



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

JIANGWEI WANG, University of Connecticut, USA

LILI SU, Northeastern University, USA

SONGYANG HAN, University of Connecticut, USA

DONGJIN SONG, University of Connecticut, USA

FEI MIAO, University of Connecticut, USA

Hybrid traffic which involves both autonomous and human-driven vehicles would be the norm of the autonomous vehicles' practice for a while. On the one hand, unlike autonomous vehicles, human-driven vehicles could exhibit sudden abnormal behaviors such as unpredictably switching to dangerous driving modes – putting its neighboring vehicles under risks; such undesired mode switching could arise from numbers of human driver factors, including fatigue, drunkenness, distraction, aggressiveness, etc. On the other hand, modern vehicle-to-vehicle (V2V) communication technologies enable the autonomous vehicles to efficiently and reliably share the scarce run-time information with each other [21]. In this paper, we propose, to the best of our knowledge, the first efficient algorithm that can (1) significantly improve trajectory prediction by effectively fusing the run-time information shared by surrounding autonomous vehicles, and can (2) accurately and quickly detect abnormal human driving mode switches or abnormal driving behavior with formal assurance without hurting human drivers' privacy.

To validate our proposed algorithm, we first evaluate our proposed trajectory predictor on NGSIM and Argoverse datasets and show that our proposed predictor outperforms the baseline methods. Then through extensive experiments on SUMO simulator, we show that our proposed algorithm has great detection performance in both highway and urban traffic. The best performance achieves detection rate of 97.3%, average detection delay of 1.2s, and 0 false alarm.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

## 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 [11, 35, 74]. 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 the autonomous vehicles in the presence of human-driven vehicles is challenging. Human-driven vehicles could exhibit sudden abnormal behaviors such as unpredictably switching to dangerous

---

This work was partially supported by NSF Grants 1932250, 1952096, and 2047354 (CAREER)..

Authors' addresses: Jiangwei Wang, jiangwei.wang@uconn.edu, University of Connecticut, 371 Fairfield Way, Department of ECE, Storrs, CT, USA, 06268; Lili Su, lsu@northeastern.edu, Northeastern University, 100 Forsyth St, Department of ECE, Boston, MA, USA, 02115; Songyang Han, songyang.han@uconn.edu, University of Connecticut, 371 Fairfield Way, Department of CSE, Storrs, CT, USA, 06268; Dongjin Song, dongjin.song@uconn.edu, University of Connecticut, 371 Fairfield Way, Department of CSE, Storrs, CT, USA, 06268; Fei Miao, fei.miao@uconn.edu, University of Connecticut, 371 Fairfield Way, Department of CSE, Storrs, CT, USA, 06268.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, or post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2378-962X/2023/9-ART \$15.00

<https://doi.org/10.1145/3616398>

driving modes. These switches might arise from factors such as fatigue, drunkenness, distraction, and aggressiveness. If not detected in a timely manner, such unannounced switches could quickly put their neighboring vehicles under serious safety threats. To the best of our knowledge, most literature on abnormal driving behavior detection has been focusing on monitoring either behavioral parameters such as eye blinking and yawning [34, 43, 60, 65, 78], or vehicular parameters such as speed variability, steering wheel angle, and steering wheel grip force [45, 83], which require placing sensors on vehicle parts like steering wheel, accelerator or brake pedal; see Section 2.3 for details. Unfortunately, human-driven vehicles in the hybrid traffic might not have the required sensor placements to collect the relevant run-time measurements. Moreover, such measurements, even if available, are privacy sensitive and may not be shared with other vehicles. When autonomous vehicles does not have the direct measurement data from on-board sensors of the human-driven vehicles, and can only observe human-driven vehicles as part of the driving environment based on autonomous vehicle sensors, it is still challenging to detect abnormal driving behavior.

On the positive side of hybrid traffic, modern vehicle-to-vehicle (V2V) communication technologies enable the autonomous vehicles to efficiently and reliably share the scarce run-time information with each other [21]. Sharing such information can be highly beneficial: the U.S. Department of Transportation (DOT) has estimated that V2V communication can address up to 82% of all crashes in the United States involving unimpaired drivers, potentially saving thousands of lives and billions of dollars [36]. In addition, navigation and control strategies based on V2V shared information can also improve both traffic efficiency and safety [31].

**Contributions:** We propose, to the best of our knowledge, the first efficient algorithm that can (1) significantly improve trajectory prediction accuracy by effectively fusing the run-time information shared by surrounding autonomous vehicles, and can (2) accurately and quickly detect abnormal human driving mode switches with formal assurance without hurting human drivers' privacy. Our algorithm consists of two major components: trajectory prediction component and switch detection component.

- On trajectory prediction: We adapt the recently proposed transformer network to the applications of connected and autonomous vehicles (CAVs). Moreover, we propose multi-encoder attention mechanism to effectively using the shared information among CAVs. Our model is named as Multi-Encoder Attention based Trajectory Predictor (MEATP). We evaluate our proposed MEATP on NGSIM [20] and Argoverse [17] datasets, the experiment results show that MEATP with information sharing outperforms the well-adopted Long-Short Term Memory (LSTM) [32] and some transformer [71] based trajectory predictors. Moreover, information sharing improves the prediction performance of MEATP by 50%.
- On switch detection: We adapt the CuSum algorithm [56] and its variant to monitor the patterns of the run-time prediction errors observing that the prediction errors before and after abnormal driving behaviors can be well captured by two different probability distributions. The choice of the CuSum algorithm is motivated by its ease of implementation and strong provably optimality guarantees for formal assurance. In this work, we consider two popular scenarios: (1) CAVs have full knowledge of both the pre-change and post-change probability distributions and (2) CAVs have full knowledge of the pre-change distribution only and partial knowledge of the post-change distribution. Note that before the change happens, the driving mode of the human driver is deemed to be normal. Hence the corresponding prediction error distribution can be efficiently computed from existing datasets. Through experiments on SUMO-generated highway and urban traffic datasets, we show that: our proposed algorithm has great detection performance in both highway and urban traffic; it can be generalized to different scenarios where we have full or partial knowledge about post-change distributions; equipped with MEATP with shared information, it is robust to the noises in observations of the surrounding autonomous vehicles. The best performance achieves detection rate of 97.3%, average detection delay of 1.2s, and 0 false alarm.

## 2 RELATED WORK

### 2.1 Trajectory prediction and transformer neural network

Trajectory prediction is challenging because of the large amount of latent variables involved such as end goals, intentions at each moment, driving styles of human drivers and interactions with other vehicles. There has been extensive progress in vehicle trajectory prediction. Existing literature is divided into three categories: physics-based model, maneuver-based model and interaction-aware model [42].

Classical dynamic or kinetic model based methods have been designed to predict the future trajectories based on the current state of the vehicles [7, 8, 50, 68, 72]. However, they are unable to predict any change in the vehicle motion caused by a particular maneuver or external factors. Deep learning techniques have been widely utilized in maneuver-based model and interaction-aware model, yielding higher accuracy and longer prediction time [44, 63]. It has also been shown that the interaction-aware model outperforms in the three classes [42]. Therefore, deep learning-based interaction-aware trajectory prediction has become the main stream in recent years. Many methods have been proposed, including (but not limit to): LSTM combined with convolutional neural network (CNN) to predict the vehicle trajectories on US highways from NGSIM dataset [23], LSTM encoder-decoder to produce the probabilistic future location of the vehicles over occupancy grid map [57], LSTM combined with dynamic geometric graph to predict the vehicle trajectories and road agent behavior [15]. In addition, attempts have been made to integrate deep learning with model predictive control to reason about the future behavior of nearby vehicles [18].

A transformer network was originally designed for natural language processing [71], recently it has also been modified to process spatial-temporal data, for traffic flow prediction [77], pedestrian trajectory prediction [29] and vehicle trajectory prediction. An encoder-decoder architecture based on multi-head attention generates the distribution of the predicted trajectories for multiple vehicles in parallel [37]. Dong et al. present a dynamic graph attention network to deal with social interactions and predict multi-modal trajectories with probability [24]. Messaoud et al. apply multi-head attention by considering a joint representation of the static scene and surrounding agents to generate future trajectories [52]. There are other works applying attention mechanism to trajectory prediction [10, 47]. These works employ attention mechanism to detect vehicles that are more likely to affect the target vehicle's trajectory and pay more attention to them. The temporal dynamics of target and surrounding vehicles' trajectories, however, are not considered.

There are two types of predictions based on the design of the neural networks: (1) One autonomous vehicle (ego vehicle) predicts its surrounding vehicles' trajectories, thus all the surrounding vehicles are target vehicles. In this case, the input contain the past trajectories of the ego vehicle and the target vehicles, the outputs are the predicted trajectories of the target vehicles [37, 57]. This type of prediction, however, doesn't take the trajectories of the vehicles that are around the target vehicle into consideration. As these vehicles interact with the target vehicle, missing their trajectories may result in inaccurate prediction. (2) One autonomous vehicle predicts one target vehicle's trajectory [3, 15, 18, 23, 24, 52]. In this case, the input contain the past trajectories of the target vehicle and its neighboring vehicles. These works assume that the trajectories of target vehicle and all its neighboring vehicles are available. However this assumption is too strong in real world, one autonomous vehicle won't be able to get all vehicles' data around target vehicle for prediction. Therefore, in this work, we propose an multi-encoder attention based predictor that fuses the shared information among connected autonomous vehicles (CAVs) to predict the human-driven vehicles' trajectories.

### 2.2 Information sharing for connected vehicles

Connected autonomous vehicles (CAVs) have been proposed and studied for a long time. Information sharing of basic safety messages (BSMs) (velocity, position, heading angle, and yaw rate) is beneficial to autonomous vehicles' learning and control approaches in scenarios such as freeways, intersections, and lane-merging [31, 41, 55, 61].

Besides BSMs, environment information captured by vision sensors (such as cameras and Lidars) is also useful to improve autonomous vehicles' decision-making, trajectory planning, and perceptions [12, 38, 53]. GPS reports from neighboring vehicles can be used to identify driving hazards [79]. It has also been shown that the V2V communications among the autonomous vehicles can improve traffic safety, traffic flow stability and throughput [25, 67, 80].

There have been concerns about the privacy when sharing the data with other vehicles. V2V and V2I communications allow information to be transmitted between vehicles for safety reasons, but they also expose the vehicle's movements and geographical location to external networks, from which people can access to locate a vehicle driver [30]. This is a serious problem with location-based data, as human traces are unique, enabling an adversary to trace movements even with limited side information [27, 28]. Also, access to the interconnected AVs' wireless network enables public and private agencies to conduct remote surveillance of AV users, which can undermine individual autonomy through psychological manipulation and intimidation [30]. Schoonmaker [64] highlights the inadequacies of protecting location-based data based on customer consent, as customers accept the terms and conditions without fully understanding them. Surveys are also being conducted to investigate public attitude towards the privacy concerns brought by sharing the data, there are low confident on how the information should be exchanged between two vehicles [2, 76]. In the US, the new SPY Car Act gives NHTSA the authority to protect the use of (and access to) driving data in all vehicles manufactured for sale in the US [66]. All vehicles must provide owners or lessees the ability to stop the data collection, except for data essential for safety and post-incident investigations, and manufacturers are prohibited from using the collected data for marketing or advertising without consent from the owners or lessees [66]. Therefore, in this work, the hybrid traffic is consisted by human-driven vehicles that cannot communicate or don't willing to communicate, and the CAVs.

