A Hybrid Deep Learning Based Visual System for In-Vehicle Safety

Rajkumar Joghee Bhojan, D. Ramyachitra, Subramanian Ganesan, Ragavi Rajkumar

Abstract—In the automotive industry, researchers, AI experts, and developers are actively pushing deep learning based approaches for In-vehicle safety. In this research paper, we propose a hybrid deep learning based visual system for providing feedback to the driver in a non-intrusive manner. We describe a hybrid SSD-RBM model for face feature identification. In this system, object detection, object tracking, and observations are processed through a full pipeline of image processing and detect the driver's movements and generate a safe and efficient action plan in real time. This in-vehicle interactive system assists drivers in regulating driving performance and avoiding hazards.

Index Terms—Computer Vision, Deep Learning, Driver Alert, In-vehicle Safety.

I. INTRODUCTION

Global Status report on road safety in 2018, the number of road traffic deaths continues to rise steadily, reaching 1.35 million in 2016. The report further says that Vehicle safety is increasingly critical to the prevention of crashes and has been shown to contribute to substantial reductions in the number of deaths and serious injuries resulting from road traffic crashes. [1]

According to the records NHTSA research, 2018 the most common problem traffic accident causes are:

- Frontal crashes, where vehicles driving in opposite direction have collided;
- Lane departure collisions, where a lane changing vehicle collided with a vehicle from an adjacent lane;
- Crashes with surrounding vehicles while parking, passing through an intersection or a narrow alley, etc.,
- Failures to see or recognize the road signalization [traffics signs and signs] and consequently cause a traffic accident due to inappropriate driving [2].
- The authors Steven Landry, et al., proposed an In-vehicle sonification [alert with no speech] prototyping tool for Driver State and performance Data.

In this paper, authors stated that automotive UI (User Interface) designers are searching for alternative avenues for delivering information to the driver that complement the mostly visual task of driving. Auditory channels provide the flexibility to display a wider variety of information to the driver without increasing the workload of driving task. It is

important to identify types of auditory displays and sonification strategies that provide integral information necessary for the driving task, and not overload the driver with unnecessary or intrusive data. Their system is intended to integrate driving performance data and driver affective state in real-time. Automobiles and in-vehicle safety systems have been improving over the past several decades [3].

Deep Learning:

Deep Learning is a specific subfield of machine learning: a new take on learning representation from data that puts an emphasis on learning successive layers of increasingly meaningful representations. In recent years, DL, RL, and deep RL methods are poised to reshape the future of ML [4]. The core concept of DL is to learn data representations through increasing abstraction levels. Almost in all levels, more abstract representations at a higher level are learned by defining them in terms of less abstract representations at lower levels. This type of hierarchical learning process is very powerful as it allows a system to comprehend and learn complex representations directly from the raw data [5], making it useful in many disciplines [6]. Several DL architectures have been reported in the literature, including deep neural network (DNN), RNN, convolutional neural network (CNN), deep auto encoder (DA), deep Boltzmann machine (DBM), deep belief network (DBN), deep residual network, deep convolutional inverse graphics network, and so on. In this research, we used Deep Learning methods such as SSD and DBM as they are efficient tools for object detection, object tracking and image segmentation of driver's movements.

II. COMPUTER VISION

The authors in “Practical Machine Learning with Python” state that Computer vision is all about the art and science of making machines understand high-level useful patterns and representations from images and video so that it would be able to make intelligent decisions similar to what a human would do upon observing its surroundings [7]. In computer vision, feedback has also played an important role in some vision tasks. For example, feedback was used to select the internal attention to achieve better object recognition performance. [8] Computer vision solutions are today in use in manned vehicles for improved safety or comfort. [9].

A. Driver Alert

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables.

Published on April 9, 2019.

