IEEE Transactions on Intelligent Vehicles, Volume 8, Issue 1, January 2023, pp. 37-47
DOI:10.1109/TIV.2022.3163458

Driver Anomaly Quantification for Intelligent Vehicles: A Contrastive Learning Approach with Representation Clustering

Zhongxu Hu, *Member, IEEE*, Yang Xing, *Member, IEEE*, Weihao Gu, Dongpu Cao, *Senior Member, IEEE*, and Chen Lv *Senior Member, IEEE*

***Abstract*—Driver anomaly quantification is a fundamental capability to support human-centric driving systems of intelligent vehicles. Existing studies usually treat it as a classification task and obtain discrete levels for abnormalities. Meanwhile, the existing data-driven approaches depend on the quality of dataset and provide limited recognition capability for unknown activities. To overcome these challenges, this paper proposes a contrastive learning approach with the aim of building a model that can quantify driver anomalies with a continuous variable. In addition, a novel clustering supervised contrastive loss is proposed to optimize the distribution of the extracted representation vectors to improve the model performance. Compared with the typical contrastive loss, the proposed loss can better cluster normal representations while separating abnormal ones. The abnormality of driver activity can be quantified by calculating the distance to a set of representations of normal activities rather than being produced as the direct output of the model. The experiment results with datasets under different modes demonstrate that the proposed approach is more accurate and robust than existing ones in terms of recognition and quantification of unknown abnormal activities.**

***Index Terms*—Driver anomaly, online quantification, continuous variable, contrastive learning, representation clustering.**

I. Introduction

INTELLIGENT driving has attracted considerable attention, and tremendous progress has been achieved in recent years[1, 2]. Autonomous vehicles have been primarily investigated to replace human drivers in order to enhance driving performance and avoid possible fatalities. However, human drivers will still play an important role in the driving task for a certain period before full driving automation is achieved. Hence, the coexistence and cooperation of humans and vehicles represents an urgent and exciting new focus for the development in-vehicle technology[3, 4]. To enhance user safety and the efficiency of collaboration with intelligent vehicles, a reliable driver monitoring system (DMS) should be further pursued to parse the state of a human driver, which is a fundamental functionality to support advanced driver assistance systems and partially automated vehicles[5–8].

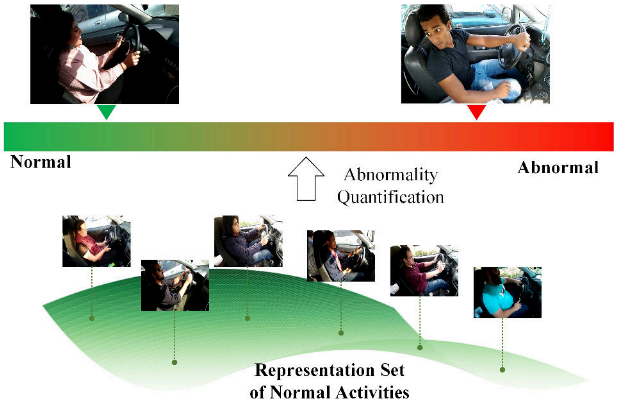

Normal
Abnormal
Abnormality Quantification
Representation Set of Normal Activities

Fig. 1: Driver anomaly quantification using a set of representations of normal activities. The demonstration can be found on the YouTube website.

Driver anomaly quantification, as a typical task of a DMS, has been studied for a long time[9–15]. Recently, many researchers have revisited this topic by leveraging the powerful representation capabilities of deep learning, leading to impressive achievements[16–24]. However, these studies usually treat it as a classification task, in which driver activity is classified into several predefined classes. Unfortunately, this approach does not provide a reasonable solution for downstream applications. For example, a shared control algorithm usually needs to allocate a continuous value ($0 \sim 1$), representing an authority weight, to the human driver in accordance with the driver's state [25–29]. From the perspective of application and control, the vehicle does not need to recognize the specific activity of the driver; it needs only to know whether the driver is in a normal or abnormal state and how close to or far away from the normal driving state the driver is. This means that the typical driver activity classification method is not suitable for constructing a human-centric intelligent driving system. Another problem is that any collected dataset has difficulty covering all possible types of abnormal human activities, and the existing datasets usually include only several typical activities, limiting the recognition capability of models trained on these datasets for previously unseen activities. Furthermore, the collection of labeled activity samples of different types is a laborious task. Therefore, this study aims to propose a method that can recognize driver anomalies, especially unknown ones,

Z. Hu and C. Lv are with School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore. E-mail: {zhongxu.hu,lyuchen}@ntu.edu.sg

Y. Xing is with Centre for Autonomous and Cyber-Physical Systems, Cranfield University, United Kingdom. E-mail: yang.x@cranfield.ac.uk

W. Gu is with Haomo.AI, Beijing, China. E-mail: guweihao@haomo.ai

D. Cao is with School of Vehicle and Mobility, Tsinghua University, China. E-mail: dp_cao2016@163.com

and quantify abnormality in the form of a continuous value rather than through discrete classification.

Contrastive learning is an emerging technique that has recently attracted considerable attention[30, 31]. In contrastive learning, a self-supervised (unsupervised) methodology is adopted to learn a general representation of a domain by minimizing the distance between similar samples and maximizing the distance between different samples. This technique provides some inspiration regarding how to alleviate the reliance on a complete anomaly dataset and quantify the abnormality. Typical contrastive learning can be regarded as self-supervised metric learning, in which data augmentation techniques are leveraged to create image pairs: two randomly augmented versions of the same image are fed into the model as a positive pair, while other images in the same batch are fed in as negative pairs. The goal of contrastive learning is to find the trade-off point that can balance alignment and uniformity on the hypersphere[32]. Alignment refers to the requirement for features extracted from similar samples to be close to each other in the projected space, whereas uniformity refers to the need for the projected features to have a uniform distribution to preserve the unique information of each one. Theoretically, better performance can be achieved in downstream applications by fine-tuning the model accordingly. Therefore, in this study, the aim is to obtain a feature extractor by leveraging contrastive learning and then use it to build a set of representation vectors corresponding to normal driver activities. Abnormal activity can then be recognized by calculating the distance from normal activities, and this distance can be used not only to distinguish anomalies but also to quantify the abnormality, as shown in Fig. 1. To this end, the alignment of normal samples should be enhanced, while the uniformity with respect to abnormal samples should be maintained. In this study, the label information is utilized to form a supervised contrastive loss for improving model performance[33]. The positive pairs may be formed not only from different versions of a same image, but also from samples with the same label. Furthermore, this study introduces the conception of clustering, such that normal activity representations are forced to cluster around the center of the normal feature set. Finally, a clustering supervised contrastive loss (C-SCL) is proposed to train the feature extractor model.

The main contribution of this study is that contrastive learning is introduced to obtain a well-established representation set of driver normal activities that can be utilized to quantify anomalies using a continuous value rather than through traditional discrete classification, with the goal of bridging the gap with downstream applications. Meanwhile, a novel C-SCL that can further cluster the representations of normal samples is proposed to improve the model representation capability.

This paper is organized as follows. Section 2 comprehensively discusses related work. Section 3 describes the proposed method and the training framework. Section 4 analyzes and discusses the experimental results. Conclusions are presented in Section 5.

