IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 18 July 2023, accepted 23 August 2023, date of publication 5 September 2023, date of current version 11 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312190

RESEARCH ARTICLE

Neuromorphic Driver Monitoring Systems: A Proof-of-Concept for Yawn Detection and Seatbelt State Detection Using an Event Camera

PAUL KIELTY[1], MEHDI SEFIDGAR DILMAGHANI[1], WASEEM SHARIFF[1,2], CIAN RYAN[2], JOE LEMLEY[2], AND PETER CORCORAN[1], (Fellow, IEEE)

[1]Department of Electronic Engineering, College of Science and Engineering, University of Galway, Galway, H91 TK33 Ireland
[2]Sensing Team, Xperi Inc., Galway, H91 V0TX Ireland

Corresponding author: Paul Kielty (p.kielty3@universityofgalway.ie)

This work was supported by the Science Foundation Ireland at ADAPT, the SFI Research Centre for Artificial Intelligence (AI)-Driven Digital Content Technology, University of Galway, through the SFI Research Centers Program under Grant 13/RC/2106_P2.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Xperi Research Ethics Committee.

ABSTRACT Driver monitoring systems (DMS) are a key component of vehicular safety and essential for the transition from semi-autonomous to fully autonomous driving. Neuromorphic vision systems, based on event camera technology, provide advanced sensing in motion analysis tasks. In particular, the behaviours of drivers' eyes have been studied for the detection of drowsiness and distraction. This research explores the potential to extend neuromorphic sensing techniques to analyse the entire facial region, detecting yawning behaviours that give a complimentary indicator of drowsiness. A second proof of concept for the use of event cameras to detect the fastening or unfastening of a seatbelt is also developed. Synthetic training datasets are derived from RGB and Near-Infrared (NIR) video from both private and public datasets using a video-to-event converter and used to train, validate, and test a convolutional neural network (CNN) with a self-attention module and a recurrent head for both yawning and seatbelt tasks. For yawn detection, respective F1-scores of 95.3% and 90.4% were achieved on synthetic events from our test set and the "YawDD" dataset. For seatbelt fastness detection, 100% accuracy was achieved on unseen test sets of both synthetic and real events. These results demonstrate the feasibility to add yawn detection and seatbelt fastness detection components to neuromorphic DMS.

INDEX TERMS Driver monitoring, drowsiness detection, event camera, computer vision, CNN, LSTM, neuromorphic sensing, seatbelt, yawn.

I. INTRODUCTION

Drowsy driving is one of the leading causes of motor accidents globally, increasing a driver's risk of an accident by a factor of 5 or more compared to when they are alert [1]. In the past decade, a great deal of study has been dedicated to the development of level 5 autonomy, or completely autonomous driving [2], [3], [4], [5], [6]. Until we reach this level, monitoring the driver for signs of drowsiness and other unsafe driving behaviours can save many lives in non-autonomous and the semi-autonomous vehicles of today. To this end, DMS utilising various technologies have been proposed. One of these systems proposed by Khan et al. [7] employs an RGB camera in an IoT-based automated system to monitor drivers and detect drowsiness. While their work and others [8] discusses blink behavior as an indicator of drowsiness, a frame rate of at least 100 frames per second (FPS) is required for accurate blink detection [9], which is

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Quan.

beyond the capabilities of the 30-60 FPS cameras typically found in current DMS.

The introduction of neuromorphic vision sensors promises a new era for DMS, by addressing a number of the hardware limitations of conventional RGB and near-infrared (NIR) systems, including low frame rate, power consumption, and low-light performance. The neuromorphic vision sensors used in event cameras are designed to mimic the visual-processing abilities of living objects by only gathering the relevant data from an observed scene. Instead of using a conventional shutter-based technique to capture an image, they report an event anytime a pixel in the sensor detects a change in brightness above a certain threshold. Each event is defined by four parameters: the timestamp, the x and y coordinates of the pixel that reported the event, and the polarity, which indicates whether an increase or decrease in brightness caused the event. As events are typically only generated by motion or changes in lighting, event cameras are extremely useful in motion analysis tasks.

These modifications also enable event cameras to offer a wider dynamic range, higher temporal resolution, and lower power consumption than conventional cameras. Events are recorded with an accuracy of one microsecond and can provide equivalent frame rates exceeding 10,000 FPS [10]. These properties, and the parameters that can be modified to control the output event streams [11] for operation in various lighting conditions, make the event cameras highly suited to the various requirements of DMS. This has already been demonstrated by Ryan et al. [12]. with an event-based DMS capable of real-time face and eye tracking, and blink detection as indicator of drowsiness. This could be combined with other symptoms of tiredness, such as yawns, for more accurate predictions of driver exhaustion levels.

Drowsiness detection is critical when considering driver safety, however there are few measures as simple and effective as the seatbelt. The risk of injury to a belted passenger is 65% lower than that of an unbelted passenger [13], and in the United States, seatbelt use was shown to reduce mortality by 72% [14]. Existing seatbelt alert systems that rely on under-seat pressure sensors are easily spoofed, and provide no assurance of if the seatbelt is correctly fastened. This makes seatbelt fastness detection another desirable feature of camera-based DMS. Systems that can recognise seatbelt use, even from surveillance footage outside the car, have been made possible with deep learning approaches [15]. These techniques typically use RGB or NIR frames, often with some form of edge detection pre-processing [16], however, a correctly calibrated event camera can similarly isolate edges and other scene elements without additional processing [11]. DMS that already utilise event cameras could incorporate seatbelt fastness detection with no added hardware costs.

This research expands on our previous work of developing a proof-of-concept event-based yawn detection system [17] and combines it with a seatbelt fastness detection algorithm. Large datasets of synthetic events were simulated for developing these algorithms, and a set of real events was collected for testing in addition to publicly available data. The network architecture designed for both tasks combines a CNN backbone with self-attention module and a recurrent head. Highly accurate models with very low inference times were achieved, allowing real-time operation of both yawn detection and seatbelt fastness detection.

The remainder of this paper is organised as follows: Section II examines related research in the spaces of yawn detection, seatbelt detection, and event based DMS. In Section III we outline our network architecture, followed by the datasets, event processing, and training details for both yawn detection and seatbelt state detection tasks. In Section IV we present our results and compare them to others in literature. Section VI contains our final conclusions of the work and its implications.

II. RELATED WORK

In this section we discuss the current literature related to the two safety features developed in this work. Yawn detection is a key indicator of driver drowsiness and the detection of seatbelt is an important component of passenger safety. By implementing and validating these two safety features this work demonstrates the general feasibility of replacing a conventional RGB or NIR based DMS that employs conventional computer vision algorithms with a fully neuromorphic DMS.