### 2.3 Abnormal human driver's behavior detection

Existing work on abnormal human driver's behavior detection is mostly restricted to monitoring driver's behavioral parameters or vehicular parameters. Specifically, there is a large body of literature that uses camera to extract the behavioral parameters such as eye closure ratio, eye blinking, head position, facial expressions, and yawning, to classify whether the driver is in fatigued driving or distracted driving mode [34, 65, 78]. A handful of work monitor vehicular parameters such as changing patterns, vehicle speed variability, steering wheel angle, and steering wheel grip force – assuming sensor placements on vehicle parts like steering wheel, accelerator or brake pedal [45, 83]. However, due to the privacy sensitive nature of the human drivers images and the detailed sensor measurements, none of these methods are applicable to our setting of hybrid traffic. In this work, we focus on a different challenge of detecting abnormal behavior based only on the observations of the CAVs.

## 3 PROBLEM DESCRIPTION

Unpredictable human driver abnormal behavior is a great obstacle to safe autonomy in hybrid traffic systems. The autonomous vehicles might also have problems during driving. There have been large numbers of work dealing with problems in perception [70], behavior planning [9, 75], and control of autonomous vehicles [40], these specific problems about autonomous vehicles are out of our scope in this paper. In this work, the challenge we focus on is: how to detect the human driver's abnormal behavior with high accuracy in a short time based on shared sensing information among CAVs, without violence of human driver's privacy. Normal driving behavior can be characterized by well-controlled speed, reasonable headways, mild accelerations, decelerations, and lane changes. In contrast, as human driver's abnormal behaviors are often caused by human factors such as drunkenness, aggressiveness, and fatigue, the resulting abnormal behaviors usually show up with unreasonable operations on vehicles, which provides useful clues for detecting them. First, regarding aggressive driving, the drivers' abnormal behavior can be reflected by observable signs from the perspective of autonomous vehicles, such as accelerating suddenly, violating the speed limit, and making frequent lane changes to surpass the other

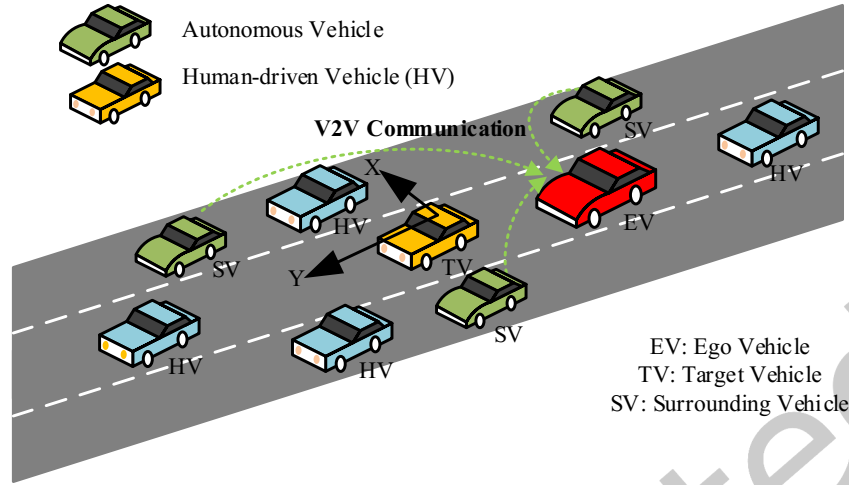


Fig. 1. Hybrid traffic with information sharing.

vehicles [6]. Second, considering the drunk and fatigued driving, when the driver is intoxicated by alcohol, or affected by tiredness, he or she is more likely to issue a sudden acceleration or deceleration due to a response delay [22, 33], therefore the velocity control can be poor. From this perspective, when we are trying to predict the trajectory of a human-driven vehicle, the distributions of the prediction errors of normal driving mode and abnormal driving mode could be quite different.

To address the above challenge, in this work, we propose an algorithm that is composed by trajectory predictor component and driving mode switch detection component to detect the abnormal driving behaviors. Our proposed algorithm takes advantages of information sharing among CAVs to boost both prediction and detection performance. In this section, we first describe the hybrid traffic system including autonomous vehicles and human-driven vehicles, then illustrate the details of information sharing that will be utilized in both trajectory prediction and abnormal behavior detection.

### 3.1 Hybrid traffic system description

The traffic is comprised by autonomous vehicles that are connected via V2V communication and human-driven vehicles. For ease of exposition, we refer to the autonomous vehicle that is doing prediction as the ego vehicle (EV), other autonomous vehicles that are within the communication range of the ego vehicle as surrounding vehicles (SV), and the human-driven vehicles is being predicted and as target vehicle (TV). Without loss of generality, we focus on one TV. It is easy to see that our algorithm works for the general multiple target vehicles setting under which we can run the algorithm in parallel for different target vehicles. Notably, the human-driven vehicles are not communicating with others. The system is illustrated in Fig. 1.

We define a stationary reference frame based on the TV. At time step  $t_0$ , the origin are set at TV's current location,  $X$  - axis is lateral direction, while the  $Y$  - axis is longitudinal direction. In the range of  $y = -30m$  to  $y = 30m$ , there are in total  $N$  neighboring vehicles that interact with the TV, including  $N_A$  autonomous vehicles and  $N_H$  human-driven vehicles. The range of +30 meters and -30 meters is proposed by Deo et al. [23] and still widely used in recent works [46, 52]. This setting has been tested and validated on Highway and Urban datasets including NGSIM and nuScenes.

### 3.2 Information shared to the ego vehicle

All surrounding vehicles' observations include point cloud data from Lidar and camera image. With the Lidar-camera fusion techniques [13], surrounding vehicles can get the locations of the nearby human-driven vehicles. At time  $t_0$ , For each surrounding vehicle, the information it shares with EV include:

- (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.

Upon receiving the shared information, EV transfers the GPS locations into the coordinates in current reference frame: for  $i \in [0, N]$  vehicles, coordinates in the past  $t_h + 1$  time steps:  $x_i(t_0 - t_h), y_i(t_0 - t_h), \dots, x_i(t_0), y_i(t_0)$ .

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 [18, 26, 81]. 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 4.3.4 and Section 5 respectively.

It's worth noting that, information sharing among CAVs has more advantages at gathering neighboring vehicles' data compared to existing prediction methods. Multi-agent prediction, usually with multi encoders have been widely used in the existing works [3, 15, 18, 23, 24, 52]. Each encoder corresponds to one vehicle's data around the target vehicle. However, these existing works assume that the ego vehicle can get all the vehicles' trajectories around the target vehicles, which can be a too strong assumption in some scenarios. When the target vehicle is driving in front of or behind the ego vehicle, shown in Fig. 1, the target vehicle will certainly block part of the ego vehicle's Lidar signal and camera view, therefore the ego vehicle won't be able to get the trajectories of the vehicles that are in front of or behind the target vehicle. The information, however, is vital to predict the target vehicle's trajectory, since the neighboring vehicles' movement can affect the trajectory of the target vehicle. For instance, if the vehicle in front slows down, the target vehicle will also need to slow down to keep a safe distance. In this work, with the shared information among CAVs, other autonomous vehicles can extract, compute, and send the information to the ego vehicle. To the best of our knowledge, we are the first to propose a multi-encoder attention based trajectory predictor that fuses the shared information among the CAVs, and utilize the prediction for our proposed abnormal driving behavior detection Algorithm 1.

## 4 ABNORMAL BEHAVIOR DETECTION FRAMEWORK

To make the discussion concrete, we present our framework in Section 4.2 – right after the presentation of our proposed trajectory prediction method (Section 4.1). Though our framework is stated w. r. t. the trajectory predictor proposed in Section 4.1 only, it works for more general predictors. Our framework is also generic in its abnormal behavior detection component. In Section 4.3, we present three different concrete instantiations to compute the crucial statistic used in the framework.

### 4.1 Multi-encoder attention based trajectory predictor (MEATP)

We propose a trajectory predictor, named as *multi-encoder attention based trajectory predictor (MEATP)* to fuse the real-time measurement data and utilize the benefit of shared sensing information among CAVs. We specify its inputs and outputs in Section 4.1.1, introduce its network architecture in Section 4.1.2, and present the loss function used in its training in Section 4.1.3.

Transformer based network has been applied in vehicle trajectory prediction[10, 24, 37, 47, 52]. They use attention mechanism to distinguish the vehicles that are more likely to affect the target vehicle's trajectory and give them more attention. In this work, our proposed multi-encoder attention layer, not only finds the more

influential vehicles by assigning the corresponding encoder larger weights, but also finds the more important time steps that affects the trajectory of the target vehicles using the multi-head attention layer in the multi-encoder attention. Therefore, our proposed model can find the important features in both spatial and temporal dimensions, whereas the existing works' attention mechanism only focus on spatial dimension.

**4.1.1 Inputs and outputs.** The inputs to the MEATP include (I.1) the trajectories of neighboring vehicles, denoted by  $C_i(t_0)$  for  $i \in [1, N]$ , and (I.2) the trajectories of the target vehicles, denoted by  $S(t_0)$ . Both (I.1) and (I.2) are over a sliding time window  $\{t_0 - t_h, t_0 - t_h + 1, \dots, t_0\}$ , formally,

$$C_i(t_0) \triangleq [c_i(t_0 - t_h), \dots, c_i(t_0 - t_h + \ell), \dots, c_i(t_0)] \quad (1)$$

with  $c_i(t_0 - t_h + \ell) \triangleq [x_i(t_0 - t_h + \ell), y_i(t_0 - t_h + \ell)]$  being the 2-dimensional position of the  $i$ -th neighboring vehicle at time slot  $t_0 - t_h + \ell$  for  $\ell = 0, \dots, t_h$ , and

$$S(t_0) \triangleq [s(t_0 - t_h), \dots, s(t_0 - t_h + \ell), \dots, s(t_0)] \quad (2)$$

with  $s(t_0 - t_h + \ell) \triangleq [x_0(t_0 - t_h + \ell), y_0(t_0 - t_h + \ell)]$  being the 2-dimensional vector that records the TV's position at time slot  $t_0 - t_h + \ell$  for  $\ell = 0, \dots, t_h$ .

