

Driver State Monitoring Using Pose Estimation: Detecting Fatigue, Stress, and Emotional States for Safer Roads

Hao Feng Chan, Dexter Sing Fong Leong, Shakir Hussain Naushad Mohamed, Wui Chung Alton Chau
School of Engineering, Deakin University, Australia

Andi Prademon Yunus
Telkom University, Indonesia

Takao Ito
Hiroshima University, Japan

Zheng Cai, Xinjie Deng, Yit Hong Choo*
Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
**Email: s221210401@deakin.edu.au, y.choo@deakin.edu.au*

Abstract

Driving under fatigue, stress, or emotional impairment poses significant risks to road safety. This paper proposes a custom pose estimation framework designed to detect driver states, such as fatigue and stress, by analyzing body posture, head pose, and gesture dynamics. Using a novel deep learning approach trained on diverse driving scenarios, the model identifies physiological and behavioral markers associated with impaired states. Unlike existing methods, this system integrates pose estimation with emotional and movement analysis, enabling robust performance in challenging conditions, including poor lighting and occlusions.

Keywords: Driving, Pose estimation, Fatigue, Deep learning, Computer vision

1. Introduction

Driver fatigue is a critical factor contributing to road accidents worldwide, impairing driver attention, decision-making capabilities, and reaction time. According to studies [1], drivers who are deprived of more than four hours of sleep are 10.2 times more likely to cause accidents. In Australia, fatigue is classified as one of the Fatal Five contributory causes of road crashes, alongside speeding, alcohol, distractions, and seatbelt non-compliance. The Transport Accident Commission (TAC) suggests that 20% of serious road crashes and 30% of fatal crashes in Australia are related to driver fatigue [2]. Fatigue-related crashes are significantly more common on rural roadways and during extended driving periods, especially at night [3]. Real-time monitoring and timely warnings can significantly reduce fatigue-related incidents and improve road safety.

Current driver fatigue detection methods can be categorized into traditional machine learning and deep learning approaches. Traditional methods, such as the Histogram Oriented Gradient (HOG) [4] feature descriptor combined with Support Vector Machines (SVM) [5], rely on visual cues like the Eye Aspect Ratio

(EAR) [6] and Mouth Aspect Ratio (MAR) [7] to identify drowsiness through eye closure and yawning. While effective in controlled environments, these methods lack adaptability to varying lighting conditions, diverse facial features, and head postures. Additionally, they are computationally less efficient and unsuitable for real-time applications in dynamic driving environments [8]. Deep learning approaches leverage advanced neural network architectures to enhance accuracy and robustness. For instance, the Residual Channel Attention Network (RCAN) [9] combines facial landmark detection with attention mechanisms to classify the states of the eyes and mouth more accurately. However, these methods require heavy computational resources and large datasets to generalize well across different driving conditions and demographics [10]. Moreover, non-intrusive fatigue detection systems remain sensitive to challenges such as poor lighting, facial occlusions (e.g., sunglasses and masks), and varying driver behaviors, which can reduce detection accuracy in real-world scenarios [11].

Recent advancements, such as lightweight driver fatigue detection methods based on facial analysis, have been proposed to address these challenges. By directly extracting the eye and mouth regions using facial key points, these methods eliminate the need for manual

parameter adjustment and adapt to changes in head posture [12]. For example, a facial ROI state recognition network (SRNet-FR) with the integration of Ghost module and SimAM achieves an accuracy of 99.03% with a parameter size of only 0.61M, meeting practical real-time requirements even under conditions of partial occlusion and low lighting [13].

In addition to visual analysis, emerging studies have explored physiological signals for fatigue detection. For example, a novel study utilized electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) to monitor driver fatigue during seven-hour simulated driving experiments [14]. Results showed a significant link between behavioral data and hemodynamic changes in the prefrontal lobe after four hours, indicating key periods of performance decline. Although limited by a small sample size, the findings align with established fatigue standards and suggest physiological methods can complement visual fatigue detection systems.

To address the challenges of driver fatigue detection, computer vision models utilizing deep learning have emerged as efficient and reliable solutions. Among these models, You Only Look Once (YOLO) [15] stands out due to its speed and accuracy. YOLO is a real-time object detection system that processes images in a single pass, making it significantly faster than other detection models that require multiple passes. YOLO has gained popularity for its versatility and robustness in various computer vision applications, from autonomous driving to surveillance and beyond [16][17]. For driver fatigue detection, YOLO can handle varying lighting, partial occlusions, and diverse behaviors, making it a practical choice. Its lightweight design, optimized for real-time applications, excels at object detection tasks while being open-sourced, enabling broad accessibility and adaptability for diverse implementations, such as detecting the fatigue level of drivers. [18].

