Journal of Robotics, Networking and Artificial Life
Vol. 11(2); December (2025), pp. 139-144
ISSN (Online): 2352-6386; ISSN (Print): 2405-9021
https://alife-robotics.org/jrnal.html

Research Article

Deep Learning Approaches for Enhancing Driver Safety Through Real-Time Pose and Emotion Recognition

Hao Feng Chan[1], Dexter Sing Fong Leong[1], Shakir Hussain Naushad Mohamed[1], Wui Chung Alton Chau[1], Andi Prademon Yunus[2], Zheng Cai[3], Xinjie Deng[3], Yit Hong Choo[3], Takao Ito[4*]
[1]*Faculty of Engineering, Deakin University, Australia*
[2]*Telkom University, Indonesia*
[3]*Institute for Intelligent Systems, Deakin University, Australia*
[4]*Hiroshima University, Japan*
Email: s221210401@deakin.edu.au, itotakao@hiroshima-u.ac.jp
[*]Corresponding Author

ARTICLE INFO

Article History

Received 12 December 2024
Accepted 07 August 2025

Keywords

Deep Learning
Driver Monitoring
Fatigue Detection
Pose Estimation
Computer Vision

ABSTRACT

Driver fatigue, feelings of emotional distress and impairment as a result of stressful events can pose significant risks of road safety. This research paper proposes a deep learning-based pose estimation system, which aims to identify unsafe driver states, by quantifying the driver's posture, head orientation, and their gesture movements. The developed model is trained in a diverse range of driving situations and identifies physiological and behavioural markers associated with fatigue driving. Unlike existing methods, it integrates pose estimation in conjunction with emotional and motion cues, allowing it to function reliably even during low-lighting or partially obscured conditions.

1. Introduction

Driver fatigue remains a critical concern influencing road safety worldwide. Fatigue impairs situational awareness, decision-making, and reaction time, leading to a higher risk of accidents [1]. Statistics highlight its impact, with drivers who sleep fewer than four hours being 10.2 times more likely to be involved in crashes. In Australia, fatigue accounts for 20% of serious traffic accidents and 30% of fatalities, as reported by the Transport Accident Commission (TAC) [2]. Long-distance rural driving further exacerbates fatigue-related incidents, emphasizing the need for effective detection and intervention systems [3].

Fatigue detection relies on two primary methodologies: classical machine learning and deep learning approaches. Classical techniques, such as Histogram Oriented Gradient (HOG) [4] and Support Vector Machines (SVM) [5], incorporate features like Eye Aspect Ratio (EAR) [6] and Mouth Aspect Ratio (MAR) [7] to identify drowsiness indicators such as eye closure and yawning. While functional in controlled environments, these methods lack adaptability to diverse conditions like variable lighting or facial occlusions, limiting their real-world applicability [8].

Deep learning advancements address challenges by leveraging high-capacity models capable of processing large datasets. Residual Channel Attention Networks (RCAN), for instance, utilise attention modules to enhance landmark detection accuracy but demand significant computational resources [9], [10], [11]. Recent innovations, such as SRNet-FR with integrated Ghost modules and SimAM, demonstrate improved adaptability and efficiency, achieving 99.03% accuracy with minimal parameters, even under challenging circumstances [12], [13].

Emerging trends incorporate physiological signals alongside visual analysis for enhanced fatigue detection accuracy. Electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) paired with behavioural observations reveal critical performance decline after extended driving periods [14]. Lightweight models like You Only Look Once (YOLO) [15] and PyTorch-based frameworks such as ResNet [16] and EfficientNet [17] offer promising solutions for real-time applications, balancing computational efficiency with robust detection capabilities under dynamic conditions [16], [17], [18], [19], [20].

This research investigates the potential for YOLO for real-time detection of driver fatigue, and compares its performance against ResNet [19] and EfficientNet [17]. ResNet and EfficientNet are only included as comparison

due to their known classification accuracy and scalability. [24].

2. Methodology

This research investigates driver fatigue detection using three efficient deep learning models: YOLO, ResNet, and EfficientNet.

2.1. You Only Look Once (YOLO)

YOLO is a compact but popular model for object detection as it offers high processing speed and strong accuracy [15]. Its single-stage architecture combines feature extraction with classification, thereby reducing computational load [18]. Speed is critical in applications involving real-time data processing, as decisions must be made quickly during real-time interactions with drivers, who at any moment could be found in dangerous situations.