The outputs of the MEATP are the distributions of the future trajectory of the TV over time window  $\{t_0 + 1, \dots, t_0 + t_f\}$ . Assume the predicted future trajectory follows bivariate Gaussian distribution [14, 23], notably, this assumption has been widely adopted in the literature [14, 15, 23, 37, 52], tested in numbers of real-world trace datasets, including NGSIM Lyft Level 5, Argoverse Motion Forecasting, and the Apolloscape Trajectory [16, 19, 51]. the output of the decoder is the bivariate Gaussian parameters at every time step in the future  $t_f$  time steps:

$$\Omega \triangleq [\Omega(t_0 + 1), \dots, \Omega(t_0 + t_f)], \quad (3)$$

where

$$\Omega(t) \triangleq [\mu(t), \sigma(t), \rho(t)] \triangleq [\mu_x(t), \mu_y(t), \sigma_x(t), \sigma_y(t), \rho(t)], \text{ for } t \in [t_0 + 1, t_0 + t_f],$$

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

**4.1.2 Network architectures.** Our network architecture is illustrated in Fig. 2. To effectively utilize the shared information, instead of the single-encoder single-decoder transformer architecture, we use a multi-encoder single-decoder architecture. Having multiple encoders enables the EV to learn the spatial and temporal features from the trajectories of the TV and of neighboring vehicles. As can be seen from Table 4 of our experimental results, compared with CS-LSTM, MEATP can more effectively extract the relevant information contained in the shared trajectories, hence can achieve smaller mean and variance of the prediction errors.

Our proposed architecture contains  $N + 1$  encoders. For ease of exposition, we index these encoders from 0 to  $N$ . Encoder 0 takes  $S(t_0)$  – the target vehicle's trajectory over time window  $\{t_0 - t_h, \dots, t_0\}$  – as its input. Encoder  $i$  (for  $i = 1, \dots, N$ ) takes  $C_i(t_0)$  – the  $i$ -th vehicle's trajectory over time window  $\{t_0 - t_h, \dots, t_0\}$  – as its input. The inputs plus the positional encoding are then sent to the next layer as queries  $Q$ , keys  $K$ , and values  $V$ . Positional encoding is just an embedding mechanism to ensure the order of the input. Following the seminal work [71], each of the encoder consists of two sub-layers. The first is a multi-head attention mechanism, and the second is a simple, positionwise fully connected feed-forward network. We use residual connection around each of the two sub-layers, followed by layer normalization. Multi-head attention layer is composed by  $h$  heads of scaled dot-product attentions, each of which is expressed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (4)$$

where  $d_k$  is the dimension of keys. The scaling factor  $\frac{1}{\sqrt{d_k}}$  is used to prevent the dot product from being numerically too large. Scaled dot-product attention allows the model to pay more attention to one time step in the input that is useful for the prediction. Multi-head attention executes the  $h$  scaled dot product attention heads in parallel.

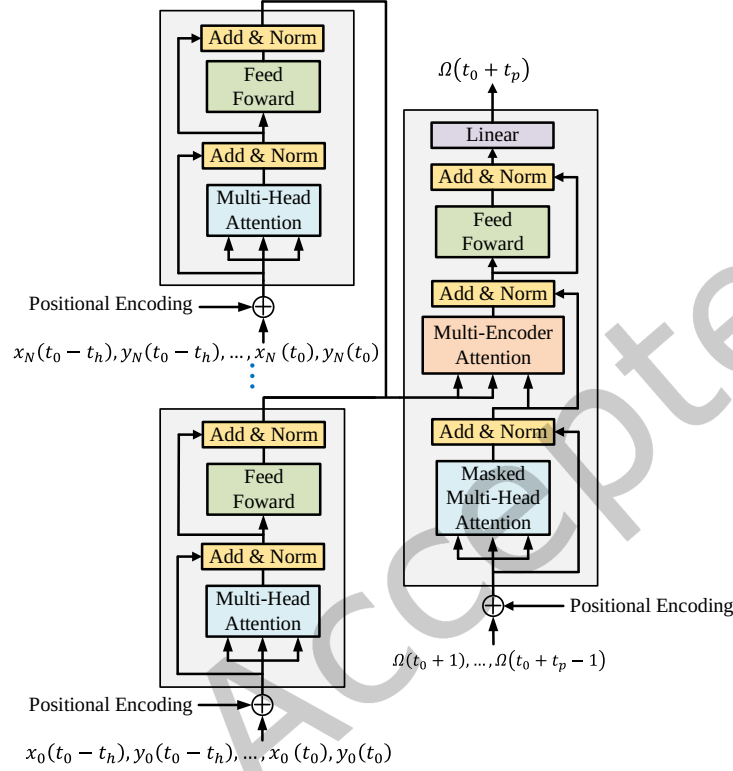


Fig. 2. MEATP network architecture. The Multi-Encoder Attention Mechanism in decoder is shown in Fig. 3

Attention outputs are concatenated and linearly transformed into the same dimension of  $Q$ . Therefore:

$$\text{Multihead}(Q, K, V) = \text{concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (5)$$

where  $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ . With the multi-head attention layer, multiple (instead of just one) time steps of historical data are processed simultaneously. Each encoder  $i$  passes its value  $V_i$  and key  $K_i$  to the multi-encoder attention layer of decoder.

Different from the encoders, at time step  $t_p$ , the decoder takes the outputs in the past  $t_p - 1$  steps as inputs, and outputs the bivariate Gaussian parameters, which represent the probability distribution of the target vehicle's future coordinates. To be noticed that, the key part in the decoder is the multi-encoder attention. As shown in Fig. 3, the multi-encoder attention in the decoder is composed by  $N+1$  multi-head attention. Query  $Q_0$ , which comes from 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 dimension of  $Q_0$ . Multi-encoder attention are expressed as:

$$\text{Multiencoder}(Q_0, K, V) = \text{concat}(M_0, \dots, M_N)W^M \quad (6)$$



where  $M_i = \text{Multihead}(Q_0, K_i, V_i)$ . The idea of multi-encoder attention comes from the road traffic interaction. Since the target vehicle 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 target vehicle, interact with  $K_i, V_i$ , the decoder learns both temporal and spatial information of the neighboring vehicles and target vehicle itself.

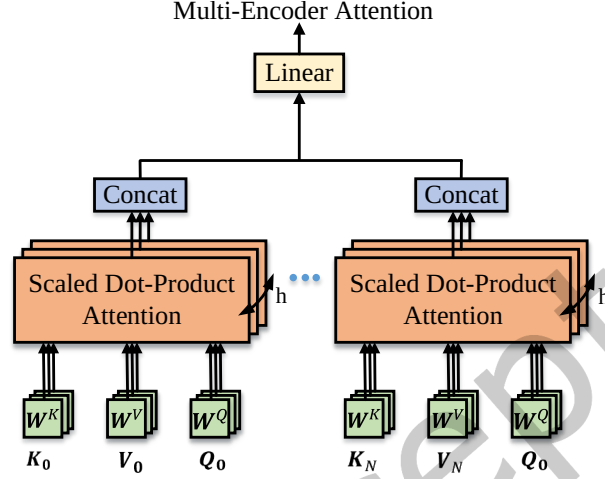


Fig. 3. Multi-encoder attention in the decoder

**4.1.3 Loss function.** Since we assume the output at each future time step follows bivariate Gaussian distribution, we train our neural network base on the weighted sum of two losses below:

$$\begin{aligned}
 L_1 &= \sum_{t=t_0+1}^{t_0+t_f} (-\log P(Z(t)|\mu(t), \sigma(t), \rho(t))), \\
 L_2 &= \sum_{t=t_0+1}^{t_0+t_f} \|Z(t) - \mu(t)\|,
 \end{aligned} \tag{7}$$

where  $Z(t) = [x_0(t), y_0(t)]$  is the true coordinates of target vehicle at future time step  $t$ .  $L_1$  is the sum (in future  $t_f$  time steps) of negative log likelihood of true trajectory given the predicted trajectory distribution.  $L_2$  is the  $l_2$  norm between true trajectory and predicted trajectory in future  $t_f$  time steps. Note that here we choose the mean on  $x$  and  $y$  axis  $\mu(t) = [\mu_x(t), \mu_y(t)]$  in output distribution as target vehicle's predicted future trajectory. By minimizing the loss function, we make the predicted trajectory stay close to the future true trajectory. To be noticed that, these two loss functions have been widely used in the exiting literature [14, 15, 23, 37, 52], it is not specifically for transformer networks, all the existing literature uses the true future trajectory in their training phase. In this work, in order to make our abnormal behavior detection algorithm work better, we want the distance between the predicted and true trajectory to be as small as possible. Therefore, the loss function  $L_2$  are assigned with larger weight. If we denote our proposed predictor as a function  $f_W$ , with  $W$  being the weight parameters, including  $W_Q, W_K, W_V$  in scaled dot-product attention layers,  $W_O$  in Multihead attention layers,  $W_M$  in Multiencoder attention layers, and the weights in the feed forward layers. we have  $[\Omega(t_0+1), \dots, \Omega(t_0+t_f)] = f_W(S(t_0), C_1(t_0), \dots, C_N(t_0))$ . Therefore, by minimizing the loss functions in training phase, we are actually doing gradient decent on the weight parameters  $W$  of our proposed model.

#### 4.2 Shared information based abnormal behavior detection algorithm

Recall that the EV is the one that executes the MEATP and monitors whether the TV is in an abnormal driving mode or not. Figuring out which statistic to monitor is highly non-trivial as the driving mode of the TV can switch to an abnormal driving mode at any *unknown* time, and the underlying prediction error distributions prior and posterior to the switch time are different. To make the discussion concrete, we use  $\gamma$  to denote this unknown switch time. 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  $S(t_0)$  and  $\{C_i(t_0)\}_{i=1}^N$ , and then passes them as inputs to the MEATP to obtain  $\mu(t_0 + 1)$ . Finally, when the true  $Z(t_0 + 1)$  is revealed, the EV computes its prediction error

$$e_{t_0+1} = ||Z(t_0 + 1) - \mu(t_0 + 1)||, \quad (8)$$

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

We use  $f$  and  $g$  to denote the distribution of  $e_n$  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  $e_n \stackrel{i.i.d.}{\sim} f$  for  $n = 1, 2, \dots, t_0$ . If  $\gamma \leq t_0$ , i.e., the TV is in the abnormal mode, then  $e_n \stackrel{i.i.d.}{\sim} g$  for  $n = t_0, \dots$ . 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  $\{e_n, n = 1, 2, \dots\}$ .** The “distance” between  $f$  and  $g$  can be captured by the Kullback-Leibler divergence  $D_{KL}(f, g)$ , defined as

$$0 < D_{KL}(f, g) \triangleq \int f(x) \log \left( \frac{f(x)}{g(x)} \right) dx < \infty. \quad (9)$$

**Definition 1** (Detection algorithm as a stopping time). A *stopping time* w.r.t. the sequence of random prediction errors  $\{e_n, n = 1, 2, \dots\}$  is a random variable  $\tau$  with the property that for each  $n$ , the event  $\{\tau = n\}$  is measurable w.r.t.  $\sigma(e_1, \dots, e_n)$  – the  $\sigma$ -algebra generated by  $e_1, \dots, e_n$ . Formally,  $\{\tau = n\} \in \sigma(e_1, \dots, e_n)$  for each  $n$ . A detection algorithm is a stopping time of random prediction errors that declare the detection of a change [73].

