Automatic Unusual Driving Event Identification for Dependable Self-Driving

SenSys '18: 16th ACM Conference on Embedded Network Sensor Systems

Presented by -
Khushwant Kaswan
241110035

CS724
SENSING COMMUNICATIONS AND NETWORKING FOR SMART WIRELESS DEVICES
IIT KANPUR

Introduction

The paper addresses the need to detect unusual driving events and corner cases from dashcam video and inertial sensors , aiming to improve the reliability of self-driving systems. Current automated driving solutions struggle with unexpected situations, as dependability in complex, real-world scenarios remains challenging motivating this research.

This paper introduces a low-cost yet scalable solution to collect such events from any dash cam equipped vehicle to take advantage of the billions of miles that humans already drive.

Challenges and Motivation

- Most existing efforts collect driving data with a small fleet of tens to hundreds of highly instrumented vehicles that are continuously operated with test drivers, but it is challenging to cover billions of miles with such a small fleet.
  Solution : Stress testing on proving grounds and in simulations
- Manual Human inspection of collected data to flag unusual driving events is one possible solution, but will require plenty of extra effort, amplify privacy concerns, and increase storage and networking overhead for collecting all data.
  Automatically identifying unusual driving events remains a challenge.
- To address this challenge, authors proposed an automatic unusual driving events identification system, which can detect unusual situations through in-vehicle algorithms and can easily be scaled for wide deployment.

Current automated Driving (NHTSA)

National Highway Traffic Safety Administration

Level 0 - No Automation
Level 1 - Driver Assistance
Level 2 - Partial Automation
Level 3 - Conditional Automation
Level 4 - High Automation
Level 5 - Full Automation

Tesla's Autopilot is classified as L2-3

MOTIVATION
To accelerate development of truly dependable L4 and L5 systems.

Testing

Testing of ’known corner’ cases through a combination of proving ground testing and simulations.

- Public Road Testing
- Closed Source Testing
- Simulation Testing

It remains unclear whether this testing with known corner cases can lead to the desired **level of robustness** or whether some unknown corner cases exist that still need to be discovered in the next billions of miles of driving.

Design Goals

- **Sense from human driven vehicles** - The unusual driving event identification system should not rely on a full self-driving sensor suite but should expect minimal infrastructure and data source such as dashcams.
- **Minimize data uploads** - The system should be able to identify relevant events while most of the rich video data remains in the vehicle and uploading only the identified event's data to the cloud instead of uploading whole video.
- **Off-the-shelf devices** - An on-board device with the sensing and computational capabilities of a high-end smartphone providing real time processing/near real-time computation with limited resources.

The design of this system will take front view videos and inertial readings as input and provide data that can be directly used as training or testing on current automatic driving components.

For systems which require other sensor input, the detected events can be used to define test cases and simulate them on testing sites to collect the necessary data for other sensors.

Unusual Events

- Sudden braking and swerving
- traffic scenarios that blind sensors,
- scenarios where key traffic participants are occluded or obscured by other objects
- unexpected movements by traffic participants
- Deer standing next to road....

