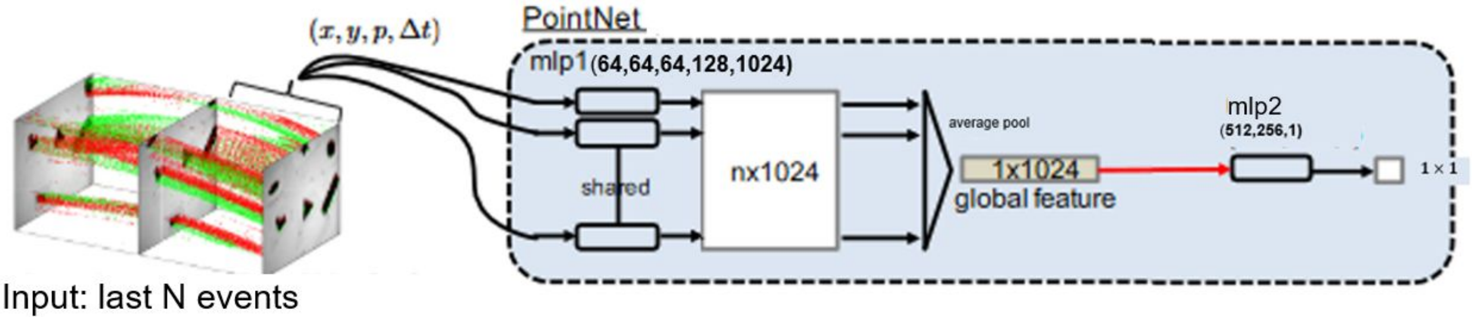
- Events = points in (x, y, t, p) dimensions



[PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al.]
[EventNet: Asynchronous Recursive Event Processing, Sekikawa et al.]

PointNet-based models

- Events = points in (x, y, p, t) dimensions
- PointNet is able to process Point Clouds (sets of points):



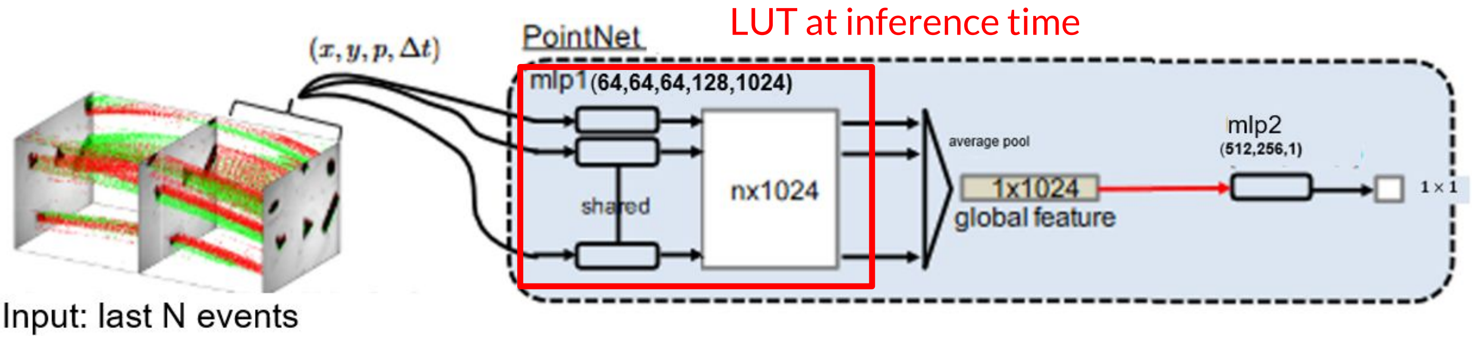
[PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al.]

[EventNet: Asynchronous Recursive Event Processing, Sekikawa et al.]

PointNet-based models

Inspired by EventNet, during inference time:

- Precompute the result of mlp1 into Look Up Table (LUT) of shape W×H×T×2
- Significantly faster than the vanilla PointNet and frame-based models



[PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al.]
[EventNet: Asynchronous Recursive Event Processing, Sekikawa et al.]

