# Towards Safe Autonomy in Hybrid Traffic: The Power of Information Sharing in Detecting Abnormal Human Drivers Behaviors

**Jiangwei Wang**[1] , **Lili Su**[2] , **Songyang Han** [1] and **Dongjin Song** [1] and **Fei Miao** [1]

[1] University of Connecticut
[2] Northeastern University

{jiangwei.wang, songyang.han, dongjin.song, fei.miao}@uconn.edu, l.su@northeastern.edu,

## Abstract

Hybrid traffic, consisting of both autonomous and human-driven vehicles, would be the norm of autonomous vehicle practice for decades. It comes with both challenges and opportunities: human-driven vehicles could exhibit sudden abnormal behaviors such as unpredictably switching to dangerous driving modes – putting its neighboring vehicles under risk; modern vehicle-to-vehicle (V2V) communication technologies enable the autonomous vehicles to efficiently and reliably share the scarce run-time information among each other. To this end, we propose, to the best of our knowledge, the first algorithm that can quickly detect the occurrence of abnormal human driving behaviors with formal assurance yet without hurting privacy. Through extensive empirical studies on existing datasets and simulators, we demonstrate that our proposed algorithm with information sharing can not only predict the trajectory more accurately but also can detect the abnormal behaviors with high accuracy (95.7%) with reasonably low detection delay (1.5s).

## 1 Introduction

Despite the rapid development of autonomous vehicles in past decades, hybrid traffic which involves both autonomous and human-driven vehicles would be a norm for a long time [Bimbraw, 2015; Huang *et al.*, 2016; Veres *et al.*, 2011]. In this work, we exploit the safety advantages raised by the extended sensing capability of autonomous vehicles through beneficial information sharing.

Enabling safe autonomy of autonomous vehicles in hybrid traffic is challenging. Normal driving behaviors can be characterized by well-controlled speed and heading features such as mild accelerations, decelerations, and lane changes. In contrast, abnormal behaviors do not exhibit those features [Hu *et al.*, 2017]. Human-driven vehicles might not maintain a normal driving mode and could *suddenly* switch to safety-threatening driving modes. Such switches usually arise from human factors such as fatigue, drunkenness, distraction, and aggressiveness. If not detected in a timely manner, such unannounced switches could quickly put their surrounding vehicles under serious safety threats. To the best of our knowledge, existing abnormal driving behavior detection designs are rather heuristic and focus on monitoring either behavioral parameters such as eye blinking and yawning [Yan *et al.*, 2016; Hu *et al.*, 2019; Shahverdy *et al.*, 2020; Lemley *et al.*, 2019; Reddy *et al.*, 2017], or vehicular parameters such as speed variability, steering wheel angle, and steering wheel grip force [Li *et al.*, 2017; Zhenhai *et al.*, 2017], which require placing sensors on vehicle parts like steering wheel, accelerator or brake pedal. Unfortunately, human-driven vehicles might not have the required sensor placements to collect the relevant run-time measurements. What's worse, such measurements, even if available, are privacy sensitive and should not be shared with other vehicles.

On the positive side of hybrid traffic, modern V2V communication technologies enable the autonomous vehicles to efficiently and reliably share the scarce run-time information with each other [Committee and others, 2009]. The U.S. Department of Transportation (DOT) has estimated that V2V communication based on DSRC can address up to 82% of all crashes in the United States involving unimpaired drivers, potentially saving thousands of lives and billions of dollars [Kenney, 2011]. In addition, navigation and control strategies based on V2V shared information can also improve both traffic efficiency and safety [Han *et al.*, 2019].

The above pros and cons of hybrid traffic naturally bring two questions: (1) How to design an algorithm that can detect the human driver's abnormal behaviors with high accuracy in a short time, without violation of human driver's privacy? (2) How will the information sharing among autonomous vehicles help to detect the abnormal behaviors? To answer these questions, We propose, to the best of our knowledge, the first efficient algorithm that can accurately and quickly detect abnormal human driving mode switches with formal assurance but does not hurt privacy.

**Contributions:**

- We propose multi-encoder attention based trajectory predictor (MEATP), which successfully utilizes the shared information among autonomous vehicles.
- To protect human driver's privacy, We develop abnormal driving behavior detection algorithm by monitoring the run time trajectory prediction error.
- Extensive experiments on existing datasets and simula-

tor show the effectiveness of our proposed detection algorithm and its robustness against noises in inputs.

## 2 Problem Description

### 2.1 Hybrid traffic system description

We refer to the autonomous vehicle that is doing trajectory prediction as the ego vehicle (EV), other connected autonomous vehicles (CAVs) that are within the communication range of the ego vehicle as surrounding vehicles (SVs), and a human-driven vehicle being predicted by the EV as a target vehicle (TV). Without loss of generality, we focus on one TV and denote other human-driven vehicles as HVs. Our algorithm works for the general multiple TVs setting by executing it for different TVs in parallel. The system is illustrated in Fig. 1. Notably, the human-driven vehicles are not communicating with others.

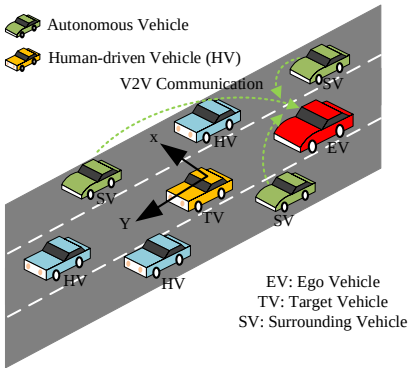


Figure 1: Hybrid traffic with information sharing.