**Remark 1.** A detection algorithm should be a stopping time otherwise it is not implementable due to lack of information, i.e.,  $\{\tau = n\} \notin \sigma(e_1, \dots, e_n)$  for some  $n$ . When used as a detection algorithm, the event  $\tau = n$  is interpreted as “a distribution change is declared at time  $n$ ”.

Given the random sequence prediction error  $\{e_n, n = 1, 2, \dots\}$ , it's hard to tell when the distribution has changed by manual inspection, instead, we compute a statistics  $W_n$  given  $\{e_n, n = 1, 2, \dots\}$ , which will be specified in Section 4.3. Once the statistic  $W_n$  exceeds some threshold  $b$ , we declare the change in distributions, i.e., abnormal driving behavior happened. We propose Algorithm 1 to illustrate how EV detects whether one TV has turned into abnormal driving mode. Notably, this algorithm can be ran in parallel to detect multiple TVs. Based on the shared information, the EV equipped with trained predictor computes the prediction errors at each time step, the statistic  $W_n$  is then updated accordingly. Once the statistic  $W_n$  exceeds the threshold, the EV declares the detection of abnormal behavior. We consider different scenarios regarding the amount of knowledge known on probability distributions  $f$  and  $g$ : (1) We have full knowledge of the pre-change and post-change distributions of  $f$  and  $g$ , (2) we have full knowledge of pre-change distribution  $f$ , but we only have partial knowledge about the post-change distribution  $g$ . Assuming full knowledge on pre-change distribution is standard in the literature on quickest change detection [73]. Beside, this assumption can be easily satisfied in our applications: before the change happens, the driving mode of the human driver is deemed to be normal. Hence the corresponding prediction error distribution can be efficiently computed from existing datasets. Based on whether we have full

knowledge about the post-change distributions or not and, if not, how much we know about it, the statistic  $W_n$  can be calculated in CuSum algorithm [56], MCuSum algorithm [5, 69], or generalized likelihood ratio test (GLRT) [49] approach. We elaborate the details of those algorithms and how to update statistic  $W_n$  in Section 4.3.

---

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

---

```

Initialize  $W_0 \leftarrow 0$ ,  $t_0 \leftarrow 0$ ,  $\mu(t_0) \leftarrow Z(t_0)$ ;
Choose CuSum/ MCuSum/ GLRT algorithm based on the knowledge of the post-change distribution;
if Using MCuSum then
    Set threshold  $b \leftarrow \log(\frac{M}{\alpha})$ ; /*  $\alpha$  is the given false alarm budget and  $M$  is the number of
    possible  $g$  */
else
    Set threshold  $b \leftarrow |\log \alpha|$ ;
end
while true do
    Receive information shared by the SVs;
    Update  $S(t_0)$  and  $\{C_i(t_0)\}_{i=1}^N$  by incorporating this newly received information;
     $\mu(t_0 + 1) \leftarrow f_W(S(t_0), \{C_i(t_0)\}_{i=1}^N)$ ; /* Compute  $\mu(t_0 + 1)$  by calling MEATP */
     $e_{t_0} \leftarrow \|Z(t_0) - \mu(t_0)\|$ ;
    Compute statistic  $W_{t_0}$ ;
    if  $W_{t_0} \geq b$  then
        Declare the detection of abnormal behavior;
        Break;
    end
     $t_0 \leftarrow t_0 + 1$ ;
end
    
```

---

**Minimax quickest change detection (QCD):** Next we state the performance metrics of the detection algorithms. Observing that in practice the prior knowledge on the switch timing of the driving modes is barely available, we do not impose any distributional assumption of  $\gamma$ . Instead, we allow  $\gamma$  to be an arbitrary and unknown value which can even vary across executions. In other words, we consider minimax QCD. The measure of the false alarm is the false alarm rate in the literature [73]:

$$\text{FAR}(\tau) = \frac{1}{E_\infty(\tau)}, \quad (10)$$

where  $E_\infty$  stands for the expectation measure that the change happens at time  $\infty$ , i.e., the change doesn't occur. To provide strong safety guarantee, we adopt the Lorden's minimax formulation [49] in which the detection delay is measured via

$$\text{WADD}(\tau) = \sup_{n \geq 1} \text{ess sup } E_n[(\tau - n)^+ | \mathbf{e}_1, \dots, \mathbf{e}_n], \quad (11)$$

where  $(\cdot)^+ = \max\{0, \cdot\}$ ,  $E_n[\cdot]$  is the expectation operator when the change occurs at time  $n$ , and  $\text{ess sup}(\cdot)$  of a scalar-valued random variable is the sup of its support.

### 4.3 Algorithms to calculate $W_n$ for abnormal behavior detection

In this section, we illustrate the abnormal behavior detection component. Given whether we have full knowledge about the post-change distributions or not, we update the statistic  $W_n$  based on CuSum algorithm, MCuSum algorithm and generalized likelihood ratio test (GLRT) algorithm. We provide details of adopting these algorithms in our problem.

**4.3.1 CuSum algorithm with full knowledge of pre-change and post-change distributions.** We first consider the best case, where we have full knowledge of the probability distributions  $f$  and  $g$ . 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 adapt the CuSum algorithm to detect the abnormal human driver behaviors.

**CuSum's algorithm** [56]: For some given  $b > 0$  (chosen later in Proposition 1),

$$\tau \triangleq \inf \{n \geq 1 : W_n \geq b\}, \quad (12)$$

where

$$W_n = \begin{cases} \max\{W_{n-1} + \log L(\mathbf{e}_n), 0\}, & \text{for } n \geq 1; \\ 0, & \text{for } n = 0. \end{cases}$$

and  $L(\mathbf{e}_n) = g_\theta(\mathbf{e}_n)/f_\phi(\mathbf{e}_n)$  is the likelihood ratio.

The intuitions behind the CuSum algorithm are as follows: Upon and after the change point  $\gamma$ , the prediction error  $\mathbf{e}_n$  follows the distribution  $g_\theta$ , and  $\mathbb{E}_{n \geq \gamma} [\log (g_\theta(\mathbf{e}_n)/f_\phi(\mathbf{e}_n))] = D_{\text{KL}}(g_\theta, f_\phi) > 0$ . Similarly, before  $\gamma$ , the prediction error  $\mathbf{e}_n$  follows the distribution  $f$ , and  $\mathbb{E}_{n < \gamma} [\log (g_\theta(\mathbf{e}_n)/f_\phi(\mathbf{e}_n))] = -D_{\text{KL}}(f_\phi, g_\theta) < 0$ . Thus, with more and more observations on the prediction errors, we expect  $\sum_{n:n \geq \gamma} \log (g_\theta(\mathbf{e}_n)/f_\phi(\mathbf{e}_n))$  to bypass the given threshold  $b$ . In a sense, the  $\max\{\cdot, 0\}$  operation in  $W_n$  has two effects: (1) it makes a tentative guess on  $\gamma$  and (2) it does the first level protection against the random fluctuation in  $\log (g_\theta(\mathbf{e}_n)/f_\phi(\mathbf{e}_n))$ . A more fine-grained protection against the random fluctuation is controlled by the choice of  $b$ ; the larger the  $b$ , the more accurate the detection yet the longer delay.

Clearly, there is a trade-off between the false alarm rate  $\text{FAR}(\tau)$  and the detection delay; if we can tolerate arbitrary  $\text{FAR}(\tau)$ , we can achieve 0 detection delay by trivially set  $\tau = 0$ . It has been proved in the literature that the CuSum algorithm is optimal [54, 62].

**Proposition 1.** [73] *The CuSum algorithm with  $b = |\log \alpha|$  for any given  $\alpha \in (0, 1)$  is first order asymptotically optimal. Furthermore, the false alarm rate and the average detection of the CuSum algorithm are bounded as follows:*

$$\text{FAR}(\tau) \leq \alpha, \text{ and } \text{WADD}(\tau) = O\left(\frac{|\log \alpha|}{D_{\text{KL}}(g, f)}\right).$$

**4.3.2 MCuSum with unknown parameters.** Recall that we assume that  $f_\phi$  and  $g_\theta$  are the gaussian distributions. Denote the mean and variance of  $f_\phi$  and  $g_\theta$  as  $(\mu_0, \sigma_0)$  and  $(\mu_1, \sigma_1)$ , respectively. Thus, we have:

$$\log L(\mathbf{e}_n) = \frac{(\mathbf{e}_n - \mu_0)^2}{2\sigma_0^2} - \frac{(\mathbf{e}_n - \mu_1)^2}{2\sigma_1^2} + \log\left(\frac{\sigma_0}{\sigma_1}\right) \quad (13)$$

In CuSum's algorithm, we assume that we have full knowledge about the pre-change distribution parameters  $\mu_0, \sigma_0$  and post-change distribution parameters  $\mu_1, \sigma_1$ . However, when the human driver switches to abnormal driving mode, it's more realistic that we cannot know the exact post-change prediction error distributions. But instead, we have some prior knowledge about the post-change distribution and estimate the parameters based on

our prior knowledge. In this section, we assume that we have full knowledge of the pre-change distribution  $f$ , the post-change distribution  $g$  with parameter  $\theta$ :

$$\theta \in \Theta = \{\theta_1, \theta_2, \dots, \theta_M\} \quad (14)$$

Therefore, the unknown post-change distribution belongs to a finite set of distributions. In this scenario, We adopt the MCuSum algorithm to detect the abnormal driver's behavior:

**MCuSum algorithm** [69]:

$$\tau_{MC} \triangleq \inf \{n \geq 1 : \max_{j \in \{1, \dots, M\}} W_n(\theta_j) \geq b\}, \quad (15)$$

where

$$W_n(\theta_j) = \begin{cases} \left[ W_{n-1}(\theta_j) + \log\left(\frac{g_{\theta_j}(\mathbf{e}_n)}{f_{\phi}(\mathbf{e}_n)}\right) \right]^+, & n \geq 1; \\ 0, & n = 0. \end{cases} \quad (16)$$

In Eq. (15),  $[x]^+ \triangleq \max\{x, 0\}$ , and, denoting  $\phi = (\mu_0, \sigma_0)$  and  $\theta_j = (\mu_j, \sigma_j)$ , the log likelihood can be written as

$$\log\left(\frac{g_{\theta_j}(\mathbf{e}_n)}{f_{\phi}(\mathbf{e}_n)}\right) = \frac{(\mathbf{e}_n - \mu_0)^2}{2\sigma_0^2} - \frac{(\mathbf{e}_n - \mu_j)^2}{2\sigma_j^2} + \log\left(\frac{\sigma_0}{\sigma_j}\right). \quad (17)$$

Thus, to detect a change when the post-change parameter is unknown,  $M$  CuSum algorithms are executed in parallel, one for each post-change parameter. A change is declared the first time a change is detected in any one of the CuSum algorithms. In our context, in order to detect the abnormal human driver behavior, we choose several possible means and standard deviations based on the prior information we have about the prediction errors after the abnormal behavior happens. Then we compute the CuSum algorithm in parallel based on each pair of the possible mean and standard deviation, we declare the detection of the abnormal behavior once the statistic evolves above the threshold in any one of the CuSum algorithm. The asymptotic optimality of the MCuSum algorithm is proved in [69].

