Driver Distraction Recognition-driven Collision Avoidance Algorithm for Active Vehicle Safety

K. B. Devika[1], Asish Bera[2], Venkata Ramani Shreya Yellapantula[1], Ardhendu Behera[2*], Yonghuai Liu[2], and Shankar C. Subramanian[1]

[1]*Department of Engineering Design, Indian Institute of Technology Madras,* India.
[2]*Department of Computer Science, Edge Hill University,* United Kingdom.
[*]beheraa@edgehill.ac.uk

***Abstract*— This paper integrates human driver factors with a model-based Collision Avoidance System (CAS) to enhance the safety of semi-autonomous vehicles. Driver Activity Recognition (DAR) through Driver Distraction States (DDS) has been used as the key component to trigger the CAS so that collisions can be averted. DDS has been generated using realistic normal driving scenarios and suitably integrated with a Full State Feedback (FSF) controller-based CAS. The integrated algorithm has been tested using a Hardware in Loop (HiL) setup, which is interfaced with the vehicle dynamics software IPG TruckMaker®. The performance of the algorithm has been evaluated for various on-road scenarios and found to be effective in avoiding rear-end collisions.**

Collision Avoidance, Convolutional Neural Network, Driver Distraction, Driver Activity Recognition, Hardware in Loop, Full State Feedback Controller.

I. Introduction

The rapid growth of the transportation sector demands safer and more efficient operation of automobiles. According to the World Health Organisation (Dec, 2018), approximately 1.35 million people die in road accidents each year and on average 3,700 people lose their lives every day on roads [1], [2]. The United Nation's 2030 agenda for sustainable development looks forward to halving the number of global deaths and injuries from road traffic accidents [3]. Technologies that enhance the safety of automotive and transportation sectors are hence the need of the hour. Human factors contribute to more than 90% of road crashes in European Union countries [4]. According to the Ministry of Road Transportation and Highways, in India, around 80% of road accidents are caused due to human errors [5]. In this context, autonomous/semi-autonomous vehicular operation has gained significant research attention.

Even though autonomous vehicles are promising solutions, considering the increase in transportation demand, complete autonomous operation of passenger vehicles still requires more technological and infrastructural developments. In this context, semi-autonomous vehicular operation could improve safety without the need for high-level infrastructure. Active safety systems like Collision Avoidance System (CAS) that would assist the driver through warnings or/and interventions can reduce vehicle crashes due to human errors [6]. CAS algorithms for the autonomous operation of passenger vehicles have thoroughly been addressed by the research community in the recent past [7]–[11]. However, semi-autonomous vehicular operation that involves switching between human driver and CAS, which is more feasible from an immediate implementation view point, has not been adequately accounted for. A collaborative driving scheme that fuses control inputs from the human driver and CAS, and allows autonomous and manual driving capabilities has been presented in [12]. The current paper attempts to integrate human factors through Driver Activity Recognition (DAR) technique with the CAS design to enhance semi-autonomous vehicle safety.

DAR is a challenging task for intelligent vehicles to monitor secondary activity that often causes Driver Distraction (DD) during driving, such as texting message over phone, talking to passenger, drinking, etc. [14], [15]. Some recent studies explore computer vision to recognise DD (a change in driver's attention from driving to secondary tasks) by analysing images or videos of a driver [2], [13], [16]–[24]. The proposed work advances these ideas further to avoid fatal consequences by integrating DAR and CAS. Motivated with a substantial progress in DAR leveraging attention mechanism [25], this work recognises the Driver Distraction State (DDS) from image/video contents by solving a binary classification task i.e., whether the driving state is a safe/attentive (DDS=0) or distracted/inattentive action (DDS=1). In this regard, this paper attempts to realise a semi-autonomous CAS, which primarily relies on the driver for collision avoidance and makes use of an automatic braking system in the event of driver inattention. DAR is used as a key component to synthesise DDS whether CAS should be triggered. The proposed framework is shown in Fig. 1.

This paper utilises a model-based approach using a Full State Feedback (FSF) controller [9], [26] that would brake the vehicle when the relative spacing and relative speed between the host vehicle and the obstacle (the lead vehicle) reduce beyond certain predefined safe limits. To realise semi-autonomous vehicular operation, the CAS has been activated based on the binary DDS output (safe or distracted, as defined above). The schematic representation of the semi-autonomous CAS operation is shown in Fig. 1b. DAR determines when CAS should be triggered and the FSF controller-based CAS implements braking when the inter-vehicular distance is below the threshold spacing. The major contributions of this study are:

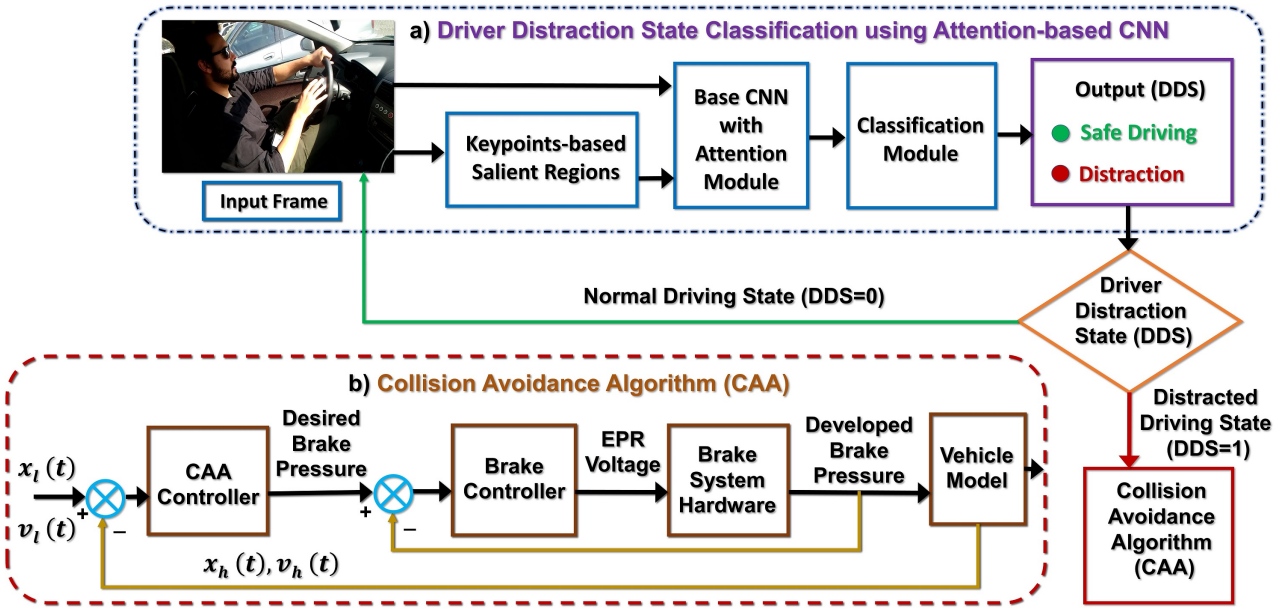

Fig. 1. Proposed integrated framework using image or video-based Driver Distraction State (DDS) to trigger the collision avoidance algorithm (CAA). a) A simplified schematic diagram of the keypoints-based attentional deep convolutional neural networks (CNN) [13] to determine driver distraction (DDS=0 or 1). The input frame extracted from a video scenario is used to localise the informative regions within the frame. Next, these salient regions (smaller to larger in size) along with the input frame are passed through the attention module to focus on the most discriminative region to make the final decision (DDS=0 or 1) by the classification module. b) The main components of the CAA are shown. CAA comes into action when DDS=1 and the algorithm detects an impending collision. Depending upon the relative speed and spacing between the vehicles, the FSF based CAA controller finds the required deceleration to stop the vehicle at a predefined standstill spacing.

- An integrated framework that uses human driver behavioural factors to trigger collision avoidance algorithm automatically, if necessary, to enhance vehicular operation through certain level of semi-autonomy.
- Image or video-driven automatic driver distraction state recognition using attention-based deep networks, which is integrated with a FSF controller-based CAS.
- Evaluation of the integrated approach through Hardware in Loop (HiL) experiments.

The rest of the paper is organised as follows. Driver activity recognition through DDS classification is described in Section II. Section III explains the vehicle dynamics model and the CAS design principle. Section IV presents HiL results and discussions. Section V concludes the paper.

II. Driver Activity Recognition (DAR)

DAR plays a key role in CAS as it decides when the CAS should be triggered. It is regarded as a sub-category of human action recognition. The advancement of Convolutional Neural Networks (CNN) facilitates robust solutions to recognise the driver distraction state [2], [15], [20], [23], [27]. Recently, a CNN via neural architecture search (NAS) is proposed to detect various types of DDS including cognitive, emotional, and sensorimotor distractions [2].

In addition with the attention mechanism [25], deep networks achieve superior performance to discriminate subtle variations in driver's secondary activities [18], [21], [22]. Attention mechanism is originally introduced in natural language processing [25], [28]. It has been recently explored in self-attention mechanism that efficiently captures long-range dependencies between image regions for visual recognition [29]. A few methods have been developed for DAR using attention mechanism [18], [21], [22]. In [18], the channel-level and spatial-level attentions are applied for DDS classification. The results in [30] imply that an LSTM (long short-term memory) network with attention can achieve better performance to estimate DDS. Regional Attention Network (RAN) [21] combines multiple attention mechanisms over different image-regions for DDS recognition. Recently, AG-Net [22] implements a keypoints-based attention mechanism for image recognition. It localises a set of semantic regions to learn contextual information from smaller to larger patches to full image. The regions are derived using the Scale Invariant Feature Transform (SIFT) [31] keypoints and a Gaussian mixture model. These informative regions are further guided by the attention mechanism to discriminate subtle variations of activities in images. It achieves state-of-the-art performance (accuracy: 94.56%) to recognise ten common driving-related activities in Distracted Driver V2 (AUC-V2) dataset [15] using ResNet-50 as a backbone [32]. Likewise, RAN [21] introduces a hybrid attention mechanism by exploring regions and achieves very competitive accuracy (94.27%) with the same backbone. In this study, driver's inattentiveness is monitored as a binary (attentive or inattentive) classification problem by adapting AG-Net [22] framework.