R. J. Bhojan is with Wipro Technologies, USA.

D. Ramyachitra is with Bharathiar University, Coimbatore, India

S. Ganesan is with the Department of Electrical and Computer Engineering, Oakland University, Rochester, USA.

R. Rajkumar is with Sharon High School, MA, USA.

Driver's visual system is mostly busy with the driving task, designing active safety systems based on other sensory channels like auditory channels. According to multiple resource theory each task has a vector that shows the amount and qualitative level of the additional resources. Assuming that driving includes visual, spatial, and manual depends on the task demands. When there are no task demands to the driver, the driver tends to go to semi-sleeping state in long run driving. Therefore, it is important to identify the driver's visual state and strategies that can provide integral information necessary for the driving task, and not overload the driver with unnecessary or intrusive data.

III. METHOD

Driver alert system components are FC, CC, and BC as shown in Figure 1

A. Architecture

Fig.1. Driver Alert Design and Components

B. Front-end Component (FC):

FC module is deployed on the edge device. As shown in the top box in Figure 1, it consists of three submodules, which are image pre-processing (e.g., blurry image detection), face detectors, and the filters based segmentation. After the image pre-processing module, an original clear image is generated for segmentation. Next, the face detector, combined with different filters to segment the original image. After segmentation, we can generate a clear and segmented image. These images will be transferred to the server via the Communication Module (introduced below) for further classification.

C. Communication Component (CC):

CC provides two channels for communication between the FC and the Back-end Component (BC). It transfers the image data from the FC to the BC via Input Channel, and it also passes the detection results from the BC to the FC via Output Channel.

D. Back-end Component (BC):

The BC module runs on the cloud server, which is configured to use OpenCV2 (an open source deep learning framework) for SSD model training and testing. OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products [10]. Then the trained model is deployed on the server and used for classifying the image. More specifically, the segmented image is first passed through our proposed Convnet, then the features are generated from the model.

IV. HYBRID APPROACH

To characterize face similarities from different aspects such as fatigued or not, we concatenate the features extracted from different face region pairs by different supervised learnings such as Decision tree and random forest methods. The resulting high-dimensional relational features are classified by Supervised Learnings for face verification. In Supervised training, we started to use a pre-trained model from AffectNet and identified facial state of the driver. The features for identifying the facial states such as, A Decision Tree (DT) is a binary classifier resembles a tree in which the features are represented by nodes and the edges leaving a node are labeled by the feature weight and leaves represent the categories. Decision Tree algorithm stops when no other selection is made [11]. The decision tree finally gives an output as either the subject is fatigued or tired and this result will be inputted to Convnet analysis.

In the Convnet, we inputted the data we want to use (images, non-performance activity, etc.) and the data gets passed through different "layers" of the net. Each layer modifies the input values to try and morph it into something useful and predictive in the model.

EigenFaces face recognizer views at all the training images of all the characters as a complex and try to deduce the components. These components are necessary and helpful and discard the rest of the images, in this way it not only extracts the essential elements from the training data but also saves memory by rejecting the less critical segments. After pre-training each Supervised learning and the ConvNet separately, the entire hybrid network is jointly optimized to further improve the accuracy.

A. RBM Learning

The Boltzmann machine proposed by Geoffrey Hinton [12] and colleagues in 1983, is a well-known example of a stochastic neural network that can learn internal representations and solve combinatorial optimization problems. The Boltzmann machine is a fully connected network comprising two-state units. It employs simulated annealing for transitioning between the possible network states. The units flip their states on the basis of the current state of their neighbors and the corresponding edge weights to maximize a global consensus function, which is equivalent to minimizing the network energy. According to authors in [13] A Restricted Boltzmann Machine is an undirected graphical model with stochastic visible variables

$k \in \{0, 1\}$ and stochastic hidden variables $h \in \{0, 1\}$, where each visible variable is connected to each hidden variable. An RBM is a variant of the Boltzmann Machine, with the restriction that the visible units and hidden units must form a bipartite graph. This restriction allows for more efficient training algorithms, in particular the gradient-based contrastive divergence algorithm [14].