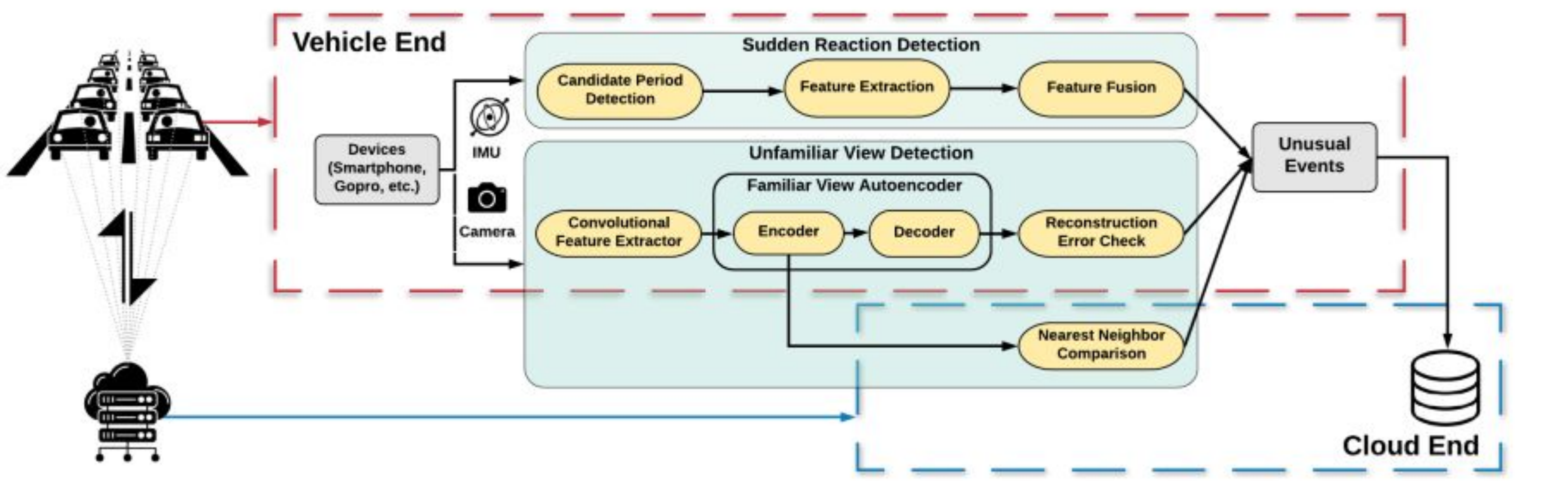
(a) Swerving to avoid a tire segment (b) Braking due to a cut-in vehicle (c) Special lightening condition (d) Complex roadside lightening condition (e) Unusual traffic and vehicle pose

System Overview

- Two-staged strategy
  - **Sudden Reaction Detection** : how a human driver reacts in terms of sudden steering and braking
  - **Unfamiliar View Detection** : how different the camera inputs are from previously observed inputs
- Off-the-shelf on-board device : it requires camera, accelerometer and gyroscope sensors, and processing capabilities to execute neural networks, and network connectivity to allow collection of data about unusual events and corner cases.

Rationale behind two-staged strategy :

- situations that surprise a human driver are more likely to also challenge an automated driving system
- one can also expect situation that does not elicit a reaction from an attentive human driver, but could confuse automated driving algorithms



Vehicle End
Sudden Reaction Detection
Candidate Period Detection
Feature Extraction
Feature Fusion
Devices (Smartphone, Gopro, etc.)
IMU
Camera
Unfamiliar View Detection
Familiar View Autoencoder
Convolutional Feature Extractor
Encoder
Decoder
Reconstruction Error Check
Nearest Neighbor Comparison
Unusual Events
Cloud End

Sudden Reaction Detection

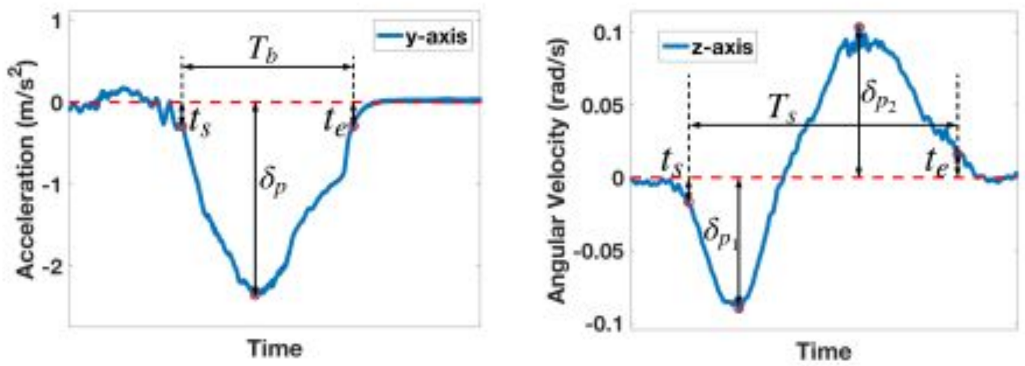
- Data is collected by the accelerometer and gyroscope of an IMU sensors.
- Sudden reaction events : hard braking or high speed swerving
- Three-stage inertial sensing detection technique
    - Candidate Period Detection
    - Feature Extraction
    - Feature Fusion

Since the pose of the IMU within the vehicle is usually unknown, the system uses coordinate alignment algorithms to project the IMU's reading from its own coordinates system to the vehicle's coordinates system.

