DOI: 10.54254/2755-2721/31/20230119

A review of techniques and methods for deep learning techniques in driver fatigue detection

Yawen Luo

School of Computing and software system, University of Melbourne, Parkville, Victoria, 3010, Australia

1548183470@qq.com

Abstract. Road accidents in which fatigue driving is a significant cause of death are responsible for many deaths worldwide. Approximately 100,000 crashes are caused by driver fatigue each year. Also, fatigue driving is responsible for about 16% of road accidents in general and more than 20% of highway accidents, so fatigue driving accounts for a large percentage of vehicle accidents. Fatigue driving detection usually uses subjective and objective methods. Subjective methods rely on analysing the driver's psychological and facial expression information, while objective methods use external devices to extract feature parameters and apply artificial intelligence algorithms. However, these methods have limitations, such as subjectivity and individual differences. Deep learning, a promising tool inspired by neural networks, offers automatic feature learning, robust pattern recognition, and high adaptability. This review explores the application of deep learning in fatigue driving detection. It examines various deep learning feature extraction methods, classification models, prediction models, and related datasets. By leveraging deep learning techniques, fatigue driving detection can achieve higher accuracy and effectiveness, providing a reliable solution to this critical road safety problem. The review concludes with recommendations and future perspectives in this area.

Keywords: Fatigue Driving, Deep Learning, Driver Monitoring, Feature Extraction, Classification, And Recognition Methods.

1. Introduction

Road injuries are one of the leading causes of human fatalities in the world. Fatigue refers to extreme exertion caused by mental or physical work, in which individuals may sleepiness, lack of energy, loss of attention, and occasionally a significant reduction in cognitive function can be seen, as well as a short decline in cognitive ability [1]. Professor J A Horne et al. found through a survey that fatigue driving is responsible for about 16% of road accidents in general and more than 20% of highway accidents [2]. To prevent fatigue driving from causing significant hazards to the safety of drivers and other road users, effective fatigue driving detection methods need to be proposed and adopted.

In Yang Hai's experiment, Yang Hai selected 166 car drivers to successfully detect driver fatigue through a subjective detection method of semi-structured interviews [3]. Although this subjective method is feasible, the results of the subjective detection method tend to be difficult to compare and analyse reliably due to possible differences in subjective perceptions of different drivers and subjective assessments of the observers and lack objectivity and consistency. Rahim et al. used an infrared heart rate sensor to detect whether the driver is driving fatigued while a pulse sensor measures the heart rate of the driver's fingers or hands [4]. The advantage of this objective detection method is that it has high reliability and accuracy and is not easily influenced by the subjective consciousness of the driver. However, at the same time, the objective detection method has individual differences for different drivers' physiological and behavioural characteristics, and some objective indicators may not be applicable to all people, which also leads to the accuracy and universality of the results.

Deep learning techniques are gradually appearing in the public eye and are achieving important results and widespread applications in various fields. To extract high-level properties from the input, deep learning employs artificial neural networks that were modelled after the biological neurons in the human brain. Deep learning, as an emerging technological tool, can overcome the limitations of traditional detection methods such as automatic feature learning, powerful pattern recognition, high adaptability, and generalization capabilities. This review will evaluate and summarize the methods and applications of deep learning in fatigue driving detection. Firstly, different deep learning feature extraction methods will be summarized, then different classification and prediction models for fatigue driving detection will be presented, and the datasets used for the corresponding experiments will also be presented in the process.

2. FEATURE EXTRACTION METHOD

2.1. Convolutional Neural Network (CNN)

Convolutional neural network (CNN) captures local features and spatial relationships in an image through a series of convolutional and pooling layers, and gradually extracts more abstract features. Weight sharing and hierarchical connectedness with automated self-training are the most crucial image aspects in CNN. During training, semi-connected and fully connected layers serve diverse functions and offer a suitable environment for error feedforward and backward propagation. The term "backward propagation" is frequently used to describe the process of information spreading backward.