A. YAWN DETECTION

Driver drowsiness is a critical factor in road accidents, and various studies have explored yawning as a key indicator for detecting drowsiness. Abtahi et al. propose a real-time system using face and mouth detection for accurate yawning measurement and drowsiness detection [18]. Omidyeganeh et al. present a computer vision-based system that significantly improves yawning detection rates by using a modified implementation of the Viola-Jones algorithm and backprojection theory [19]. Knapik and Cyganek introduce a novel approach utilising thermal imaging for driver fatigue recognition based on yawning, demonstrating high efficacy in both laboratory and real car environment [20]. Yang et al. propose a subtle facial action recognition method for yawning detection, utilising a 3D deep learning network and a keyframe selection algorithm to distinguish yawning from similar facial actions [21]. Liu et al. design a multimodal fatigue detection system that combines eye and yawn information, achieving a high accuracy rate of up to 95% in detecting drowsiness [22]. Kumari et al. develop a real-time drowsiness and yawn detection system using Python and the Dlib model, based on eye closure and yawn frequency, to minimise fatigue-related vehicle accidents [1]. Dehankar et al. propose a non-invasive driver drowsiness and yawning detection system using computer vision techniques and a Raspberry Pi microcontroller, achieving rapid fatigue detection within a few seconds [23]. Alshaqaqi et al. introduce a driver drowsiness detection system that computes the eye aspect ratio and lip distance to determine drowsiness and yawning, aiming to reduce accidents caused by driver fatigue [24]. Melvin et al. propose a novel approach based on facial motion identification using convolutional neural networks, addressing challenges in accurate yawning recognition in real-world driving conditions [25]. These studies collectively contribute to the development of effective driver drowsiness detection systems by leveraging yawning as a prominent indicator, aiming to enhance transportation safety and mitigate accidents caused by drowsy driving.

To the best of the authors' knowledge, there is only one prior study that focuses directly on event-based driver yawn detection [17]. This work, in part, serves as an extension of that study, aiming to further explore the potential of neuromorphic sensing techniques for yawn detection in driver drowsiness. Our research utilises a neuromorphic vision system, leveraging event camera technology, to analyse the entire facial region and capture yawning behaviours. This provide a complementary indicator of tiredness to enhance. A dataset comprising 952 video clips and corresponding neuromorphic image frames is constructed and used for training and testing a CNN with self-attention and a recurrent head.

Event-based yawn detection offers several advantages, including the ability to capture micro-facial movements that indicate the onset of yawning. By focusing on specific yawn events, rather than continuous monitoring, this approach reduces computational requirements and enhances the accuracy of detection. Additionally, event-based yawn detection enables the identification of subtle variations in yawning patterns, allowing for a more refined analysis of driver drowsiness levels. This innovative approach holds great potential in improving the effectiveness and efficiency of driver drowsiness monitoring systems, ultimately contributing to enhanced road safety.

B. SEATBELT DETECTION

Seatbelt detection is a crucial task in the automotive industry to ensure driver and occupant safety. Current technology primarily focuses on buckling detection, but proper seatbelt routing detection to ensure the seatbelt is safely routed through the body to protect the wearer, remains a challenge.

Baltaxe et al. [26] addressed the problem of marker-less vision-based detection of improper seatbelt routing. They trained deep neural networks using a large database of images and achieved high accuracy in classifying seatbelt routing scenarios. This work contributes to improving automotive safety by reducing injuries caused by improperly routed seatbelts. Chun et al. [27] proposed NADS-Net, a light architecture for driver and seatbelt detection using convolutional neural networks. Their architecture, based on the feature pyramid network backbone, showed optimal performance for driver/passenger state detection tasks.

Authors in [16] presented a classification model for driver seatbelt status detection based on image analysis from a vehicle's in-cabin camera. They utilised a YOLO neural network and a two-step approach to detect the main part of the belt and its corner. The model achieved accurate classification of belt fastness, including cases where the belt is fastened behind the human body. Naik et al. [28] proposed a technique using convolutional neural networks (CNN) to detect driver's seatbelt usage. Their ConvNet achieved higher accuracy compared to other classification algorithms and demonstrated the potential for reducing accidents caused by non-compliance with seatbelt usage.

Authors in [29] focused on the automatic vertical height adjustment of incorrectly fastened seatbelts using deep learning. They evaluated three CNN architectures and found that DenseNet121 achieved the highest classification accuracy. Their proposed system provides a solution for ensuring correct seatbelt positioning, thereby enhancing driver and passenger safety in fleet vehicles. Hosseini and Fathi [15] proposed a deep learning-based system for detecting vehicle occupancy and driver's seatbelt status. Their method employed a combination of pre-trained ResNet34 and power mean transformation layers, achieving high accuracy in detecting occupants and seatbelt violations. The proposed system demonstrates promising performance compared to state-of-the-art methods.

Madake et al. [30] addressed seatbelt detection for assisted driving scenarios. They proposed a real-time system using a combination of FAST key point detection, BRIEF method, and Decision Trees. Their algorithm showed high classification accuracy, considering practical constraints such as dynamic environments, illumination variations, and low-quality images. Authors in [31] presented an efficient and lightweight model for seatbelt detection on mobile devices. They pruned the SSD MobileNet V2 model and utilised the LSD linear segment detection multipoint fitting algorithm to enhance detection performance. Their model outperformed existing methods, demonstrating its practicality for mobile-based seatbelt detection. Upadhyay et al. [32] proposed a real-time seatbelt detection system using the YOLO deep learning model. They emphasised the importance of monitoring seatbelt fastening in automobiles and addressed the limitations of existing algorithms. Their YOLO-based model achieved accurate seatbelt detection, contributing to automotive safety by ensuring proper seatbelt usage.

Although there are many prior research works relating to seat belt detection, we believe that this work is the first to explore the potential of event cameras to monitor and verify seatbelt state and fastening activity. More specifically we are interested in the potential for neuromorphic sensing to better evaluate the correct completion of the fastening/unfastening process. Due to potential differences in how event cameras features can be leveraged for the prediction of fastening/unfastening actions against a stationary fastened/unfastened seatbelt, this paper investigates them

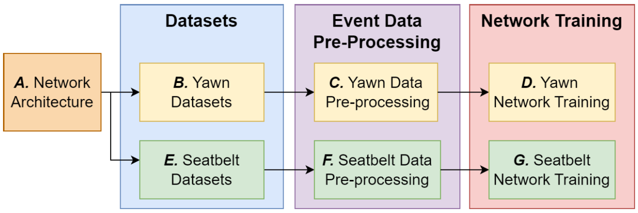
Datasets
Event Data Pre-Processing
Network Training
A. Network Architecture
B. Yawn Datasets
C. Yawn Data Pre-processing
D. Yawn Network Training
E. Seatbelt Datasets
F. Seatbelt Data Pre-processing
G. Seatbelt Network Training

FIGURE 1. Overview of methodology structure.

as two distinct tasks to inform which method should be prioritised.