II. Related Work

The existing works for driver activity basically treat it as a classification problem, then it can be tackled by the efficient deep learning approach[34–36]. A commonly used input is an in-cabin image, and many convolutional neural network (CNN)-based approaches have been proposed from different perspectives[11, 13, 37]. [9] presented an ensemble of four CNN models to handle different parts of the driver, including the face, hands, and body, to recognize driver activity. [38] proposed an attend and guide network to classify driver behavior by obtaining the spatial structures of images through the identification of semantic regions and their spatial distributions. [39] concatenated three CNN models to construct a hybrid framework for detecting distracted driver behavior. [22] leveraged the generative adversarial network approach to augment a collected dataset with new training samples and trained a CNN model to recognize driver distraction. [24] proposed a coarse temporal attention network by exploiting spatiotemporal attention to model driver activity, utilizing an attention mechanism to generate high-level action-specific contextual information. [40] adopted a multitask learning approach and constructed triplets of images to improve the performance of vision-based driver distraction recognition. The image triplets were used to force networks to explore global information.

To improve model performance, additional information may be utilized. [18] proposed a bidirectional posture–appearance interaction network that utilizes skeleton data to enhance the model performance, and the proposed model was verified on a collected bus driver behavior dataset. Furthermore, [16] utilized semantic contextual cues in addition to skeleton data to improve the recognition accuracy by modeling the pairwise relation between body joint configurations and objects of interaction to capture structural information. [20] utilized multistream inputs and proposed a dedicated CNN model to handle them, which has the form of a tight ensemble architecture to improve the robustness of the model. [7] leveraged a smartphone to monitor driver behavior by using the camera and other built-in sensors, including the accelerator, gyroscope, Global Positioning System (GPS) receiver, and microphone.

Some studies have attempted to reduce the computation time. [19] proposed a new CNN-based driver activity recognition model by decreasing the filter size to reduce the size of the model. [21] proposed a lightweight CNN model with an octave-like convolution mixed block that uses pointwise convolution to expand the feature maps into two sets of branches. [23] utilized the neural architecture search technique, which can automatically search for the optimal model architecture, to build a fine-grained detection method for driver distraction. [8] adopted the depthwise separable convolution approach to build a lightweight CNN model for driver activity recognition.

In addition to data collected by vision-based sensors, driving data collected in different modes have also been adopted to recognize driver behavior. [17] leveraged multimodal driving data and adopted a stacked long short-term memory (LSTM) network architecture with an attention mechanism to detect driver distraction. [28] also used driving data as input, recognizing driver distraction by comparing normal driving parameters against those obtained while performing a secondary task and employing an effective fuzzy logic algorithm. [41] utilized the steering angle and velocity to recognize anomalous driving behavior by using a CNN model.

In contrast to existing works, this study aims to obtain a continuous value for quantifying the anomalies by leveraging the contrastive learning approach with representation clustering. Some studies have predefined different driver activities into three distraction severity levels with corresponding scores by using the fuzzy logic approach, as in the work of [29]. However, the problem is still treated as a multiclass classification task, and the distraction value is discrete. The previous work most closely related to ours is [42], which also utilized a supervised contrastive loss for model training. Relative to that work, this study introduces the clustering conception to enhance the alignment of normal samples and improve the representation performance of the model.

III. Methodology

In this section, we will introduce the proposed contrastive learning approach for quantifying driver activity anomaly and the novel C-SCL for representation clustering.

A. Contrastive Learning for Driver Anomaly Quantification

Driver activities can basically be classified into two categories: normal and abnormal. Abnormal activities are usually uncertain, whereas normal activities are definite. As a result, we can build a representation set for normal activities and then calculate the distance from a given activity sample to the set of normal samples, and the obtained continuous distance can indicate the abnormality of the given sample. Doing so requires a model with good feature representation capabilities, and the emerging contrastive learning approach provides us with some inspiration on how to do this.

Contrastive learning can be regarded as a self-supervised metric learning technique that can be used to learn a general representation of a domain without labels by teaching the model which samples are similar or different. In practice, a combination of two random augmentations (cropping, rotation, noise, dropout, etc.) can be applied to each image in a dataset to create a corresponding image pair $(x_{i'}, x_{i''})$, which are then fed into a deep learning model $f_{\theta}(\cdot)\colon \chi \to \mathbb{R}^d$ to extract the corresponding representation vectors, and the goal is to train the model to learn that the two samples in each pair are similar because they are essentially different versions of the same image. This goal can be abstractly described as follows:

$$d(f_{\theta}(x_{i'}), f_{\theta}(x_{i''})) << d(f_{\theta}(x_i), f_{\theta}(x_j)) \tag{1}$$

The above equation means that the model needs to reduce the distances between the embedding vectors of positive pairs while increasing the distances between negative pairs, and various studies have proposed different distance metric approaches for this purpose[43, 44]. In self-supervised contrastive learning, a positive pair consists of differently augmented versions of the same image, while a negative pair consists of different images. In supervised metric learning, a positive pair consists of images with the same label, whereas a negative pair consists of images with different labels. In particular, the commonly used InfoNCE loss [45] function in self-supervised contrastive learning is defined as follows:

$$\pounds_{cl} = -\sum_{i} log \frac{exp((f_{\theta}\hat{(}x_{i'}) \cdot f_{\theta}\hat{(}x_{i''}))/\tau)}{\sum_{k=1}^{2N} \mathbb{1}[k \neq i'] \cdot exp((f_{\theta}\hat{(}x_{i'}) \cdot f_{\theta}\hat{(}x_{k}))/\tau)} \tag{2}$$

where x_k is an augmented version of some sample in the dataset $\chi \in [N]$, $\cdot$ denotes the inner (dot) product, $f\hat{(}\cdot)$ denotes the normalization of a vector, and τ is a temperature hyperparameter.

The goal of contrastive learning is to project normalized representation vectors onto a hypersphere where the distribution of the projected vectors has the characteristics of alignment and uniformity simultaneously. Alignment refers to the requirement for the extracted representations of similar samples to be close to each other, whereas uniformity refers to the need for the representations to have a uniform distribution to preserve the unique information of each one. Theoretically, the feature representations can be optimized in this way. However, the uniformity characteristic in self-supervised learning will drive the model to maximize the distance between different images $(x_{i'}, x_k)$ even if $x_{i'}$ and x_k are both normal activity samples. This goes against our purpose, which is to cluster all normal samples together. Therefore, this study still leverages label information to improve the model performance, and our goal is for samples of normal activity to gather close to each other while remaining separated from samples of abnormal activity, which is essential for obtaining a reasonable continuous abnormality value.

B. Novel Supervised Contrastive Loss for Representation Clustering

The typical self-supervised contrastive loss considers only samples from the same image as positive pairs, while the remainder of the training batch is treated as negative pairs. As a result, various normal and abnormal samples will be equally distributed across the hypersphere, and each normal sample may be located nearby one or more abnormal samples, which is in contrast to our purpose. Therefore, the contrastive loss is modified to include label information such that a positive pair consists of samples with the same label rather than different versions of the same image, whereas a negative pair consists of samples with different labels, as follows:

$$\pounds_{scl} = -\sum_{i} \frac{1}{|P(i)|} \sum_{j \in P(i)} log \frac{exp(\frac{(f_{\theta}\hat{(}x_{i}) \cdot f_{\theta}\hat{(}x_{j}))}{\tau})}{\sum_{k=1}^{2N} \mathbb{1}[k \neq i] \cdot exp(\frac{(f_{\theta}\hat{(}x_{i}) \cdot f_{\theta}\hat{(}x_{k}))}{\tau})} \tag{3}$$

where $P(i)$ denotes samples that have the same label as x_i. The modified supervised contrastive loss drives the feature extractor to align embedding representations with the same label, resulting in a clustering of the representation space that is more robust than the original one.