In addition to YOLO, PyTorch-based models such as ResNet and EfficientNet have emerged as strong benchmarks for fatigue detection. ResNet, with its skip connections, mitigates the vanishing gradient problem, allowing deep networks to capture intricate patterns like eye closure and head nodding [19]. EfficientNet, using a compound scaling approach, optimizes performance and computational efficiency, making it suitable for resource-constrained environments [20]. Both models have been successfully applied in detecting fatigue indicators like yawning and reduced alertness, providing a different perspective to YOLO [21][22].

This study evaluates YOLO for real-time driver fatigue detection, benchmarking it against ResNet [23] and EfficientNet [20]. ResNet and EfficientNet were included for comparison due to their known classification accuracy and scalability, as they effectively utilize residual connections and compound scaling approaches [23][24].

2. Methodology

This study evaluates driver fatigue detection using three lightweight models: YOLO, ResNet, and EfficientNet.

2.1. You Only Look Once (YOLO)

YOLO is a lightweight, widely used object detection framework known for its high-speed inference and accuracy [15]. Its single-stage architecture integrates feature extraction and classification, optimizing both computational efficiency and model performance [25]. This design eliminates the need for multiple passes through the data, significantly reducing latency compared to multi-stage models, making it particularly suitable for fast decision-making in safety-critical applications like driver monitoring. YOLO excels in challenging conditions, such as low-light environments or partial occlusions, its grid-based detection approach ensures effective feature extraction even from limited or obscured visual data [26]. YOLOv11 is chosen for its latest version and varied configurations: YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11, and YOLOv11x [25]. For this study, YOLOv11 was selected for its balance of computational efficiency and detection accuracy.

2.2. ResNet

ResNet50 was chosen for this study because its 50-layer structure with residual connections effectively captures detailed patterns in facial expressions and poses, while mitigating the vanishing gradient problem [23]. It offers better feature representation than ResNet18 and ResNet34 without the high computational demands of deeper models like ResNet101 and ResNet152. This makes it suitable for detecting fatigue-induced behaviors in driver monitoring systems.

2.3. EfficientNet

EfficientNet_B0 was selected for this study due to its lightweight architecture and computational efficiency. As the baseline model in the EfficientNet family, it uses fewer layers, narrower channels, and processes images at a resolution of 224×224 pixels, making it ideal for resource-constrained environments like embedded systems or edge devices [20]. Its efficiency is suitable for driver fatigue detection applications requiring real-time monitoring and quick decisions [27], enabling continuous processing of cues like yawning or eye closure without straining system resources.

2.4. Data Processing and Training

To compare and verify performance between the models, a standardized dataset was utilized throughout the study. A publicly accessible dataset from Roboflow [28] was used and reorganized into two categories: safe driving and dangerous driving, as shown in Fig. 1 and Fig. 2, respectively.

The dataset consists of 3544 images and is divided into three subsets: 70% (2496 images) for training, 20% (713 images) for validation, and 10% (335 images) for testing. This method in machine learning defines dataset partitions, ensuring balanced and fair distribution for models. It provides ample training data while maintaining enough validation and test samples to consistently assess model performance. Mathematically, this is represented as:

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

where N is the total number of images, N_{train} , N_{val} , N_{test} are the number of training, validation, and testing samples, respectively. The dataset split evenly represents both classes to prevent bias during training and evaluation [29].



Fig. 1 Safe Driving

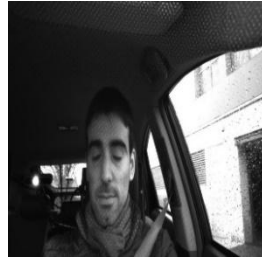


Fig. 2 Dangerous Driving

YOLOv11, ResNet, and EfficientNet were trained on the same dataset for 25 epochs to ensure fair comparison. The models classified images into two categories using pose estimation and visual features. YOLOv11s used a lightweight setup with a batch size of 32, image size of 640 pixels, and two data loading workers for efficiency. Validation measured precision, recall, and F1-score, with optional prediction saving for further analysis.