**Proposition 2.** [69] *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 and the average detection of the MCuSum algorithm are bounded as follows:*

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

**4.3.3 Generalized likelihood ratio test with unknown parameters.** In this subsection, we further relax our requirements on the post-change distributions. We assume that we know the values  $\mu_0, \sigma_0$ , and we only have partial prior information about the distribution  $\mu_1, \sigma_1$ :

$$\mu_1 - \mu_0 \geq v_m, \sigma_1 - \sigma_0 \geq \delta_m$$

Then we can no longer use the recursive form to update  $W_n$  since we don't know  $\mu_1, \sigma_1$ . We adopt the generalized likelihood ratio test based approach, we see the sum of log likelihood from  $k$  to  $n$  as a function of unknown parameter  $\theta$ ,

$$S_k^n(\theta) = \sum_{i=k}^n \log L(e_i) = \sum_{i=k}^n \log\left(\frac{g_\theta(e_i)}{f_\phi(e_i)}\right) \quad (18)$$

then we have double maximization:

$$W_n = \max_{1 \leq k \leq n} \sup_{\theta} S_k^n(\theta) \quad (19)$$

Therefore, we can adopt the generalized likelihood ratio test (GLRT) to test the abnormal driving behavior:

**GLRT algorithm** [49]:

$$\tau \triangleq \inf \{n \geq 1 : W_n \geq b\}, \quad (20)$$

where

$$W_n = \max_{1 \leq k \leq n} \sup_{\theta} S_k^n(\theta)$$

Next we elaborate how to deal with the double maximization in statistic  $W_n$ . since we know the minimum magnitude change of the parameter  $\theta_1$ , we have:

$$W_n = \max_{1 \leq k \leq n} \sup_{\theta: \mu_1 - \mu_0 \geq v_m > 0, \sigma_1 - \sigma_0 \geq \delta_m > 0} S_k^n(\theta) \quad (21)$$

since

$$\frac{(e_n - \mu_0)^2}{2\sigma_0^2} - \frac{(e_n - \mu_1)^2}{2\sigma_1^2} + \log\left(\frac{\sigma_0}{\sigma_1}\right) > \frac{(e_n - \mu_0)^2}{2\sigma_0^2} - \frac{(e_n - \mu_1)^2}{2\sigma_0^2} + \log\left(\frac{\sigma_0}{\sigma_1}\right)$$

we can use the cumulative sum:

$$S_k^n = \frac{\mu_1 - \mu_0}{\sigma_0^2} \sum_{i=k}^n (e_i - \frac{\mu_1 + \mu_0}{2} + \log(\frac{\sigma_0}{\sigma_1})) \quad (22)$$

let us introduce  $v = \mu_1 - \mu_0$ ,  $\delta = \sigma_1 - \sigma_0$ . then  $W_n$  can be written as :

$$W_n = \max_{1 \leq k \leq n} \sup_{\theta: v \geq v_m > 0, \delta \geq \delta_m > 0} \sum_{i=k}^n \left[ \frac{v(e_i - \mu_0)}{\sigma_0^2} - \frac{v^2}{2\sigma_0^2} + \log\left(\frac{\sigma_0}{\sigma_0 + \delta}\right) \right] \quad (23)$$

in such case, the constrained maximization over  $\theta_1$  is:

$$W_n = \max_{1 \leq k \leq n} \sum_{i=k}^n \left[ \frac{\hat{v}_k(e_i - \mu_0)}{\sigma_0^2} - \frac{\hat{v}_k^2}{2\sigma_0^2} + \log\left(\frac{\sigma_0}{\sigma_0 + \hat{\delta}}\right) \right] \quad (24)$$

where:

$$\hat{v}_k = \left( \frac{\sum_{i=k}^n |e_i - \mu_0|}{n - k + 1} - v_m \right)^+ + v_m, \hat{\delta} = \delta_m \quad (25)$$

**4.3.4 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. However, the measurement noises have no impacts on  $\gamma$ . Our algorithm can still detect the change in the distributions even with the noises. For instance, considering MCuSum algorithm, with the noises in the inputs to MEATP, the pre-change distribution of the prediction error changes from  $f_\phi$  to  $f_{\phi'}$ , where  $\phi' = (\mu'_0, \sigma'_0)$ , similarly, post-change distribution changes from  $g_{\theta_j}$  to  $g_{\theta'_j}$ , where  $\theta'_j = (\mu'_j, \sigma'_j)$ . For each  $\theta'_j$ , the log likelihood becomes:

$$\log \left( \frac{g_{\theta'_j}(e_n)}{f_{\phi'}(e_n)} \right) = \frac{(e_n - \mu'_0)^2}{2(\sigma'_0)^2} - \frac{(e_n - \mu'_j)^2}{2(\sigma'_j)^2} + \log\left(\frac{\sigma'_0}{\sigma'_j}\right). \quad (26)$$

## 5 EXPERIMENTS

### 5.1 Trajectory prediction

**5.1.1 Experiment details.** In this section, we use NGSIM [20] and Argoverse 1 motion forecasting[17] datasets for evaluation. NGSIM dataset consists of trajectories of freeway (US-101 and I-80) traffic sampled at frequency 10Hz over 45 minutes. Argoverse dataset is a curated collection of 324,557 scenarios with each 5 seconds long. Each scenario contains the 2D, birds-eye-view centroid of each tracked object sampled at 10 Hz. We train our model using Adam optimizer with learning rate of 0.01. 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. For NGSIM dataset, We use 3s of historical trajectories to predict the trajectories in future 5s with sample frequency of 5Hz. For Argoverse dataset, the length of historical trajectory is 2s and the length of predicted trajectory is 3s. 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. For trajectory prediction, we use the standard metrics followed by prior trajectory prediction approaches[15, 23, 37]:

- Root mean square error (RMSE) of the predicted trajectories with respect to the true future trajectories.
- Average displacement error (ADE): The average RMSE of all the predicted positions and real positions during the prediction window.
- Final displacement error (FDE): The RMSE distance between the final predicted positions at the end of the predicted trajectory and the corresponding true location.

**5.1.2 Prediction results.** We compare our methods with baselines below:

- S-LSTM [3]: This model uses a fully connected social pooling layer to deal with the LSTM encoder output, and generates a unimodal distribution for future coordinates.
- CS-LSTM [23]: This model devises a convolutional social pooling layer to process the LSTM encoder output, and generates a unimodal distribution for future coordinates.
- MHATP [37]: This method uses multi-head attention based model for probabilistic vehicle trajectory prediction.
- TraPHic [14]: This approach uses spatial attention based pooling to perform trajectory prediction of road agents in dense and heterogeneous traffic.
- MATF-D and MATF-GAN [82]: A Multi-agent tensor fusion network for contextual trajectory prediction.
- CAM and ATGLOBAL-CAM-NF [58]: A cross-agent attention model with understanding of the scene contexts for trajectory prediction.

We evaluate our model 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 the vehicles that are around it. While with information sharing, we assume that the EV is able to get the TV's neighboring vehicles' historical trajectories and feed them into encoders.

Based on the prediction results in Table 1 and Table 2 we show that:

- Our proposed MEATP with shared information has better prediction performance compared with LSTM based models and transformer based models on NGSIM dataset. This shows the advantages of our multi-encoder attention mechanism at encoding both spatial and temporal features. Information sharing among the CAVs largely improves the prediction performance.
- Our proposed MEATP with shared information outperforms most baseline methods on Argoverse dataset. It shows slightly worse prediction performance compared with GLOBAL-CAM-NF [58]. This is because GLOBAL-CAM-NF encodes the scene context, which requires the bird's eye view of the entire driving

environment. However, this information may not be available in real world when predicting the trajectories. Compare to this method, our model doesn't need the scene contexts for prediction.

Table 1. **Prediction results on NGSIM dataset.** Root mean square error (RMSE) over 5 seconds of prediction horizon for models are compared. Our Proposed model MEATP without shared information has better prediction performance than the baselines. Our MEATP with information sharing significantly decreases the RMSE values, at 4s and 5s, the RMSE values are less than half of baselines.

Evaluation Metric	Prediction Horizon (s)	S-LSTM	CS-LSTM	MHAPTP	MEATP w/o shared information	MEATP w shared information
RMSE (m)	1	0.65	0.61	0.55	0.70	0.51
	2	1.31	1.27	0.60	1.03	0.76
	3	2.16	2.09	1.12	1.26	0.85
	4	3.25	3.10	-	1.59	1.05
	5	4.55	4.37	-	2.17	1.36

Table 2. **Prediction results on Argoverse dataset.** ADE and FDE of models are compared. Our Proposed model MEATP with shared information has better prediction performance than most methods.

Evaluation Metric	CS-LSTM	MATF-D	MATF-GAN	TraPHic	CAM	GLOBAL-CAM-NF	MEATP w shared information
ADE (m)	1.39	1.35	1.26	1.04	1.13	0.80	1.13
FDE (m)	2.57	2.48	2.31	3.08	2.50	1.25	2.07

## 5.2 Human driver abnormal behavior detection

**5.2.1 Experiment details.** In this section, we use open source simulator Simulation of Urban Mobility (SUMO) [48] to generate the dataset. We first construct a highway scenario: a highway with 5 lanes, 1000m length. The traffic volume is 8000 Veh/Hour. Without loss of generality, We then set another urban traffic scenario. We construct a city street with 5 lanes, 1000m length. Two intersections with traffic lights are set 300m and 600m away from the start point of the street. Each branch road contains 4 lanes, with 2 lanes in each direction. Regarding the traffic lights, each of them are set to green light status 80% of one cycle, where one cycle is two minutes. The Traffic volumes are 7000 Veh/Hour on the main road and 200 Veh/Hour on each branch road. The total simulation time is one hour in both scenarios. Notably, we set the traffic volume according to the real world traffic report [1]. The vehicles are simulated using Krauß car following model [39]. 1000 vehicles are switched to abnormal mode once they pass a random location on the road. In our experiment, abnormal behaviors include unusual speed control, larger accelerations, decelerations, frequent lane-changing, and small headways, etc. They differ from the normal driving behaviors with altered parameters as shown in Table 3. Vehicle trajectories are collected at frequency of 10Hz in one hour with label of whether the vehicle has turned into abnormal driving mode. Both highway and urban traffic datasets are divided into training and testing set by the ratio of 7:3.