To further align normal activity samples, the clustering conception is introduced by minimizing the distances from

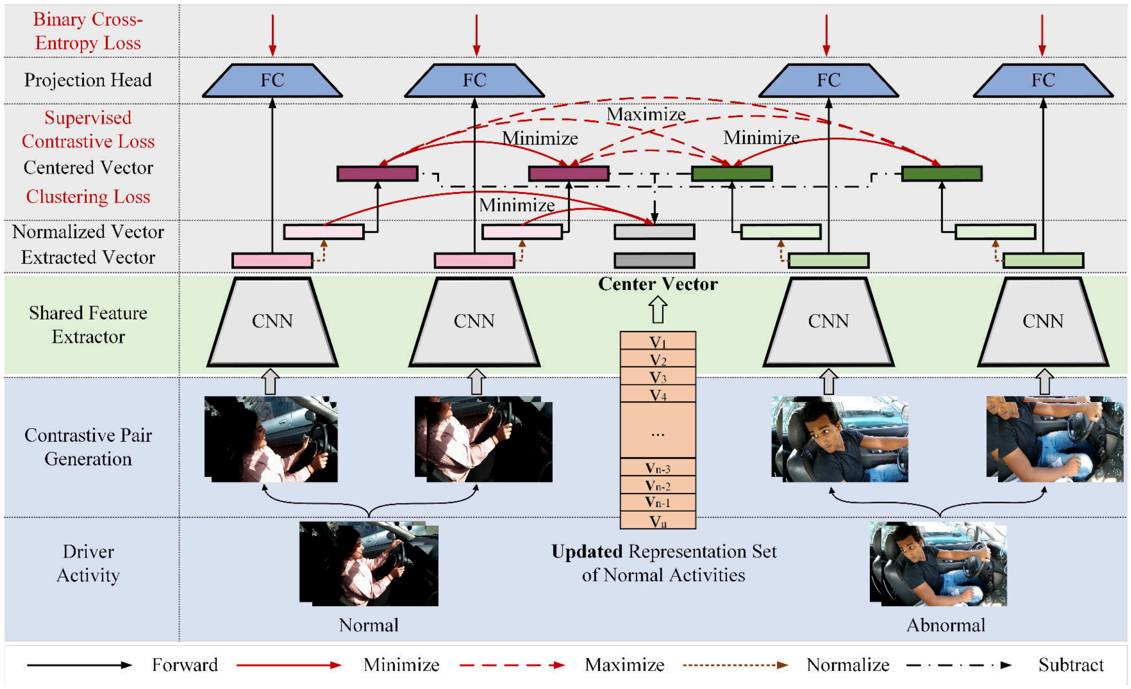


Fig. 2: Overview of the proposed contrastive learning approach with representation clustering for driver anomaly quantification. In practice, the input can be either a single image or sequential frames.

the embedding representation vectors of normal activities to their center vector. Thus, we translate the original center of the hypersphere to the center of the set of representation vectors of normal activities, and the contrastive loss is further modified as follows:

$$\pounds_{cscl} = -\sum_{i} \frac{1}{|P(i)|} \sum_{j \in P(i)} log \frac{exp(\frac{(c_i \cdot c_j)}{\tau})}{\sum_{k=1}^{2N} \mathbb{1}[k \neq i] \cdot exp(\frac{(c_i \cdot c_k)}{\tau})}$$
$$s.t.\ c_i = \frac{f_\theta(\hat{x_i}) - cent}{\left\| f_\theta(\hat{x_i}) - cent \right\|}, cent = \frac{1}{|N(i)|} \sum_{i} f_\theta(\hat{x_i}) \tag{4}$$

where $N(i)$ denotes the set of representation vectors of normal activities and $|N(i)|$ denotes the corresponding cardinality.

The contrastive loss utilizes the angle between two embedding vectors as the distance metric, which forms a uniform distribution around the representation hypersphere upon normalization of the vectors. The normalized center of the feature set of normal activities is still located within the hypersphere. As is well known, the geometric theorem regarding the angle at the circumference is that an angle at the circumference of a circle is equal to half the angle at the center subtended by the same arc. The angle change in the translated hypersphere, which is centered on the mean vector of the normal sample feature set, will be amplified in the original hypersphere. As a result, the modified contrastive loss further forces abnormal samples farther from normal ones. Finally, the training loss function can be described as follows:

$$\begin{aligned} \pounds &= \pounds_{cscl} + \pounds_{c} + \pounds_{bce} \\ &= -\sum_{i} \frac{1}{|P(i)|} \sum_{j \in P(i)} log \frac{exp((c_i \cdot c_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}[k \neq i] \cdot exp((c_i \cdot c_k)/\tau)} \\ &+ \sum_{i \in N(i)} c_i + \sum_{i} [y_i z(x_i) + (1 - y_i)(1 - z(x_i))] \end{aligned} \tag{5}$$

where $z(x)$ denotes the output of the projection head and y_i denotes the label of the corresponding sample. The proposed loss function, which includes the modified supervised contrastive loss, the representation clustering loss, and the binary cross-entropy loss, can force normal activity representations to cluster together, far away from abnormal ones, which can enhance the availability of using the continuous distance to quantify the abnormality.

C. The Architecture of the Proposed Approach

The final architecture of the proposed approach is shown in Fig. 2, and the pseudo-code of the training strategy can be found in Algorithm 1. In the training phase, each image (or set of sequential frames) is randomly augmented to obtain two different samples. Then, these samples are fed into the same feature extractor to obtain their representation vectors. The extracted vectors are first normalized, then subtracted with the center vector of the representation set of normal activity samples; then, the results are used to calculate the supervised contrastive loss and the clustering loss. Meanwhile, the non-normalized extracted vectors are fed into a projection head with a fully connected (FC) layer, which is trained with a binary cross-entropy loss, for class prediction. The

Algorithm 1 Training strategy for the proposed contrastive learning approach with representation clustering.

```
for Number of epochs do
    for Number of normal driver activity samples in the training set do
        Sample an image of normal activity $x_i$ from dataset $E$.
        Calculate the representation of the selected normal activity sample using the model $f_\theta(\cdot)$.
        Append the normalized version of the extracted vector to the normal activity feature set $N$.
    end for
    Update the center vector $cent$ of the feature set $N$ as follows: $\frac{1}{|N(i)|}\sum_i f_\theta(\hat{x_i})$
    for Number of training iterations do
        Sample $m$ images of normal driver activity $x_{n_1}, ..., x_{n_m}$ with labels $y_{n_1}, ..., y_{n_m}$ and $n$ images of abnormal driver activity $x_{a_1}, ..., x_{a_n}$ with labels $y_{a_1}, ..., y_{a_n}$ from dataset $E$.
        Update the entire model with the updated center vector $cent$ as follows:
        $\nabla\theta[\frac{1}{m+n}\sum_{i=1}^{m+n} \pounds_{sccl}(x_i, cent) + \frac{1}{m+n}\sum_{i=1}^{m+n} \pounds_{bce}(x_i, y_i) + \frac{1}{m}\sum_{i=1}^{m} \pounds_c(x_{n_i}, cent)]$
    end for
end for
Output: Updated feature extractor model and representation set of normal activities
```