III. METHODOLOGY

In this section we present the details on our network design, our collection of video datasets and the subsequent generation of synthetic events, followed by the tailored preprocessing of our event data for our different tasks, and finally the training details of our various models. Fig. 1 gives an overview of this section's structure. All of the data used in this paper was collected with informed consent and in compliance with ethical guidelines.

A. NETWORK ARCHITECTURE

The possible manifestations of yawns are frequently oversimplified in yawn-detection literature, where mouth openness is often assumed to be the only relevant feature. This is unreliable when assessed over individual frames or short time windows, as there is a risk of false positive predictions when the mouth is open for speech or laughter. An additional flaw, which is extremely challenging to solve in these systems, is not handling the common case where a person reflexively covers their mouth with their hand when yawning. Some approaches also monitor the openness of the eyes, but there is little consideration of other possible cues that often accompany a yawn, such as the hand over the mouth or large stretches of the upper body and arms. For this reason, our proposed yawn detector does not use facial landmarks or other deliberately programmed features to make a prediction. Instead, we rely on CNN components to learn the relevant features from the full input images, with a recurrent structure that can track how these features change over time in a yawn.

Similar principles can be applied when designing a network to predict seatbelt state. When viewing an individual frame from a video of someone buckling their seatbelt, there is no information on the direction of motion or previous states, and so it can be easily confused with an unbuckling action, whereas a sequence of frames makes is much easier to identify. Additionally, a fastened seatbelt does not typically undergo a lot of motion when the wearer is sitting still. For event cameras this can result in moments with very little information on the seatbelt. By extending the input sequence, we provide more time to gather information on the seatbelt to obtin for a more reliable prediction.

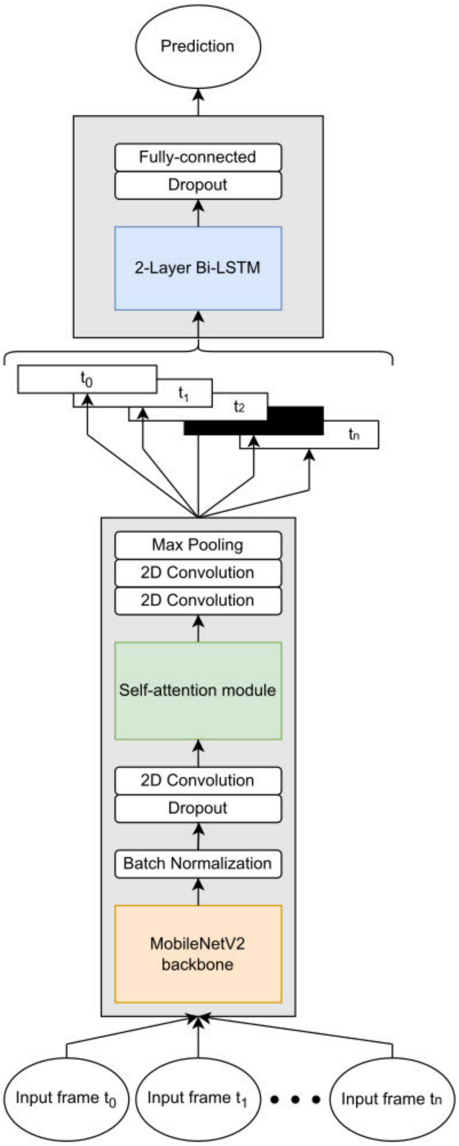

FIGURE 2. Our proposed network for yawn and seatbelt detection.

Fig. 2 gives a high-level overview of the model architecture designed for this paper. The MobileNetV2 network is used for feature extraction of the input frames. In their paper, Sandler et al. [33] demonstrate the impressive performance of MobileNetV2 as a feature extractor with an efficient, lightweight architecture. The model we used was pretrained on the ImageNet dataset [34]. After this initial feature extraction, batch normalisation and channel reduction by 2D convolution are applied to prepare the features for a self-attention module. Recent years have seen self-attention introduced to many CNN tasks for its ability to contextualise and apply a weighting to input features, with only a small computational cost. The self-attention module in our proposed network is implemented according to [35]. Fig. 3 gives the expanded diagram of this module.

When the attended feature maps are generated for every frame of the input sequence, they are stacked and passed

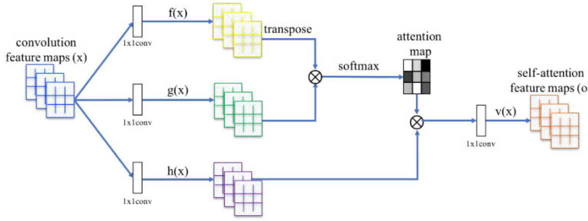

FIGURE 3. An expanded diagram of the self-attention module by [35].

to the recurrent head of the network. This is comprised of a 2 stacked bi-directional LSTM layers [36]. The LSTM's ability to retain information over a longer temporal range aligns with our needs, as both yawns and seatbelt state can be more accurately predicted over longer sequences. This also grants flexibility for inputs of different lengths, which can prove particularly useful when creating frames from a sequence of event data, as there is a vast range of possible representations which can vary the number of frames. For each frame sequence, the final output of the LSTM layers is flattened and passed to a fully connected layer to generate the output prediction. A dropout step was added before this fully connected layer to address potential overfitting on the training data.

B. YAWN DETECTION—DATASETS

A non-public industry dataset for driver drowsiness was collected by recording participants in a driving simulator at fixed times over a 24hr period. These participants were required to not consume any stimulants 12 hours prior or throughout the acquisition. They were also required to stay awake from the start time of 8AM until the acquisition was completed. The recordings used for yawn detection took place at 5PM, 2AM, and 5AM. Each of these recordings contains one hour of video captured with a Logitech Brio camera positioned behind the steering wheel. The audio of each session was also recorded and later annotated by a team within Xperi with the start time and duration of all yawns. The yawn audio annotations can be mapped to the RGB frames as all of the frames are timestamped, however, because of differences between the audible and visible cues of yawns, we cannot map frame-precise labels. For example, many yawns are mostly silent for while the mouth is opening and only become audible for a large exhale at the end. Using these audio timestamps will only label the frames for the audible portion. Variations in individual yawns causes an irregular misalignment between the audio annotations and desired frame annotations. With no feasible method to achieve frame-wise labels, the dataset and network were designed to assign a single label to a sequence of frames, denoting if a yawn occurred anywhere in the clip. The yawn audio labels have a mean duration of 4.03s with a standard deviation of 2.26s and a maximum of 9.63s. The timestamps of each yawn sample were extended to 10s and the frames over this new duration were extracted to create each sample in our RGB video yawn dataset. This duration guarantees that the entire yawn was captured in the frames, despite the misalignment of the audible and visible yawn components. A set of non-yawn sequences was created by adding a 10s offset to the end of each yawn sequence and saving another 10s of frames, provided that these frames did not collide with a subsequent yawn sequence. The video was specified to be collected at 30 frames per second giving an expected 300 frames in each sample, however, the frame rate was typically lower in the AM sessions due to an increased exposure time for the darker scenes. In the final set of RGB yawn sequences, 48.6% had fewer than the 300 frames. The frame counts of these shorter sequences only have a mean and standard deviation of 203.63 frames and 48.52 frames respectively.