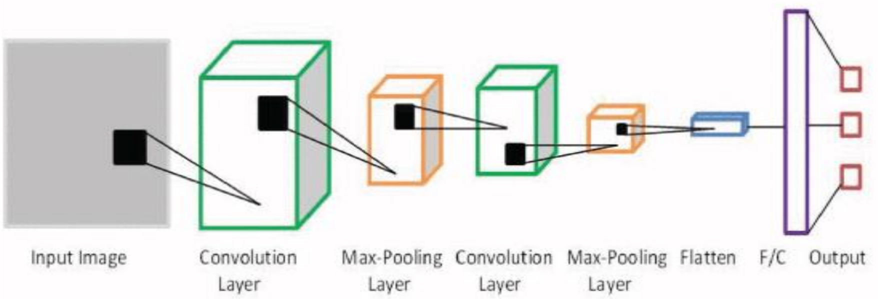

Figure 1. Schematic diagram of convolutional neural network [5].

The application of CNN for image feature extraction in eye-dot map (EOG)-based fatigue driving detection was explored in a study by Xuemin Zhu et al. The model employed two convolutional layers to extract relevant features from the EOG signal, and experiments were conducted using an EOG dataset containing 22 fatigue experimental subjects. In the experiments of Xue-Min Zhu et al. [6] the size of the EOG dataset used was relatively small, which may have limited the generalization ability and reliability of the model. Also, in this experiment, only a comparison with manual ad hoc feature extraction methods was performed, which may be biased and not comprehensive enough in assessing the performance of CNN in fatigue driving detection. To make the detection results more accurate, Zhao Zuopeng et al. used a multi-task cascaded convolutional network (MTCNN) architecture for face facial feature extraction and feature point localization in their study of driving fatigue monitoring methods [7]. In this experiment, MTCNN combines CNN face detection and feature point localization algorithms to provide more accurate face regions and feature point locations, thus improving the accuracy of feature extraction and robustness of feature extraction. Compared with the CNN model, MTCNN improves the

DOI: 10.54254/2755-2721/31/20230119

recognition accuracy of key regions and states, thus overcoming some limitations and restrictions of using CNN only.

The CNN produces satisfactory recognition results in the facial feature extraction segment of detecting fatigued drivers by capturing local features and spatial relationships in the image through a series of convolutional and pooling layers, and gradually extracting more abstract features. After that, more and more people cross-use CNN with other techniques so that some limitations and restrictions of CNN can be better circumvented.

2.2. VGG1-6 Model

VGG 16 is a pre-trained CNN that has been trained on a large-scale image dataset beforehand. Typically, researchers or institutions train a deep convolutional neural network on a huge image dataset (e.g., ImageNet) to enable it to learn rich image feature representation capabilities. These pre-trained CNN models perform well on tasks such as image classification, target detection, and image feature extraction, and have become the base model for many computer vision fields.

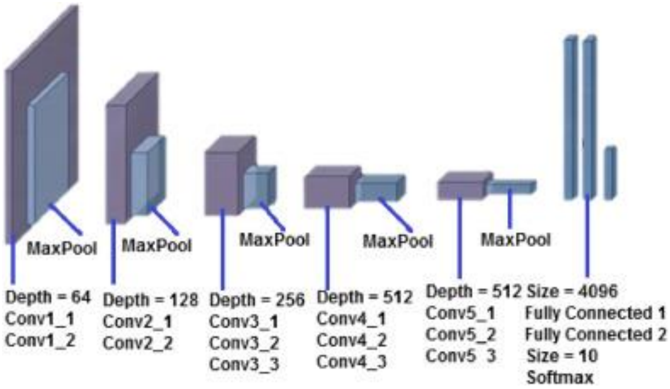

Figure 2. VGG-16 model [7].