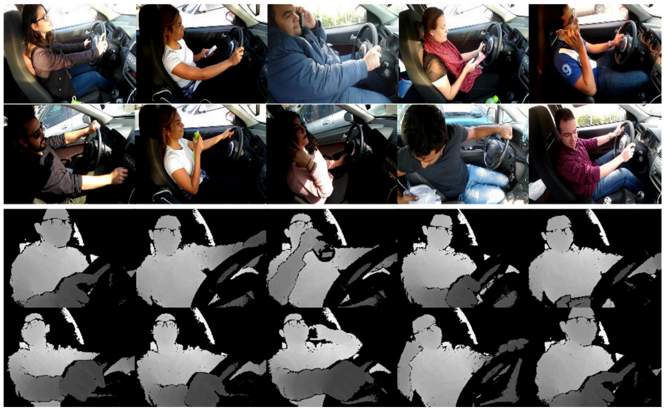


Fig. 3: Two driver activity datasets: AUC (top two rows) and DAD (bottom two rows).

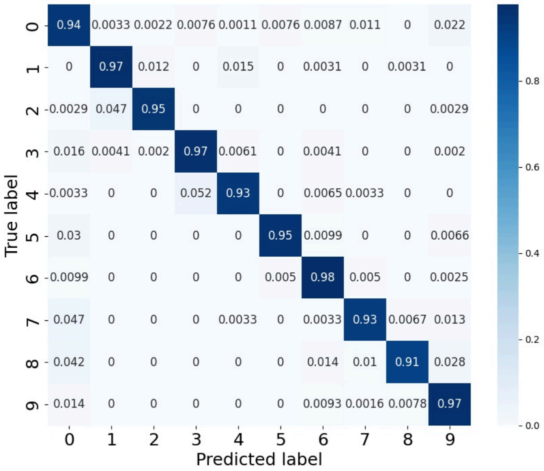


Fig. 4: Confusion matrix of the feature extractor used for recognizing multiple driver activities in the AUC test set. The value ranging from $0-9$ represents the provided label index of different activity classes in the original AUC dataset. 0 indicates that diver state is safe.

normal activity representation set is updated after each epoch of training. In the testing phase, the representation vector of all normal activity samples in the training set is firstly calculated by using the trained feature extractor, then the representation set can be obtained. For the extracted vector of each test sample, their distances from all vectors in the representation set are calculated. Finally, the minimal distance that is found by the k-nearest neighbor (KNN) algorithm [46] is utilized as the continuous abnormality value to quantify anomalies. This relies on the trained feature extractor, which can cluster the normal representations while separating the abnormal representations.

IV. Experiment

A. The Datasets

Datasets collected in two different modes are utilized to evaluate the proposed method: the American University in Cairo (AUC) Distracted Driver Dataset [9] and the Driver Anomaly Detection (DAD) dataset [42], as shown in Fig. 3.

The AUC dataset is a commonly used dataset for driver activity recognition. The authors of this dataset recruited 31 participants, of whom 22 were male and 9 were female, from 7 different countries. The driver activity data were collected in 4 different cars and are classified into ten classes, such as drinking, adjusting the radio, driving in a safe posture, fiddling with hair or makeup, reaching behind, talking to passengers, and talking on a cell phone. In these experiments, the input from the AUC dataset consists of a single frame for consistency with existing works to ensure fair comparisons. In this study, the AUC dataset is chosen as the main benchmark to evaluate the proposed method because it is commonly used.

The DAD dataset was collected from a driving simulator. Two depth cameras were mounted on top of the vehicle and in front of the driver. The front camera recorded the driver's head and body and the visible parts of the hands, while the top camera focused on the driver's hand movements. The dataset also includes infrared images that are synchronized with the depth maps. In total, 31 subjects were asked to drive in the simulator while performing normal and abnormal activities, and each subject was assigned to either the training set or the test set. The training set consists of 25 subjects, each performing six normal driving activities and eight abnormal driving activities. Meanwhile, the test set contains 6 subjects, each performing 6 normal driving activities and 24 abnormal driving activities, of which 16 do not appear in the training set. This requires that the model can recognize previously unseen abnormal activities, which is aligned with the purpose of this study. It should be noted that only the front depth images are utilized in this study. In these experiments, the input from the DAD dataset consists of 16 frames of activity video, which is also consistent with existing work.

These two datasets are highly complementary in terms of their data modes (color and depth) and input forms (single images and sequential frames). They both contain many different types of abnormal activities and thus can be conveniently used to test the ability of a model to handle previously unseen activities. These characteristics will be beneficial to us in comprehensively evaluating the proposed method.

B. *Evaluation of the Proposed Contrastive Learning Approach*

1) Feature Extractor: For the AUC dataset, ResNet18 [47] is utilized as the feature extractor and the baseline model, and the multiclass learning capability is first verified to ensure a fair evaluation of the proposed method. Accordingly, the number of final output dimensions of the FC layer of ResNet18 is modified to 10 for consistency with the AUC dataset. Then, the cross-entropy loss function is utilized to train the model with the stochastic gradient descent (SGD) optimizer with an initial learning rate of $1e \times 10^{-4}$. The confusion matrix of the evaluation results can be found in Fig. 4, which clearly indicates that the overall classification accuracy is high. In the results of various classes, the samples of the *Reach Behind* (labeled as 8) class easily tends to be recognized as the *Talking Passenger* (labeled as 9) because they are similar in some ways, resulting in the relatively low accuracy in its classification, as well as the *Talk Right* (labeled as 4) and the *Hair Makeup* (labeled as 7). We also compare this model with other state-of-the-art methods, as shown in Tab. I. The utilized modified ResNet18 can outperform most of the state-of-the-art methods by leveraging only the raw image rather than the other extra complementary features. The VGG-16 modified by the [48] has a slightly higher accuracy using more parameters. This study does not focus on elaborately designing a classification model, whereas proposing an efficient contrastive learning framework. In general, the comparison shows that the modified ResNet18 can achieve the state-of-the-art performance. The experimental results indicate that ResNet18 has an outstanding ability to recognize multi-class driver activities and thus is suitable to be used as the feature extractor. To handle sequential frames, the 3D Resnet18 network [49] is chosen as the feature extractor for the DAD dataset. Because the DAD dataset is designed for abnormal activity recognition, the learning capability of 3D Resnet18 is directly verified in the next comparison.

2) Clustering Supervised Contrastive Loss: Driver anomaly quantification can first be considered as a binary classification task to evaluate the availability of the obtained abnormality value. Therefore, an intuitive method is to use an FC layer as a binary classifier to recognize the extracted features while adopting only the binary cross-entropy loss function to train the model. In this study, this approach is adopted as the baseline model, with the output of the FC layer serving as the classification metric. The baseline model also considers only the $\pounds_{bce}$ loss, without the contrastive loss or the clustering loss. To evaluate the recognition capability for previously unseen activities, the impact of different numbers of training samples is investigated while using the same whole test set.

TABLE I: Comparison of multiclass driver activity recognition based on the AUC dataset.