Besides, low pass filters are used to remove the noise from the raw IMU's reading caused by vehicle vibrations and bad road conditions.

Candidate Period Detection

First stage of our detection mechanism to identify candidate periods in terms of braking and swerving events



(a) Braking Event (b) Swerving Event

Feature Extraction and Fusion

- Features used : Amplitude , Derivatives , Duration and Fused Feature.
- Amplitude : the sudden braking events and swerving events could be detected based on a threshold δ value
- derivative of acceleration to represent the urgency of a braking event
- derivatives of gyroscope reading during swerving events to identify urgent swerving events
- Unusual event often happens in a short moment, use the event duration to detect the urgency of unusual events.
- Feature Fusion :

$$f_{fusion} = \sum_i w_i f_i \qquad w_i = \frac{n_{f_i}}{\sum_i n_{f_i}}$$

UNFAMILIAR VIEW DETECTION

Two unfamiliar view detection methods :
(i) in-vehicle detector based on autoencoder reconstruction error
(ii) joint in-vehicle and cloud detector based on the autoencoder embedded vectors

Two main components : Familiar View Autoencoder and CNN Feature Extractor

An autoencoder is a neural network that is trained to encode the input into a set of low dimensional representations, which can be decoded to reconstruct an output that is nearly identical to its input. Two main parts, an encoder function $\phi(x)$ and a decoder that produce a reconstruction $x' = \psi(\phi(x))$.
An autoencoder is designed to simply learn the set $\psi(\phi(x)) = x$ and minimize minimize reconstruction loss ($L(x, x')$).

$$L(x, x') = \|x - x'\|^2 = \|x - \psi(\phi(x))\|^2$$

UNFAMILIAR VIEW DETECTION

Self Steering System implemented using CNN Feature Extractor : An self steering system maps raw pixels from a single front-facing camera directly to steering commands using a convolutional neural network.

Based on the Implementation, Given a front-facing camera image X (66x200x3), it will be firstly processed by 5 convolutional layers **c**, and then fed to 3 fully connected layers **f** for steering angle prediction.

$$\theta_{error} = \theta - \hat{\theta} = \theta - f(c(X))$$

CNN layers : First 3 layers are using 2x2 stride and 5x5 Kernel size and last two layers are non strided with a 3x3 Kernel size