### 2.2 Information shared to the ego vehicle

With Lidar point cloud data and camera images, and Lidar-camera fusion techniques [Caltagirone *et al.*, 2019; Yin *et al.*, 2021; Qin *et al.*, 2019], each SV can get the GPS locations of the nearby human-driven vehicles. At current system time $t_0$, the information each SV shares with the EV includes

(1) its own GPS locations in the past $t_h$ time steps,

(2) locally sensed GPS locations of the nearby human-driven vehicles in the past $t_h$ time steps.

In real world applications, it's unrealistic for the EV to get accurate GPS locations of all the human-driven vehicles only from the observations of the CAVs. There will be noises in the measurement due to a variety of factors such as the limitation of on board sensors, the error of object detection and tracking algorithms, etc [Yin *et al.*, 2021; Fayyad *et al.*, 2020; Cho *et al.*, 2019]. In this work, we discuss two cases: first is an ideal case, where (1) and (2) are all accurate GPS locations without noise, second is a more realistic case, where (1) are accurate and (2) are noisy. We will elaborate and validate our abnormal behavior detection algorithm against noises in inputs in Section 3 and Section 4 respectively.

As illustrated in Fig. 1, without information sharing, it's impossible for the EV to get the trajectory information of the vehicles that are in front of the TV. Yet, such information is important for prediction of the TV as those vehicles interact with the TV and influence its future trajectories.

## 3 Abnormal Behavior Detection Framework

In this section, we first present multi-encoder attention based trajectory predictor (MEATP) in Section 3.1. Based on that, in Section 3.2, we elaborate on the shared information based abnormal behavior detection algorithm (Algorithm 1). We illustrate the details of calculating the statistic in Section 3.3.

### 3.1 Multi-encoder attention based trajectory predictor

To effectively utilize the shared information in predicting the trajectory of the TV, we propose a **multi-encoder attention based trajectory predictor (MEATP)**, illustrated in Fig. 2. Our model is based on the sequence to sequence transformer [Vaswani *et al.*, 2017] but with multiple encoders for fusing shared information. Through experiments, we will show that our proposed architecture effectively extracts the relevant spatial temporal information contained in the shared trajectories and significantly reduces both the mean and variance of the prediction errors.

**Inputs and outputs**

**Inputs:** We define the reference frame according to the TV with $X$-axis as the lateral direction and the $Y$-axis as the longitudinal direction, shown in Fig. 1. At time step $t_0$, the origin is the TV's current location. In the range of $y \in [-30m, 30m]$, let $N$ be the total number of vehicles that interact with the TV, including $N_A$ autonomous vehicles and $N_H$ human-driven vehicles. Upon receiving shared information, the EV transforms the GPS locations into the coordinates of its current reference frame. The inputs to the multi-head attention encoders are composed of two parts. The **first** part are the trajectories of neighboring vehicles that interact with the TV, $i \in \{1, \dots, N\}$:

$$\boldsymbol{C}_i(t_0) \triangleq [\boldsymbol{c}_i(t_0 - t_h), \cdots, \boldsymbol{c}_i(t_0 - t_h + \ell), \cdots, \boldsymbol{c}_i(t_0)] \quad (1)$$

with $\boldsymbol{c}_i(t_0 - t_h + \ell) \triangleq [(x_i(t_0 - t_h + \ell), y_i(t_0 - t_h + \ell))]$ being the 2-dimensional coordinates of the $i$–th neighboring vehicle at time $(t_0 - t_h + \ell)$ for $\ell = 0, \cdots, t_h$, Note that, the coordinates of the $N_H$ human driven vehicles can be noisy as we have explained in Section 2. The **second** part is the trajectory of the TV over a sliding time window $\{t_0 - t_h, t_0 - t_h + 1, \cdots, t_0\}$.

$$\boldsymbol{S}(t_0) = [\boldsymbol{s}(t_0 - t_h), \dots, \boldsymbol{s}(t_0 - t_h + \ell), \dots, \boldsymbol{s}(t_0)] \quad (2)$$

with $\boldsymbol{s}(t_0 - t_h + \ell) \triangleq [x_0(t_0 - t_h + \ell), y_0(t_0 - t_h + \ell)]$ being the 2-dimensional vector that records the TV's coordinates.

**Output:** Distributions of the future trajectory of the TV over time window $t \in \{t_0 + 1, \cdots, t_0 + t_f\}$. Assume the predicted future trajectory follows bivariate Gaussian distribution [Chandra *et al.*, 2019; Deo and Trivedi, 2018], the output of the decoder is the bivariate Gaussian parameters at every time step in the future $t_f$ time steps: $\boldsymbol{\Omega} = [\boldsymbol{\Omega}(t_0 + 1), \dots, \boldsymbol{\Omega}(t_0 + t_f)]$,

$$\boldsymbol{\Omega}(t) = [\boldsymbol{\mu}(t), \boldsymbol{\sigma}(t), \rho(t)] = [\mu_x(t), \mu_y(t), \sigma_x(t), \sigma_y(t), \rho(t)],$$

$(\mu_x(t), \sigma_x(t))$ and $(\mu_y(t), \sigma_y(t))$ are the mean and standard deviation (SD) in $x$-axis and $y$-axis, respectively, and $\rho(t)$ is the corresponding correlation-coefficient.