The public YawDD dataset [37] was also used for testing our yawn detector. This dataset is comprised of videos taken from both the dashboard and rear-view mirror of a car. The rear-view mirror videos were not included in our experiments as our proposed system uses a camera behind the steering wheel, making the camera position of the dashboard videos more appropriate for testing.

C. YAWN DETECTION—EVENT SIMULATION AND PRE-PROCESSING

A frequent blocker in neuromorphic vision research is a lack of large-scale public datasets. This has led to the development of event simulators such as V2E [38]. This enables the synthesis of realistic events from RGB video using the differences between successive frames. By including each frame's timestamp at simulation time, we can ensure the simulated events are distributed over the time span of the source frame pair. This is particularly important in our yawn dataset where the framerate slows as the scene gets darker. The RGB frames were cropped to a 500 × 500 area containing the face before simulation. Our event data are initially saved as lists of individual events in text format but to use this data in CNNs and other image-based systems, it must first be represented in a 2D array or frame. This is typically achieved by accumulating a group of events and summing the positive and negative events at each pixel location to create a 2D frame [10]. When transforming an event recording into frames with this technique, the decision of how many events should be accumulated per frame must be carefully considered. The two most common approaches are to accumulate events over a fixed duration or accumulate a fixed number of events for each frame. The former method of grouping the events by a fixed duration is useful in tasks that could benefit from the temporal information in a sequence of frames as the generated frames will have fixed time spacing, much like conventional video formats. However, this approach is prone to generating frames with few events if there is little motion in the scene over the fixed duration. The alternative approach of forming each frame from a fixed number of events gives some assurance of a minimum amount

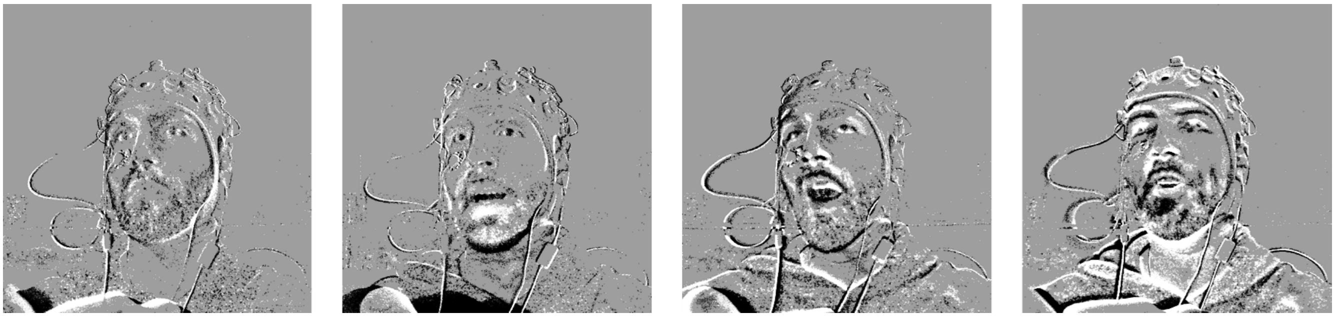

FIGURE 4. Sample event frames from a yawn sequence where the mouth is always visible.

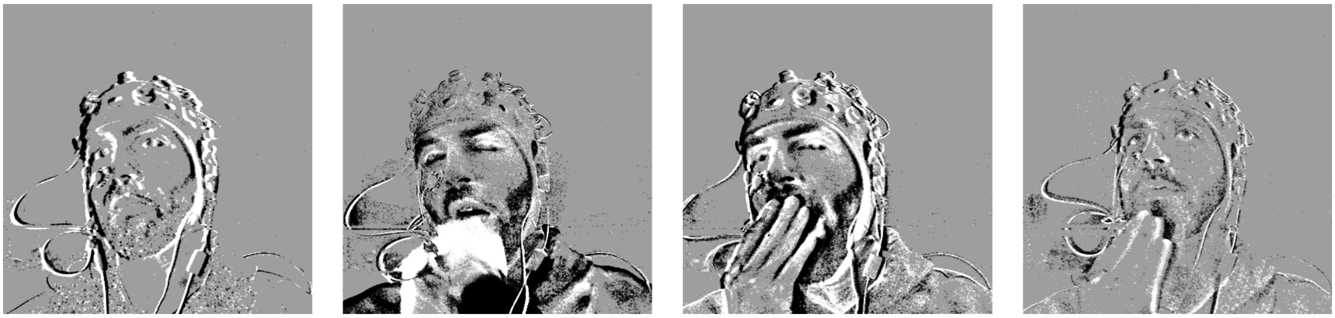

FIGURE 5. Sample event frames from yawn sequences where the mouth is covered by a hand while yawning.

of spatial information in each frame, at the loss of much of this temporal information.

We hypothesise that the fixed duration method is more applicable for yawn detection. This yields frames at fixed rate, much like a conventional camera's output, and the temporal information carried in a sequence of these frames can be useful when identifying yawns from other actions such as speech, due to differences in the rate of mouth motion. The choice of this event frame duration should be informed by the requirements of the underlying task. Accumulating events over a long period risks an aliasing effect, where speech frames could appear as one long mouth open sequence if insufficiently sampled. On the other hand, using too short a period can yield many frames with low spatial information. For our final yawn dataset, each frame is generated by accumulating events over a duration of 0.1s, resulting in frame sequences of 100 frames at 10FPS. This reduction from the 30FPS of the source data has 3 primary justifications: (1) A higher frame frequency is unnecessary to distinguish a yawn from speech. (2) With fewer than 300 frames in many RGB sequences, accumulating an equal or greater number of event frames would require an additional interpolation step, otherwise a freezing effect occurs in the event videos due to several frames showing the same motion. (3) A reduction from 300 to 100 frames for each sample carries a significant speedup to network training. The event frames' pixel values are clipped to ±10 and then normalised between 0,255. The 37 subjects in our simulated event yawn dataset were split into three sets for training, validation, and testing. The breakdown of each set is shown in Table 1. There is no overlap of subjects between the three sets. Sample event frames from two yawn sequences are shown in Fig. 4 and Fig. 5 The former has the mouth fully visible throughout, but the latter shows the mouth covered by the subject's hand.