Anshul Pintoz et al.'s work used a deep dataset of 10,000 photos that were gathered for training. The authors' average degree of correctness was 93.3%. [8]. This shows that VGG-16 has excellent results and accuracy in extracting the driver's facial features among detecting fatigue driving. However, for this experiment, the model does not consider the effects of motion blur and illumination conditions on the model, which may lead to degradation of classification performance and have a significant impact on accuracy when used in real life, as well as a negative impact on feature extraction and classification. Guangzhe Zhao et al. to further investigate whether VGG16 can be used in the detection of fatigue driving They prepared the NTHU-Drowsy Driver Detection (NTHU-DDD) video dataset and a homemade dataset and used the VGG16 model to extract the eye and mouth features respectively. In the home-made dataset, the accuracy of VGG16 validation was 91.4%, while in the NTHU-DDD dataset, it was 91.88%. [9]. This shows that VGG16 has excellent performance in the application of measuring driver fatigue. It can achieve high accuracy and safety to extract the driver's facial features for fatigue detection.

In general, the network structure of the VGG16 model is relatively deep, requiring high computational resources and time for the training and inference process. This increases the complexity of the model and the difficulty of training. Also, the VGG16 model has many parameters, especially in the fully connected layers, which increases the risk of overfitting, especially when the training data is small. To cope with overfitting, methods such as regularization techniques or data augmentation may need to be used. Compared with CNN models, VGG16, as a deep CNN model, has high accuracy and expressiveness in image recognition and computer vision tasks. It effectively captures features of images through deep network structures and small-sized convolutional kernels and can perform migration learning using pre-trained models.

2.3. AlexNet

AlexNet can distinguish more than a thousand items in a flash because of its neural network that has been trained beforehand with the ImageNet database The ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), started by ImageNet in 2010, is a contest for object recognition. Krizhevsky et al. won first place in the ILSVRC in 2012 with the lowest error rate of 15.3%. [10].

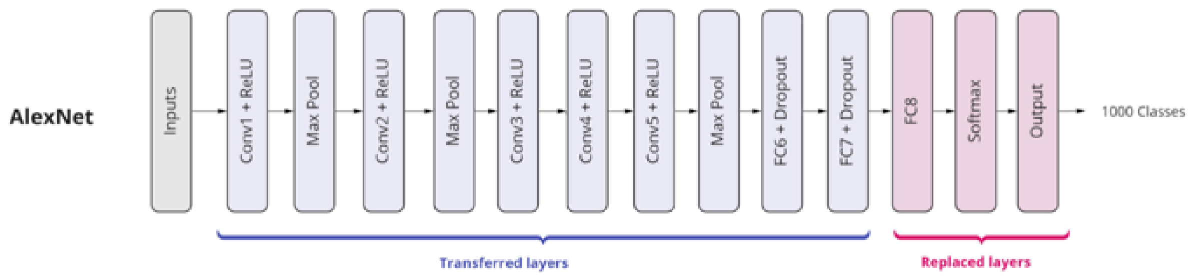

Figure 3. AlexNet model [11].

The dataset used in the study by Salma Anber et al. was a driver drowsiness detection dataset collected in the computer vision laboratory of the National Turinwa University (NTHU). This implementation used the pre-trained neural network AlexNet. The model was not tested on a dataset with real driving conditions in the study by Salma Anber et al. This could lead to a reduction in the generalization ability of the model: a wide variety of outcomes can occur in the real world and the model does not take these factors into account during the training phase, which could lead to a reduction in its generalization ability.

AlexNet outperforms rivals like rectified linear units (ReLUs), which enable the training of big models six times faster than conventional Tanh units. Additionally, the model training time is greatly accelerated by using numerous GPUs. Despite its excellent accuracy, VGG16 requires a lot of memory because it contains 138 million parameters, which is more than twice as many as AlexNet, which only has 62 million. When compared to VGG16, AlexNet offers better characteristics in terms of speed, accuracy, and scale because it requires the least processing power of any other transport network in terms of the number of floating-point operations (FLOPs) needed to run forward pass.