A. Driver Distraction State (DDS)

DDS can be caused by several ways and in different forms such as visual, auditory, cognitive, etc., [2], [16], [33]. In [34], seven different distractions (e.g., mirror checking, texting using a phone, etc.) are considered for experiments. In [19], eight different states are defined as distractions. Ten

driving states are selected in AUC-V2 [15] (image-based) dataset, and example images are shown in Fig. 2 (top-row). These ten unique driving activities are: C0: driving safely, C1: texting right-side, C2: talking over phone right-side, C3: texting left-side, C4: talking over phone left-side, C5: operating radio, C6: drinking, C7: reaching behind, C8: hair and makeup, and C9: talking to passenger. Moreover, this dataset follows a driver-wise split, i.e., the driver distribution is mutually exclusive for training (36 drivers; 12,555 images) and testing (8 drivers; 1,923 images) purposes, and thereby represents a more challenging and realistic driving scenario [17]. The drivers are selected across seven countries, and more details about the dataset are given in [15].

Recently, in addition with the C0-C9 (AUC-V2) driving activities, 3 more activities (reaching side, hands free and switch gear) are included in the video-based Driver Monitoring Dataset (DMD) [23]. It uses similar view-point of the camera and driver's secondary activities with respect to AUC-V2. DMD represents the RGB, depth, and IR videos from 3 cameras capturing face, body, and hands of 37 drivers. However, a lite version of this dataset is available for research, representing the actions of 5 drivers [24]. More details about DMD are given in [23].

B. Driving Scenarios

Videos showing a natural driving scenario involving normal driving as well as some secondary activities are not widely available (except the lite version of DMD), to the best of the authors' knowledge. Many of the existing driving datasets are created in a simulated and/or laboratory environment for the sake of feasibility, efficiency and low cost, focusing on the image/video-based recognition of various distraction categories by exploring deep learning techniques. Due to the unavailability of such realistic driving scenarios publicly, this work combines the attentiveness and distractions of a driver to simulate a driving video imitating a naturalistic manner for a longer duration ($>$ 60 seconds (s)). Recently, a shorter video-clip (e.g. duration 3 s) representing ten driving activities is evaluated on AUC-V2 dataset in [17]. Inspired by this, new driving scenarios with the aforesaid alternate driving characteristics are created in this work using the AUC-V2 dataset (Table I). Each video represents a specific driving action, performed by a particular driver (shown in Fig. 2). The video-clips of a person involved in different actions are randomly selected to create a longer video sequence, illustrating a normal driving situation followed by other random combination of secondary activities (C1-C9), and safe driving. For example, a generated scenario is composed with a driver involved in a secondary activity such as texting or talking on the phone, and followed by other distractions like drinking and talking to passengers, etc., and finally normal driving activity. In this way, different scenarios that are random combinations of a few distractions are created, but, not all distracted activities are selected in a single scenario from a realistic perspective. However, DAR is regarded as a binary classification, therefore, it has trivial impact in selecting the specific distraction category (i.e. fine-grained) in this study.

TABLE I

DETAILS OF THE DDS IN DRIVING SCENARIO USING AUC-V2 WITH RESPECT TO TIME AND CLASSIFICATION ACCURACY (%)

Time (s)	Action	Ground-truth	Accuracy (%)
0.0 - 16.4 s	safe-driving	DDS=0	100.0
16.5 - 52.0 s	distracted	DDS=1	96.6
52.1 - 62.0 s	safe-driving	DDS=0	100.0

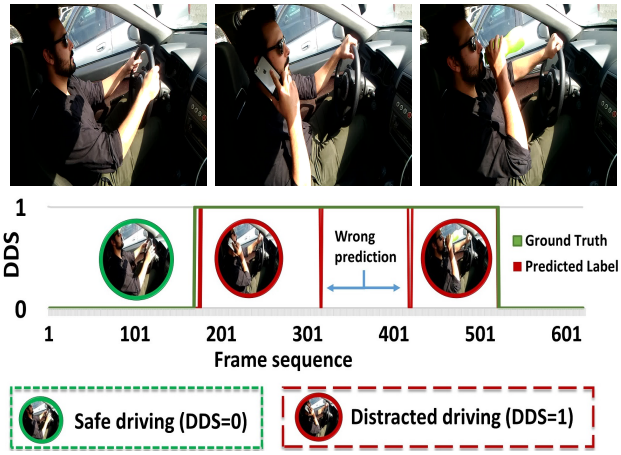

Fig. 2. Example images from [15] showing driver's state. Left: safe driving (DDS=0); mid (talking over phone) and right (drinking): distracted (DDS=1).