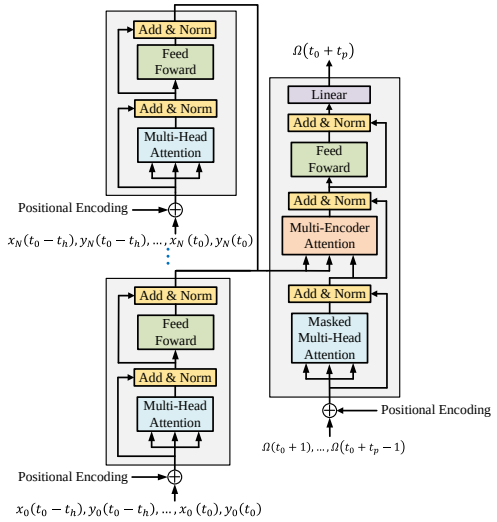


Figure 2: **MEATP: multi-encoder single-decoder architecture.** The Multi-Encoder Attention Mechanism in decoder is shown in Fig. 3

### Model architectures

Our proposed architecture contains $N+1$ encoders: one encoder takes the TV's trajectory $\boldsymbol{S}(t_0)$ as its input; the remaining $i \in \{1, \ldots, N\}$ encoders corresponds to one of the $N$ neighboring vehicle that interacts with the TV. Although the trajectories of human-driven vehicles can be noisy, the way we use the encoder-decoder structure is similar to the idea of denoising autoencoders [Vincent *et al.*, 2010] and thus can make MEATP be robust to the noisy inputs. The inputs plus the positional encoding, are sent to the next layer as queries $Q$, keys $K$, and values $V$. Each encoder consists of two sub-layers: a multi-head attention layer and a position-wise fully connected feed-forward network [Vaswani *et al.*, 2017]. Multi-head attention layer is composed by $h$ heads of scaled dot-product attentions:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V,$$

where $d_k$ is the dimension of keys. Attention outputs are concatenated and linearly transformed into the same dimension of Q:

$$\text{Multihead}(Q, K, V) = \text{concat}(\text{head}_1, \ldots, \text{head}_h)W^O, \quad (3)$$

$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$. The encoders pass the output as values and keys, which contains the weighted information of historical data, to the multi-encoder attention layer of decoder. TV encoder outputs $(K_0, V_0)$; $i$–th encoder corresponds to $i$–th vehicle and outputs $(K_i, V_i)$.

At time step $t_p$, decoder takes the outputs in the past $(t_p - 1)$ steps as input, and outputs the bivariate Gaussian parameters of the probability distribution of the TV's future coordinates. In decoder, we develop **multi-encoder attention** to utilize the shared information. As shown in Fig. 3, the multi-encoder attention is composed by $N+1$ multi-head attention. Query $Q_0$ of the decoder interacts with each pair of keys and values $K_i, V_i$ in a multi-head attention. The outputs of the multi-head attention are concatenated and linearly transformed into the same dimension as $Q_0$. Multi-encoder attention is expressed as:

$$\text{Multiencoder}(Q_0, K, V) = \text{concat}(M_0, \ldots, M_N)W^M, \quad (4)$$

where $M_i = \text{Multihead}(Q_0, K_i, V_i)$. Since the TV is continuously interacting with its neighboring vehicles, the past trajectories of the neighboring vehicles, and itself, together influence its future trajectory. By letting query $Q_0$, which corresponds to the TV, interacts with $K_i, V_i$, the decoder learns both temporal and spatial information of the neighboring vehicles and the TV.

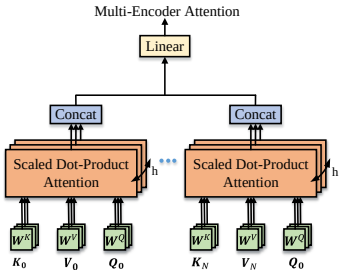


Figure 3: Multi-encoder attention

### Loss function

Formally, we denote our proposed predictor as a function $f_W$, with $W$ being all the trained parameters, and $f_W(\boldsymbol{S}(t_0), \boldsymbol{C_1}(t_0), \cdots, \boldsymbol{C_N}(t_0)) = [\boldsymbol{\Omega}(t_0+1), \ldots, \boldsymbol{\Omega}(t_0+t_f)]$ – recalling that the output of each prediction step follows a bivariate Gaussian distribution. Our MEATP is trained based on the following weighted sum:

$$L_1 = \sum_{t=t_0+1}^{t_0+t_f}(-\log P(\boldsymbol{Z}(t)|\boldsymbol{\mu}(t), \boldsymbol{\sigma}(t), \rho(t))),$$
$$L_2 = \sum_{t=t_0+1}^{t_0+t_f} ||\boldsymbol{Z}(t) - \boldsymbol{\mu}(t)||,$$

where $\boldsymbol{Z(t)} = [x_0(t), y_0(t)]$ is the true coordinates of the TV at the prediction time $t$, $L_1$ is the negative log likelihood of the true trajectories given the predicted trajectory distributions over time steps $\{t_0+1, \cdots, t_0+t_f\}$, and $L_2$ is the total $l_2$ deviation of the true trajectories from the predicted means. In our experiments, we choose the mean on $X$ and $Y$ axis $\boldsymbol{\mu}(t) = [\mu_x(t), \mu_y(t)]$ in output distribution as TV's predicted future trajectory.