The authors describe facial emotion recognition fundamentally identifies emotion which shapes how humans' self-control and reaction based on the situation as well as the environment to which they belong. This research further explains that Skybiometry is considered to be a state of the art in recognizing and detecting facial expressions. Authors Mollahosseini, et al., made a dataset "AffectNet" for facial expressions [15]. AffectNet serves as the largest database of facial expressions, valence, and arousal represented in two different emotion models. With the help of evaluation metrics, deep neural network baselines can perform better than the conventional learning methods [16]. In this research paper, we used AffectNet dataset for identifying driver's facial identification (especially fatigues) while the driver is driving. Manually labeling datasets with object masks is extremely time consuming [17].

B. Single Shot Detector (SSD):

The authors in [18] describe that Single Shot Multibox Detector (SSD) discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. The authors further illustrate that SSD at prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, this computer vision network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. As SSD network works on a single network principle, it completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation. Hence the single network makes SSD easy to train and integrate into systems that require a detection component. The authors in [18] further reassure that Experimental results on the PASCAL VOC, COCO, and ILSVRC datasets confirm that SSD has competitive accuracy to methods that utilize an additional object proposal step and is much faster, while providing a unified framework for both training and inference [18]. The architecture of SSD is given in the below Fig. 2.

Fig.2. SSD Architecture

V. MODEL DEVELOPMENT

A. Camera

In this research, a camera is mounted facing front side of the driver's seat nearby dashboard so that it will capture driver's movement in real time. The camera is used mostly for object detection, recognition and tracking for example, driver's eye movement, mouth yawning, face movement and body movement.

B. Object Detection

Object detectors have become significantly more accurate and gained important new capabilities [19]. As stated by Shaoshan Liu, et al [20] in recent years, we have seen the rapid development of vision-based deep-learning technology, which enables highly accurate object detection and tracking. The Convolution Neural Network (CNN), a deep neural network (DNN) that is widely used in object-recognition tasks, has a four-layer evaluation process:

- The convolution layer contains a variety of filters to extract features from the input image. Because each filter contains a set of learnable parameters that will be derived after the training stage, the convolution layer can be the most computationally intensive layer in the CNN.
- The activation layer has mechanisms for deciding whether to activate the target neuron.
- The pooling layer reduces the representation's spatial size to reduce the number of its parameters and consequently the required computation.
- The fully connected layer contains neurons with full connections to all activations in the pooling layer.

The authors in [21] are further stated that Prior works have shown that with a large amount of annotated data, convolutional neural networks (CNNs) can be trained to achieve a super-human performance for various visual recognition tasks. As tremendous efforts are dedicated into the discovery of effective network architectures and training methods for further advancing the performance, it is also important to investigate into effective approaches for data annotation as data annotation is essential but expensive.

In this research, we used a single shot detector for an easy and lightweight model. The SSD network works on feedforward convolutional network which make fixed -sized bounding boxes and scores based the object detection. As shown in Fig. 3, driver's face and movements are captured in real time using the camera mounted on the dashboard.

Fig.3 Object Detection – The Subject is in driving state.

C. Object Tracking

When the subject (Driver) is fatigued or in a semi-sleep state, there is a chance of falling either in front side or sideways as shown in Figure 4. In this type of situation, the

Algorithm for Driver's alert will be used for body tracking. The SSD algorithm is used to capture driver's movements continuously without any disturbances. The main goal of our algorithm is to train an object detector that takes video image as input. To train such an object detector, the training and validation images of the detector are annotated with a bounding box per object and its category. Such an annotation is commonly seen in public datasets including PASCAL VOC and MS COCO Object tracking aims to automatically track the moving object's trajectory and send the results to decision making. The main goal is to ensure that the vehicle does not collide with a moving object, whether a vehicle or a person crossing the road.

Fig. 4 Object Tracking - The subject is in a semi-sleep state.