Next, a given video scenario V, the frames are sampled at a time interval $t = 100$ ms (10 fps) to extract a set of images N from V, i.e., $I = \{I_t | t = 1 \dots N\}$ with the associated labels $y = \{y_t | t = 1 \dots N\}$. *Safe driving* (C0) is labelled with 0, and all the secondary activities (C1-C9) are grouped together as *distraction* and labelled with 1. However, the original image-labels as provided with the dataset [15] are followed for initial categorisation of the driver's ten fine-grained actions in simulating a longer video scenario. Next, the fine-to-coarse modeling of those extracted frames (i.e., depicting the distracted actions) is performed as a simplified binary classification (shown in Fig. 1a). The attention-driven CNN model classifies a given frame I_t in the end as attentive driving (DDS=0) or inattentive driving (DDS=1) state.

According to the method in [22], this work adopts a similar implementation protocol in selecting the hyper-parameters, data augmentation, and data distribution (i.e., training samples of AUC-V2 dataset). A simple transfer-learning approach with fine-tuning on the target dataset is implemented for testing the driving scenario V. The frame-set (containing 620 images in V), which is obtained from a specific driver's behavior (as described earlier), is used for experiment. Thus, the current frame-set is a subset of AUC-V2 test-set. The details about the frames as extracted with respect to a given time interval are given in Table I. It is evident that the model can classify the safe-driving state (DDS=0) with 100% accuracy, and the prediction accuracy of distractions alone (DDS=1) is 96.60%. The attentional deep network provides an average accuracy of 98.06% for the

TABLE II

DETAILS OF THE DDS IN DRIVING SCENARIO USING DMD WITH RESPECT TO TIME AND CLASSIFICATION ACCURACY (%)

Time (s)	Action	Ground-truth	Accuracy(%)
0.0-17.1, 30.5-40.1 57.5-63.4, 65.8-68.6 72.1-75.2 s	safe-driving	DDS=0	98.69
17.1-30.4, 40.2-57.4 63.5-65.7, 68.7-72.0 s	distracted	DDS=1	98.75

Fig. 3. Example images from [23] showing driver's state. Left: safe driving (DDS=0); mid (talking to passenger) and right (drinking): distracted (DDS=1).

binary classification using the ResNet-50 backbone CNN. A pictorial representation of the simulated scenario is shown at the bottom-row of Fig. 2. Mainly, the frames are misclassified during the transition from safe/attentive driving state to a distraction/inattentive state, i.e., from DDS=0 to DDS=1 at around 16 s. There is also wrong prediction momentarily at around 31 s and 42 s, which are shown in red in Fig. 2. DDS=1 is initially detected at 16.5 s which effectively triggers the CAS. Thus, the later wrong predictions do not affect adversely in the performance to avoid a collision. Hence, initial detection of DDS=1 plays a vital role to initiate the CAS. The correct classification states are overlapped with the ground-truth labels (shown in green) in Fig. 2. We have also tested the scenario created with DMD [23] using transfer learning which is actually trained on one dataset (AUC-V2) and tested on other dataset (DMD). In this cross-dataset testing, the model performs very well ($\approx$ 98.7%) in recognising different states (DDS=0 or 1) of the driver. The performance is given in Table II. Sample images of the DMD are shown in Fig. 3. We have simulated a scenario for 1.14 minutes from actual RGB video of a driver. This longer scenario is a combination of safe driving followed by distraction for multiple times and the duration of each sub-interval is given in Table II. The average accuracy of these sub-intervals are reported here (Table II). A similar trend is observed as stated above in misclassifying the frames during the transition from distracted state to safe-driving state and vice-versa. For example, the frame representing the transition state at 65.8 s is misclassified. Therefore, a similar characteristic during classifying the frame sequence is observed in driving scenarios, generated using both AUC-V2 and DMD datasets. It simply considers that every distraction is equally important to be recognised to trigger urgently the collision avoidance algorithm. This is a safe and conservative strategy for vehicle maneuver, where the trigger of the brake does not mean that a collision is happening, but be more cautious to avoid the case that it is too late to trigger the brake. The DAR module is subsequently integrated with the CAS for driving safety enhancement, described next.

III. COLLISION AVOIDANCE SYSTEM (CAS)

This work utilises a model-based CAS, which makes use of a FSF controller to calculate the deceleration required for collision avoidance. The longitudinal vehicle model that has been used for CAS design is described next.