Method	Input	Accuracy (%)
AlexNet [9]	Raw Image	93.65
	Face & Hands	86.68
InceptionV3 [9]	Raw Image	95.17
	Face & Hands	90.88
MVE [50]	Image & Face & Hands & Skin	95.77
GA-WE [50]	Image & Hands & Face & Skin	95.98
Fusion [51]	Multiple Images	92.36
DenseNet [52]	Raw Image & Body Pose	94.20
Modified VGG-16 [48]	Raw Image	96.31
Modified Resnet18	Raw Image	95.01

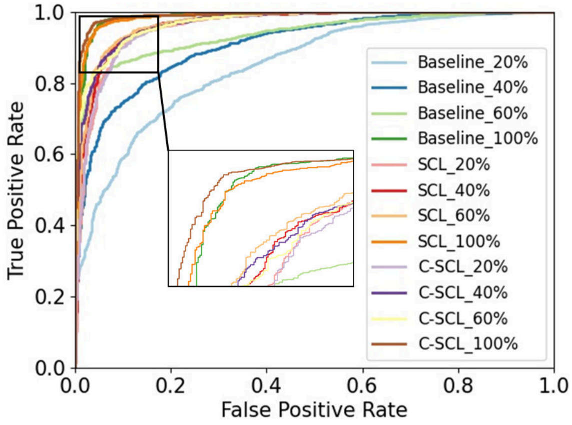
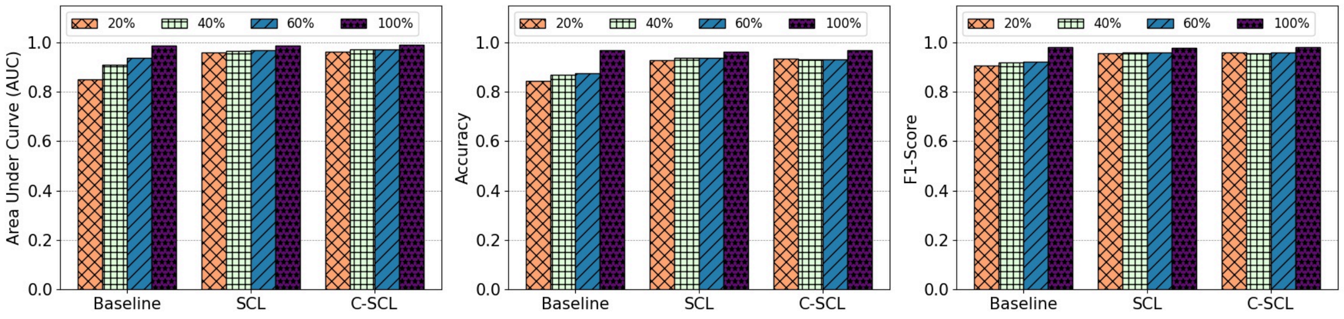


Fig. 5: Comparison of the receiver operating characteristic (ROC) curves of the different approaches trained on different numbers of abnormal activity samples from the AUC dataset

For the AUC dataset, the training set includes 1 normal driver activity and 9 abnormal activities. In this experiment, separate models are trained on the 1 normal activity in combination with 2, 4, 6, and all abnormal activities, with the same hyperparameter configuration, where the input is resized to (224×224), the batch size is 256, the learning rate is $1e \times 10^{-4}$ with the SGD optimizer, and the temperature coefficient τ of the contrastive loss is 0.25. The experimental results are evaluated in terms of various metrics, as shown in Fig. 5, Fig. 6 and Tab. II. For the *Accuracy* and the *F1-Score*, the best result is found by traversing different thresholds. In this comparison, the used feature extractor is Resnet18 for all approaches, while the *Baseline* method, as above described, utilizes the final binary output to recognize the abnormal samples. SCL denotes the supervised contrastive loss without the center modification and the clustering loss, and C-SCL denotes the proposed loss function. With both the SCL and the C-SCL, the abnormal samples can be identified by calculating the minimum distance from the normal activity representation of the training set rather than the direct binary output, and the



(a)

(b)

(c)

Fig. 6: Comparison of the different approaches trained on different numbers of abnormal activity samples from the AUC dataset.

TABLE II: Comparison of the different approaches trained on different numbers of abnormal activity samples from the AUC dataset.

Method	Abnormal Number	AUC (%)	Accuracy (%)	F1-Score (%)
Baseline	20%	84.93	84.30	90.57
	40%	90.94	86.80	91.84
	60%	93.78	87.47	92.18
	100%	98.81	96.97	98.07
SCL	20%	96.00	92.92	95.57
	40%	96.56	93.61	95.98
	60%	96.92	93.59	95.97
	100%	98.62	96.27	97.63
C-SCL	20%	96.07	93.36	95.86
	40%	97.09	93.18	95.71
	60%	97.17	93.22	95.76
	100%	99.01	96.90	98.03

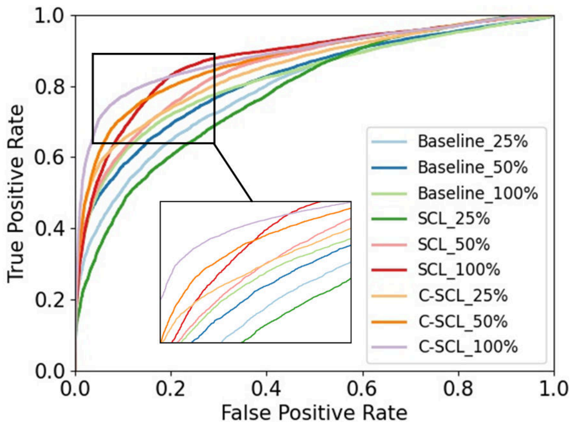
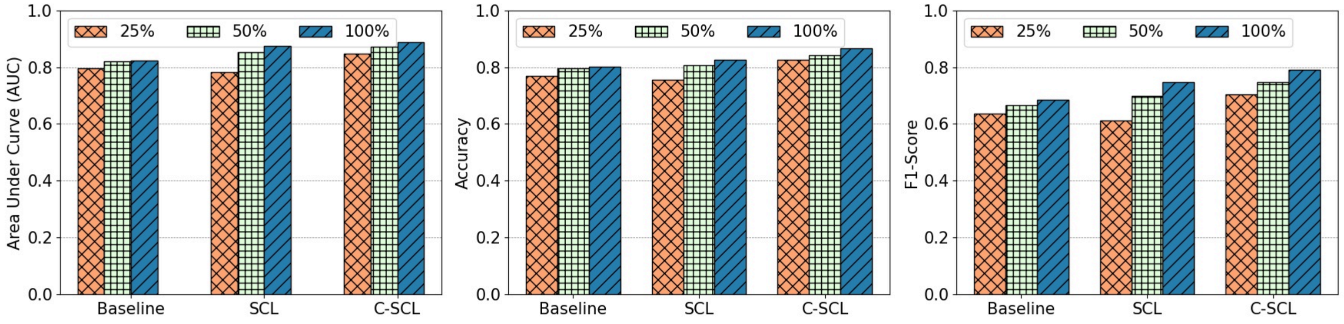


Fig. 7: Comparison of the ROC curves of the different approaches trained on different numbers of abnormal activity samples from the DAD dataset.

minimum distance is calculated by using the KNN algorithm with the hyper-parameter 2. A comparison of the results shows that different approaches achieve equivalent performance when the models are trained with all abnormal samples, and even the baseline model slightly outperforms the SCL method at this point. However, the contrastive learning approach significantly outperforms the baseline approach when training is conducted with partial abnormal samples, where the performance of the baseline drops drastically. This demonstrates that the contrastive learning approach is more robust in recognizing unseen abnormal activities. Further, the proposed C-SCL leveraging representation clustering can improve the model performance without requiring additional computations compared to the SCL.