3. CLASSSIFICATION AND RECOGNITION METHOD

3.1. Long term Memory Network (LSTM)

Long-term short-term memory (LSTM) networks are a special kind of recurrent neural networks (RNN). Each block contains recurrently connected memory units and three multiplication units, input, output, and ignore gates, which control the flow of information within the memory block. LSTM can better capture long-term dependencies in sequential data by introducing memory units and gating mechanisms.

The YawDD dataset and the self-built dataset were chosen among Long Chen et al. to verify the performance of the LSTM method. It may have a detrimental effect among practical applications for nighttime detection. Meanwhile Ming-Zhou Liu et al. used LSTM model to detect driver's fatigue state and obtained an accuracy of 99.78% with the fatigue detection algorithm of LSTM through a homemade dataset, i.e., 600 driver images [12].

In conclusion, the LSTM model performs quite well at identifying drivers who are fatigued. Sequential data can be handled by the model. Normal neural networks have pieces that are independent of one another, as well as independent inputs and outputs. Because of this, the output of the first moment has no bearing on the output of the second. And the gradient dispersion problem in the long-term sequential RNN model is effectively solved, so the LSTM model has faster learning speed and higher accuracy.

DOI: 10.54254/2755-2721/31/20230119

3.2. *Multimodal Fusion*

Multimodal fusion refers to the process of integrating and fusing information from different perceptual modalities (e.g., image, speech, text, etc.). In multimodal scenarios, data from different modalities have different feature representations and information expressions, and by fusing them together, more comprehensive, accurate and rich information can be provided.

Xiaomin Li suggested a neural network technique that was trained using SEED-VIG, a public dataset from Shanghai Jiao Tong University, and was based on multimodal fusion of EEG and PFC electrograms. The testing outcomes demonstrated that multimodal fusion, with an accuracy of 98.3%, has greater tiredness detection identification than single modality [13]. This experiment demonstrates that multimodality which helps the application of fatigue detection system in driver driving process. However, there are many methods of multimodal extraction of various physiological signal features among this experiment, and the extraction operation is tedious, and there may be better options.

Multimodal feature fusion methods can significantly improve the accuracy of fatigue detection. With a rich information representation that can contain information from different perceptual modalities, more comprehensive, accurate and rich information can be provided by fusing these data. Also has complementary information, the data of different modalities are often complementary, fusion of multi-modal data can make up for the limitations of single modal data and improve the robustness and accuracy of the model.

4. Public datasets for driver fatigue monitoring

In addition to the previously mentioned datasets, a more detailed selection of datasets is provided for future researchers of deep learning solutions for detecting fatigued driving:

Three datasets, AFW [14], FDDB [15] and PASCAL FACE [16], are widely used for recognizing human facial features. The AFW dataset has 205 images containing 473 labelled faces. The annotations for each face include a rectangular bounding box, 6 landmarks and pose angles, and the experimental test shows the best performance of 97.2% AP. the FDDB dataset contains 2845 images with 5171 annotated faces, and the experimental result shows the highest recall of 91.74%. the PASCAL FACE dataset contains 851 images and 1341 annotated faces with the best result of 92.11% AP. These datasets perform comprehensively and superiorly on the face analysis task, with results having low bias and limited ability to adapt to face appearance and background variations, despite the small data size.

A suitable dataset for drowsiness modelling is DROZY [17], which contains rich drowsiness-related data. In the DROZY dataset, 14 subjects performed three consecutive 10-min psychomotor vigilance tests (PVT). The total sleep deprivation time between the first and third PVT was approximately 28-30 hours. Recent experimental studies (e.g., by Caio Bezerra Souto Maior et al.) have used the DROZY dataset, which did not provide alerts in the alert state but averaged 16.1 alerts in the drowsy state, with an accuracy of 94.44% [18].