D. Model Training

The SSD training will be used to handle multiple object categories. In order to deploy AI-based solutions to Driver Alert, we need to follow three major building blocks: data preparation, model generation, and model deployment. Data preparation focuses on getting data ready for training and testing neural networks, covering topics such as data recording, ground truth labeling, big data storage, etc. Model generation involves developing network architectures, training the networks, and evaluating the trained models. A model is considered "trained" if the difference between its outputs and the corresponding ground truth labels (the expected outputs)

E. Decision-making:

At this stage, a decision is made by the "Driver Alert" and it starts functioning as per the conditions. The first condition will be a speaker based alert to the driver and the second condition will be shaking the backrest of the driver's seat.

VI. EXPERIMENTS

In our research, as shown in the flowchart (Fig. 5), video images are captured by a camera which is mounted on the Vehicle's dashboard.

Fig.5. Flow Chart for Driver Alert

The algorithm developed for Driver's Alert keeps watching driver's movement by both object detection and object tracking methods. If the driver is fatigued, it checks for next condition where "Yawning" is happening, if this condition is also happening, the algorithm looks eye movement and body movements by the detection method. The trained model works as if an agent sitting in front of the driver and make the decision. When Driver's Alert predicts that there is a need for alert, it will invoke a voice-based alert or it will shake the driver's seat.

VII. CONCLUSION

Though there are so many researchers have studied the fatigue detection, fundamental problems is to solve avoid road hazards. This paper has proposed a new hybrid SSD-RBM model for driver's face identification and verification. The model learns directly and jointly extracts relational visual features from eye movement and body movement under the supervision of face identities. Both feature extraction and recognition stages are unified under a single deep network architecture and RBM. The components are jointly optimized for the target of face verification and finally communicated to Driver Alert component.

ACKNOWLEDGMENT

Authors would like to thank project stakeholders for their valuable support in collecting the data. Author Rajkumar is immensely thankful to Dr. D. Ramyachitra, who provided insight and expertise that greatly assisted this research paper. He extends his thanks to Dr. Subramaniam Ganesan who has spent a lot of his time for correcting this research paper and his constant encouragement and support throughout this work. He also extends his gratitude to Ragavi Rajkumar for her support in this paper.

REFERENCES

- [1]. Global Status Report on Road Safety 2018.
- [2]. <https://www.nhtsa.gov/research-data>
- [3]. Steven Landry, Myounghoon, Pavlo, Jan Hammerschmidt, "Report on the in-vehicle auditory interactions workshop_ taxonomy, challenges, and approaches", AutomotiveUI, 2015.
- [4]. V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.
- [5]. Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
- [6]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [7]. Sarkar, Dipanjan, "Practical Machine Learning with Python - A Problem-Solver's Guide to Building Real-World Intelligent Systems", 2016, Apress publications.
- [8]. Haiguang Wen Kuan Han, Junxing Shi, Yizhen Zhang, Eugenio Culurciello, Zhongming Liu, "Deep Predictive Coding Network for Object Recognition", *International Conference on Machine Learning (ICML)*, 2018.
- [9]. Reinhard Klette, "Computer Vision in Vehicles - A Brief Introduction", *Technical Report*, February 2014, ResearchGate.
- [10]. <https://opencv.org/about.html>
- [11]. Rajkumar Joghee Bhojan, Ramyachitra, Vivekanandan, Subramaniam Ganesan, "A Machine Learning Based Approach for Detecting Non-Deterministic Tests and its Analysis in Mobile Application Testing", *International Journal of Advanced Research in Computer Science*, ISSN No. 0976-5697, pp. 220-223.
- [12]. Geoffrey Hinton, "A Practical Guide to Training Restricted Boltzmann Machines", 2010.
- [13]. D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, pp. 106–154, 1962.
- [14]. Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis and Effychios Protopapadakis, "Deep Learning for Computer Vision: A Brief Review", *Computational Intelligence and Neuroscience*, Vol 2018, Hindawi.
- [15]. Ali Mollahosseini et al. AffectNet: A Database for facial expression, Valence, and Arousal Computing in the Wild Department of Electrical and Computer Engineering, University of Denver, Denver, CO, 80210, 2015.
- [16]. Mirafre R Prospero, Edson B Lagmayo, Anndee Christian L Tumalak, "Skybiometry and AffectNet on Facial Emotion Recognition Using Supervised Machine Learning Algorithms", *ICCCV 2018*, Singapore.
- [17]. David Acuna, Huan Ling, Amlan Kar, Sanja Fidler, "Efficient Interactive Annotation of Segmentation Datasets with Polygon-RNN++", *CVPR*, 2018.
- [18]. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Chen-Yang Fu, Alexander C. Berg, "SSD: Single Shot MultiBox Detector", *arXiv*, 2016.
- [19]. Ronghang Hu, Piotr Dollar, Kaiming He, Trevor Darrell, Ross Girshick, "Learning to Segment Every Thing", *Computer Vision and Pattern Recognition*, 2018.
- [20]. Shaoshan Liu, Jie Tang, Zhe Zhang, Jean-Luc, "Computer Architectures for Autonomous Driving", *Computer*, Volume: 50, Issue: 8, 2017, IEEE Computer Society, DOI:10.1109/MC.2017.3001256.
- [21]. Chieh-Chi Kao, Teng-Yok Lee, Pradeep Sen, Ming-Yu Liu, "Localization-Aware Active Learning for Object Detection", *arXiv:1801.0512v1*, Jan 2018