TABLE 1. Distribution of our event yawn dataset partitions.

	Train set	Valid set	Test set	Total
Subject counts	21	8	8	37
Event frame counts	65,200	15,000	15,000	95,200
Yawn sequence counts	331	75	75	481
Non-yawn sequence counts	321	75	75	471

The YawDD dash videos were converted from 30FPS RGB video to 10FPS event video following the same process as our custom yawn dataset. The start and stop frames of the yawns were annotated and 100 frame sequences were extracted with the yawn frames centered. Non-yawn sequences were also saved from the frames between yawns. This totaled to 12,300 synthetic event frames, containing 78 yawn sequences and 45 non-yawn sequences.

D. YAWN DETECTION—TRAINING DETAILS

The yawn training sequences were augmented to achieve better generalisation. This includes rotating 50% of sequences within ±10°, mirroring about a vertical axis, and cropping to squares of randomised size and position (within some limits to ensure the full face is still visible). The augmentations were only randomised between sequences, so each frame in a sequence had identical transformations applied. All frames were downsampled to 256 × 256 using pixel area relation before input to the network. The network was trained for 100 epochs with a batch size of 5. The initial learning rate of 10^{-4} was halved every 10 epochs. Binary cross entropy loss was calculated between the predicted and actual labels of each sequence in the validation set. The dropout probability was set to 0.1.

E. SEATBELT STATE DETECTION—DATASETS

Another non-public in-cabin industry dataset was used for our seatbelt detection algorithm. Using a near-infrared (NIR) camera in the rear-view mirror position of a car, various subjects were recorded fastening and unfastening their seatbelts repeatedly. The video frames were labelled by the following classes:

1) The subject's seatbelt is fastened.
2) The subject's seatbelt is unfastened.
3) The subject is fastening their seatbelt.
4) The subject is unfastening their seatbelt.

The wide field of view lens of the camera captured both the driver and passenger seat. Both seats were given distinct labels of the seatbelt state. These videos were split into crops of the driver's seat and crops of the passenger seat, and the passenger seat crops were mirrored horizontally to have similar perspective and seatbelt direction to the driver's seat crops. The network then has a simpler task predicting on the cropped images rather than requiring both seats to be considered separate features. Knowing the camera position is fixed, the same cropping and mirroring can be carried out as required at inference. In this paper, the term "static classes" refers to 0 and 1, and "transition classes" refers to 2 and 3.

F. SEATBELT STATE DETECTION—EVENT SIMULATION AND PRE-PROCESSING

The seatbelt state classification task poses unique challenges in choosing an approach to accumulate frames, as the seatbelt is relatively stationary once fastened/unfastened and generates few events, but the fastening/unfastening actions generate a comparatively huge number of events. Both previously described methods (fixed duration and fixed event count) were tested, but neither were fully suitable. It proved too difficult to find a fixed duration large enough to keep a stationary seatbelt sufficiently visible without significantly reducing the number of frames for capturiung the fastening/unfastening actions. Alternatively, using a fixed event count was also unreliable in keeping the seatbelt visible as there is no guarantee that the events contain relevant information. The event count was often saturated by unrelated movements such as head motion or the background changing outside the car window. Specifying a number of events large enough to keep the seatbelt visible in all of these cases is impractical, as just one frame can span a huge time period when the rate of events is low.

A customised approach was developed for the final iteration of the seatbelt dataset. Each frame was required to reach a minimum number of events, but only within a rectangle bounding the subject's torso to minimise frames generated from irrelevant motion in the scene. Additionally, each frame was required to span a minimum duration of 200ms to prevent the generation of a proportionally huge number of transition frames, which have a much higher rate of events over the torso region than the static classes. This can also be thought of as capping the frame rate to 5FPS. This hybrid approach produced frames with much more reliable seatbelt visibility, as demonstrated in Fig. 6 where the fixed counts/duration were specified so each method generates 75 frames of the same "Seatbelt Fastened" clip. In the full 75 frames, the seatbelt was visible in (a) 27%, (b) 71%, and (c) 93%.

The events used to create Fig. 6 are from a set of real events that were collected for testing of the network. A Prophesee EVK4 event camera was mounted beside the rear-view mirror of a driving simulator and focused on the driver's seat. Subjects were asked to fasten and unfasten their seatbelt at random intervals throughout each recording. These videos were labelled manually with the same 4 classes as the NIR dataset, but with the start and stop of each class defined by event timestamps instead of frames. This initial test dataset was limited to 6 subjects to validate this proof-of-concept use case. Table 2 gives breakdown of the final event seatbelt dataset by class.

G. SEATBELT STATE DETECTION—TRAINING DETAILS

For seatbelt state detection, four distinct models were developed with our same network structure:

1) SEATBELT ON VS. SEATBELT OFF (STATIC CLASSES)

A binary classifier trained on just the static classes to directly assess the potential to predict on event data with little seatbelt motion.

2) FASTENING VS. UNFASTENING CLIPS (TRANSITION CLASSES)

A binary classifier trained on just the transition classes to assess if predicting the changing state of the seatbelt is more reliable than using the static seatbelt in event data.

3) COMBINED STATIC CLASSES VS. COMBINED TRANSITION CLASSES

In a real-world deployment of a seatbelt state detector, all 4 classes must be handled. This necessitates another binary model for a preliminary filtering to determine if an input sequence of frames should be passed to the static model (1) or transition model (2) to refine the prediction.

4) 4-CLASS MODEL

Trained with all 4 classes of our synthetic seatbelt dataset to handle all states in a single model. The classes are all considered independently by this network, so each frame sequence is predicted as containing only one class, and the previous state does not inform new predictions.

TABLE 2. Breakdown of seatbelt dataset partitions.