1) Longitudinal Vehicle Model: This work considers straight-line braking scenarios. The equation of motion representing the longitudinal motion of a vehicle is

$$Ma(t) = -F_{bf}(t) - F_{br}(t) - R_a(t) - R_{rf}(t) - R_{rr}(t), \quad (1)$$

where, F_{bf} and F_{br} represent the braking force developed at the front and rear tyre-road interface, respectively. R_a is the aerodynamic drag force, R_{rf} and R_{rr} are the rolling resistance forces corresponding to front and rear wheels of the vehicle. The weight of the vehicle is represented by W. Acceleration due to gravity, acceleration, and mass of the vehicle are represented by g, a, and M, respectively.

The load on the front wheels (W_f) and the rear wheels (W_r) are given by

$$W_f(t) = \frac{W}{L}\left(L_r - H\frac{a(t)}{g}\right),$$
$$W_r(t) = \frac{W}{L}\left(L_f + H\frac{a(t)}{g}\right), \quad (2)$$

where, L_f and L_r, represent the longitudinal distance of the front axle and rear axle from the center of gravity (C.G.) of the vehicle, respectively, and $L = L_f + L_r$, H is the height of the C. G. of the vehicle. The maximum longitudinal brake force that the tyres can sustain is given by

$$F_{xf}(t) = \mu_{max} W_f(t),$$
$$F_{xr}(t) = \mu_{max} W_r(t), \quad (3)$$

where, μ_{max} is the maximum tyre-road friction coefficient.

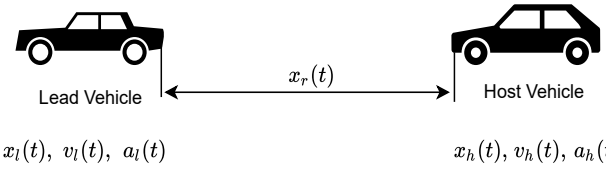

Fig. 4. The variables used in the Collision Avoidance Algorithm (CAA).

2) Collision Avoidance Algorithm (CAA): The variables used in CAA design are presented in Fig. 4. Here, x_l, v_l and $a_l(t)$ represent lead vehicle position, speed and acceleration, respectively. Corresponding host vehicle variables are x_h, v_h and a_h. In order to detect any impending threat, the relative spacing $(x_r(t))$ and the relative longitudinal speed $(v_r(t))$ between the two vehicles are used. The position, and the speed of the vehicle are assumed to be available.

The spacing between the vehicles, x_r is given by

$$x_r(t) = x_l(t) - x_h(t). \quad (4)$$

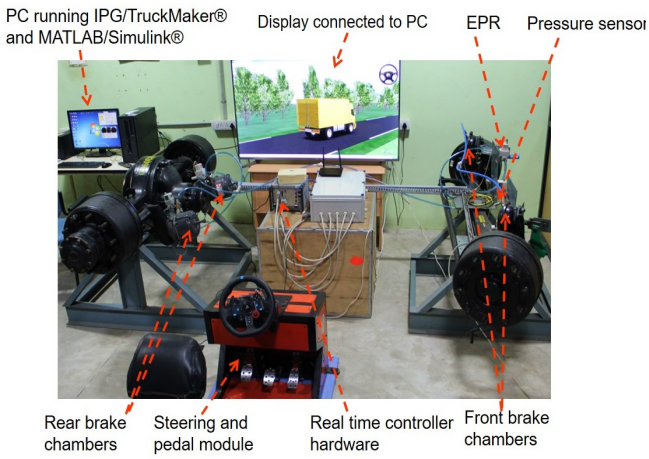

Fig. 5. HiL experimental setup.

The relative longitudinal speed between the lead vehicle and the host vehicle is given by

$$v_r(t) = v_l(t) - v_h(t). \tag{5}$$

The desired safe distance (s_d) that is to be maintained between the vehicles is

$$s_d(t) = hv_h(t) + s_o. \tag{6}$$

Here, h denotes the time headway and s_o represents the standstill distance between the two vehicles when the host vehicle comes to rest. Now, the spacing error between the vehicles is denoted by

$$e(t) = x_r(t) - s_d(t). \tag{7}$$

The controller for collision avoidance should work in such a way that, the desired spacing and zero relative longitudinal speed between the vehicles have to be maintained. In order to incorporate this aspect, the variable $\delta(t)$ is defined as

$$\delta(t) = e(t) + hv_r(t) = x_r(t) - (hv_h(t) - hv_r(t) + s_o). \tag{8}$$

Automatic braking would happen when $\delta(t) < 0$.

A Full State Feedback (FSF) controller has been designed in [26] to obtain the desired control input for collision avoidance. The control structure $u(t)$ is given by

$$u(t) = a_h(t) = k_1 v_r(t) + k_2 \delta(t). \tag{9}$$

Here, k_1 and k_2 are the feedback gains. The tuning procedure for these gain selection has been detailed in [26]. From $u(t)$, the brake force required for collision avoidance can be obtained using Eq. (1). In this algorithm, brakes would be applied only when the control input, $u(t)$ takes negative values (deceleration).