ResNet and EfficientNet models were fine-tuned on the dataset, with final layers adjusted for two output classes. Training used the Adam optimizer (learning rate 0.001) and cross-entropy loss. Images were resized to 224x224 pixels, with a batch size of 32.

The dataset preparation and split aimed to capture and minimize overfitting in the dataset, while the validation and test sets provided unbiased metrics for evaluating generalization abilities of the model. Performance metrics such as precision, recall, and F1-score were planned for evaluation post-training to measure the effectiveness of each model in detecting driver states defined mathematically as:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

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

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. These formulations are standard in evaluating classification models, providing clear criteria for assessing performance. In the context of driver state detection, such metrics are crucial for determining the effectiveness of models in identifying various driver behaviors [30].

The primary evaluation metric for this study is mean Average Precision (mAP), complemented by classification accuracy, precision, recall, and F1-score. These metrics provide an assessment of the model performance in detecting driver states. All models underwent training and evaluation under identical conditions to ensure a fair comparison. This setup is an assessment of the capability of different models in classifying driver states, with a focus on their computational efficiency, accuracy, and feasibility for real-time applications.

3. Results and Discussion

3.1. Simulation and Performance Analysis

In this section, a comparative analysis of the YOLOv11, ResNet, and EfficientNet models is presented for the task of driver fatigue detection. The experiments were executed on a system equipped with an NVIDIA GeForce RTX 4060 GPU, an Intel Core i7-12560H CPU operating at 2.50 GHz, and 16 GB of RAM.

The comparative performance of YOLOv11, ResNet, and EfficientNet is summarized in Table 1. Performance metrics such as precision, recall, and F1-score were evaluated post-training to assess the effectiveness of each model in detecting driver states.

Table 1. Comparative performance Metrics of YOLOv11, Resnet and EfficientNet

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

YOLOv11 outperformed ResNet and EfficientNet in terms of accuracy and F1-score. This may be attributed to its single-stage detection framework, which integrates feature extraction and classification which results in its

precise predictions with minimal latency. In contrast, ResNet and EfficientNet, despite their advanced architectures, showed lower recall and F1-scores. This is likely due to their reliance on sequential processing, smaller input image sizes (224×224 vs. YOLOv11 640×640), and higher sensitivity to input resizing. These factors reduced their ability to preserve spatial details crucial for distinguishing fine-grained features, particularly in complex scenarios such as detecting the Dangerous Driving class.

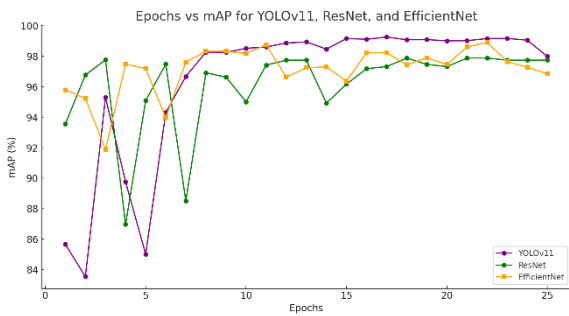


Fig. 3 Epochs vs mAP for YOLOv11, ResNet and EfficientNet

Fig. 3 illustrates the training and validation accuracy curves for all three models. YOLOv11 accuracy stabilized earlier, indicating faster convergence compared to ResNet and EfficientNet. Comparing Fig. 4, Fig. 5 and Fig. 6 their confusion matrix reveal that YOLOv11 achieved the highest true positive rates for both driving classes and exhibited fewer false negatives in the Dangerous Driving class, which is critical for safety applications. ResNet and EfficientNet, on the other hand, had slightly higher false negatives, potentially due to architectural limitations in their ability to process detailed spatial features.

Interestingly, as seen in Table 1, the mean Average Precision (mAP) scores for ResNet and EfficientNet were lower than those for YOLOv11. While the high accuracy and F1-scores of ResNet and EfficientNet suggest strong classification ability, their mAP scores reflect a limitation in balancing precision and recall across varying confidence thresholds. This can be attributed to a combination of factors: the smaller input sizes used for ResNet and EfficientNet led to a loss of critical spatial features, and their sequential processing architectures may struggle with optimizing precision and recall across all confidence thresholds. In contrast, YOLOv11 grid-based detection mechanism and end-to-end design allow it to preserve spatial details effectively and maintain high performance across varying thresholds.