	Per-class event frame count				Total
	Seatbelt on	Seatbelt off	Fastening	Unfastening	
Training	7,958	5,357	3,344	2,393	19,052
Validation	1,146	983	1,088	749	3,966
Testing (of which are real events)	1,128 (426)	853 (310)	919 (675)	468 (307)	3,368 (1,718)
Total	10,232	7,193	5,351	3,610	26,386

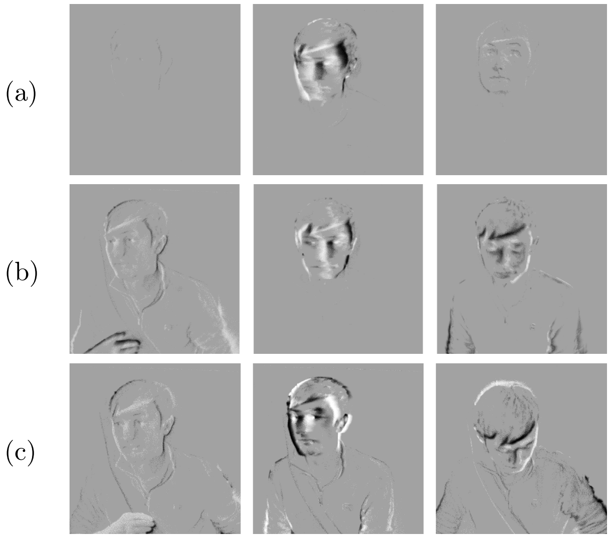

FIGURE 6. A "Seatbelt fastened" clip that was converted from events to frames using (a) a fixed time period, (b) a fixed event count, and (c) a fixed event count over the torso region with a minimum duration, so that each method yields 75 frames total. Of these 75 frames, the seatbelt was visible in (a) 27%, (b) 71%, and (c) 93%.

All 4 seatbelt models were trained with the same parameters. A fixed sequence length of 10 frames per sample was chosen. A longer sequence gives more robust predictions, but reduces the number of samples for developing the network. These samples were also augmented by random cropping but ensuring the torso is visible, before downsampling to 256 × 256. The binary model was trained with a batch size of 15 sequences and the learning rate of 10^{-4} was halved every 5 epochs. The dropout probability was set to 0.2.

IV. RESULTS

A. YAWN DETECTION

The 10 epochs with the lowest validation loss were tested on (a) our test set and (b) the simulated YawDD dash set. The best model with determined by the highest mean F1 score on both sets. The precision, recall, and F1 score of this model on each dataset partition are listed in Table 3. Running on an NVIDIA GeForce RTX 2080Ti GPU, an Intel i7-9700K CPU, and 32GB of RAM, the average inference times were measured at 0.44s per 100 frame sequence. Each of these sequences corresponds to 10s of data, granting a large cushion for real-time inference with more limited hardware.

TABLE 3. Results of our best yawn detection model tested on all of our synthetic event sets.

Dataset	Precision	Recall	F1
Train	97.0%	98.1%	97.6%
Valid	90.4%	100%	94.9%
Test (ours)	95.9%	94.7%	95.3%
YawDD	89.9%	91.0%	90.4%

TABLE 4. Comparison of yawn detection methods by performance on YawDD dataset.

Method	Precision	Recall	F1
Dong et al [39]	100%	67%	80.2%
Zhang et al. [40]	87.1%	88.6%	87.8%
Akrout and Mahdi [41]	82%	83%	82.5%
Omidyeganeh et al. [19]	70%	70%	70%
Ours	89.9%	91%	90.4%

TABLE 5. Results of our model tested on all of our event seatbelt test datasets.

Seatbelt Model	Test set F1-score	
	Simulated	Real
(1) 2-class (static)	100%	100%
(2) 2-class (transition)	100%	88.5%
(3) 2-class (static vs. transition)	90.8%	84.7%
(4) 4-class	96.3%	85.4%

Our method is compared to several other yawn detection methods in Table 4 by their results on the YawDD dataset. We have achieved a high level of performance, surpassing these methods with no YawDD data present in the training or validation sets.

The attended feature maps output by the self-attention module can be resized to visualise the areas that are more heavily weighted by the network in each frame. The visualisations in Fig. 7 demonstrate how our network sees the face, and in particular the mouth, as the most important features for yawn detection, even when the mouth is covered by a hand as in Fig. 7 (a).

B. SEATBELT DETECTION

The performance of each seatbelt model is given in Table 5 by macro-averaged F1 score on both the synthetic and real test sets. The binary static model (1) proved the most capable, correctly predicting the seatbelt on/off state in all unseen

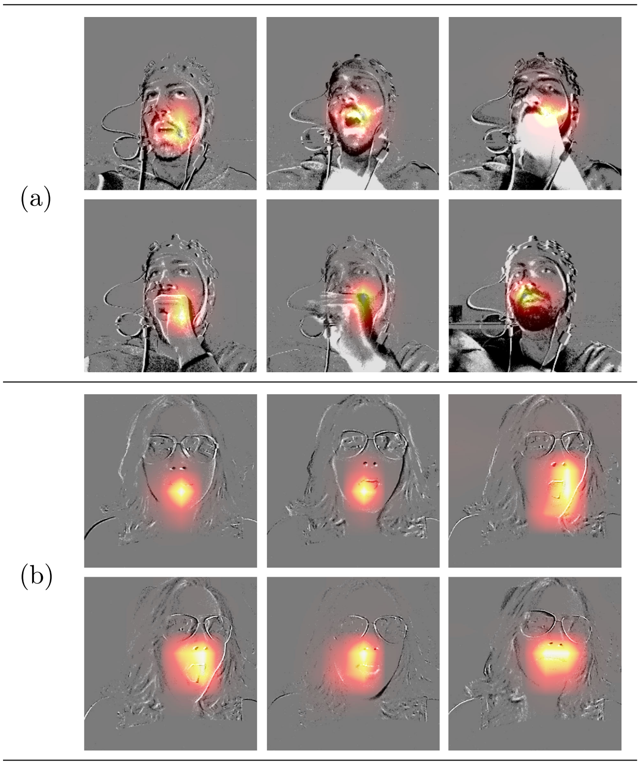

FIGURE 7. Visualised attention maps of yawn frames generated from (a) our test set and (b) the simulated YawDD dataset.

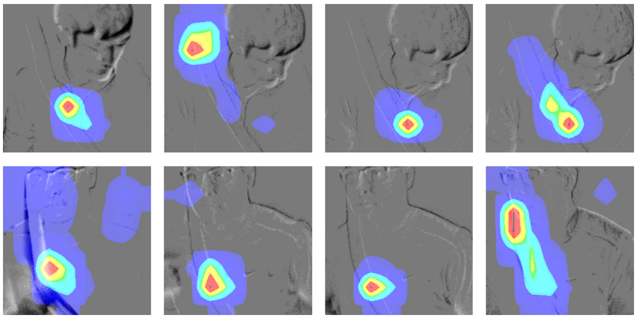

FIGURE 8. Visualised attention maps of seatbelt frames generated from real events in the test set.

test sequences, both real and synthetic. The binary transition model (2) also achieved perfect accuracy on the simulated test set, but the noticeable difference in performance on the real data indicates overfitting on the synthetic events. These results indicate that the resting state of the seatbelt is more reliable for prediction than tracking the transition states. To select which of these two model to used for an input frame sequence in a practical DMS, a model, the third binary model (3) is needed. This was surprisingly the lowest performing of all models, despite having an objectively easier task than the 4-class model (4), which uses exactly the same data but categorizes them more precisely. This result, combined with high training accuracy on model (3), reveals more overfitting to be the cause of the lower performance. Both models (3) and (4) are fed full videos and so must precisely select the frames from the continuous sequence where the state changes (e.g. from "fastening" to "fastened"), while models (1) and (2) are given discrete sequences and do not need such a fine demarcation of states, which contributes to the difference in accuracy.