## 3.2 Abnormal human-driver's behavior detection

Recall that the EV is the one that executes the MEATP and monitors whether the TV is in an abnormal driving mode or not. We use $\gamma$ to denote the unknown time where TV turns into abnormal driving mode. We consider the most challenging scenario wherein no prior information on $\gamma$ is available.

Recall that $t_0$ is the variable that indicates the current system time; its value increases by 1 as each time goes by. At each time, the EV first receives the shared GPS locations from its SVs, transforms them into $\boldsymbol{S}(t_0)$ and $\{\boldsymbol{C_i}(t_0)\}_{i=1}^N$, and then passes them as inputs to the MEATP to obtain $\boldsymbol{\mu}(t_0+1)$. Finally, when the true $\boldsymbol{Z}(t_0+1)$ is revealed, the EV computes its prediction error

$$\boldsymbol{e}_{t_0+1} = ||\boldsymbol{Z}(t_0+1) - \boldsymbol{\mu}(t_0+1)||, \quad (5)$$

As $t_0$ increases over time, via the above process, the EV computes a sequence of prediction errors $\{\boldsymbol{e}_n, n = 1, 2, \cdots\}$.

We use $f$ and $g$ to denote the distribution of $\boldsymbol{e}_n$ when $n < \gamma$ and $n \geq \gamma$, respectively. Clearly, if $\gamma > t_0$, i.e., the TV's driving mode has not switched yet, the TV is currently in a normal mode with $\boldsymbol{e}_n \overset{i.i.d.}{\sim} f$ for $n = 1, 2, \cdots, t_0$. If $\gamma \leq t_0$, i.e., the TV is in the abnormal mode, then $\boldsymbol{e}_n \overset{i.i.d.}{\sim} g$ for $n = t_0, \cdots$. At any time, the EV is interested in knowing whether a mode switch has occurred or not and wants to detect such switch as soon as possible under a given false alarm budget. **Therefore we formulate the problem of detecting abnormal human driver behaviors as detecting the change in distributions of the sequence of random prediction errors $\{\boldsymbol{e}_n, n = 1, 2, \cdots\}$.**

**Definition 1** (Detection algorithm as a stopping time)**.** A *stopping time* of the sequence of random prediction errors $\{\boldsymbol{e}_n, n = 1, 2, \cdots\}$ is a random variable $\tau$ that the event $\{\tau = n\}$ is measurable of $\sigma(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ for each $n$. When used as a detection algorithm, the event $\tau = n$ is interpreted as *"a distribution change is declared at time $n$"* [Veeravalli and Banerjee, 2014].

Given $\{\boldsymbol{e}_n, n = 1, 2, \cdots\}$, it's hard to tell when the distribution has changed by manual inspection, instead, we compute a statistic $W_n$ given $\{\boldsymbol{e}_n, n = 1, 2, \cdots\}$; the computation of $W_n$ is specified in Section 3.3. Once the statistic $W_n$ exceeds some carefully calibrated threshold $b$, we declare the change in distributions, i.e., abnormal driving behavior happened. Our algorithm is formally described in Algorithm 1. Notably, this algorithm can be ran in parallel to detect multiple TVs.

**Algorithm 1:** Abnormal Behavior Detection based on Shared Information

Initialize $W_0 \leftarrow 0$, $t_0 \leftarrow 0$, $\mu(t_0) \leftarrow Z(t_0)$;
Set threshold $b \leftarrow \log(\frac{M}{\alpha})$;
/* $\alpha$ is the given false alarm budget and $M$ is the number of possible $g$ */
**while** *true* **do**
  Receive information shared by the SVs;
  Update $\boldsymbol{S}(t_0)$ and $\{\boldsymbol{C}_i(t_0)\}_{i=1}^N$ by incorporating this newly received information;
  $\boldsymbol{\mu}(t_0+1) \leftarrow f_W(\boldsymbol{S}(t_0), \{\boldsymbol{C}_i(t_0)\}_{i=1}^N)$;
  /* Compute $\boldsymbol{\mu}(t_0+1)$ by calling MEATP */ $\boldsymbol{e}_{t_0} \leftarrow ||\boldsymbol{Z}(t_0) - \boldsymbol{\mu}(t_0)||$ ;
  Compute statistic $W_{t_0}$ using MCuSum algorithm;
  **if** $W_{t_0} \geq b$ **then**
    Declare the detection of abnormal behavior;
    **Break;**
  **end**
  $t_0 \leftarrow t_0 + 1$;
**end**

## 3.3 The update of $W_n$ and the choice of $b$

We first focus on the ideal case where the locally sensed (at each SV) GPS locations of the nearby human-driven vehicles are noiseless. Then we study the more realistic setting where the sensed locations are noisy.

