Title

Author 1 [1] Author 2 [1] Author 3 [2] Author 4 [2] Author 5 [1] Author 6 [3]

[1]University of Warwick

[2]Institution 1

[3]Institution 2

Sun 18th Jun, 2023

Outline

Introduction

Modern *driver monitoring systems* (DMSs) in Level-2+ self-driving-enabled cars aim to enhance safety by estimating drivers' readiness levels for driving and enabling safe control handovers when necessary.

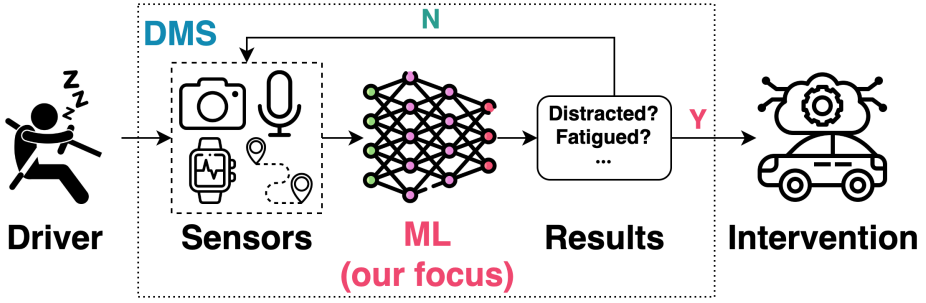

Fig. 1: A simplified illustration of a DMS.

These systems usually rely various sensors, which may be deployed at different in-car locations, to comprehensively monitor drivers' states, e.g.,

- **RGB**: optical details.
- **Depth**: 3D information.
- **Infrared**: thermal information.
- **ECG**: heart rates.
- **Audio**: speech and sound.

Hence, modern DMSs are *multimodal* (and *multiview*).

WARWICK

Our work specifically focuses on *driver action recognition*, which involves classifying drivers' actions into *normal driving* and several *non-driving-related activities* (NDRAs), e.g., texting and drinking.

(a) Top IR **(b)** Top Depth **(c)** Front IR **(d)** Front Depth

Fig. 2: Sample frames from the DAD dataset [1].

WARWICK

Our contributions in this paper are as follows:

1. We propose a novel robust *multiview multimodal* DMS for driver action recognition that leverages feature-level fusion through masked *multi-head self-attention* (**MHSA**).
2. We manually annotated the anomalies in DAD dataset with 9 fine-grained classes of non-driving-related activities (NDRAs).
3. We conduct extensive experiments on the DAD dataset to compare different fusion strategies, assess the significance of individual views/modalities, and evaluate the efficacy of patch masking in enhancing MHSA's robustness against view/modality collapses. Results show that our MHSA-based DMS achieves state-of-the-art performance with an AUC-ROC score of 97.0%.

Related Work

WARWICK

- AUC-DD [2] is the first public dataset for DMSs. It was collected using an RGB camera from a single side view and thus have some limitations.

Fig. 3: A sample from the AUC-DD dataset [2] illustrating that RGB is not robust to illumination changes.

WARWICK

- Later databases [1], [3]–[5] have incorporated additional views and modalities to address these issues.
  - For example, top and front views have also been introduced to capture the driver's hand and head movements amongst other movements.
  - Regarding modalities, IR and depth have also become popular, as they can provide thermally based features and geometry information, which are complementary to the optical details from RGB.
- Among these datasets, we benchmark our models on DAD [1], the only one designed for SAE L2+ with open-set recognition: its test set contains extra classes of NDRAs in addition to those in the training split.

WARWICK

Various multiview multimodal DMSs have also been proposed with different emphases:

- Kopuklu *et al.* [1] proposed a novel learning framework based on contrastive learning.
- Ortega *et al.* [4] and Su *et al.* [6] proposed to leverage Conv-LSTM structures.
- Only Shan *et al* [7] proposed a feature-level modality fusion method, but it has several drawback:
  - Features are pooled before fusion, which leads to the loss of semantic information.
  - Its fusion module has the additional task of handling the temporal dimension.

Thanks!

References I

[1] O. Köpüklü, 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.

[2] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa, "Real-time distracted driver posture classification," in *Neural Information Processing Systems (NIPS 2018), Workshop on Machine Learning for Intelligent Transportation Systems*, Dec. 2018.

[3] M. Martin, A. Roitberg, M. Haurilet, *et al.*, "Drive&act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2801–2810.

[4] J. D. Ortega, N. Kose, P. Cañas, *et al.*, "Dmd: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis," in *European Conference on Computer Vision*, Springer, 2020, pp. 387–405.

[5] I. Jegham, A. B. Khalifa, I. Alouani, and M. A. Mahjoub, "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3mdad," *Signal Processing: Image Communication*, vol. 88, p. 115 960, 2020.

[6] L. Su, C. Sun, D. Cao, and A. Khajepour, "Efficient driver anomaly detection via conditional temporal proposal and classification network," *IEEE Transactions on Computational Social Systems*, 2022.

References II

[7] G. Shan, Q. Ji, and Y. Xie, "Multi-view vision transformer for driver action recognition," in *2021 6th International Conference on Intelligent Transportation Engineering (ICITE 2021)*, Springer, 2022, pp. 962–973.