Table 3. **Parameters of different driving behavior types.** On longitudinal direction: vehicles in abnormal driving mode have larger acceleration and deceleration abilities, nonusual maximum speed, smaller minimum gaps to leading vehicles, and less perfect driving (sigma denotes the driver imperfection in Krauß car following model, the larger the more imperfection). On lateral direction: vehicles in abnormal driving mode are less willing to perform cooperative lane changing (larger lcCooperative), they will tend to change lane more frequently to gain high speed (larger lcSpeedGain), and less perfect in lane changing (larger lcSigma).

Parameters	Normal driving behavior	Abnormal driving behavior
accel ( $m/s^2$ )	2.6	7
decel ( $m/s^2$ )	4.5	8
miniGap ( $m$ )	2.5	1.0
sigma	0.1	0.8
maxSpeed (highway) ( $m/s$ )	30	20 or 45
maxSpeed (urban) ( $m/s$ )	16	7 or 25
speedFactor	1.0	1.2
lcCooperative	1.0	0.1
lcSpeedGain	1.0	5.0
lcSigma	0.1	0.8

Table 4. **Distributions of prediction errors on highway dataset.** MEATP with shared information has the smallest mean and the smallest standard deviation. With all the three predictors, the mean and the standard deviation of the prediction errors on abnormal vehicles are larger than those on normal vehicles.

Parameters	Prediction horizon (s)	CS-LSTM		MEATP w/o shared information		MEATP w shared information	
		Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Mean (m)	1	0.89	1.43	0.68	1.41	0.59	1.07
	2	1.30	1.96	0.85	1.72	0.73	1.54
	3	2.71	3.59	1.23	2.23	0.89	1.93
SD (m)	1	0.93	1.52	1.00	1.35	0.62	1.36
	2	1.68	2.50	1.47	2.37	0.79	1.98
	3	1.94	3.73	1.57	3.56	0.88	3.17

**5.2.2 Detection results.** We train our proposed MEATP and the baseline CS-LSTM separately on the normal vehicles' trajectories in SUMO-generated highway and urban traffic datasets. We apply the trained predictors to testing set, then compute the mean and standard deviations of the prediction errors in future 3s. The distributions of the prediction errors on highway and urban traffic are shown in table 4 and table 5 respectively. It can be clearly seen that the means and standard deviations of prediction errors on the abnormal vehicles are larger than those of normal vehicles. This is consistent with our expectation, since the prediction will be less accurate when the abnormal vehicles have more unreasonable behaviors, such as sudden accelerations and decelerations, frequent lane changes, etc. What's more, our proposed MEATP with shared information has smallest mean and standard deviation among three predictors.

Table 5. Distributions of prediction errors on urban traffic dataset.

Parameters	Prediction horizon (s)	CS-LSTM		MEATP w/o shared information		MEATP w shared information	
		Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Mean (m)	1	1.09	1.36	0.78	1.29	0.67	0.95
	2	1.62	2.00	0.96	1.65	0.83	1.42
	3	2.78	3.43	1.32	2.40	1.01	1.83
SD (m)	1	0.88	1.73	0.89	1.31	0.69	1.35
	2	1.71	2.84	1.56	2.66	1.02	1.88
	3	2.89	4.60	2.48	3.72	1.13	2.92

Based on the trained predictors and the probability distributions, we apply proposed algorithm 1 to detect the abnormal behaviors. For each target vehicle, at every time step, we use the trained predictor to predict the trajectory in future 5s, then compute the prediction error  $e_n$  based on the true trajectory and predicted trajectory. After that,  $W_n$  is computed based on  $e_n$  and compared with threshold  $b$  to detect the change point. For each target vehicle, the inference runtime of our detection algorithm equipped with MEATP is 4.8 ms per time step. To be noticed that, depending on whether we have full knowledge of the post-change distribution  $g$ , we evaluate the algorithm 1 based on three different methods: CuSum, MCuSum and GLRT algorithm in highway and urban traffic datasets respectively. We add four levels of Gaussian noise to the neighboring 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 [4, 59, 81], level 0 means no noise in inputs. The detection results are shown in Table 6 to Table 8. Notably, ADD represents average detection delay in unit of sample. We first summarize our findings based on the results and analyze them separately in the following parts. Our findings:

- Our proposed algorithm 1, equipped with well trained predictor, has shown great detection performance in both highway and urban traffic scenarios.
- Information sharing among CAVs helps increase the detection rate, lower the false alarms and average detection delay.
- our proposed algorithm 1 can be generalized to different scenarios: (1) we have full knowledge about the pre-change distribution  $f$  and post-change distribution  $g$ ; (2) We have full knowledge about the pre-change distribution  $f$  and only partial knowledge about the post-change distribution  $g$ . In both scenarios, algorithm 1 shows remarkable detection performance. When having least information about the post-change distributions, algorithm 1 equipped with MEATP with shared information still achieves: detection rate 91.0%, average detection delay 24.9 samples, 16 false alarms in 300 vehicles.
- Equipped with MEATP with shared information, our detection algorithm is robust to the noises in observations of the surrounding autonomous vehicles.

Fig. 4 and Fig. 5 show the statistic evolution given the prediction errors of an abnormal vehicle and a normal 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.

**Detection results based on CuSum algorithm.** In first scenario, we have full knowledge of the pre-change distribution  $f$  and post-change distribution  $g$ . We update statistics  $W_n$  based on CuSum algorithm, the detection results are shown in table 6. It can be seen that, when having full knowledge of the pre-change and post-change distributions, our proposed algorithm 1 equipped with the well trained predictor can detect the most of the abnormal driving behavior within a short time in both highway and urban traffic scenarios. Even using the

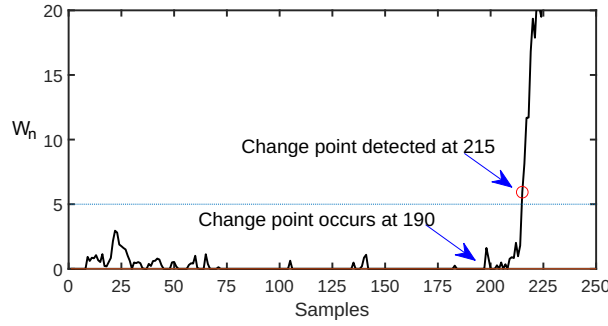


Fig. 4. **Abnormal behavior statistics.** Threshold  $b = 5$ . The abnormal vehicle changes from normal driving mode to aggressive driving mode at time step 190. At time step 215, the statistic  $W_n$  exceeds the threshold, thus the detection of abnormal behavior is declared. Detection delay is 25 samples in this case.

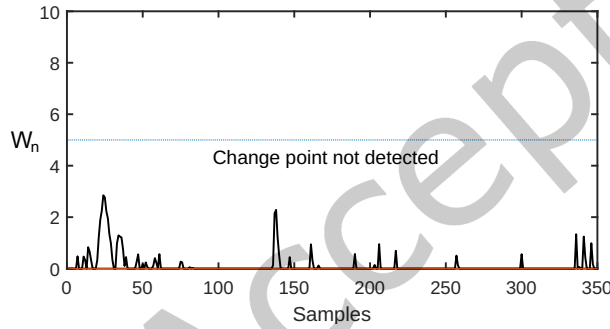


Fig. 5. **Normal behavior statistics.** Threshold  $b = 5$ . Since  $W_n$  never exceeds the threshold, the change point, namely the abnormal behavior is not detected.

baseline CS-LSTM as predictor, our proposed algorithm 1 can achieve around 89.0% of detection rate with ADD of 18.7 samples. Compared with the baseline predictor CS-LSTM, our proposed predictor MEATP with shared information outperforms among three predictors, it achieves detection rate of 97.6%, 12.3 samples of ADD, and zero false alarms in highway traffic scenario. Because the shared information enables the predictor to learn more about the interactions between target vehicles and its neighboring vehicles, the prediction performance is much better. Based on the small means and standard deviations of the prediction errors, the detection algorithm is more likely to detect the changes in the distributions, less likely to declare false alarm, rendering highest detection rate, smallest false alarms and detection delay. Compared with detection results in highway traffic datasets, the detection results in urban traffic have similar detection rate, yet more false alarms, this may be caused by larger variance in prediction errors in urban traffic dataset. Notably, the detection delay in unit of second is proportional to the sampling rate. The average detection delay of 12.3 samples with 10 Hz sampling rate rendering the detection delay to be around 1.2s. If we want to detect the abnormal human driver's behavior as soon as possible, we can increase the sampling rate. Thus, our proposed abnormal behavior detection algorithm is generic, the detection delay can be improved by improving the hardware sampling capabilities.

By analysis on CuSum algorithm based detection results, we show that:

- When having full knowledge of pre-change and post-change probability distributions, our proposed detection algorithm 1, equipped with well trained predictor, is able to detect most of the abnormal behaviors in short time with small amount of false alarms.
- Our algorithm 1 has remarkable detection performance in both highway and urban traffic scenarios. In highway traffic, the best detection performance is: 0 false alarms, detection rate of 97.6%, and ADD 12.3 samples. in urban traffic, the best detection performance is: detection rate 96.3%, ADD 18.4 samples , and 6 false alarms in 300 vehicles.
- The shared information among autonomous vehicles increases the detection rate and lowers the detection delay, while maintains least false alarms.

Table 6. Detection Results on SUMO dataset based on CuSum

Models	Parameters	highway traffic noise level					urban traffic noise level				
		0	1	2	3	4	0	1	2	3	4
CS-LSTM	detected	267	264	263	239	226	271	262	258	225	223
	false alarm	7	16	26	31	59	20	22	35	59	67
	ADD	18.7	31.3	30.4	28.7	25.5	30.1	25.1	32.6	33.4	31.7
	detection rate	89.0%	88.0%	87.7%	79.7%	75.3%	90.3%	87.3%	86.0%	75.0%	74.3%
MEATP w/o shared information	detected	282	278	270	261	250	277	273	269	270	259
	false alarm	7	16	21	33	42	13	17	25	28	35
	ADD	22.0	25.4	24.3	25.5	30.1	25.5	30.3	28.2	27.3	27.5
	detection rate	94.0%	92.7%	90.0%	87.0%	83.3%	92.3%	89.6%	91.0%	90.0%	86.3%
MEATP w shared information	detected	293	291	290	288	285	289	284	286	286	283
	false alarm	0	1	1	2	4	6	10	8	8	11
	ADD	12.3	18.3	17.1	19.6	23.2	18.4	13.6	14.8	15.2	18.7
	detection rate	97.6%	97.0%	96.7%	96.0%	95.0%	96.3%	94.7%	95.3%	95.3%	94.3%

**Detection results based on MCuSum algorithm.** In second scenario, we have full knowledge of the pre-change distribution  $f$ , the unknown post-change distribution  $g$  with parameter  $\theta$  belongs to a finite set of distributions:  $\theta \in \Theta = \{\theta_1, \theta_2, \dots, \theta_M\}$ . Regarding the true post-change parameters  $\mu_1, \sigma_1$ , we add two small magnitude (less than 0.3) gaussian noise to each of them separately, meaning  $M = 4$  in this case. The detection results based on MCuSum algorithm are shown in table 7. The overall detection performance are slightly worse than the first scenario where we have full knowledge of distributions  $f$  and  $g$ . This is consistent with our expectations. Since we only have partial knowledge about the post-change distributions  $g$ , we can only use the possible set  $\theta \in \Theta = \{\theta_1, \theta_2, \dots, \theta_M\}$ . We perform  $M$  CuSum algorithms in parallel, at every time step, we choose the maximum  $W_n(\theta)$  to see whether it exceeds the threshold.