**Noiseless HV location measurements**

In this setting, we choose the threshold $b$ and the update rule based on the MCuSum algorithm [Tartakovsky and Polunchenko, 2008] – an asymptotically optimal QCD algorithm. For ease of exposition, we assume that $f$ and $g$ are gaussian distributions; they can be parameterized by their means and standard deviations as $f_\phi$ and $g_\theta$, where each of $\phi$ and $\theta$ is a tuple of mean and standard deviation. We assume that we have full knowledge of the pre-change distribution $f_\phi$ as it can be learned from historical data. In contrast, we assume that we only have partial knowledge of $g_\theta$ with parameter $\theta \in \Theta = \{\theta_1, \theta_2, \ldots \theta_M\}$. Incomplete knowledge of $g_\theta$ arises from the fact that when the driver switches into an abnormal driving mode, he is more likely to take unreasonable actions like frequent lane changing, sudden acceleration and deceleration, etc. We treat the set $\Theta$ as finite as in practice we can do discretization on the parameter space. In our context, we choose several possible means and standard deviations based on the prior information we have about the prediction errors after the abnormal behavior happens. The above uncertainty in $g_\theta$ can be efficiently handled by the MCuSum algorithm, detailed as follows.

**MCuSum algorithm** [Tartakovsky and Polunchenko, 2008]:

$$\tau_{MC} \triangleq \inf\{n \geq 1 : \max_{j \in \{1, \cdots, M\}} W_n(\theta_j) \geq b\}, \qquad (6)$$

where

$$W_n(\theta_j) = \begin{cases} \left[W_{n-1}(\theta_j) + \log(\frac{g_{\theta_j}(\boldsymbol{e}_n)}{f_\phi(\boldsymbol{e}_n)})\right]^+, & n \geq 1; \\ 0, & n = 0. \end{cases} \qquad (7)$$

In Eq. (6), $[x]^+ \triangleq \max\{x, 0\}$, and, denoting $\phi = (\mu_0, \sigma_0)$ and $\theta_j = (\mu_j, \sigma_j)$, the log likelihood can be written as

$$\log \frac{g_{\theta_j}(\boldsymbol{e}_n)}{f_\phi(\boldsymbol{e}_n)} = \frac{(\boldsymbol{e}_n - \mu_0)^2}{2\sigma_0^2} - \frac{(\boldsymbol{e}_n - \mu_j)^2}{2\sigma_j^2} + \log(\frac{\sigma_0}{\sigma_j}). \qquad (8)$$

Under the MCuSum algorithm, $W_n(\theta_j)$ for $j = 1, \cdots, M$ are updated in parallel in each time step. A change is declared upon the first time at least one $W_n(\theta_j)$ hits $b$. It turns out that we can set $b = \log \frac{M}{\alpha}$ for some given false alarm budget. Asymptotic optimality of the MCuSum algorithm with this choice of $b$ is proved in literature.

**Proposition 1.** *[Tartakovsky and Polunchenko, 2008] The MCuSum algorithm with $b = \log \frac{M}{\alpha}$ for any given $\alpha \geq 0$ is first order asymptotically optimal. Furthermore, the false alarm rate (FAR) and the average detection delay (WADD)*[1] *of $\tau_{MC}$ are bounded as follows:*

$$FAR(\tau_{MC}) \leq \alpha, \text{ and } WADD^\theta(\tau_{MC}) \lesssim \left(\frac{|\log \alpha|}{D_{KL}(g_\theta, f_\phi)}\right) \text{ as } \alpha \to 0, \forall \theta \in \Theta,$$

*where $D_{KL}(g_\theta, f_\phi)$ is the Kullback-Leibler divergence between $g_\theta$ and $f_\phi$.*

[1] Two key performance metrics of a detection algorithm in Lorden's minimax QCD formulation [Lorden and others, 1971].

**Noisy HV location measurements**
When SVs share noisy human driven vehicles' GPS locations to the EV, due to the difference in the inputs to our proposed MEATP, the distribution of prediction error changes accordingly. The pre-change distribution changes from $f_\phi$ to $f_{\phi'}$, where $\phi' = (\mu_0', \sigma_0')$, post-change distribution changes from $g_{\theta_j}$ to $g_{\theta_j'}$, where $\theta_j' = (\mu_j', \sigma_j')$. Note that the measurement noises have no impacts on $\gamma$. Thus, we can still use the MCuSum algorithm. For each $\theta_j'$, the log likelihood becomes:

$$\log(\frac{g_{\theta_j'}(\boldsymbol{e}_n)}{f_{\phi'}(\boldsymbol{e}_n)}) = \frac{(\boldsymbol{e}_n - \mu_0')^2}{2\left(\sigma_0'\right)^2} - \frac{(\boldsymbol{e}_n - \mu_j')^2}{2\left(\sigma_j'\right)^2} + \log(\frac{\sigma_0'}{\sigma_j'}). \quad (9)$$

# 4 Experiments

## 4.1 Trajectory prediction

1) **Experiment details**: In this section, we first train our model on NGSIM US-101 and I-80 dataset [Colyar and Halkias, 2007]. It consists of trajectories of freeway (US-101 and I-80) traffic sampled at frequency 10Hz over 45 minutes. The dataset is split into training and testing set by ratio of 7:3. We train our model using Adam with learning rate of 0.01. We assign $L1$ and $L2$ loss with weight 0.3 and 0.7 respectively. The dimension of the model, also known as number of features, is 16. The number of heads is 8. For the feed forward layer, it contains a linear layer of size (16, 32), a Relu Layer, and another linear layer of size (32,16) in sequential. We use past 3s of trajectories to predict the trajectories in future 5s with sample frequency of 5Hz. Therefore, $t_h = 16$ and $t_f = 25$. In our experiments, we use a server configured with Intel Core i9-10900X processors and four NVIDIA RTX2080Ti GPUs. Our experiments are performed on Python 3.6.0, PyTorch 1.6.0, and CUDA 11.0.