YOLO performs well even in suboptimal settings, such as low light, and when some degree of facial obstruction is present. Its grid method of detection ensures that even with little visual information, the visual content can still be processed effectively [21]. This study utilizes the most recent version, YOLOv11, which has the best baseline detection in terms of accuracy and processing speed, so that is the model selected for this study.

2.2. ResNet

The chosen model for this study is ResNet50, because of its 50 layers structure with residual connections that allow the network to learn complex features including facial expressions, posture, all while addressing the vanishing gradient issue [19]. It is also able to deliver stronger feature extraction compared with smaller models like ResNet18 and ResNet34, and less computationally intensive than ResNet101 or ResNet152. Overall, ResNet50 strikes an appropriate balance making it a feasible option for recognizing fatigue-related cues in the performance of driver behaviour monitoring applications.

2.3. EfficientNet

EfficientNet_B0 is selected for the study because of its lightweight and streamlined architecture, and strong computational efficiency. The EfficientNet family of models has EfficientNet_B0 as the first entry point. The design has fewer layers, narrower channels, and image processed with a resolution of 224×224. These properties make EfficientNet_B0 ideal for distributed resource-constrained environments, such as edge devices or embedded devices [17]. Furthermore, its efficiency and architecture is suitable for real-time driver fatigue detection, enabling continuous tracking of indicators like yawning and eye closure without overloading system resources [20].

2.4. Data Processing and Training

To ensure a fair and consistent performance assessment across the models, a uniform dataset was used throughout the study. A publicly available dataset from Roboflow [22], is selected, and then altered to have two different categories: safe driving and dangerous driving, as shown in Figure 1 and Figure 2.

In this study, safe driving is defined by behaviours such as a steady head posture, no signs of yawning, and open eyes. In contrast, dangerous driving includes indicators such as yawning, tilting of the head or eye closure, which are visually labelled based on these specific behavioural markers. This distinction ensures that the dataset captures critical fatigue-related cues that are easily identifiable during real-time monitoring.

The Roboflow dataset contains a total of 3,544 images. This dataset is then separated into three groups: 10% (335 images) for testing, 20% (713 images) for validation and finally, 70% (2,496 images) used for training. This commonly used strategy in machine learning aims to balance the data ratio in each partition, allowing sufficient data for training while ensuring enough data remains for unbiased validation and evaluation of the models. This dataset balancing can be represented mathematically as:

$$N_{total} = N_{train} + N_{val} + N_{test}$$

Where N_{total} represents the total number of images, N_{train} , N_{val} , N_{test} are the number of training, validation, and test samples respectively. The dataset split represents both classes (safe and dangerous driving) to prevent any bias during training and evaluations [23].

Figure 1 User Driving Safely

Figure 2 User Driving Dangerously

To ensure that all models are fairly evaluated, each model, YOLOv11, ResNet, and EfficientNet, are trained under the same dataset for a total of 25 epochs. Each of the models is split into two classes, safe driving, and dangerous driving by utilizing pose estimation and visual features. YOLOv11 used a batch size of 32, with image size adjusted to keep it lightweight. It has a batch size of 32, and an image size of 640, including 2 data loading workers to maximize efficient training. Each of the models are validated to generate performance metrics. The metrics include precision, recall, and F1-score. with optional prediction saving for further analysis.

The ResNet and EfficientNet models are both fine-tuned, with their final layers restructured to classify the two driver

states. The training was conducted using the Adam optimizer, with a constant learning rate of 0.001. The model updates were guided by cross-entropy loss. All images were resized to 224×224 pixels to fit the architectural input constraints. A batch size of 32 was used during training to balance stability of learning and computational requirements.

The dataset is structured to help mitigate overfitting and maintain consistent model performance. An appropriate training-validation-testing split was utilized to deliver accurate, unbiased assessments of how well each model generalizes unseen data. Tests were quantified by post-training by measuring the models' precision, recall, and F1 scores, which serve as standardized metrics for measuring the models' performance in determining the driver states. Each metric is mathematically defined as follows:

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

These calculations are standard in classification tasks as they provide a sensible way to measure a model's performance. When applied to driver state detection, these metrics are particularly valuable in determining how well a model can distinguish the accuracy of behavioural conditions behind the wheel [24].