We again visualise the attended feature maps to verify the network has learned to find appropriate features in the input frames. Fig. 8 gives a sample of this on 2 sequences from the real event set, and depicts the networks tendency to heavily weight the regions containing the seatbelt.

V. CONCLUSION

In this article we provide proof of concept methods for both yawn detection and seatbelt state detection with event cameras using lightweight deep learning models. This includes further evidence of the efficacy of synthetic event data in developing neuromorphic algorithms that can generalize to real data. Recent months have seen neuromorphic research trend away from frame-based approaches in favour of sparse representations, but this paper demonstrates how frames can efficiently compress event data for tasks with lesser time requirements. Our yawn detection algorithm offers superior performance to typical keypoint-based methods by accounting for associated motions of the upper body and handling the frequent cases where the mouth is occluded by a hand. Event cameras are typically employed for their fast response times and motion analysis qualities, but with the models developed for continuous monitoring of the seatbelt - even while stationary for long periods - we demonstrate how event data can be manipulated to satisfy a diverse set of requirements for assorted tasks. The proposed neuromorphic event-based algorithms for detecting yawns and seatbelt state fill a research gap and offers promising potential for advanced driver-assistance systems and intelligent safety features.

A. FUTURE WORK

Future work will seek to improve the seatbelt algorithms by considering the fixed order of states. In particular, the 4 class model should weight future predictions based on the current predicted state. For example, given the current state is "seatbelt fastened", the network should have the knowledge that "unfastening" must follow. Further collection of real event data and sourcing more public datasets of seatbelt states and yawns are planned to greatly expand our research. Additionally, the deployment of these models will be investigated within the limitations of embedded hardware typically found in DMS. All models in this article use the same architecture, granting scope for extremely efficient deployment.

ACKNOWLEDGMENT

The authors recognize that the data, equipment, and expertise provided by Xperi Inc., were vital in enabling this research.

REFERENCES

[1] S. Kumari, K. Akanksha, S. Pahadsingh, and S. Singh, "Drowsiness and yawn detection system using Python," in *Proc. Int. Conf. Commun., Circuits, Syst.* Singapore: Springer, 2021, pp. 225–232.

[2] P. Ghorai, A. Eskandarian, Y.-K. Kim, and G. Mehr, "State estimation and motion prediction of vehicles and vulnerable road users for cooperative autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 16983–17002, Oct. 2022.

[3] K. Kuru and W. Khan, "A framework for the synergistic integration of fully autonomous ground vehicles with smart city," *IEEE Access*, vol. 9, pp. 923–948, 2021.

[4] D. Miculescu and S. Karaman, "Polling-systems-based autonomous vehicle coordination in traffic intersections with no traffic signals," *IEEE Trans. Autom. Control*, vol. 65, no. 2, pp. 680–694, Feb. 2020.

[5] K. Kuru, "Conceptualisation of human-on-the-loop haptic teleoperation with fully autonomous self-driving vehicles in the urban environment," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 448–469, 2021.

[6] M. I. Pereira, R. M. Claro, P. N. Leite, and A. M. Pinto, "Advancing autonomous surface vehicles: A 3D perception system for the recognition and assessment of docking-based structures," *IEEE Access*, vol. 9, pp. 53030–53045, 2021.

[7] M. A. Khan, T. Nawaz, U. S. Khan, A. Hamza, and N. Rashid, "IoT-based non-intrusive automated driver drowsiness monitoring framework for logistics and public transport applications to enhance road safety," *IEEE Access*, vol. 11, pp. 14385–14397, 2023.

[8] E. Perkins, C. Sitaula, M. Burke, and F. Marzbanrad, "Challenges of driver drowsiness prediction: The remaining steps to implementation," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 2, pp. 1319–1338, Feb. 2023.

[9] A. Picot, A. Caplier, and S. Charbonnier, "Comparison between EOG and high frame rate camera for drowsiness detection," in *Proc. Workshop Appl. Comput. Vis. (WACV)*, Dec. 2009, pp. 1–6.

[10] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, Jan. 2022.

[11] M. S. Dilmaghani, W. Shariff, C. Ryan, J. Lemley, and P. Corcoran, "Control and evaluation of event cameras output sharpness via bias," *Proc. SPIE*, vol. 12701, pp. 455–462, Jun. 2022.

[12] C. Ryan, B. O'Sullivan, A. Elrasad, A. Cahill, J. Lemley, P. Kielty, C. Posch, and E. Perot, "Real-time face & eye tracking and blink detection using event cameras," *Neural Netw.*, vol. 141, pp. 87–97, Sep. 2021.

[13] N. Fouda Mbarga, A.-R. Abubakari, L. N. Aminde, and A. R. Morgan, "Seatbelt use and risk of major injuries sustained by vehicle occupants during motor-vehicle crashes: A systematic review and meta-analysis of cohort studies," *BMC Public Health*, vol. 18, no. 1, p. 1413, Dec. 2018.

[14] C. S. Crandall, L. M. Olson, and D. P. Sklar, "Mortality reduction with air bag and seat belt use in head-on passenger car collisions," *Amer. J. Epidemiol.*, vol. 153, no. 3, pp. 219–224, Feb. 2001.

[15] S. Hosseini and A. Fathi, "Automatic detection of vehicle occupancy and driver's seat belt status using deep learning," *Signal, Image Video Process.*, vol. 17, no. 2, pp. 491–499, Mar. 2023.

[16] A. Kashevnik, A. Ali, I. Lashkov, and N. Shilov, "Seat belt fastness detection based on image analysis from vehicle in-abin camera," in *Proc. 26th Conf. Open Innov. Assoc. (FRUCT)*, Apr. 2020, pp. 143–150.

[17] K. Paul, M. S. Dilmaghani, C. Ryan, J. Lemley, and P. Corcoran, "Neuromorphic sensing for yawn detection in driver drowsiness," in *Proc. 15th Int. Conf. Mach. Vis. (ICMV)*, Jun. 2023, pp. 1–12.

[18] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2011, pp. 1–4.

[19] M. Omidyeganeh, S. Shirmohammadi, S. Abtahi, A. Khurshid, M. Farhan, J. Scharcanski, B. Hariri, D. Laroche, and L. Martel, "Yawning detection using embedded smart cameras," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 3, pp. 570–582, Mar. 2016.