We compare our methods with baseline: LSTM with convolutional social pooling (CS-LSTM) [Deo and Trivedi, 2018]. This model devises a convolutional social social pooling layer to process the LSTM encoder output, generates a unimodal distribution for future coordinates. We evaluate our proposed MEATP in two modes, with and without shared information. When there is no information sharing, we assume that the EV can only get the trajectories of vehicles that are around it through its own sensors. While with information sharing, we assume that EV is able to get the TV's neighboring vehicles' historical trajectories by fusing the shared information and feed them into encoders.

2) **Results**: We use root mean square error (RMSE) in unit of meter in future 5s to compare our proposed method with the baselines. In Table 1 we show that: (1) Our model has smaller RMSE values compared with baselines when there is no shared information. (2) Our proposed multi-encoder attention successfully fuses the shared information, encodes the spatial and temporal features of TV and all the neighboring vehicles, the prediction performance is improved by more than $50\%$ compared with baseline CS-LSTM.

## 4.2 Human driver abnormal behavior detection

1) **Experiment details**: We use open source simulator Simulation of Urban Mobility (SUMO) [Lopez *et al.*, 2018] to generate the highway and urban traffic datasets. The settings include: (1) Highway scenario: a 1000 m highway with 5 lanes. (2) Urban scenario: a 1000m city street with 5 lanes, two traffic lights at 300m and 600m from starting point, each of them are set to green light status $80\%$ of one cycle, where one cycle is two minutes. (3) 8000 vehicles in both scenarios running end to end, total time of traffic flow is 1 hour. 1000 vehicles are switched to abnormal driving mode once they pass a certain location on the road. They differ from the normal driving behaviors with altered parameters as shown in Table 2. Note that, sigma denotes the driver imperfection in Krauß car following model [Krauß, 1998], the larger the more imperfection. Larger lcSigma represents less perfect in lane changing. Larger lcCooperative means vehicles are less willing to perform cooperative lane changing, larger lcSpeedGain means vehicles tend to change lane more frequently to gain high speed. Vehicle trajectories are collected at frequency of 10Hz of in one hour with label of whether the vehicle is abnormal.

| Models | Prediction horizon (s) | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| CS-LSTM | 0.61 | 1.27 | 2.13 | 3.21 | 4.37 |
| MEATP w/o shared information | 0.84 | 0.99 | 1.18 | 1.39 | 2.16 |
| MEATP w shared information | 0.51 | 0.89 | 1.09 | 1.30 | 1.72 |

Table 1: Trajectory prediction results

| Parameters | Normal driving behaviors | Abnormal driving behaviors |
|---|---|---|
| maxAccel $(m/s^2)$ | 2.6 | 7 |
| maxDecel $(m/s^2)$ | 4.5 | 8 |
| miniGap $(m)$ | 2.5 | 1.0 |
| sigma | 0.1 | 0.8 |
| maxSpeed (highway) $(m/s)$ | 30 | 50 |
| maxSpeed (urban) $(m/s)$ | 15 | 30 |
| speedFactor | 1.0 | 1.2 |
| lcCooperative | 1.0 | 0.1 |
| lcSpeedGain | 1.0 | 5.0 |
| lcSigma | 0.1 | 0.8 |

Table 2: Parameters of different driving behaviors

2) **Results**: We train our proposed MEATP and the baseline CS-LSTM on the normal vehicles' trajectories in SUMO highway traffic dataset. We define the prediction error as the mean error between true trajectories and predicted trajectories. We add four levels of Gaussian noise to the human driven vehicles' coordinates, with mean and SD being (0.3m, 0.2m), (0.3m, 0.4m), (0.6m, 0.2m), (0.6m, 0.4m) respectively based on the exiting 3D object detection and tracking algorithms [Yin *et al.*, 2021; Qin *et al.*, 2019; Asvadi *et al.*, 2016], level 0 means no noise in inputs. We apply the trained predictors to vehicles' trajectories, then compute the mean and SDs of the prediction errors in future 3s. The distributions of the prediction errors on normal vehicles

| Models | Prediction horizon (s) | noise level 0 | | 1 | | 2 | | 3 | | 4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| CS-LSTM | 1 | 0.76 | 0.54 | 2.18 | 1.43 | 2.25 | 1.37 | 2.27 | 1.62 | 2.31 | 1.54 |
| | 2 | 1.4 | 0.97 | 2.38 | 1.56 | 2.55 | 1.85 | 3.03 | 2.34 | 3.10 | 2.14 |
| | 3 | 2.23 | 1.56 | 3.28 | 2.44 | 3.49 | 2.83 | 4.07 | 3.72 | 4.86 | 3.57 |
| MEATP w/o shared information | 1 | 0.66 | 0.85 | 0.76 | 0.95 | 0.97 | 1.21 | 1.10 | 1.28 | 1.20 | 1.65 |
| | 2 | 0.94 | 1.19 | 1.10 | 1.45 | 1.44 | 2.00 | 1.61 | 2.18 | 1.65 | 2.25 |
| | 3 | 1.40 | 1.79 | 1.65 | 2.15 | 2.13 | 2.86 | 2.34 | 2.41 | 2.41 | 3.21 |
| MEATP w shared information | 1 | 0.52 | 0.62 | 0.52 | 0.53 | 0.54 | 0.57 | 0.57 | 0.58 | 0.57 | 0.58 |
| | 2 | 0.78 | 0.77 | 0.81 | 0.78 | 0.82 | 0.77 | 0.85 | 0.81 | 0.86 | 0.82 |
| | 3 | 1.05 | 0.98 | 1.09 | 0.96 | 1.10 | 0.98 | 1.15 | 1.04 | 1.18 | 1.05 |