Mean Average Precision (mAP) and additional evaluation metrics including accuracy, precision, recall, and F1-score were the principal evaluation metrics in this study. All metrics that are used to provide an evaluation of each model's ability to correctly detect driver states under various conditions. The mAP is a standard metric that averages precision across multiple thresholds, giving precise model performance results. Models are also compared in terms of their computational efficiency, with considerations for real-time processing in driver monitoring systems. All models used the same experimental setup for training and testing. This evaluation framework outlines the potential strengths of each model with respect to classification accuracy, computational cost, and ability to operate at real-time speeds in a driver monitoring device.

3. Results and Discussion

3.1. Simulation and Performance Analysis

In this section, a comparative analysis of YOLOv11, ResNet, and EfficientNet models are presented for the task of fatigue detection. All models are trained under identical conditions, to ensure a fair evaluation. The experiments were conducted on a computer with an NVIDIA GeForce RTX 4060 GPU, an Intel Core i7-12560H CPU at 2.50 GHz, and 16 GB of RAM, reflecting a typical high-performance computing setup.

Table 1 provides a summary of the performance of YOLOv11, ResNet, and EfficientNet in detecting states of alert or fatigue. Key performance metrics such as precision, recall, and F1-score are measured and recorded on the test data after training to assess each model's ability in performing this task.

Table 1 A Comparison of the Performance Metrics of Resnet, EfficientNet, and YOLOv11

Metrics	Resnet	EfficientNet	YOLOv11
Loss	0.68	0.54	0.28
Accuracy	96.85%	97.48%	98.9%
Precision	97.0%	97.4%	98.7%
Recall	96.7%	97.5%	98.8%
F1-Score	96.8%	97.5%	98.8%
mAP (Final)	97.74%	96.85%	98.09%

Compared to ResNet and EfficientNet, which are also evaluated in this study, YOLOv11 produced the best accuracy and F1-score. This advantage is likely a result of YOLOv11's one-stage detection model which extracts features and classifies data at the same time in one single forward pass, resulting in predictions that are fast and highly accurate with minimal time difference. ResNet and EfficientNet, although both very sophisticated designs, produced lower recall and F1-scores.

One of the factors that may have contributed to the varying performance of the models is the input resolution. YOLOv11 for instance was trained with an image resolution of 640 x 640, allowing it to retain more spatial information. Compared to both ResNet and EfficientNet, which used a lower resolution of 224 x 224. While a lower resolution reduced computational load, it also limited the models' ability to capture finer details, which may have affected detection accuracy, particularly in cases of fatigue or subtle driving behaviours.

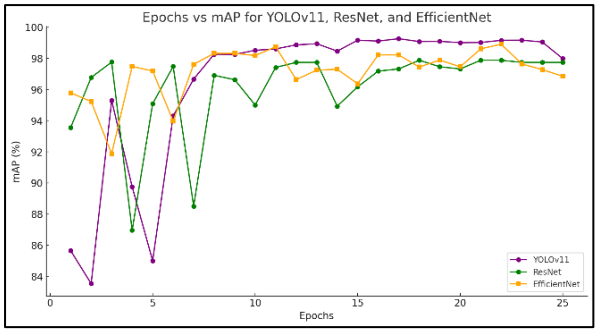

Figure 3 Epochs vs mAP for YOLOv11, ResNet and EfficientNet

The training and validation accuracy plots for all three models are illustrated in Figure 3. YOLOv11 exhibited earlier stabilization in accuracy, which potentially indicates that its convergence was faster than that of ResNet and EfficientNet models. As indicated in the confusion matrices in Figure 4 (YOLOv11) , Figure 5 (ResNet) and Figure 6 (EfficientNet), YOLOv11 had the highest true positive rates for both driving classes, and had the fewest false negatives in the Dangerous Driving class, an important consideration for safety-sensitive applications. ResNet and EfficientNet had slightly more false negatives, which may be related to design limitations that restricted the models' ability to capture limited spatial detail.