Experiment Metrics

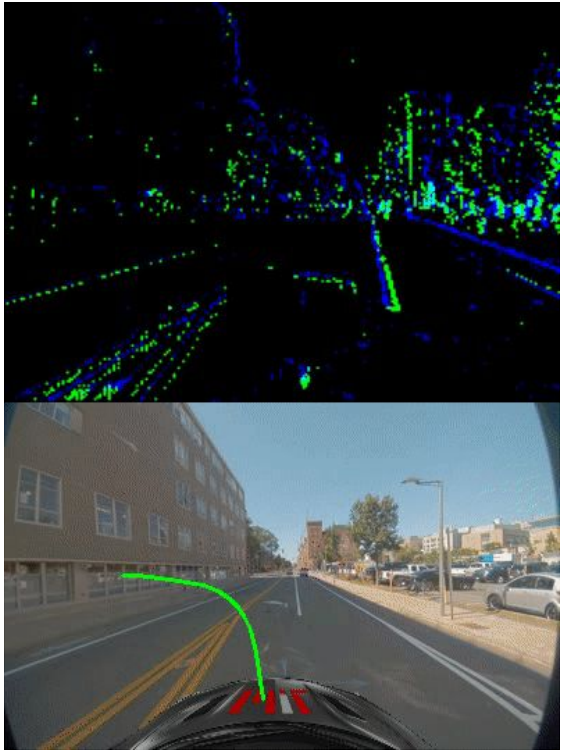
Given ground truth value α and prediction value $\hat{\alpha}$

- Rooted Mean Square Error (RMSE) = $\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{\alpha}_j - \alpha_j)^2}.$

- Expected Variance (EVA) = $1 - \frac{\mathrm{Var}(\hat{\alpha} - \alpha)}{\mathrm{Var}(\alpha)}.$

Experiment Dataset

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



Experiment Result

Comparison between Frame-based and PointNet-based Models

	Frame-based	PointNet-based (with fixed N=5000)
EVA	0.193	0.144
RMSE (m^-1)	0.00657	0.00722

Experiment Result

Comparison between Frame-based and PointNet-based Models

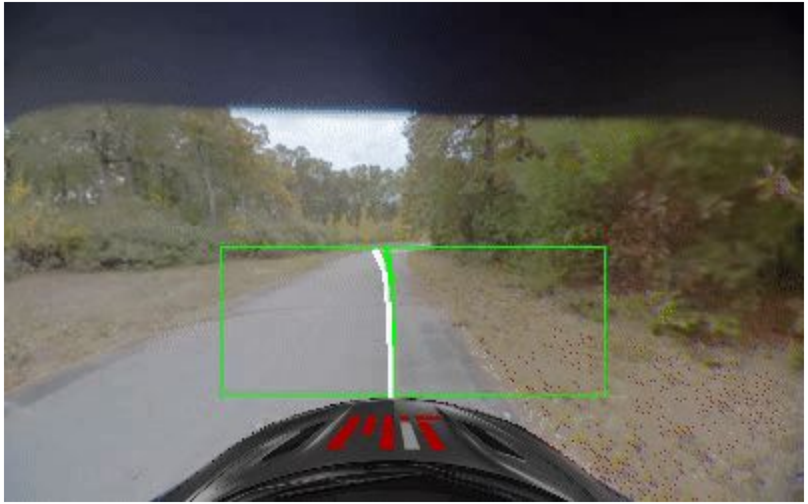
	Frame-based	PointNet-based (with fixed N=5000)
EVA	0.193	0.144
RMSE (m^-1)	0.00657	0.00722

EVA result of PointNet-based models trained and validated on different number of points

train\valid	N=1000	N=2000	N=4000	N=10000
N=1000	0.104	0.105	0.095	0.054
N=2000	0.111	0.116	0.113	0.078
N=4000	0.109	0.125	0.148	0.146
N=10000	0.029	0.039	0.060	0.122