There is also the YawDD dataset [19], which provides 322 videos covering a wide range of driving situations, including male and female drivers, wearing/not wearing glasses, and normal driving/yawning while driving. Each yawning video in this dataset contains a yawning event. The yawn detection system in Figure 1 was evaluated on the YawDD dataset, and the results showed a yawn detection accuracy of 92% and a false alarm rate of 13%. In contrast, the YawDD dataset has a richer number of samples and diverse variation conditions compared to the DROZY dataset, however, this also makes it potentially relatively difficult to train models on this dataset.

5. Outlook and development suggestions

Deep learning models can be further optimized and improved to enhance the accuracy and robustness of fatigue driving detection. At the same time, explore adaptive learning methods so that the model can perform personalized fatigue detection based on individual differences among drivers and under different conditions. In addition, go ahead and try to combine with other systems and technologies, such as autonomous driving word count, in-vehicle systems, etc., to achieve real-time monitoring and feedback. When signs of driver fatigue are detected, the driver is alerted to take measures, such as rest or alternate driving, by means of internal vehicle alarms and vibrating seats, thus enhancing driver safety awareness. Many experiments go more to use some idealized data sets, thus ignoring the complex real-life environment, so in subsequent tests, more attempts can be made to detect fatigue driving in complex environments, such as bad weather, night driving and tunnel driving conditions, to further research and improve fatigue driving detection methods. Solve the problems of light changes and background interference to improve the robustness and adaptability of the model. In addition, this paper mainly focuses on driver fatigue detection. Fatigue detection has many application scenarios, for example, whether students are tired, whether company employees feel tired, etc. These areas can be followed up by using transfer learning methods, driver's driving fatigue detection and migration recognition models to complete fatigue detection in other areas.

6. Conclusion

Fatigue driving is one of the major causes of road accidents and poses a serious threat to the safety of drivers and other road users. In recent years, deep learning techniques have made significant progress in the field of fatigue driving detection, offering new possibilities to improve detection accuracy and reliability. In this review, a variety of deep learning models and methods are considered. For feature extraction, models such as convolutional neural networks (CNN), and pre-simulated CNNs are introduced. For classification and testing methods, multimodal fusion, long short-term memory (LSTM), is introduced to implement fatigue driving detection. CNN can be used for image feature extraction to extract key point features such as eyes and mouth for fatigue detection by analysing the driver's facial image. LSTM, on the other hand, can capture time series information and convert single frame feature values into time fatigue feature sequences to determine driver’s fatigue status more accurately. Multimodal fusion technology combines multiple information sources such as facial expressions, voice, and physiological signals to provide more comprehensive and reliable fatigue driving detection results. Datasets such as AFW, FDDB, and PASCAL FACE provide a solid foundation for the field of face analysis, while the DROZY and YawDD datasets provide valuable resources for fatigue detection research. Although each dataset has its strengths and limitations, together they contribute to the development of deep learning in this field. Overall, deep learning technologies have great potential to provide more accurate, reliable, and practical solutions in the field of fatigue driving detection and play an important role in preventing traffic accidents caused by fatigue driving. Through continuous research and innovation, we expect to establish a safer driving environment and protect the lives and properties of drivers and other road users.

References

[1] S. M. Marcora, W. Staiano, and V. Manning, “Mental fatigue impairs physical performance in humans,” Journal of Applied Physiology, vol. 106, no. 3, pp. 857–864, Mar. 2009, doi: https://doi.org/10.1152/japplphysiol.91324.2008.
[2] J. A. Horne and L. A. Reyner, “Sleep related vehicle accidents,” BMJ, vol. 310, no. 6979, pp. 565–567, Mar. 1995, doi: https://doi.org/10.1136/bmj.310.6979.565.
[3] H. Abdul Rahim, A. Dalimi, and H. Jaafar, “Detecting Drowsy Driver Using Pulse Sensor,” Jurnal Teknologi, vol. 73, no. 3, Mar. 2015, doi: https://doi.org/10.11113/jt.v73.4238.
[4] R. Hooda, V. Joshi, and M. Shah, “A comprehensive review of approaches to detect fatigue using machine learning techniques,” Chronic Diseases and Translational Medicine, Aug. 2021, doi: https://doi.org/10.1016/j.cdtm.2021.07.002.
[5] J. Koushik, “Understanding Convolutional Neural Networks,” arXiv:1605.09081 [stat], May 2016, Available: https://arxiv.org/abs/1605.09081
[6] X.-M. Zhu, W.-L. Zheng, B.-L. Lu, X. Chen, S. Chen, and C. Wang, “EOG-based drowsiness detection using convolutional neural networks,” International Joint Conference on Neural Network, Jul. 2014, doi: https://doi.org/10.1109/ijcnn.2014.6889642.