As demonstrated in Table 1, YOLOv11 achieved a mean Average Precision (mAP) that surpassed both ResNet and EfficientNet. ResNet and EfficientNet performed well in terms of accuracy and F1-score but lacked in mAP due to their challenges in balancing precision and recall at a range of confidence levels. This performance gap can be associated with their lower input resolution which led to loss of important spatial features and limited their sequential architecture when adapting over different thresholds. YOLOv11, on the other hand, utilizes a grid-based detection approach and an end-to-end design to retain spatial detail and shows consistent high confidence performance for all detection cases.

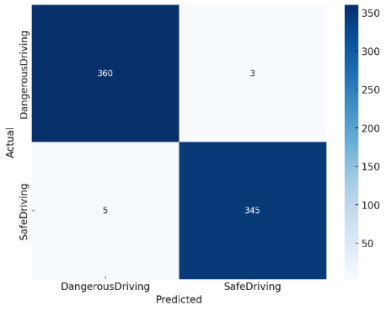

Figure 4 YOLOv11 Output Confusion Matrix

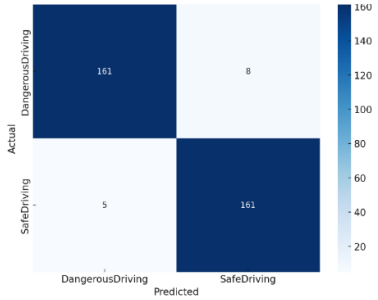

Figure 5 ResNet Output Confusion Matrix

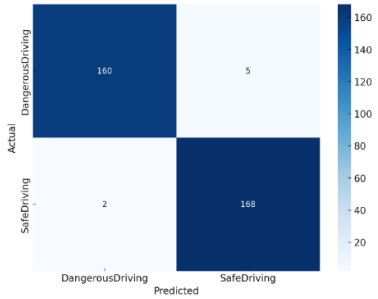

Figure 6 EfficientNet Output Confusion Matrix

4. Conclusion

The comparative investigation presented in this research establishes the potential of deep learning models for improving driver monitoring systems. Although YOLOv11 generally produced the top performance metrics across all categories, its superiority should be interpreted relative to a balance of speed, accuracy, and input resolution. Moreover, ResNet and EfficientNet provided viable options that consume fewer resources, suggesting that the model selection should be based on the intended application and the limits of the system to which it will be integrated. Beyond performance values, this study has established the relevance of assessing models in the context of systems integration. The expected latency in using the model, suitability for either embedded hardware or computers, and the model's resistance to errors presented by the environment like occlusion or low light will determine the actual performance of the 'model' in the field. Ethical concerns raised by the approved resolution modes, such as driver privacy and driver data utilization, will also need to be path followed as these systems move closer to their utilization. Future research will include the expansion to multi-class behaviour classifications, temporal data analysis, and would benefit from investigating hybrid models that include both visual and biometric data streams for detecting driver state to address the safety gap between human driving and autonomous driving.

References

1. Sagberg, F., Road accidents caused by drivers falling asleep. Accident analysis & prevention, 1999. 31(6): p. 639-649.