IV. Results and Discussions

The performance of the presented DDS integrated with CAS scheme has been evaluated through Hardware-in-Loop (HiL) experiments. The HiL setup used in this study is shown in Fig. 5. The setup consists of a 4X2 axle assembly of a 16-tonne truck which uses an Electro-Pneumatic Regulator (EPR). The hardware is interfaced with a high fidelity vehicle dynamic simulation software, IPG TruckMaker®, through which virtual vehicles, the road and collision scenarios have been simulated. The FSF-based collision avoidance algorithm has been programmed in MATLAB/Simulink®. It is then imported to a real-time controller (IPG Xpack4), which is interfaced with the IPG/TruckMaker® and the brake system hardware.

A. DDS and CAS Integration

In this work, the DDS output (Section II) is used to decide when to trigger the designed CAS. The integrated framework has been designed in such a way that whenever DDS is 1 and the CAA detects a potential threat, autonomous braking would be applied to stop the vehicle. Here, DDS states corresponding to the time series data for 62 s (provided using 620 images) are used in the experiments. The window between 16.5 s and 52 s (refer to Table I) during which DDS=1 has been used to introduce a collision threat. Hence, the HiL experiment is implemented from the 16^{th} second of the DDS data. The integrated algorithm was implemented on a vehicle of 16,200 kg, moving with a higher initial speed (90 km/h) on a dry (friction coefficient = 0.8) straight road. The efficacy of the integrated algorithm has been tested using two standard test maneuvres such as lead vehicle at rest and vehicle following scenarios, which are common test cases for evaluating a Collision Avoidance Algorithm. Lead vehicle at rest which can be experienced when the host vehicle observes any static object or a vehicle parked on the road. Vehicle following situation that happens when the lead vehicle suddenly begins to decelerate until it stops. This is a common scenario due to traffic condition or adverse weather condition. Both of these scenarios can cause fatal consequence or forward collision if the driver of the host vehicle is inattentive due to distractions.

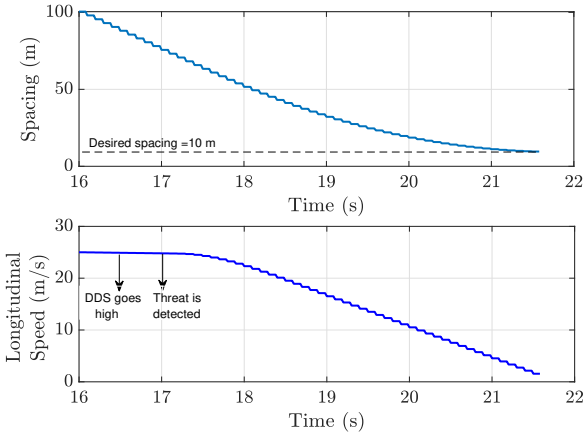

Fig. 6. Lead vehicle at rest case - Spacing and longitudinal speed profiles.

B. Lead Vehicle at Rest Scenario

In this test case, the host vehicle is assumed to be travelling at a speed of 25 m/s and it approaches toward a lead vehicle that is stationary. The DDS state goes high at 16.5 s of the time series data. The CAA detects the stationary vehicle as

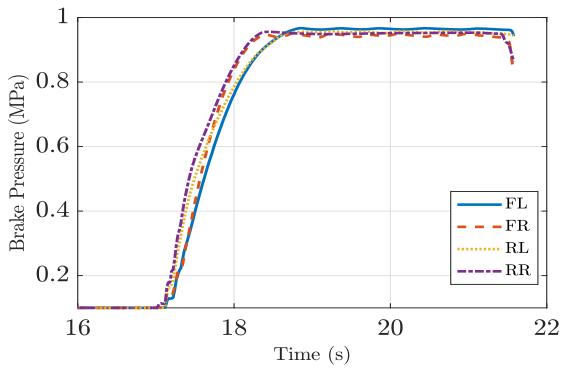

Fig. 7. Lead vehicle at rest case - Brake pressure (absolute).

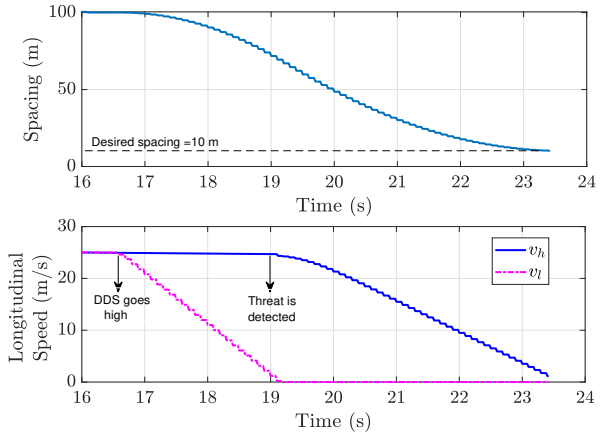

Fig. 8. Vehicle following case - Spacing and longitudinal speed profiles.