AUTHOR BIOGRAPHIES

Rajkumar Joghee Bhojan is a seasoned Architect and a recognized thought leader with Wipro Technologies. He has two decades of professional experience both in academics and IT industries. He has executed projects in diverse geographies including Australia, India, and the USA. He holds M.Sc., in Physics, MCA. and M. Phil (CS). He received "Best Thought Leader"

award from Wipro Technologies. He is a research Scholar and his research areas include Deep Learning, AI and Software Engineering. He has published and presented technical papers in international conferences and journals. He is a patent holder and he has filed a couple of Patents in USPTO. He has worked as a Solution Architect, Corporate Trainer, QA Manager, Scrum Master, and Senior Test Consultant. He spoke at IEEE Forums, ATAGG, STPCon, Google Tech Talks, etc.

Dr. D. Ramyachitra is working as an Assistant Professor in Bharathiar University, Coimbatore. She has 18 years of experience. She has published more than 50 articles in reputed journals and presented papers in more than 50 National and International conferences. Her area of interest includes Grid computing, Image processing and Data mining.

Dr. Subramaniam Ganesan, is a Professor in the department of Electrical and Computer Engineering, Oakland University, Rochester, MI 48309, USA. He has over 30 years of teaching and research experience in Digital Computer systems. He was the chair of the CSE department from 1991 to 98. He has published over 100 journal papers, more than 200 papers in conference proceedings, and 3 books. He published a book on Java in 2003. He developed a custom DSP board with software for his DSP book. He is a senior member of IEEE, IEEE Computer Society Distinguished Visiting Speaker, IEEE Region 4 technical activities member and Fellow of ISPE. He received Life time Achievement award from ISAM, Lloyd L. Withrow Distinguished Speaker award from SAE, Best Teacher award from ASEE, and Oakland University. He has organized many international conferences. He is the editor in chief of an International Journal of Embedded system and Computer Engineering and International journal of Sensors and applications. He is the session organizer on "Systems engineering" at SAE world congress for the past 15 years. His research interests are in Real time system, parallel architectures and computer systems, Automotive embedded systems security and signal processing.

Ragavi Rajkumar is a senior student from Sharon High School, MA, USA. She has received "Student of the Term" and "Seal of Bilingualism in Hindi" awards from the school. Her academic interests are Data science and Biotechnology.