Our Question

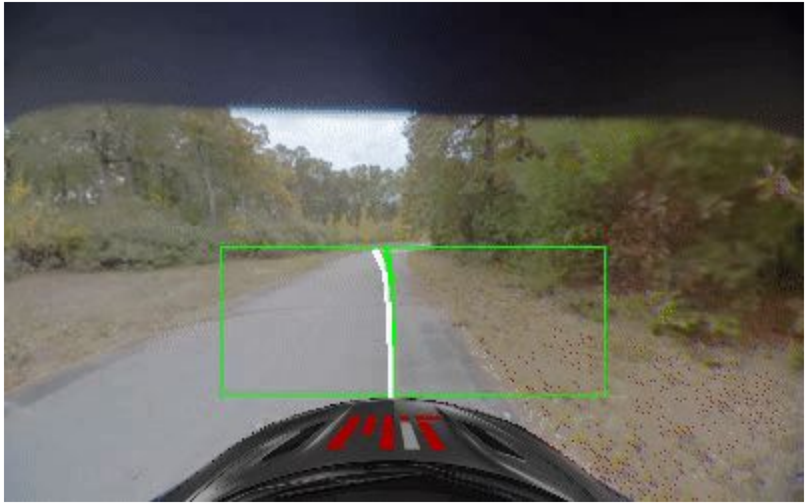
Can these models actually drive a car?



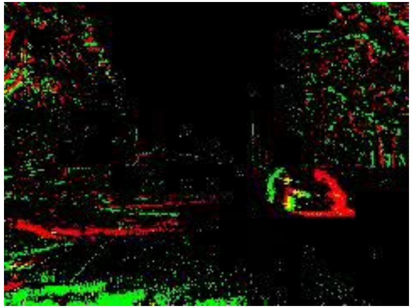
Our Question

Can these models actually drive a car?

The model may cheat by predicting the **motion of the car** rather than **learning the steering wheel angle**!