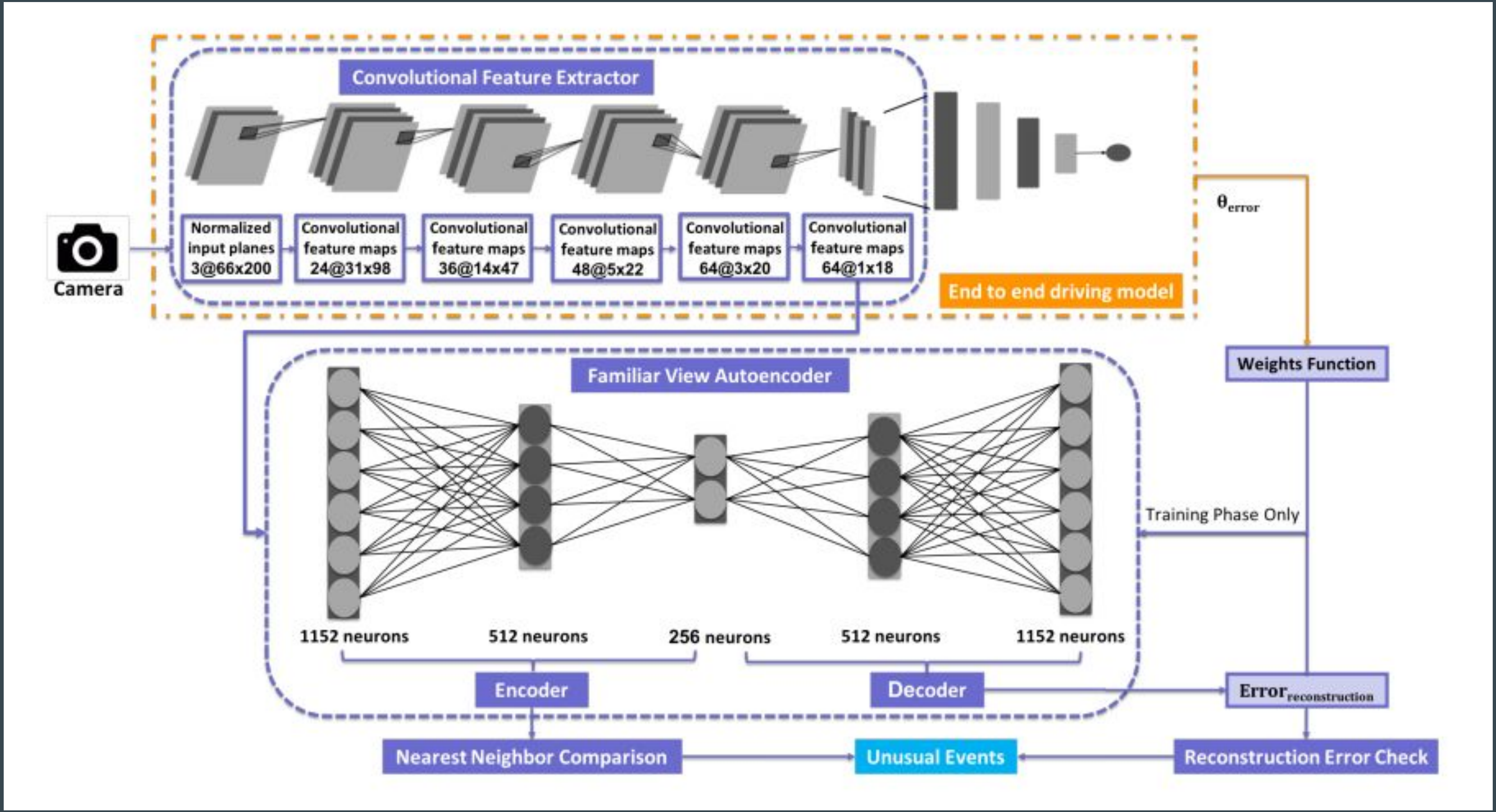
UNFAMILIAR VIEW DETECTION

Sample Weighted Loss Function :

the initial weights for each training sample : $$w = \frac{1}{\log_{scalar_1}(\|\theta_{error}\| * scalar_2 + bias)}$$

θerror is the steering prediction error
scalar 1 and scalar 2 are used to control the range and density of weights w
bias is the value we set to keep the value in the valid domain of log function

Based on the weights definition, the loss function for familiar view autoencoder is defined as $$L(x, x') = w * \|c(x) - \psi(\phi(c(x)))\|^2 + L_2$$



Joint in-vehicle and cloud detector system design

UNFAMILIAR VIEW DETECTION

In-vehicle detector leverages the outputs from the decoder of familiar view autoencoder to estimate the distance between an input sample and the whole training set by evaluating the reconstruction error. Will have relatively lower accuracy but much less computation cost as it only need perform one time pass of the neural Network.

Joint in-vehicle and cloud detector takes the outputs from the encoder part of familiar view autoencoder to check the distance between the input sample and its nearest neighbours by encoding training samples into the same space. Such nearest neighbours comparison requires iterating pairwise distance evaluations on all the known samples, thus takes much more time but can produce relatively higher accuracy.