[7] Z. Zhao, N. Zhou, L. Zhang, H. Yan, Y. Xu, and Z. Zhang, "Driver Fatigue Detection Based on Convolutional Neural Networks Using EM-CNN," Computational Intelligence and Neuroscience, vol. 2020, p. e7251280, Nov. 2020, doi: https://doi.org/10.1155/2020/7251280.

[8] S. Masood, A. Rai, A. Aggarwal, M. N. Doja, and M. Ahmad, "Detecting distraction of drivers using Convolutional Neural Network," Pattern Recognition Letters, Jan. 2018, doi: https://doi.org/10.1016/j.patrec.2017.12.023.

[9] S. Anber, W. Alsaggaf, and W. Shalash, "A Hybrid Driver Fatigue and Distraction Detection Model Using AlexNet Based on Facial Features," Electronics, vol. 11, no. 2, p. 285, Jan. 2022, doi: https://doi.org/10.3390/electronics11020285.

[10] G. Zhao, Y. He, H. Yang, and Y. Tao, "Research on fatigue detection based on visual features," IET Image Processing, vol. 16, no. 4, pp. 1044–1053, Apr. 2021, doi: https://doi.org/10.1049/ipr2.12207.

[11] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, Apr. 2015, doi: https://doi.org/10.1007/s11263-015-0816-y.

[12] M. Liu, X. Xu, J. Hu, and Q. Jiang, "Real time detection of driver fatigue based on CNN-LSTM," IET Image Processing, Nov. 2021, doi: https://doi.org/10.1049/ipr2.12373.

[13] X. Li, "Fatigue Detection Based on Multimodal Fusion Neural Network," IEEE Xplore, Feb. 01, 2023. https://ieeexplore.ieee.org/abstract/document/10145407/authors#authors (accessed Jun. 26, 2023).

[14] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," *IEEE Xplore*, Jun. 01, 2011. https://ieeexplore.ieee.org/abstract/document/5995316 (accessed May 20, 2021).

[15] V. Jain and E. Learned-Miller, "FDDB: A Benchmark for Face Detection in Unconstrained Settings." Accessed: Oct. 18, 2019. [Online]. Available: http://vis-www.cs.umass.edu/fddb/fddb.pdf

[16] J. Yan, X. Zhang, Z. Lei, and S. Z. Li, "Face detection by structural models," *Image and Vision Computing*, vol. 32, no. 10, pp. 790–799, Oct. 2014, doi: https://doi.org/10.1016/j.imavis.2013.12.004.

[17] Q. Massoz, T. Langohr, Clémentine François, and J. Verly, "The ULg multimodality drowsiness database (called DROZY) and examples of use," *ORBi (University of Liège)*, Mar. 2016, doi: https://doi.org/10.1109/wacv.2016.7477715.

[18] C. B. S. Maior, M. J. das C. Moura, J. M. M. Santana, and I. D. Lins, "Real-time classification for autonomous drowsiness detection using eye aspect ratio," *Expert Systems with Applications*, vol. 158, p. 113505, Nov. 2020, doi: https://doi.org/10.1016/j.eswa.2020.113505.

[19] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD," *Proceedings of the 5th ACM Multimedia Systems Conference on - MMSys '14*, 2014, doi: https://doi.org/10.1145/2557642.2563678.