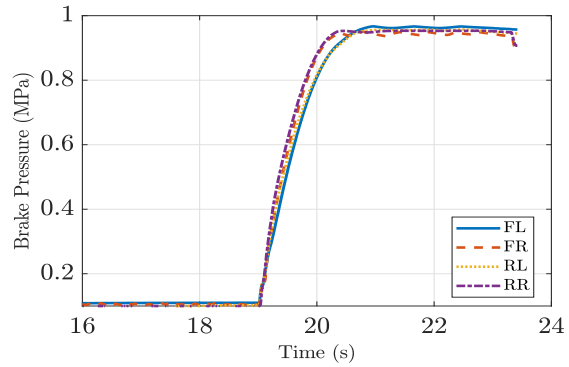

Fig. 9. Vehicle following case - Brake pressure (absolute).

a threat at 17 s and since the DDS is still high, automatic braking is triggered and the host vehicle is brought to rest. The plots showing the spacing between the vehicles and the longitudinal speed are shown in Fig. 6. The CAA has been designed such that the standstill spacing (s_o) between the vehicles is 10 m. It can be observed from the spacing profile that, the standstill spacing of 10 m has been maintained when the host vehicle came to rest. The corresponding brake chamber pressure profiles (control inputs) for all the four wheels are shown in Fig. 7. In this plot, FL, FR, RL and RR correspond to front-left, front-right, rear-left, and rear-right brake chambers, respectively. It is clear from Fig. 7 that the brake chamber pressure profiles reach the maximum value during this period i.e., from the threat detection (17 s) until the host vehicle comes to rest ($\approx$ 21.5 s). It is also evident that the host vehicle can successfully maintain the desired safe distance within 5 s since the threat is detected (Fig. 6). A similar result is also observed in the next test-scenario.

C. Vehicle Following Scenario

In this test case, the host vehicle is assumed to follow a moving lead vehicle. Both vehicles are assumed to have the same initial speed of 25 m/s, maintaining an initial distance of 100 m approximately between the two vehicles. Now, the lead vehicle has started decelerating at 16.5 s of the time series data. The DDS is also high at 16.5 s. The CAA detects the threat at 19 s and automatic braking is activated to decelerate the host vehicle. The corresponding spacing and speed profiles are shown in Fig. 8. In this test scenario also, the integrated algorithm could ensure the desired standstill spacing of 10 m. The corresponding brake chamber pressures for all the four wheels are shown in Fig. 9. The characteristics of brake chamber pressure profiles of the host vehicle from the threat detection (19 s) to come rest ($\approx$ 23.5 s) are similar to Fig. 7, as stated earlier.

In summary, the HiL experimental results on both benchmark scenarios show that the collision can be prevented by stopping the host vehicle maintaining the desired safe distance (s_o =10 m) which is constrained by the headway offset. This safe distance measure is considered as an assessment criterion of the proposed technique to avert a collision. The results clearly imply the practical feasibility of integrating DDS data with CAA to trigger autonomous braking, if necessary, for safer semi-autonomous vehicular operation.

V. Conclusion

A novel integrated approach using the human driving behavioral states and collision avoidance algorithm has been presented in this paper. The proposed approach could enhance the safety in intelligent transportation systems, especially for semi-autonomous vehicles. The distraction state of a driver is automatically assessed using an attention-driven CNN to initiate the CAS. The HiL experimental results demonstrated that the proposed system can be an effective measure to avert rear-end collision on different road scenarios using different states of the driver's in-vehicle activities.

The current study proposed a methodology to integrate CAS and DDS recognition and the efficacy of this methodology has been tested by evaluating two common test maneuvers for CAS such as lead vehicle at rest and vehicle following scenarios. However, realistic driving situations often involve more complex scenarios such as lane changing maneuvers, high traffic density operation, and also inclement environmental conditions. These would be explored as an interesting further enhancement of this integrated technique to evaluate its wider applicability. For example, the road

curvature and traffic conditions can influence the speed and acceleration of the vehicle. Also, other fine-grained distraction states would be incorporated and the computational efficiency of this approach would be analysed as another possible future direction.

ACKNOWLEDGEMENTS

This research was supported by the UKIERI-DST grant CHARM (UKIERI-2018-19-10). The authors thank the Ministry of Skill Development and Entrepreneurship, Government of India, for funding through the grant EDD/14-15/023/MOLE/NILE.

REFERENCES

[1] [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[2] J. Chen, Y. Jiang, Z. Huang, X. Guo, B. Wu, L. Sun, and T. Wu, "Fine-grained detection of driver distraction based on neural architecture search," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[3] [Online]. Available: https://sdgs.un.org/2030agenda.

[4] S. Kockum, R. Örtlund, A. Ekfjorden, and P. Wells, "Volvo Trucks Safety Report 2017," Volvo Trucks Accident Research Team, Gothenburg, Sweden, 2017.

[5] "Road Accidents in India - 2019," Ministry of Road Transport & Highways, New Delhi, India, July 2020.