By analysis on MCuSum algorithm based detection results, it shows that:

- There is a trade off between the knowledge we have about post-change distributions and abnormal driving behavior detection performance.
- When we have full knowledge about pre-change distribution  $f$ , and the unknown post-change distribution  $g$  belongs to a finite set, our proposed algorithm 1 still has remarkable detection performance in both

Table 7. Detection Results on SUMO dataset based on MCuSum

Models	Parameters	highway traffic noise level					urban traffic noise level				
		0	1	2	3	4	0	1	2	3	4
CS-LSTM	detected	263	258	241	225	213	265	260	224	219	207
	false alarm	18	32	42	49	60	17	29	51	65	71
	ADD	37.3	24.0	16.7	14.5	11.7	12.5	15.9	16.8	11.6	12.8
	detection rate	87.7%	86.0%	80.3%	75.0%	71.0%	88.3%	86.7%	74.7%	73.0%	69.0%
MEATP w/o shared information	detected	271	267	262	255	253	263	257	253	251	228
	false alarm	12	23	23	26	31	14	21	25	27	57
	ADD	33.9	27.9	30.2	28.6	27.3	20.7	23.7	27.2	27.6	25.7
	detection rate	90.3%	89.0%	87.3%	85.0%	84.3%	87.6%	85.6%	84.3%	83.7%	76.0%
MEATP w shared information	detected	291	288	287	284	283	287	283	283	281	278
	false alarm	4	6	8	10	10	6	10	11	11	15
	ADD	14.9	18.2	23.7	22.5	20.1	16.0	19.8	18.6	18.1	19.5
	detection rate	97.0%	96.0%	95.7%	94.7%	94.3%	95.7%	94.3%	94.3%	93.6%	92.6%

Table 8. Detection Results on SUMO urban dataset based on GLRT

Models	Parameters	highway traffic noise level					urban traffic noise level				
		0	1	2	3	4	0	1	2	3	4
CS-LSTM	detected	252	233	234	220	201	243	229	231	218	195
	false alarm	24	32	36	56	74	26	32	44	75	75
	ADD	27.1	27.5	23.7	24.5	26.3	24.3	29.0	28.7	28.2	29.3
	detection rate	84.0%	77.7%	78.0%	73.3%	67.0%	81.0%	76.3%	77.0%	72.7%	65.0%
MEATP w/o shared information	detected	271	259	260	258	248	268	254	253	250	245
	false alarm	18	23	27	27	38	25	32	34	37	50
	ADD	22.5	25.7	20.1	22.6	25.0	31.4	33.2	31.7	32.1	33.5
	detection rate	90.3%	86.3%	86.7%	86.0%	82.7%	89.3%	84.7%	84.3%	83.3%	81.7%
MEATP w shared information	detected	287	286	283	280	279	283	283	280	278	273
	false alarm	5	6	7	11	13	10	11	13	14	16
	ADD	21.4	20.1	21.8	23.3	23.9	23.0	20.4	21.0	20.2	24.9
	detection rate	95.7%	95.3%	94.3%	93.3%	93.0%	94.3%	94.3%	93.3%	92.7%	91.0%

highway and urban traffic scenario. Best detection performance based on MEATP with shared information. In highway traffic: detection rate 97.0%, ADD 14.9 samples, 4 false alarms in 300 vehicles; in urban traffic: detection rate 95.7%, ADD 16 samples, 6 false alarms in 300 vehicles.

**Detection results based on GLRT algorithm.** In the third scenario, we have full knowledge of the pre-change distribution  $f$  with parameter  $\phi(\mu_0, \sigma_0)$ , for the unknown post-change distribution  $g$  with parameter  $\theta(\mu_1, \sigma_1)$ , we only know the minimum magnitude change compared with pre-change parameters:  $\mu_1 - \mu_0 \geq \nu_m, \sigma_1 - \sigma_0 \geq \delta_m$ . The detection results based on GLRT algorithm are shown in table 8. As can be seen, since we have even less information about the post-change distributions, the detection performance slightly drop compared with second scenario.

By analysis on GLRT algorithm based detection results, we show that:

- When we have full knowledge of the pre-change distribution parameter  $\phi$ , and we only know that the minimum magnitude difference between post-change parameter  $\theta$  and pre-change parameter  $\phi$ , our proposed algorithm 1 still shows great detection performance. MEATP with shared information in highway traffic: detection rate 95.7%, ADD 21.40 samples, 5 false alarms in 300 vehicles; in urban traffic: detection rate 94.3%, ADD 23.0 samples, 10 false alarms in 300 vehicles.

## 6 CONCLUSION

This paper proposes abnormal human driving behavior detection algorithm for CAVs based on shared sensing information in hybrid traffic system. This work, to the best of our knowledge, is the first efficient algorithm that can accurately and quickly detect abnormal human driving mode switches based on CAVs sensing data without using in-vehicle sensing data that may hurt human-driver privacy. We first propose a multi-encoder attention based interaction-aware trajectory prediction model called MEATP. Based on the predictor MEATP, we further develop an abnormal behavior detection method. Through extensive experiments on both public dataset and simulator, We show that (1) our proposed MEATP predictor outperforms the baselines; (2) our proposed algorithm detects abnormal behaviors with remarkable high accuracy (the best performance achieves detection rate of 97.3%) and low detection delay; (3) shared information boosts the performance of both trajectory prediction and abnormal behavior detection.

## REFERENCES

- [1] [n. d.]. Caltrans. <https://dot.ca.gov/programs/traffic-operations/census/traffic-volumes>. Accessed: 2022-02-18.
- [2] Nedaa Baker Al Barghuthi and Huwida Said. 2019. Readiness, Safety, and Privacy on Adopting Autonomous Vehicle Technology: UAE Case Study. In *2019 Sixth HCT Information Technology Trends (ITT)*. IEEE, 47–52.
- [3] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 961–971.
- [4] Alireza Asvadi, Pedro Girão, Paulo Peixoto, and Urbano Nunes. 2016. 3D object tracking using RGB and LIDAR data. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 1255–1260.
- [5] Taposh Banerjee and Venugopal V Veeravalli. 2014. Data-efficient quickest change detection with unknown post-change distribution. In *2014 IEEE International Symposium on Information Theory*. IEEE, 741–745.
- [6] Tobias Bär, Dennis Nienhüser, Ralf Kohlhaas, and J Marius Zöllner. 2011. Probabilistic driving style determination by means of a situation based analysis of the vehicle data. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 1698–1703.
- [7] Alexander Barth and Uwe Franke. 2008. Where will the oncoming vehicle be the next second?. In *2008 IEEE Intelligent Vehicles Symposium*. IEEE, 1068–1073.
- [8] Thomas Batz, Kym Watson, and Jurgen Beyerer. 2009. Recognition of dangerous situations within a cooperative group of vehicles. In *2009 IEEE Intelligent Vehicles Symposium*. IEEE, 907–912.
- [9] Andrew Best, Sahil Narang, Daniel Barber, and Dinesh Manocha. 2017. Autonovi: Autonomous vehicle planning with dynamic maneuvers and traffic constraints. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2629–2636.
- [10] Manoj Bhat, Jonathan Francis, and Jean Oh. 2020. Trajformer: Trajectory prediction with local self-attentive contexts for autonomous driving. *arXiv preprint arXiv:2011.14910* (2020).
- [11] Keshav Bimraw. 2015. 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)*, Vol. 1. IEEE, 191–198.

- [12] Noam Buckman, Alyssa Pierson, Sertac Karaman, and Daniela Rus. 2020. Generating Visibility-Aware Trajectories for Cooperative and Proactive Motion Planning. In *ICRA 2020*. IEEE, 3220–3226.
- [13] Luca Caltagirone, Mauro Bellone, Lennart Svensson, and Mattias Wahde. 2019. LIDAR–camera fusion for road detection using fully convolutional neural networks. *Robotics and Autonomous Systems* 111 (2019), 125–131.
- [14] Rohan Chandra, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. 2019. Taphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8483–8492.
- [15] Rohan Chandra, Tianrui Guan, Srujan Panuganti, Trisha Mittal, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. 2020. Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms. *IEEE Robotics and Automation Letters* 5, 3 (2020), 4882–4890.
- [16] M. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays. 2019. Argoverse: 3D Tracking and Forecasting with Rich Maps. In *Proceedings of (CVPR) Computer Vision and Pattern Recognition*. 8740 – 8749.
- [17] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. 2019. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8748–8757.
- [18] Kyunghoon Cho, Timothy Ha, Gunmin Lee, and Songhwai Oh. 2019. 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)*. IEEE, 2076–2081.
- [19] James Colyar and John Halkias. 2007. Us highway 101 dataset. *Highway Administration (FHWA), Tech. Rep* (2007), 07–030.
- [20] James Colyar and John Halkias. 2007. Us highway I-80 dataset. *Highway Administration (FHWA), Tech. Rep* (2007), 07–030.
- [21] DSRC Committee et al. 2009. Dedicated short range communications (DSRC) message set dictionary. *SAE Standard J 2735* (2009), 2015.
- [22] Drew Dawson and Kathryn Reid. 1997. Fatigue, alcohol and performance impairment. *Nature* 388, 6639 (1997), 235–235.
- [23] Nachiket Deo and Mohan M Trivedi. 2018. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1468–1476.
- [24] Bo Dong, Hao Liu, Yu Bai, Jinbiao Lin, Zhuoran Xu, Xinyu Xu, and Qi Kong. 2021. Multi-modal trajectory prediction for autonomous driving with semantic map and dynamic graph attention network. *arXiv preprint arXiv:2103.16273* (2021).
- [25] Daniel J Fagnant and Kara Kockelman. 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice* 77 (2015), 167–181.
- [26] Jamil Fayyad, Mohammad A Jaradat, Dominique Gruyer, and Homayoun Najjaran. 2020. Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors* 20, 15 (2020), 4220.
- [27] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. 2014. De-anonymization attack on geolocated data. *J. Comput. System Sci.* 80, 8 (2014), 1597–1614.
- [28] Matthew Gillespie. 2016. Shifting automotive landscapes: Privacy and the right to travel in the era of autonomous motor vehicles. *Wash. UJL & Pol’y* 50 (2016), 147.
- [29] Francesco Giuliani, Irtiza Hasan, Marco Cristani, and Fabio Galasso. 2021. Transformer Networks for Trajectory Forecasting. In *2020 25th International Conference on Pattern Recognition (ICPR)*. 10335–10342. <https://doi.org/10.1109/ICPR48806.2021.9412190>
- [30] Dorothy J Glancy. 2012. Privacy in autonomous vehicles. *Santa Clara L. Rev.* 52 (2012), 1171.
- [31] S. Han, J. Fu, and F. Miao. 2019. Exploiting Beneficial Information Sharing Among Autonomous Vehicles. In *2019 IEEE 58th Conference on Decision and Control (CDC)*. 2226–2232.
- [32] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [33] Jie Hu, Li Xu, Xin He, and Wuqiang Meng. 2017. Abnormal driving detection based on normalized driving behavior. *IEEE Transactions on Vehicular Technology* 66, 8 (2017), 6645–6652.
- [34] Yaocong Hu, Mingqi Lu, and Xiaobo Lu. 2019. Driving behaviour recognition from still images by using multi-stream fusion CNN. *Machine Vision and Applications* 30, 5 (2019), 851–865.
- [35] WuLing Huang, Kunfeng Wang, Yisheng Lv, and FengHua Zhu. 2016. Autonomous vehicles testing methods review. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 163–168.
- [36] J. B. Kenney. 2011. Dedicated Short-Range Communications (DSRC) Standards in the United States. *Proc. IEEE* 99, 7 (July 2011), 1162–1182. <https://doi.org/10.1109/JPROC.2011.2132790>
- [37] Hayoung Kim, Dongchan Kim, Gihoon Kim, Jeongmin Cho, and Kunsoo Huh. 2020. Multi-head attention based probabilistic vehicle trajectory prediction. In *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1720–1725.
- [38] Seong-Woo Kim and Wei Liu. 2015. The impact of cooperative perception on decision making and planning of autonomous vehicles. *IEEE Intell. Transp. Syst. Mag.* 7, 3 (2015), 39–50.
- [39] Stefan Krauß. 1998. Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics. (1998).
- [40] Sampo Kuutti, Richard Bowden, Yaochu Jin, Phil Barber, and Saber Fallah. 2020. A survey of deep learning applications to autonomous vehicle control. *IEEE Transactions on Intelligent Transportation Systems* 22, 2 (2020), 712–733.