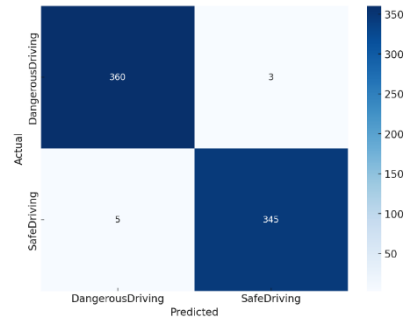


Fig. 4 Confusion Matrix for YOLOv11

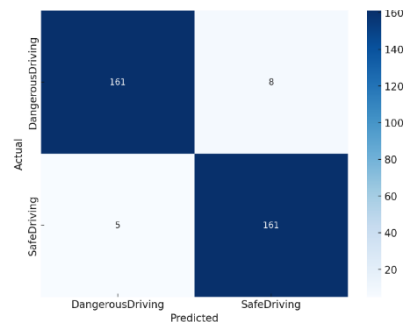


Fig. 5 Confusion Matrix for ResNet

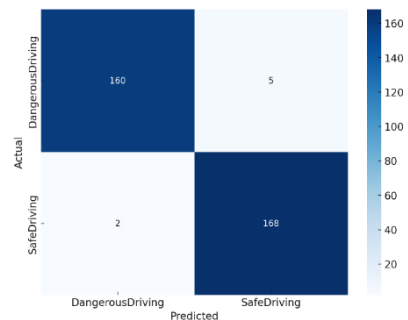


Fig. 6 Confusion Matrix for EfficientNet

4. Conclusion

This study compared the performance of YOLOv11, ResNet, and EfficientNet for driver fatigue detection using metrics such as accuracy, precision, recall, and F1-score. YOLOv11 outperformed the other models, achieving the highest accuracy (98.9%) and F1-score (98.8%), attributed to its single-stage detection framework and larger input resolution (640×640). ResNet and EfficientNet, while achieving competitive results with F1-scores of 96.8% and 97.5%, respectively, were limited by smaller input sizes (224×224) and higher sensitivity to image resizing, reducing their effectiveness in distinguishing complex features. These findings highlight YOLOv11 suitability for real-time driver monitoring systems due to its notable classification performance and computational efficiency. Future work shall focus on testing these models in more diverse and

real-world driving scenarios to further validate their applicability.

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. Valuthotiyil 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 Overview of the Key Architectural Enhancements*. ArXiv, 2024. abs/2410.17725.
16. Redmon, J., et al., *You Only Look Once: Unified, Real-Time Object Detection*. 2016. 779-788.
17. Jiang, P., et al., *A Review of Yolo Algorithm Developments*. Procedia Computer Science, 2022. 199: p. 1066-1073.
18. Jones, A., *Cutting-Edge Detection of Fatigue in Drivers: A Comparative Study of Object Detection Models*. 2024.
19. Guan, Z., *Research Progress on the Development of Fatigue Driving Detection Based on Deep Learning*. Theoretical and Natural Science, 2024. 52: p. 128-136.
20. 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.
21. Alghanim, M., et al., *A Hybrid Deep Neural Network Approach to Recognize Driving Fatigue Based on EEG Signals*. International Journal of Intelligent Systems, 2024. 2024.
22. Khan, T., G.A.-O. Choi, and S.A.-O. Lee, *EFFNet-CA: An Efficient Driver Distraction Detection Based on Multiscale Features Extractions and Channel Attention Mechanism*. LID - 10.3390/s23083835 [doi] LID - 3835. (1424-8220 (Electronic)).
23. 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.
24. Tajane, K., et al., *EffiConvRes: An Efficient Convolutional Neural Network with Residual Connections and Depthwise Convolutions*. 2023. 1-5.
25. Alif, M.A.R., *YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems*. Cornell University, 2024.
26. 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.
27. Lin, C., et al., *Efficient and accurate compound scaling for convolutional neural networks*. Neural Networks, 2023. 167: p. 787-797.
28. Model, D. *Driver Detection Dataset*. 2024 [cited 2025 January 16]; Open Source Dataset]. Available from: <https://universe.roboflow.com/drowsy-model/driver-detection-ajfzd>.
29. Zhuo, S., et al. *on YOLOv5 and Dlib*. 2023. Singapore: Springer Nature Singapore.
30. 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 behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling 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.

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.