A similar comparison is also conducted on the DAD dataset, with the different approaches being trained on 2, 4, and 8 abnormal activities. For the DAD dataset, the input consists of 16 front-view depth maps with dimensions of (112×112). Therefore, the used feature extractor is the 3D Resnet18 for handling the sequential input in this comparison. The training hyperparameters are the same as those used on the AUC dataset. The experimental results are shown in Fig. 7 and Fig. 8. A comparison of the different approaches yields similar conclusions to those found on the AUC dataset. It is worth noting that the training set of the DAD dataset contains only 8 abnormal activities, whereas the test set contains 16 additional abnormal activities. Therefore, the performances of the different approaches when trained on all training samples are still not equivalent, unlike the case of the AUC dataset. The performance of the model trained using the proposed C-SCL approach on only two abnormal activities is competitive with that of the model trained using the baseline approach on all eight abnormal activities in the training set. This finding further demonstrates that contrastive learning is more robust than the baseline approach in recognizing unknown abnormal activities. Furthermore, the proposed C-SCL can more significantly improve the model performance on the DAD dataset compared to the AUC dataset.

The proposed contrastive learning approach with representation clustering has been evaluated on datasets corresponding to two different modes. The experimental results demonstrate

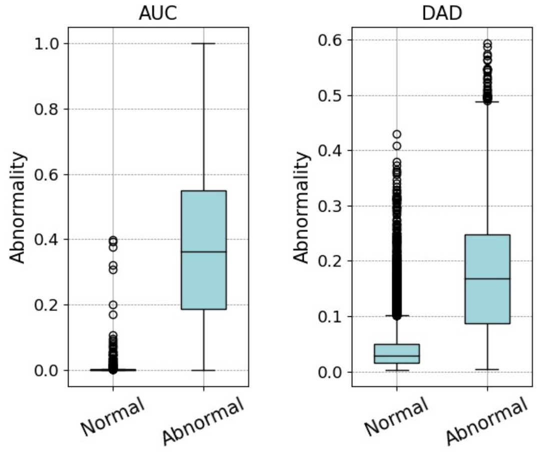


(a)

(b)

(c)

Fig. 8: Comparison of the different approaches trained on different numbers of abnormal activity samples from the DAD dataset.



(a)

(b)

Fig. 9: Abnormality scores of normal and abnormal samples calculated using the proposed method on the AUC and DAD test sets.

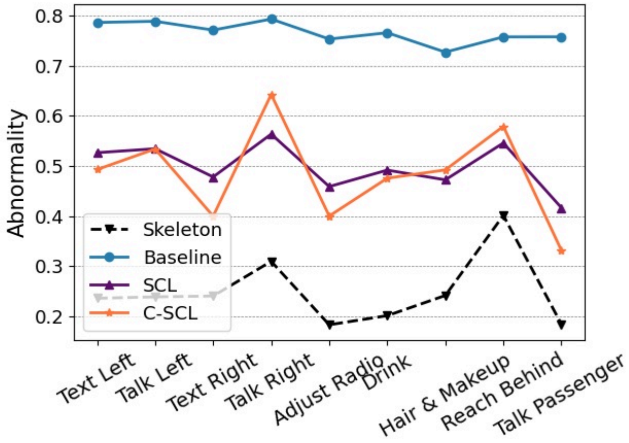


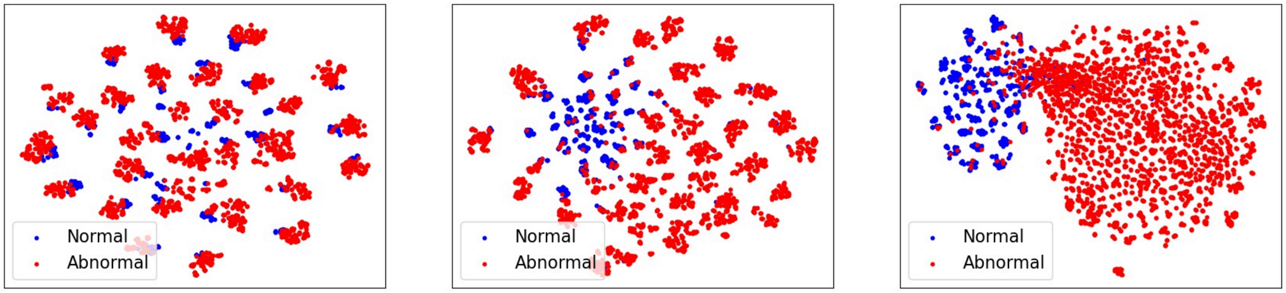
Fig. 10: Comparison of the mean abnormality scores of different activities as calculated with different approaches on the AUC test set.

the efficiency and feasibility of its obtained abnormality value for recognizing abnormal activities; in particular, it is robust in recognizing unknown anomalies.

C. Continuous Abnormality Quantification Evaluation

The core point of this study is to obtain a continuous abnormality value that can quantify the driver anomalies. The preceding experiments have demonstrated that the obtained abnormality value used proposed method can efficiently recognize abnormal activities in particular previously unseen anomalies. Another important characteristic we wish to achieve is that the model output can correctly reflect the degree of abnormality. The distributions of the abnormality values calculated using the proposed method on the test sets of the two datasets used in this study are shown in Fig. 9. First, the results verify that the proposed method can clearly split normal and abnormal samples, with the exception of only a few samples and outliers. The abnormality values of the normal samples are small and close together, especially on the AUC dataset, while the abnormality values of abnormal ones are diversely distributed and separated from the normal ones. The results also reflect the greater diversity of the AUC dataset (which consists of RGB images) compared to the DAD dataset (which consists of depth maps) from the perspective of the distribution of the data domain. The distinction between the normal and abnormal samples reveals that the proposed method provides a loose range for selecting a reasonable threshold and demonstrates its robustness in determining the threshold.

To further evaluate the rationality of the abnormality values for different kinds of abnormal activities, the skeleton poses of the drivers in the AUC dataset are obtained using a human body pose detector. The obtained skeleton keypoints are normalized and reshaped into a vector to calculate the distances of different abnormal activities from normal ones. The mean abnormality values of the various abnormal activities calculated with the different approaches are shown in Fig. 10, where the abnormality values calculated from the



(a) Baseline

(b) SCL

(c) C-SCL

Fig. 11: Visualization of the features extracted using different approaches on the AUC test set, generated by leveraging the t-SNE algorithm. The results show that the proposed C-SCL has the best clustering characteristics.

skeleton keypoints are adopted as the reference and conform to intuitive knowledge. The *Reach Behind* activity shows the largest pose change amplitude, followed by the *Talk Right* activity due to the camera being mounted on the right, which causes the Talk Right activity to be more obvious from the captured images than the Talk Left activity. Fig. 10 shows that the results obtained with the proposed C-SCL are most consistent with the skeleton results, while the baseline model has difficulty reflecting the differences between the different abnormal activities. This shows that the output of the proposed method yields more intuitive results that can better reflect the degree of abnormality.