[6] K. D. Kusano and H. C. Gabler, "Safety Benefits of Forward Collision Warning, Brake Assist, and Autonomous Braking Systems in Rear-End Collisions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1546–1555, Dec 2012.

[7] Y. Xiang, S. Huang, M. Li, J. Li, and W. Wang, "Rear-end collision avoidance-based on multi-channel detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3525–3535, 2019.

[8] S. Cheng, L. Li, H.-Q. Guo, Z.-G. Chen, and P. Song, "Longitudinal collision avoidance and lateral stability adaptive control system based on mpc of autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 6, pp. 2376–2385, 2019.

[9] V. R. S. Yellapantula, K. B. Devika, and S. C. Subramanian, "Communication latency and speed-dependent minimum time headway for connected heavy road vehicle collision avoidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4739–4748, 2019.

[10] C. Chen, X. Liu, H.-H. Chen, M. Li, and L. Zhao, "A rear-end collision risk evaluation and control scheme using a bayesian network model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 264–284, 2018.

[11] J. Funke, M. Brown, S. M. Erlien, and J. C. Gerdes, "Collision avoidance and stabilization for autonomous vehicles in emergency scenarios," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 4, pp. 1204–1216, 2016.

[12] D. Tran, J. Du, W. Sheng, D. Osipychev, Y. Sun, and H. Bai, "A human-vehicle collaborative driving framework for driver assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3470–3485, 2018.

[13] A. Behera, Z. Wharton, A. Keidel, and B. Debnath, "Deep cnn, body pose and body-object interaction features for drivers' activity monitoring," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[14] T. Ersal, H. J. Fuller, O. Tsimhoni, J. L. Stein, and H. K. Fathy, "Model-based analysis and classification of driver distraction under secondary tasks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 692–701, 2010.

[15] H. M. Eraqi, Y. Abouelnaga, M. H. Saad, and M. N. Moustafa, "Driver distraction identification with an ensemble of convolutional neural networks," *Journal of Advanced Transportation*, vol. 2019, 2019.

[16] A. Aksjonov, P. Nedoma, V. Vodovozov, E. Petlenkov, and M. Herrmann, "Detection and evaluation of driver distraction using machine learning and fuzzy logic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2048–2059, 2019.

[17] Z. Wharton, A. Behera, Y. Liu, and N. Bessis, "Coarse temporal attention network (cta-net) for driver's activity recognition," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1279–1289.

[18] W. Wang, X. Lu, P. Zhang, H. Xie, and W. Zeng, "Driver action recognition based on attention mechanism," in *2019 6th International Conference on Systems and Informatics (ICSAI)*. IEEE, 2019, pp. 1255–1259.

[19] M. Wollmer, C. Blaschke, T. Schindl, B. Schuller, B. Farber, S. Mayer, and B. Trefflich, "Online driver distraction detection using long short-term memory," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 574–582, 2011.

[20] B. Baheti, S. Gajre, and S. Talbar, "Detection of distracted driver using convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1032–1038.

[21] A. Behera, Z. Wharton, Y. Liu, M. Ghahremani, S. Kumar, and N. Bessis, "Regional attention network (ran) for head pose and fine-grained gesture recognition," *IEEE Transactions on Affective Computing*, 2020.

[22] A. Bera, Z. Wharton, Y. Liu, N. Bessis, and A. Behera, "Attend and guide (ag-net): A keypoints-driven attention-based deep network for image recognition," *IEEE Transactions on Image Processing*, vol. 30, pp. 3691–3704, 2021.

[23] J. D. Ortega, N. Kose, P. Cañas, M.-A. Chao, A. Unnervik, M. Nieto, O. Otaegui, and L. Salgado, "Dmd: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis," in *European Conference on Computer Vision*. Springer, 2020, pp. 387–405.

[24] P. Cañas, J. D. Ortega, M. Nieto, and O. Otaegui, "Detection of distraction-related actions on dmd: An image and a video-based approach comparison." in *VISIGRAPP (5: VISAPP)*, 2021, pp. 458–465.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[26] V. Rajaram and S. C. Subramanian, "Design and hardware-in-loop implementation of collision avoidance algorithms for heavy commercial road vehicles," *Vehicle System Dynamics*, vol. 54, no. 7, pp. 871–901, 2016.

[27] C. Ryan, F. Murphy, and M. Mullins, "End-to-end autonomous driving risk analysis: A behavioural anomaly detection approach," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[28] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[29] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 7354–7363.

[30] S. M. Kouchak and A. Gaffar, "Using bidirectional long short term memory with attention layer to estimate driver behavior," in *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2019, pp. 315–320.

[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[33] A. El Khatib, C. Ou, and F. Karray, "Driver inattention detection in the context of next-generation autonomous vehicles design: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4483–4496, 2020.

[34] Y. Xing, C. Lv, Z. Zhang, H. Wang, X. Na, D. Cao, E. Velenis, and F.-Y. Wang, "Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 95–108, 2018.