Our Question

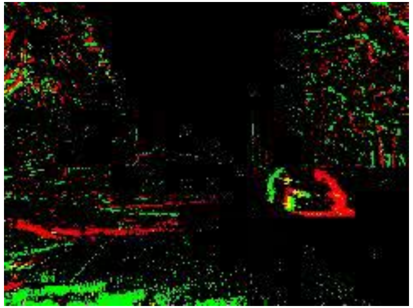


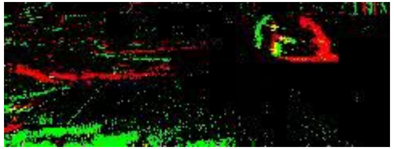
Let's look at our data again

1) Many events are irrelevant

Our Question

Let's look at our data again

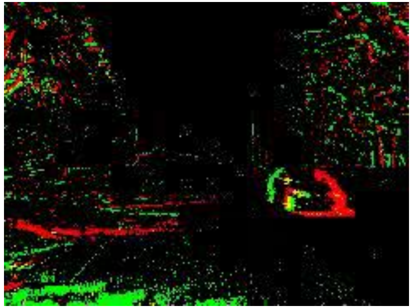




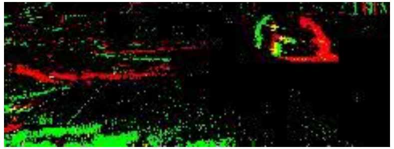
1) Many events are irrelevant

Region Of Interest (ROI) cropping

Our Question



Let's look at our data again



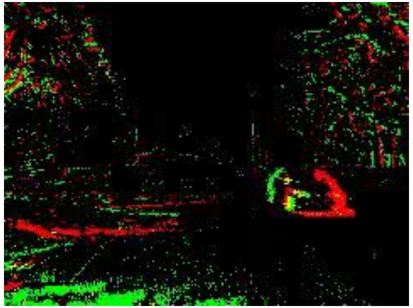
1) Many events are irrelevant

Region Of Interest (ROI) cropping

2) Event polarity gives away the motion of the car

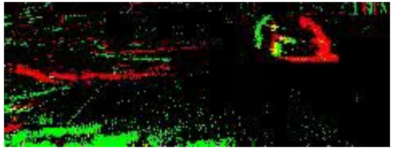
Our Question

Let's look at our data again



1) Many events are irrelevant

Region Of Interest (ROI) cropping



2) Event polarity gives away the motion of the car

Ignore the event polarity

Ablation Studies

Ablation studies using Frame-based model in last experiment

	EVA
Original Data	0.19
Data with ROI cropping	0.09
Data with polarity ignored	0.13
Data with polarity ignored and ROI cropping	0.09

Our contribution

- Sensor Integration on MIT Autonomous Vehicle

Our Contribution

- Sensor Integration on MIT Autonomous Vehicle
- Introducing PointNet-based Model for event-based driving that:
    - processes directly on raw events
    - is fast in inference time

Our Contribution

- Sensor Integration on MIT Autonomous Vehicle
- Introducing PointNet-based Model for event-based driving that:
    - processes directly on raw events
    - is fast in inference time
- Evaluation of PointNet-based Model on real world driving data

Our Contribution

- Sensor Integration on MIT Autonomous Vehicle
- Introducing PointNet-based Model for event-based driving that:
    - processes directly on raw events
    - is fast in inference time
- Evaluation of PointNet-based Model on real world driving data
- Ablation Studies

Thank you! Questions?

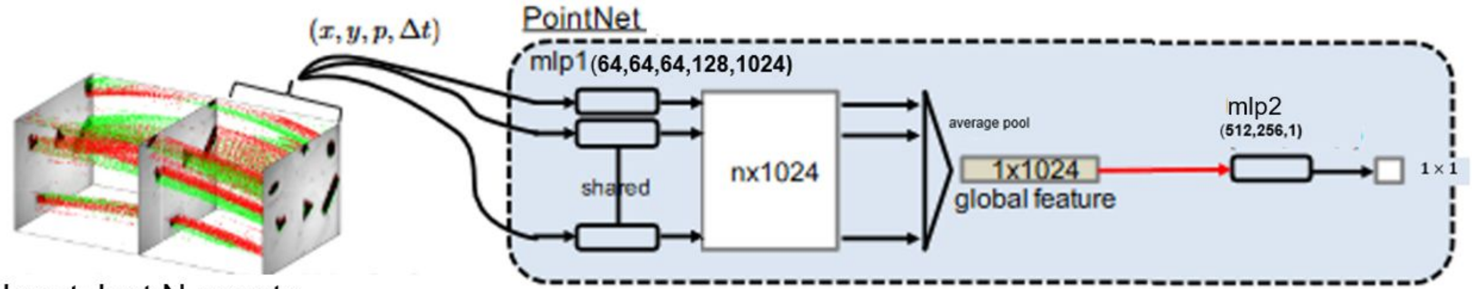
- My mentors: Dr. Igor Gilitschenski and Alexander Amini
- Prof Daniela Rus, Distributed Robotics Lab, MIT CSAIL
- MIT PRIMES
- My parents

PointNet-based models

- Events = points in (x, y, t, p) dimensions
- PointNet is able to process Point Clouds (sets of points):

$$f(\{x_1, \dots, x_n\}) \approx g(h(x_1), \dots, h(x_n)), \qquad (1)$$

where $f : 2^{\mathbb{R}^N} \to \mathbb{R}$, $h : \mathbb{R}^N \to \mathbb{R}^K$ and $g : \underbrace{\mathbb{R}^K \times \cdots \times \mathbb{R}^K}_{n} \to \mathbb{R}$ is a symmetric function.



Input: last N events

[PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al.]
[EventNet: Asynchronous Recursive Event Processing, Sekikawa et al.]