Table 3: Distributions of prediction errors on highway dataset

| Models | Parameters | noise level 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| CS-LSTM | detected | 265 | 260 | 174 | 120 | 97 |
| | false alarm | 17 | 19 | 122 | 176 | 201 |
| | ADD (s) | 1.25 | 1.59 | 1.68 | 1.16 | 1.28 |
| | detection rate | 88.3% | 86.7% | 58.0% | 40.0% | 32.3% |
| MEATP w/o shared information | detected | 263 | 257 | 253 | 251 | 228 |
| | false alarm | 14 | 21 | 25 | 27 | 57 |
| | ADD (s) | 2.07 | 2.37 | 2.72 | 2.76 | 2.57 |
| | detection rate | 87.6% | 85.6% | 84.3% | 83.7% | 76.0% |
| MEATP w shared information | detected | 287 | 283 | 283 | 281 | 278 |
| | false alarm | 6 | 10 | 11 | 11 | 15 |
| | ADD (s) | 1.59 | 1.98 | 1.86 | 1.81 | 1.95 |
| | detection rate | 95.7% | 94.3% | 94.3% | 93.6% | 92.6% |

Table 4: Detection Results on SUMO highway dataset

| Models | Parameters | noise level 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| CS-LSTM | detected | 263 | 258 | 221 | 215 | 166 |
| | false alarm | 18 | 42 | 77 | 82 | 127 |
| | ADD (s) | 3.73 | 2.40 | 1.67 | 1.45 | 1.17 |
| | detection rate | 87.7% | 86.0% | 73.7% | 71.7% | 55.3% |
| MEATP w/o shared information | detected | 271 | 267 | 262 | 255 | 253 |
| | false alarm | 16 | 23 | 23 | 26 | 31 |
| | ADD (s) | 3.39 | 2.79 | 3.02 | 2.86 | 2.73 |
| | detection rate | 90.3% | 89.0% | 87.3% | 85.0% | 84.3% |
| MEATP w shared information | detected | 291 | 288 | 287 | 284 | 283 |
| | false alarm | 4 | 6 | 8 | 10 | 10 |
| | ADD (s) | 2.49 | 1.82 | 2.37 | 2.25 | 2.01 |
| | detection rate | 97.0% | 96.0% | 95.7% | 94.7% | 94.3% |

Table 5: Detection Results on SUMO urban dataset

in SUMO highway traffic dataset, are shown in Table 3 with all the data in unit meter. It can be seen that the prediction performance of MEATP with shared information is barely effected by the noises in inputs.

Based on the trained predictors and the probability distributions, we apply the MCusum algorithm to 300 vehicles that turns into abnormal driving mode in both highway and urban traffic datasets. Fig. 4 shows the statistic evolution given the prediction errors of an abnormal vehicle. Notably, the change point, which corresponds to the distribution change, is the point where driver switches from normal driving mode to abnormal driving mode. Overall Detection results are shown in Table 4 and Table 5. Notably, ADD represents average detection delay.

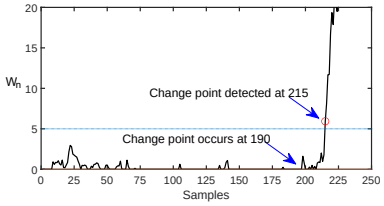


Figure 4: **Abnormal behavior statistics.** Threshold $b = 5$. Detection delay is 25 samples (2.5s) in this case.

By comparing the detection results, it shows that:

- Our proposed detection algorithm, equipped with well trained predictor, can detect the abnormal driving behaviors with $95.7\%$ detection rate, 1.59s detection delay in highway dataset, and $97.0\%$ detection rate, 2.49s detection delay in urban dataset.
- Information sharing among CAVs improves the detection rate by $10\%$ and detection delay by $25\%$.
- Equipped with MEATP with shared information, our detection algorithm is robust to the noises in observations of the surrounding autonomous vehicles.

## 5 Conclusion

This paper proposes a shared information based abnormal behavior detection algorithm. We first propose a multi-encoder attention trajectory prediction (MEATP) model. Based on the predictor, we further develop an abnormal behavior detection method. Through extensive experiments on public dataset and simulator, We show that (1) our proposed predictor outperforms the baselines; (2) our proposed algorithm detects abnormal behaviors with remarkable high accuracy and low detection delay; (3) shared information boosts the performance of prediction and detection. (4) Our proposed MEATP and detection algorithm based on it shows robustness against the noises in the shared information from surrounding autonomous vehicles.

## References


[Asvadi *et al.*, 2016] Alireza Asvadi, Pedro Girao, Paulo Peixoto, and Urbano Nunes. 3d object tracking using rgb and lidar data. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1255–1260. IEEE, 2016.