[20] M. Knapik and B. Cyganek, "Driver's fatigue recognition based on yawn detection in thermal images," *Neurocomputing*, vol. 338, pp. 274–292, 2019.

[21] H. Yang, L. Liu, W. Min, X. Yang, and X. Xiong, "Driver yawning detection based on subtle facial action recognition," *IEEE Trans. Multimedia*, vol. 23, pp. 572–583, 2021.

[22] D. Liu, C. Zhang, Q. Zhang, and Q. Kong, "Design and implementation of multimodal fatigue detection system combining eye and yawn information," in *Proc. IEEE 5th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2020, pp. 65–69.

[23] V. Dehankar, P. Jumle, and S. Tadse, "Design of drowsiness and yawning detection system," in *Proc. 2nd Int. Conf. Electron. Renew. Syst. (ICEARS)*, Mar. 2023, pp. 1585–1589.

[24] B. Alshaqaqi, A. S. Baquhaizel, M. E. A. Ouis, M. Boumehed, A. Ouamri, and M. Keche, "Driver drowsiness detection system," in *Proc. 8th Int. Workshop Syst., Signal Process. their Appl. (WoSSPA)*, May 2013, pp. 151–155.

[25] J. S. R. Melvin, B. Rokesh, S. Dheepajyothieshwar, and K. Akila, "Driver yawn prediction using convolutional neural network," in *Proc. IoT, Cloud Data Sci.*, Feb. 2023, pp. 268–276.

[26] M. Baltaxe, R. Mergui, K. Nistel, and G. Kamhi, "Marker-less vision-based detection of improper seat belt routing," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 783–789.

[27] S. Chun, N. H. Ghalehjegh, J. Choi, C. Schwarz, J. Gaspar, D. McGehee, and S. Baek, "NADS-Net: A nimble architecture for driver and seat belt detection via convolutional neural networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2413–2421.

[28] D. Rao, "Driver's seat belt detection using CNN," *Turkish J. Comput. Math. Educ.*, vol. 12, pp. 776–785, Apr. 2021.

[29] A. Ş. Şener, I. F. Ince, H. B. Baydargil, I. Garip, and O. Ozturk, "Deep learning based automatic vertical height adjustment of incorrectly fastened seat belts for driver and passenger safety in fleet vehicles," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 236, no. 4, pp. 639–654, Mar. 2022, doi: 10.1177/09544070211025338.

[30] J. Madake, S. Yadav, S. Singh, S. Bhatlawande, and S. Shilaskar, "Vision-based driver's seat belt detection," in *Proc. Int. Conf. Advancement Technol. (ICONAT)*, Jan. 2023, pp. 1–5.

[31] Y. Zang, B. Yu, and S. Zhao, "Lightweight seatbelt detection algorithm for mobile device," *Multimedia Tools Appl.*, vol. 2023, pp. 1–15, Mar. 2023.

[32] A. Upadhyay, B. Sutrave, and A. Singh, "Real time seatbelt detection using YOLO deep learning model," in *Proc. IEEE Int. Students' Conf. Electr., Electron. Comput. Sci. (SCEECS)*, Feb. 2023, pp. 1–6.

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[35] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 7354–7363.

[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[37] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD: A yawning detection dataset," in *Proc. 5th ACM Multimedia Syst. Conf.*, Mar. 2014, pp. 24–28.

[38] Y. Hu, S.-C. Liu, and T. Delbruck, "v2e: From video frames to realistic DVS events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1312–1321.

[39] B.-T. Dong, H.-Y. Lin, and C.-C. Chang, "Driver fatigue and distracted driving detection using random forest and convolutional neural network," *Appl. Sci.*, vol. 12, no. 17, p. 8674, Aug. 2022.

[40] W. Zhang and J. Su, "Driver yawning detection based on long short term memory networks," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–5.

[41] B. Akrout and W. Mahdi, "Yawning detection by the analysis of variational descriptor for monitoring driver drowsiness," in *Proc. Int. Image Process., Appl. Syst. (IPAS)*, Nov. 2016, pp. 1–5.

PAUL KIELTY received the B.E. degree in electronic and computer engineering from the University of Galway, in 2021. He is currently pursuing the joint Ph.D. degree with the University of Galway and the ADAPT SFI Research Centre. His research interest includes deep learning methods with neuromorphic vision, with particular interest in driver monitoring tasks.

MEHDI SEFIDGAR DILMAGHANI received the B.Sc. degree in electronics engineering from the University of Tabriz, in 2012, and the M.Sc. degree in electronics engineering from KNTU, in 2016. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronics Engineering, University of Galway, under the Hardiman Scholarship. During his M.Sc. studies, he had focus on electronic implementation of signal processing algorithms and wavelets. He is also a Research and Development Intern with Xperi Inc. His research interests include deep learning, computer vision, and neuromorphic sensors (event cameras).

WASEEM SHARIFF received the B.E. degree in computer science from the Nagarjuna College of Engineering and Technology (NCET), in 2019, and the M.Sc. degree in computer science, specializing in artificial intelligence from the National University of Ireland Galway (NUIG), in 2020. He is currently pursuing the Ph.D. degree with the University of Galway, under the IRC Employment Ph.D. Program. He is also a Research and Development Engineer with Xperi Inc., based in Galway. His research interest includes machine learning for computer vision applications, with a particular emphasis on automotive driver monitoring applications.

CIAN RYAN received the M.Sc. degree in computational finance and the Ph.D. degree from the University of Limerick, in 2015 and 2020, respectively. He is currently a Staff Machine Learning Engineer with Xperi Corporation based in Galway. His research interests include machine learning and deep learning methods applied to computer vision.

JOE LEMLEY received the B.S. degree in computer science and the master's degree in computational science from Central Washington University, in 2006 and 2016, respectively, and the Ph.D. degree from the National University of Ireland Galway. He is currently a Principal Research and Development Engineer and the Manager of Xperi Inc., Galway. His field of work is machine learning using deep neural networks for tasks related to computer vision. His current research interests include computer vision and signal processing for the driver monitoring systems.

PETER CORCORAN (Fellow, IEEE) currently holds the Personal Chair of Electronic Engineering with the College of Science and Engineering, National University of Galway, Ireland. He is also an IEEE Fellow recognized for his contributions to digital camera technologies, notably in-camera redeye correction and facial detection. He was the Co-Founder in several start-up companies, notably FotoNation, now the Imaging Division, Xperi Corporation. He has over 600 technical publications and patents, over 100 peer-reviewed journal articles, 120 international conference papers, and he is a co-inventor of more than 300 granted U.S. patents. He is a member of the IEEE Consumer Electronics Society for over 25 years. He is the Editor-in-Chief and the Founding Editor of *IEEE Consumer Electronics Magazine*.