Evaluation

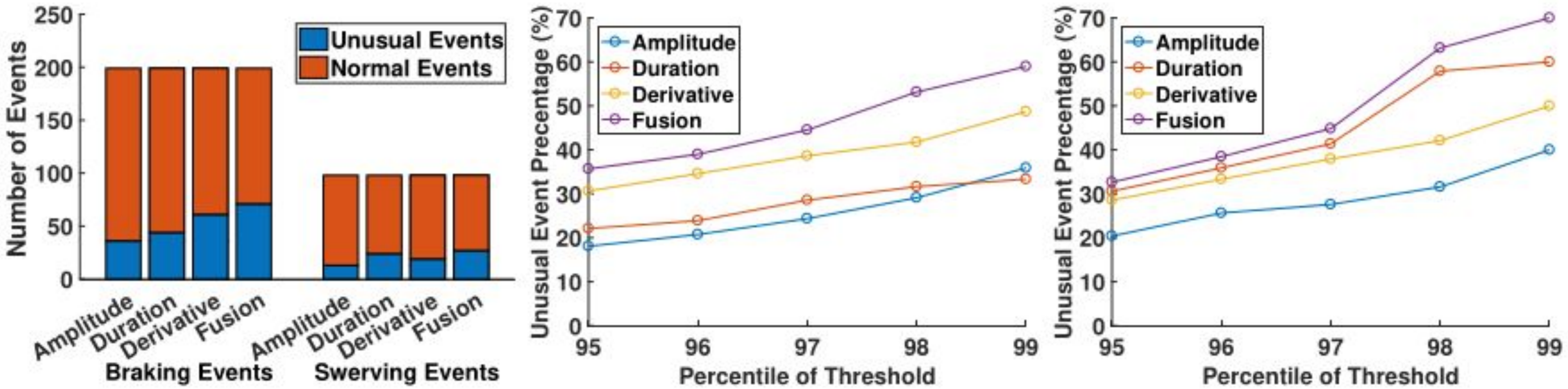
The sudden reaction detection is evaluated with a 120-hour dataset collected in Los Angeles.

Dataset from Udacity end-to-end driving challenge is used train and evaluate unfamiliar view detection model.

Why?? The 120-hour dataset collected does not have accurate driver's steering angle while driving and the steering angle is necessary for unfamiliar view detection.

Accuracies of the proposed methods are compared with the Baseline Strawman solutions.

Evaluation results of Sudden Reaction Detection



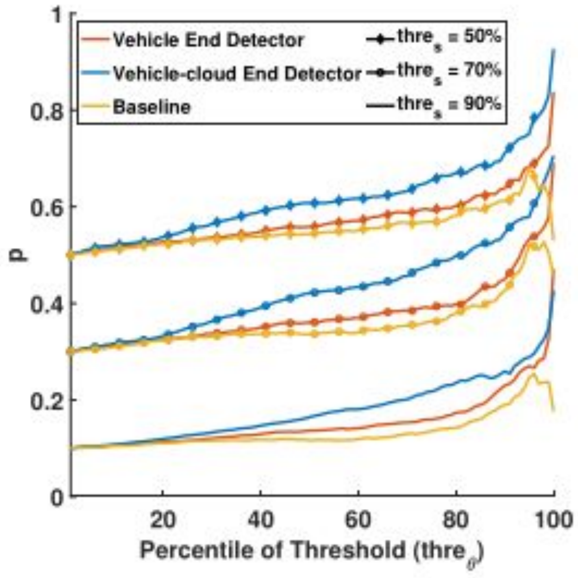
(a) Unusual Events Distribution

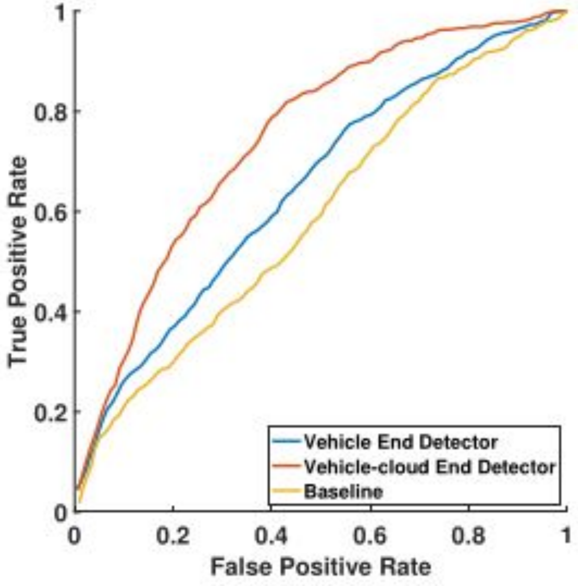
(b) Braking Events Detection Accuracy

(c) Swerving Events Detection Accuracy

Evaluation results for Unfamiliar view Detection

$$p = \frac{\text{number of detected events whose } \theta_{error} > thre_\theta}{\text{number of detected events}}$$





Evaluation compared to Baseline

Methods	Proposed Method (%)	Baseline (%)
Sudden Reaction Detection for Braking Events	53.16	29.11
Sudden Reaction Detection for Swerving Events	63.16	31.58
Vehicle End Unfamiliar View Detection	71.43	64.53
Vehicle-Cloud End Unfamiliar View Detection	80.30	64.53