[Bimbraw, 2015] Keshav Bimbraw. Autonomous cars: Past, present and future a review of the developments in the last century, the present scenario and the expected future of autonomous vehicle technology. In *2015 12th international conference on informatics in control, automation and robotics (ICINCO)*, volume 1, pages 191–198. IEEE, 2015.

[Caltagirone *et al.*, 2019] Luca Caltagirone, Mauro Bellone, Lennart Svensson, and Mattias Wahde. Lidar–camera fusion for road detection using fully convolutional neural networks. *Robotics and Autonomous Systems*, 111:125–131, 2019.

[Chandra *et al.*, 2019] Rohan Chandra, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8483–8492, 2019.

[Cho *et al.*, 2019] Kyunghoon Cho, Timothy Ha, Gunmin Lee, and Songhwai Oh. Deep predictive autonomous driving using multi-agent joint trajectory prediction and traffic rules. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2076–2081. IEEE, 2019.

[Colyar and Halkias, 2007] James Colyar and John Halkias. Us highway i-80 dataset. *Highway Administration (FHWA), Tech. Rep*, pages 07–030, 2007.

[Committee and others, 2009] DSRC Committee et al. Dedicated short range communications (dsrc) message set dictionary. *SAE Standard J*, 2735:2015, 2009.

[Deo and Trivedi, 2018] Nachiket Deo and Mohan M Trivedi. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1468–1476, 2018.

[Fayyad *et al.*, 2020] Jamil Fayyad, Mohammad A Jaradat, Dominique Gruyer, and Homayoun Najjaran. Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors*, 20(15):4220, 2020.

[Han *et al.*, 2019] S. Han, J. Fu, and F. Miao. Exploiting beneficial information sharing among autonomous vehicles. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 2226–2232, 2019.

[Hu *et al.*, 2017] Jie Hu, Li Xu, Xin He, and Wuqiang Meng. Abnormal driving detection based on normalized driving behavior. *IEEE Transactions on Vehicular Technology*, 66(8):6645–6652, 2017.

[Hu *et al.*, 2019] Yaocong Hu, Mingqi Lu, and Xiaobo Lu. Driving behaviour recognition from still images by using multi-stream fusion cnn. *Machine Vision and Applications*, 30(5):851–865, 2019.

[Huang *et al.*, 2016] WuLing Huang, Kunfeng Wang, Yisheng Lv, and FengHua Zhu. Autonomous vehicles testing methods review. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 163–168. IEEE, 2016.

[Kenney, 2011] J. B. Kenney. Dedicated short-range communications (dsrc) standards in the united states. *Proceedings of the IEEE*, 99(7):1162–1182, July 2011.

[Krauß, 1998] Stefan Krauß. Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics. 1998.

[Lemley *et al.*, 2019] Joseph Lemley, Anuradha Kar, Alexandru Drimbarean, and Peter Corcoran. Convolutional neural network implementation for eye-gaze estimation on low-quality consumer imaging systems. *IEEE Transactions on Consumer Electronics*, 65(2):179–187, 2019.

[Li *et al.*, 2017] Zuojin Li, Shengbo Eben Li, Renjie Li, Bo Cheng, and Jinliang Shi. Online detection of driver fatigue using steering wheel angles for real driving conditions. *Sensors*, 17(3):495, 2017.

[Lopez *et al.*, 2018] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using sumo. In *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018.

[Lorden and others, 1971] Gary Lorden et al. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971.

[Qin *et al.*, 2019] Zengyi Qin, Jinglu Wang, and Yan Lu. Monogrnet: A geometric reasoning network for monocular 3d object localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8851–8858, 2019.

[Reddy *et al.*, 2017] Bhargava Reddy, Ye-Hoon Kim, Sojung Yun, Chanwon Seo, and Junik Jang. Real-time driver drowsiness detection for embedded system using model compression of deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 121–128, 2017.

[Shahverdy *et al.*, 2020] Mohammad Shahverdy, Mahmood Fathy, Reza Berangi, and Mohammad Sabokrou. Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications*, 149:113240, 2020.

[Tartakovsky and Polunchenko, 2008] Alexander G Tartakovsky and Aleksey S Polunchenko. Quickest changepoint detection in distributed multisensor systems under unknown parameters. In *2008 11th International Conference on Information Fusion*, pages 1–8. IEEE, 2008.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

[Veeravalli and Banerjee, 2014] Venugopal V Veeravalli and Taposh Banerjee. Quickest change detection. In *Academic Press Library in Signal Processing*, volume 3, pages 209–255. Elsevier, 2014.

[Veres *et al.*, 2011] Sandor M Veres, Levente Molnar, Nick K Lincoln, and Colin P Morice. Autonomous vehicle control systems—a review of decision making. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 225(2):155–195, 2011.

[Vincent *et al.*, 2010] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010.

[Yan *et al.*, 2016] Chao Yan, Frans Coenen, Yong Yue, Xiaosong Yang, and Bailing Zhang. Video-based classification of driving behavior using a hierarchical classification system with multiple features. *International Journal of Pattern Recognition and Artificial Intelligence*, 30(05):1650010, 2016.

[Yin *et al.*, 2021] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11784–11793, 2021.

[Zhenhai *et al.*, 2017] Gao Zhenhai, Le DinhDat, Hu Hongyu, Yu Ziwen, and Wu Xinyu. Driver drowsiness detection based on time series analysis of steering wheel angular velocity. In *2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pages 99–101. IEEE, 2017.