D. Visualization of the Proposed Approach

To further understand the proposed method, the features extracted using the different approaches on the AUC test set are visualized using the t-distributed stochastic neighbor embedding (t-SNE) algorithm[53]. The results can be found in Fig. 11. This figure clearly shows that contrastive learning can cluster the normal samples, while the features of normal samples and abnormal samples as extracted by the baseline model are entangled. Compared with the SCL approach, the proposed C-SCL leverages the clustering concept and translates the center vector, resulting in more significant distinctions between the normal and abnormal samples. This characteristic is beneficial for quantifying driver anomaly with a continuous value and improving the model performance.

V. Conclusion

Driver anomaly quantification is a fundamental task for understanding a driver's state and building a human-centric intelligent driving system. Existing studies have basically treated it as a classification task, for which outstanding performance can be achieved by leveraging the capabilities of deep learning. However, there are two potential problems that need to be further investigated. The first is that it is impossible for any collected dataset to cover all types of activities, and the existing datasets contain only several typical activities, which will limit the recognition capabilities of models trained on these datasets for unseen activities. The second problem is that the typical driver activity classification is not a natural solution for the downstream applications, in particular, most shared control or decision algorithms need a continuous value that can indicate the driver's state rather than a discrete classification result. To overcome these problems and bridge the gap, this study revisits it and applies the contrastive learning approach. The aim of this study is to train a feature extractor with good representation capabilities and build a set of representation vectors of normal driver activities based on the trained extractor. In the testing phase, the anomalous nature of a sample can be quantified by calculating the distance from its features to the predefined representation set rather than being indicated by the direct output of the model. The calculated continuous distance can be used not only to distinguish anomalies but also to indicate the degree of abnormality. This approach requires the distribution of the extracted features to exhibit alignment and uniformity simultaneously. Therefore, this study introduces the clustering conception to enhance the representation capabilities of the model, and the proposed C-SCL can be used to further cluster normal activity samples while separating samples with different labels without additional computation. Comprehensive experiments are conducted on datasets corresponding to two different modes. The experimental results show that the proposed contrastive learning approach with representation clustering is more robust in recognizing previously unseen abnormal activities and that its continuous output can better indicate the degree of abnormality than the output of the baseline method relative to the variation in the skeleton information. Combined with the visualization of the features extracted using different approaches, the experimental comparisons demonstrate that the proposed method is efficient and feasible in quantifying the anomalies, and the obtained continuous abnormality value is reasonable.

Acknowledgment

This work was supported in part by A*STAR Grant (No. W1925d0046) of Singapore, National Key Research and the Alibaba Group, through the Alibaba Innovative Research Program and the Alibaba–Nanyang Technological University Joint Research Institute (No. AN-GC-2020-012).

References

[1] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, 2020.

[2] J. Wu, Z. Huang, C. Huang, Z. Hu, P. Hang, Y. Xing, and C. Lv, "Human-in-the-loop deep reinforcement learning with application to autonomous driving," *arXiv preprint arXiv:2104.07246*, 2021.

[3] P. Hang, C. Lv, C. Huang, Y. Xing, and Z. Hu, "Cooperative decision making of connected automated vehicles at multi-lane merging zone: A coalitional game approach," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[4] P. Hang, C. Huang, Z. Hu, Y. Xing, and C. Lv, "Decision making of connected automated vehicles at an unsignalized roundabout considering personalized driving behaviours," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4051–4064, 2021.

[5] Y. Xing, C. Lv, X. Mo, Z. Hu, C. Huang, and P. Hang, "Toward safe and smart mobility: Energy-aware deep learning for driving behavior analysis and prediction of connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[6] Z. Hu, C. Lv, P. Hang, C. Huang, and Y. Xing, "Data-driven estimation of driver attention using calibration-free eye gaze and scene features," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1800–1808, 2022.

[7] A. Kashevnik, I. Lashkov, and A. Gurtov, "Methodology and mobile application for driver behavior analysis and accident prevention," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 6, pp. 2427–2436, 2020.

[8] B. Baheti, S. Talbar, and S. Gajre, "Towards computationally efficient and realtime distracted driver detection with mobilevgg network," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 4, pp. 565–574, 2020.

[9] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa, "Real-time distracted driver posture classification," *arXiv preprint arXiv:1706.09498*, 2017.

[10] Z. Hu, Y. Hu, J. Liu, B. Wu, D. Han, and T. Kurfess, "3d separable convolutional neural network for dynamic hand gesture recognition," *Neurocomputing*, vol. 318, pp. 151–161, 2018.

[11] T. K. Chan, C. S. Chin, H. Chen, and X. Zhong, "A comprehensive review of driver behavior analysis utilizing smartphones," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 10, pp. 4444–4475, 2019.

[12] Z. Hu, Y. Hu, J. Liu, B. Wu, D. Han, and T. Kurfess, "A crnn module for hand pose estimation," *Neurocomputing*, vol. 333, pp. 157–168, 2019.

[13] J. Wang, W. Chai, A. Venkatachalapathy, K. L. Tan, A. Haghighat, S. Velipasalar, Y. Adu-Gyamfi, and A. Sharma, "A survey on driver behavior analysis from in-vehicle cameras," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[14] Z. Hu, Y. Hu, B. Wu, J. Liu, D. Han, and T. Kurfess, "Hand pose estimation with multi-scale network," *Applied Intelligence*, vol. 48, no. 8, pp. 2501–2515, 2018.

[15] Z. Hu, Y. Xing, C. Lv, P. Hang, and J. Liu, "Deep convolutional neural network-based bernoulli heatmap for head pose estimation," *Neurocomputing*, vol. 436, pp. 198–209, 2021.

[16] 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*, pp. 1–8, 2020.

[17] S. Monjezi Kouchak and A. Gaffar, "Detecting driver behavior using stacked long short term memory network with attention layer," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3420–3429, 2021.

[18] M. Tan, G. Ni, X. Liu, S. Zhang, X. Wu, Y. Wang, and R. Zeng, "Bidirectional posture-appearance interaction network for driver behavior recognition," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021.

[19] B. Qin, J. Qian, Y. Xin, B. Liu, and Y. Dong, "Distracted driver detection based on a cnn with decreasing filter size," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2021.

[20] C. Zhang, R. Li, W. Kim, D. Yoon, and P. Patras, "Driver behavior recognition via interwoven deep convolutional neural nets with multi-stream inputs," *IEEE Access*, vol. 8, pp. 191 138–191 151, 2020.

[21] P. Li, Y. Yang, R. Grosu, G. Wang, R. Li, Y. Wu, and Z. Huang, "Driver distraction detection using octave-like convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2021.

[22] C. Ou and F. Karray, "Enhancing driver distraction recognition using generative adversarial networks," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 3, pp. 385–396, 2020.

[23] 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*, vol. 22, no. 9, pp. 5783–5801, 2021.

[24] Z. Wharton, A. Behera, Y. Liu, and N. Bessis, "Coarse temporal attention network (cta-net) for driver's activity recognition," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 1278–1288.