2. Siskind, V., et al., Risk factors for fatal crashes in rural Australia. Accident Analysis & Prevention, 2011. 43(3): p. 1082-1088.
3. Casey, G.J., T. Miles-Johnson, and G.J. Stevens, Heavy vehicle driver fatigue: Observing work and rest behaviours of truck drivers in Australia. Transportation Research Part F: Traffic Psychology and Behaviour, 2024. 104: p. 136-153.
4. Tamba, M., et al., Classification of Autism Histogram of Oriented Gradient (HOG) Feature Extraction with Support Vector Machine (SVM) Method. 2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA), 2023: p. 144-148.
5. Agrawal, A., R. Gupta, and N.R.S. Jebaraj, Advancing Support Vector Machines for Automated Medical Image Diagnosis. 2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC), 2024: p. 1-7.
6. Ram, D., D.J.V.S. Koushik, and H. Pavan, Drowsiness Detection using EAR (Eye Aspect Ratio) by Machine Learning. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, 2024. 08: p. 1-13.
7. Sri Mounika, T.V.N.S.R., et al. Driver Drowsiness Detection Using Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and Driver Distraction Using Head Pose Estimation. in ICT Systems and Sustainability. 2022. Singapore: Springer Nature Singapore.
8. Florez, R., et al., A Real-Time Embedded System for Driver Drowsiness Detection Based on Visual Analysis of the Eyes and Mouth Using Convolutional Neural Network and Mouth Aspect Ratio. Sensors, 2024. 24(19): p. 6261.
9. Ye, M., et al., Driver fatigue detection based on residual channel attention network and head pose estimation. Applied Sciences, 2021. 11(19): p. 9195.
10. Gowda, M.S., et al. A Multimodal Approach to Detect Driver Drowsiness. in 2024 5th International Conference on Circuits, Control, Communication and Computing (I4C). 2024.
11. Valuthottiyil Shajahan, T., B. Srinivasan, and R. Srinivasan, Real-Time Fatigue Monitoring System in Diverse Driving Scenarios. 2024. 124-127.
12. Jiang, H. and W. Xu. A lightweight method for fatigue driving detection based on facial analysis. in 2024 36th Chinese Control and Decision Conference (CCDC). 2024.
13. Alansari, M., et al., GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations. IEEE Access, 2023. 11: p. 35429-35446.
14. Li, T., et al., Advancements in Fatigue Detection: Integrating fNIRS and Non-Voluntary Attention Brain Function Experiments. Sensors, 2024. 24(10): p. 3175.
15. Khanam, R. and M. Hussain, YOLOv11: An Ox`verview of the Key Architectural Enhancements. ArXiv, 2024. abs/2410.17725.
16. Guan, Z., Research Progress on the Development of Fatigue Driving Detection Based on Deep Learning. Theoretical and Natural Science, 2024. 52: p. 128-136.
17. Careem, R., M.G. Md Johar, and A. Khatibi, Deep neural networks optimization for resource-constrained environments: techniques and models. Indonesian Journal of Electrical Engineering and Computer Science, 2024. 33: p. 1843.
18. Alif, M.A.R., YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems. Cornell University, 2024.
19. Prajwal, T.S. and I.A. K, A Comparative Study Of RESNET-Pretrained Models For Computer Vision, in Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing. 2023, Association for Computing Machinery: Noida, India. p. 419–425.
20. Lin, C., et al., Efficient and accurate compound scaling for convolutional neural networks. Neural Networks, 2023. 167: p. 787-797.
21. Shovo, S., et al., Advancing low-light object detection with you only look once models: An empirical study and performance evaluation. Cognitive Computation and Systems, 2024. 6: p. 119-134.
22. Model, D. Driver Detection Dataset. 2024 [cited 2025 January 16]; Open Source Dataset]. Available from: https://universe.roboflow.com/drowsy-model/driver-detection-ajfzd.
23. Zhuo, S., et al. Driver State Monitoring System Based on YOLOv5 and Dlib. 2023. Singapore: Springer Nature Singapore.
24. Wang, J. and Z. Wu, Driver distraction detection via multi-scale domain adaptation network. IET Intelligent Transport Systems, 2023. 17(9): p. 1742-1751.

Authors Introduction

Mr. Hao Feng Chan

He is currently an undergraduate student majoring in Bachelor of Mechatronics Engineering (Honours) at Deakin University, Australia.

Mr. Shakir Hussain Naushad Mohamed

He is currently an undergraduate student majoring in Bachelor of Electrical and Electronics Engineering (Honours) at Deakin University, Australia.

Mr. Dexter Sing Fong Leong

He is currently an undergraduate pursuing Bachelors of Mechatronic Engineering (Honours) in Deakin University, Australia.

Mr. Chau Wui Chung Alton

He is currently an undergraduate student in Bachelor of Mechanical Engineering (Honours) at Deakin University, Australia.

Mr. Zheng Cai

He is currently a PhD candidate at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI). His research interests include multiobjective optimisation algorithms such as metaheuristic algorithms and evolutionary algorithms for scheduling problems. He is also exploring the integration of machine learning with optimisation algorithms.

Ms Deng Xinjie

She is pursuing a PhD in Information Technology at the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, with her research centered on creating lightweight deep learning algorithms for computer vision applications.

Andi Prademon Yunus, Ph.D.

He is an Assistant Professor at Telkom University, and he received his PhD in Engineering from Mie University, Japan. His research focuses on applied and fundamental machine learning for motion and behaviour computing. He also collaborates with industry partners to develop AI-based tools for language modelling and image analytics.

Dr. Yit Hong Choo

He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI), supported by the Rail Manufacturing Cooperative Research Centre (RMCRC). His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Advanced Science and Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.