~82% increase accuracy improvement vs Strawman solutions for sudden reaction detection
80% accuracy for cloud based unusual visual view detection

Usefulness of Unusual Events

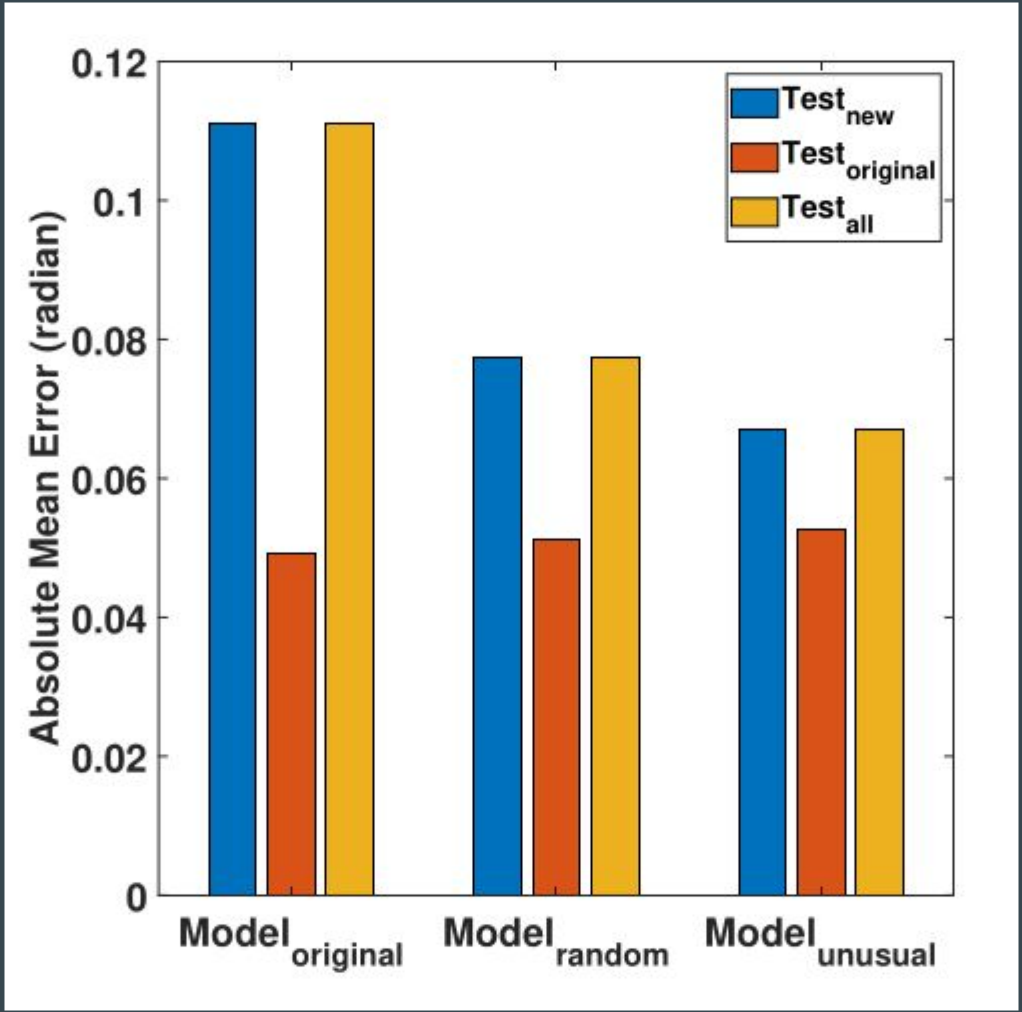
Created a new Testing dataset using another dataset also provided by Udacity.

We use 80% of the samples to train an end to end driving model Model-original and the rest of 20% samples as the test set Test-original.

Among this dataset, first 10,000 samples are picked as a sample pool and train Model-new and next 1000 samples as Test-new.

Similarly two another strategy where we randomly selects new samples from the sample pool Model-random and the other one utilizes in-vehicle detector to pick unfamiliar views Model-unusual.

From the results based on the overall test set Test-all , Model-unusual still shows the best performance than the other two, which shows the detected unusual events are able to increase model's overall performance and robustness on different roads and traffic conditions.



Thank You