[25] C. Huang, C. Lv, P. Hang, Z. Hu, and Y. Xing, "Human machine adaptive shared control for safe driving under automation degradation," *IEEE Intelligent Transportation Systems Magazine*, 2021.

[26] C. Huang, P. Hang, Z. Hu, and C. Lv, "Collision-probability-aware human-machine cooperative planning for safe automated driving," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 9752–9763, 2021.

[27] C. Huang, C. Lv, P. Hang, Z. Hu, and Y. Xing, "Human machine adaptive shared control for safe driving under automation degradation," *IEEE Intelligent Transportation Systems Magazine*, 2021.

[28] 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.

[29] A. Fasanmade, Y. He, A. H. Al-Bayatti, J. N. Morden, S. O. Aliyu, A. S. Alfakeeh, and A. O. Alsayed, "A fuzzy-logic approach to dynamic bayesian severity level classification of driver distraction using image recognition," *IEEE Access*, vol. 8, pp. 95 197–95 207, 2020.

[30] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[31] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," *arXiv preprint arXiv:2003.04297*, 2020.

[32] T. Wang and P. Isola, "Understanding contrastive representation learning through alignment and uniformity on the hypersphere," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9929–9939.

[33] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *arXiv preprint arXiv:2004.11362*, 2020.

[34] Z.-X. Hu, Y. Wang, M.-F. Ge, and J. Liu, "Data-driven fault diagnosis method based on compressed sensing and improved multiscale network," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 4, pp. 3216–3225, 2020.

[35] K. Zhou, C. Yang, J. Liu, and Q. Xu, "Dynamic graph-based feature learning with few edges considering noisy samples for rotating machinery fault diagnosis," *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2021.

[36] C. Yang, K. Zhou, and J. Liu, "Supergraph: Spatial-temporal graph-based feature extraction for rotating machinery diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 4, pp. 4167–4176, 2022.

[37] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, "Toward human-centered automated driving: A novel spatiotemporal vision transformer-enabled head tracker," *IEEE Vehicular Technology Magazine*, pp. 2–9, 2022.

[38] 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.

[39] C. Huang, X. Wang, J. Cao, S. Wang, and Y. Zhang, "Hcf: A hybrid cnn framework for behavior detection of distracted drivers," *IEEE Access*, vol. 8, pp. 109 335–109 349, 2020.

[40] D. Liu, T. Yamasaki, Y. Wang, K. Mase, and J. Kato, "Tml: A triple-wise multi-task learning framework for distracted driver recognition," *IEEE Access*, vol. 9, pp. 125 955–125 969, 2021.

[41] 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*, vol. 22, no. 3, pp. 1650–1662, 2021.

[42] O. Kopuklu, J. Zheng, H. Xu, and G. Rigoll, "Driver anomaly detection: A dataset and contrastive learning approach," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 91–100.

[43] M. Kaya and H. Ş. Bilge, "Deep metric learning: A survey," *Symmetry*, vol. 11, no. 9, p. 1066, 2019.

[44] F. Cakir, K. He, X. Xia, B. Kulis, and S. Sclaroff, "Deep metric learning to rank," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1861–1870.

[45] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, "Learning deep representations

by mutual information estimation and maximization," *arXiv preprint arXiv:1808.06670*, 2018.

[46] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE transactions on systems, man, and cybernetics*, no. 4, pp. 580–585, 1985.

[47] 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.

[48] 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.

[49] K. Hara, H. Kataoka, and Y. Satoh, "Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?" in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 6546–6555.

[50] 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.

[51] M. Alotaibi and B. Alotaibi, "Distracted driver classification using deep learning," *Signal, Image and Video Processing*, vol. 14, no. 3, pp. 617–624, 2020.

[52] A. Behera and A. H. Keidel, "Latent body-pose guided densenet for recognizing driver's fine-grained secondary activities," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2018, pp. 1–6.

[53] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

Zhongxu Hu received a mechatronic Ph.D. degree from the Huazhong University of Science and Technology of China, in 2018.

He was a senior engineer at Huawei. He is currently a research fellow within the Department of Mechanical and Aerospace Engineering of Nanyang Technological University in Singapore. His current research interests include human-machine interaction, computer vision, and deep learning applied to driver behavior analysis and autonomous vehicles in multiple scenarios.

Dr. Hu served as a Lead Guest Editor for Computational Intelligence and Neuroscience, an Academic Editor/Editorial Board for Automotive Innovation, Journal of Electrical and Electronic Engineering, and is also an active reviewer for IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Industrial Electronics, IEEE Intelligent Transportation Systems Magazine, Journal of Intelligent Manufacturing, and Journal of Advanced Transportation et al.

Yang Xing received his Ph. D. degree from Cranfield University, UK, in 2018. He is currently a Lecture with Centre for autonomous and cyber-physical systems, Department of Aerospace, Cranfield University. Before joining Cranfield, he was a Research Associate with the Department of Computer Science, University of Oxford, UK, from 2020 to 2021, and a Research Fellow with the School Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore, from 2019 to 2020. His research interests include machine learning, human behavior modeling, intelligent multi-agent collaboration, and intelligent/autonomous vehicles. He received the IV2018 Best Workshop/Special Issue Paper Award.

Dr. Xing serves as a Guest Editor for IEEE Internet of Thing, IEEE Intelligent Transportation Systems Magazine, and Frontiers in Mechanical Engineering. He is also an active reviewer for IEEE Transactions on Intelligent Transportation Systems, Vehicular Technology, Industrial Electronics, and IEEE/ASME Transactions on Mechatronics, etc.

Weihao Gu graduated from Beijing Jiaotong University, major in computer science. He is the co-founder and CEO of HAOMO.AI, a top-tier startup company in the field of autonomous driving. His current research focuses on autonomous driving, artificial intelligence, and big data. He joined Baidu in 2003, and held roles including chief architect for MP3 and video search, head of speech recognition technology, deputy general manager of Baidu map, and general manager of the company's intelligent car division. He led his team to create the first self-driving map in China. He has also developed the first low-cost assisted driving solution and low-cost automatic parking solution in China. He is a pioneer and innovator in the field of autonomous driving in China.

Dongpu Cao is an associate professor with School of Vehicle and Mobility, Tsinghua University, China. He was the director of the Driver Cognition and Automated Driving Laboratory at the University of Waterloo. He received his Ph.D. degree from Concordia University, Montreal, Quebec, Canada, in 2008. His research interests include vehicle dynamics and control, driver cognition, automated driving, and parallel driving, where he has contributed more than 150 publications and one U.S. patent.

Chen Lv is currently an assistant professor at Nanyang Technology University, Singapore. He received a Ph.D. degree at the Department of Automotive Engineering, Tsinghua University, China in 2016. He was a joint Ph.D. researcher at EECS Dept., University of California, Berkeley, and a research fellow at Cranfield University, UK. His research focuses on advanced vehicles and human-machine systems, where he has contributed over 100 papers and obtained 12 granted China patents.

2022-03-30

Driver anomaly quantification for intelligent vehicles: a contrastive learning approach with representation clustering

Hu, Zhongxu

IEEE

Hu Z, Xing Y, Gu W, et al., (2023) Driver anomaly quantification for intelligent vehicles: a contrastive learning approach with representation clustering. IEEE Transactions on Intelligent Vehicles, Volume 8, Issue 1, January 2023, pp. 37-47

https://doi.org/10.1109/TIV.2022.3163458