- [41] J. Lee and B. Park. 2012. Development and Evaluation of a Cooperative Vehicle Intersection Control Algorithm Under the Connected Vehicles Environment. *IEEE Trans. Intell. Transp. Syst.* 13, 1 (March 2012), 81–90. <https://doi.org/10.1109/TITS.2011.2178836>
- [42] Stéphanie Lefèvre, Dizan Vasquez, and Christian Laugier. 2014. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH journal* 1, 1 (2014), 1–14.
- [43] Joseph Lemley, Anuradha Kar, Alexandru Drimborean, and Peter Corcoran. 2019. Convolutional neural network implementation for eye-gaze estimation on low-quality consumer imaging systems. *IEEE Transactions on Consumer Electronics* 65, 2 (2019), 179–187.
- [44] Jiachen Li, Hengbo Ma, and Masayoshi Tomizuka. 2019. Conditional generative neural system for probabilistic trajectory prediction. *arXiv preprint arXiv:1905.01631* (2019).
- [45] Zuojin Li, Shengbo Eben Li, Renjie Li, Bo Cheng, and Jinliang Shi. 2017. Online detection of driver fatigue using steering wheel angles for real driving conditions. *Sensors* 17, 3 (2017), 495.
- [46] Lei Lin, Weizi Li, Huikun Bi, and Lingqiao Qin. 2021. Vehicle trajectory prediction using LSTMs with spatial-temporal attention mechanisms. *IEEE Intelligent Transportation Systems Magazine* (2021).
- [47] Yicheng Liu, Jinghui Zhang, Liangji Fang, Qinhong Jiang, and Bolei Zhou. 2021. Multimodal motion prediction with stacked transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7577–7586.
- [48] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. 2018. Microscopic Traffic Simulation using SUMO. In *The 21st IEEE International Conference on Intelligent Transportation Systems. IEEE Intelligent Transportation Systems Conference (ITSC)*. <https://elib.dlr.de/124092/>
- [49] Gary Lorden et al. 1971. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics* 42, 6 (1971), 1897–1908.
- [50] Panagiotis Lytrivis, George Thomaidis, and Angelos Amditis. 2008. Cooperative path prediction in vehicular environments. In *2008 11th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 803–808.
- [51] Yuxin Ma, Xinge Zhu, Sibao Zhang, Ruigang Yang, Wenping Wang, and Dinesh Manocha. 2019. Trafficpredict: Trajectory prediction for heterogeneous traffic-agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 6120–6127.
- [52] Kaouther Messaoud, Nachiket Deo, Mohan M Trivedi, and Fawzi Nashashibi. 2021. Trajectory prediction for autonomous driving based on multi-head attention with joint agent-map representation. In *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 165–170.
- [53] Aaron Miller and Kyungzun Rim. 2020. Cooperative Perception and Localization for Cooperative Driving. In *ICRA 2020*. IEEE, 1256–1262.
- [54] George V Moustakides et al. 1986. Optimal stopping times for detecting changes in distributions. *Annals of Statistics* 14, 4 (1986), 1379–1387.
- [55] Teddy Ort, Liam Paull, and Daniela Rus. 2018. Autonomous vehicle navigation in rural environments without detailed prior maps. In *ICRA 2018*. IEEE, 2040–2047.
- [56] Ewan S Page. 1954. Continuous inspection schemes. *Biometrika* 41, 1/2 (1954), 100–115.
- [57] Seong Hyeon Park, ByeongDo Kim, Chang Mook Kang, Chung Choo Chung, and Jun Won Choi. 2018. Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture. In *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1672–1678.
- [58] Seong Hyeon Park, Gyubok Lee, Jimin Seo, Manoj Bhat, Minseok Kang, Jonathan Francis, Ashwin Jadhav, Paul Pu Liang, and Louis-Philippe Morency. 2020. Diverse and admissible trajectory forecasting through multimodal context understanding. In *European Conference on Computer Vision*. Springer, 282–298.
- [59] Zengyi Qin, Jinglu Wang, and Yan Lu. 2019. Monogrnet: A geometric reasoning network for monocular 3d object localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 8851–8858.
- [60] Bhargava Reddy, Ye-Hoon Kim, Sojung Yun, Chanwon Seo, and Junik Jang. 2017. 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*. 121–128.
- [61] J. Rios-Torres and A. A. Malikopoulos. 2017. A Survey on the Coordination of Connected and Automated Vehicles at Intersections and Merging at Highway On-Ramps. *IEEE Trans. Intell. Transp. Syst.* 18, 5 (May 2017), 1066–1077. <https://doi.org/10.1109/TITS.2016.2600504>
- [62] Ya'acov Ritov. 1990. Decision theoretic optimality of the CUSUM procedure. *The Annals of Statistics* (1990), 1464–1469.
- [63] Abbas Sadat, Mengye Ren, Andrei Pokrovsky, Yen-Chen Lin, Ersin Yumer, and Raquel Urtasun. 2019. Jointly learnable behavior and trajectory planning for self-driving vehicles. *arXiv preprint arXiv:1910.04586* (2019).
- [64] Joshua Schoonmaker. 2016. Proactive privacy for a driverless age. *Information & Communications Technology Law* 25, 2 (2016), 96–128.
- [65] Mohammad Shahverdy, Mahmood Fathy, Reza Berangi, and Mohammad Sabokrou. 2020. Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications* 149 (2020), 113240.
- [66] Araz Taeihagh and Hazel Si Min Lim. 2019. Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport reviews* 39, 1 (2019), 103–128.
- [67] Alireza Talebpour and Hani S Mahmassani. 2016. Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transportation Research Part C: Emerging Technologies* 71 (2016), 143–163.
- [68] Han-Shue Tan and Jihua Huang. 2006. DGPS-based vehicle-to-vehicle cooperative collision warning: Engineering feasibility viewpoints. *IEEE Transactions on Intelligent Transportation Systems* 7, 4 (2006), 415–428.



- [69] Alexander G Tartakovsky and Aleksey S Polunchenko. 2008. Quickest changepoint detection in distributed multisensor systems under unknown parameters. In *2008 11th International Conference on Information Fusion*. IEEE, 1–8.
- [70] Jessica Van Brummelen, Marie O'Brien, Dominique Gruyer, and Homayoun Najjaran. 2018. Autonomous vehicle perception: The technology of today and tomorrow. *Transportation research part C: emerging technologies* 89 (2018), 384–406.
- [71] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762* (2017).
- [72] Harini Veeraraghavan, Nikos Papanikolopoulos, and Paul Schrater. 2006. Deterministic sampling-based switching Kalman filtering for vehicle tracking. In *2006 IEEE Intelligent Transportation Systems Conference*. IEEE, 1340–1345.
- [73] Venugopal V Veeravalli and Taposh Banerjee. 2014. Quickest change detection. In *Academic Press Library in Signal Processing*. Vol. 3. Elsevier, 209–255.
- [74] Sandor M Veres, Levente Molnar, Nick K Lincoln, and Colin P Morice. 2011. 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 (2011), 155–195.
- [75] Junqing Wei, Jarrod M Snider, Tianyu Gu, John M Dolan, and Bakhtiar Litkouhi. 2014. A behavioral planning framework for autonomous driving. In *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE, 458–464.
- [76] Jingwen Wu, Hua Liao, and Jin-Wei Wang. 2020. Analysis of consumer attitudes towards autonomous, connected, and electric vehicles: A survey in China. *Research in transportation economics* 80 (2020), 100828.
- [77] Mingxing Xu, Wenrui Dai, Chunmiao Liu, Xing Gao, Weiyao Lin, Guo-Jun Qi, and Hongkai Xiong. 2021. Spatial-Temporal Transformer Networks for Traffic Flow Forecasting. *arXiv:2001.02908 [eess.SP]*
- [78] Chao Yan, Frans Coenen, Yong Yue, Xiaosong Yang, and Bailing Zhang. 2016. Video-based classification of driving behavior using a hierarchical classification system with multiple features. *International Journal of Pattern Recognition and Artificial Intelligence* 30, 05 (2016), 1650010.
- [79] Siqian Yang, Cheng Wang, Hongzi Zhu, and Changjun Jiang. 2019. APP: augmented proactive perception for driving hazards with sparse GPS trace. In *Proceedings of the twentieth ACM international symposium on mobile Ad Hoc networking and computing*. 21–30.
- [80] Lanhang Ye and Toshiyuki Yamamoto. 2019. Evaluating the impact of connected and autonomous vehicles on traffic safety. *Physica A: Statistical Mechanics and its Applications* 526 (2019), 121009.
- [81] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. 2021. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11784–11793.
- [82] Tianyang Zhao, Yifei Xu, Mathew Monfort, Wongun Choi, Chris Baker, Yibiao Zhao, Yizhou Wang, and Ying Nian Wu. 2019. Multi-agent tensor fusion for contextual trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12126–12134.
- [83] Gao Zhenhai, Le DinhDat, Hu Hongyu, Yu Ziwen, and Wu Xinyu. 2017. 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)*. IEEE, 99–101.